CLONING AND CHARACTERISATION OF THE *WAXY* GENES OF WHEAT AND INVESTIGATION OF THEIR POTENTIAL APPLICATIONS

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

by

LIULING YAN

M. Sc. (Agronomy)

Centre for Bioprocessing and Food Technology

Victoria University of Technology

Melbourne

AUSTRALIA

May 1999

DECLARATION

I hereby declare that all work carried out in this thesis was performed by me while I was enrolled as a student for the degree of Doctor of Philosophy in the Centre for Bioprocessing and Food Technology, Victoria University of Technology. To the best of my knowledge, this work has not been submitted in whole or part for any other degree or diploma in any University and no material contained in this thesis has been previously written or published by another person, except where due reference is made in the text.

Liuling Yan

May, 1999

ACKNOWLEDGEMENTS

I am greatly indebted to my principal supervisor, Dr Mrinal Bhave, who has provided invaluable supervision for the design of this project and has made a major contribution to the success of this thesis through her wealth of knowledge in molecular biology.

I am very grateful to Professor Bob Fairclough, my co-supervisor, who has provided his warm supervision and advice, strong support and continuous encouragement for my studies, and critical reading of the manuscript.

I would like to express my thanks to Professor Margaret Britz, the former Director of the Centre for Bioprocessing and Food Technology, for offering me this opportunity for my study.

I would like to convey many thanks to Professor Rudi Appels, at Plant Industry, Commonwealth Scientific and Research Organization (CSIRO), Australia, for his valuable advice for the project, his invitation to work at his lab at Canberra for a few weeks and his providing the sequence of the *waxy* 'D' gene for comparison. Many thanks are to convey to Dr Sadequr Rahman and Dr Matthew Morell for their advice and help during my work in Canberra.

I would like to acknowledge Dr Sarah Fraser for warm help and discussions, Nikola Popovik for his help and assistance in protein work, Matthew, Helen and Hema and all the lab friends for smiling faces, wonderful friendship and enthusiastic assistance.

My sincere thanks go to Victoria University of Technology for awarding me with the Overseas Postgraduate Research Scholarships (OPRS) and University Overseas Student Scholarships throughout this study. I am thankful to the Centre for Bioprocessing and Food Technology for partial financial support, and to all staff in the School of Life Sciences and Food Technology, the Postgraduate Studies Unit and library who have contributed to this project.

Finally, I would like to express my special thanks from the bottom of my heart to my wife, Yan Wang, and my lovely daughter, Jing Yan, for their understanding and full support during the years when this work has taken almost all of my attention and time.

I wish to give my thanks to my parents, brothers and sisters for their love and support for many years.

SUMMARY

Wheat starch consists of amylose and amylopectin. The appropriate end-use of starch is dependent on an important property of starch, the ratio of amylose:amylopectin. Amylose is synthesized by the waxy protein, a predominant enzyme in the endosperm, the granule-bound starch synthase (GBSSI). A great number of efforts are being made worldwide to produce the waxy wheat with amylose-free starch by breeding programs or biotechnological approaches. Molecular genetic information for the different *waxy* genes of wheat will be of great benefit for such programs. This project has the objective of cloning and characterising the individual *waxy* genes from different diploid progenitors and various genomes of polyploid wheat. The waxy proteins encoded by the *waxy* genes of the different types of wheat were analysed.

The investigation of proteins embedded within the starch granules indicated that a single approximately 59 kDa waxy protein was present in the proposed diploid progenitors, and a doublet of waxy protein bands with slight differences in molecular weight occurred in the tetraploid *T. turgidum* (AABB) and hexaploid *T. aestivum* (AABBDD). However, the waxy proteins of the tetraploid *T. timopheevi* (A^tA^tGG) and the hexaploid *T. zhukovskyi* ($A^tA^tA^zA^zGG$) could not be separated into bands of different molecular weights. The protein profile of the 75 kDa soluble starch synthases (SSS) and the 85 kDa starch branching enzymes (SBE) appeared invariable on SDS-PAGE for different types of diploid, tetraploid and hexaploid wheat. In the higher molecular weight starch granule protein (SGP) region, it was found that two SGP bands (110 kDa and 100 kDa) occurred in the tetraploid *T. turgidum* and three SGP bands (110 kDa, 105 kDa and 100 kDa) occurred in the hexaploid *T. aestivum*. However, both the tetraploid *T. timopheevi* and the hexaploid *T. zhukovskyi* were found to contain two SGP bands (a 110 kDa and a 95 kDa). Only one single band, 100 kDa, 105 kDa, 100 kDa or 95 kDa, was detected in the SGP region in the proposed diploid progenitors. This is the first report of a 95 kDa protein in starch granules in endosperm of wheat.

The genomic copies of genes covering the region from the translation initiation codon to termination codon and the partial 5' untranslated leader were cloned and sequenced

from diploid *T. monococcum* and *T. speltoides*. These two genes consisted of 11 exons and 10 introns, the same as in the corresponding section of the barley *waxy* gene but different from the rice, maize and potato *waxy* gene in the number of introns and exons. On the basis of comparison of the two *waxy* genes, the conserved sequences in exons were utilised to design primers to clone the partial *waxy* genes from different types of diploid and polyploid wheat. The variable regions in introns 4 and 5 in these two genes were utilised to distinguish between the different *waxy* loci.

The two mature waxy proteins deduced from the respective *waxy* genes of *T. monococcum* and *T. speltoides* were found to have the same number of amino acids but slightly different calculated molecular weights and isoelectric points. These differences in physical properties of the different waxy proteins resulted from only 3% amino acid substitutions and most probably accounted for the electrophoretic differences in the two proteins revealed by SDS-PAGE. The predicted secondary structures and the number of glycosylation sites also appeared to be variable in the different predicted waxy proteins of the genus *Triticum*. The altered secondary structures of these wheat waxy proteins could be suggested to relate to the different expressions of the waxy proteins or their unequal contributions to amylose synthesis.

A series of partial *waxy* genes have been cloned from the *wx-A* and *wx-B* loci of *T. turgidum*, the *wx-A*^{t} and *wx-G* of *T. timopheevi*, the *wx-A*, *wx-B* and *wx-D* loci of *T. aestivum*, the *wx-A*^{t}, *wx-A*^{z} and *wx-G* loci of *T. zhukovskyi*, and the single *waxy* locus of each of the proposed diploid progenitors of the different genomes of the polyploid wheat. The length and sequence of intron 4 were observed to be highly variable in the two or three *waxy* loci of the polyploid wheat. The identification of the specific sequences in the variable region of intron 4 makes it possible to distinguish between the A, B, D, G, A^t or A^z genomes at the DNA sequence level.

Specific sequences of intron 4 representing individual *waxy* loci and a phylogenetic tree based on the sequence identities of the various partial *waxy* genes were analysed to determine the evolutionary relationship between different genomes of the various diploid and polyploid wheat. *T. urartu*, *T. speltoides* and *T. tauschii* were identified as the possible donors of A, B and D genomes, respectively, to *T. aestivum*,

supporting the current theory for wheat evolution (Friebe and Gill, 1996). This thesis has provided molecular evidence that *T. monococcum* might be one donor of *T. timopheevi* and thus of *T. zhukovskyi.* However, the sequence of intron 4 of the cultivated *T. monococcum* tested in this study was different from that of the third set of genomes of *T. zhukovskyi*, a putative man-made hybrid of *T. monococcum* and *T. timopheevi*. This result suggests that intraspecies polymorphism existed in *T. monococcum* or that a wild *T. monococcum* (*T. monococcum* ssp *boeoticum*) would be the progenitor of the third set of genome *of T. zhukovskyi*. Owing to the fact that all the possible G genome donors had an identical length of intron 4, the progenitor of G genome remained uncertain in this study.

Preliminary work was carried out to develop molecular markers for the three homoeologous *waxy* genes. Variations in the *waxy* gene sequences, certain unique restriction enzyme sites and lengths of certain PCR products were proposed to represent the *wx-A*, *wx-B* and *wx-D* loci of common wheat. Normal, null-A and null-B lines of common wheat were found to have identical length of fragments in the exon 5-intron 5-exon 6 region of the three *waxy* genes. The results thus suggested that all three structural genes existed in the null lines analysed and that the *waxy* gene in these was not deleted completely.

Unequal expressions of different waxy proteins in *T. turgidum* and *T. aestivum* were detected in SDS-PAGE. The sequence comparison of the three *waxy* genomic fragments from common wheat and a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991) also suggested that a possible alternative transcriptional event might be involved in the expression of the *waxy* gene in wheat.

PUBLICATIONS

1. Liuling Yan, Mrinal Bhave, Bob Fairclough, Christine Konik, Sadequr Rahman and Rudi Appels (1999) The genes encoding granule-bound starch synthases at the *waxy* loci of the A, B and D progenitors of common wheat. Genome (in press).

2. Liuling Yan, Bob Fairclough and Mrinal Bhave (1999) Molecular basis of the electrophoretic differences of the waxy proteins in wheat (submitted).

3. Liuling Yan, Bob Fairclough and Mrinal Bhave (1999) Specific sequences of the *waxy* loci from A, B, D, G, A^t and A^z genomes and their potential applications in wheat phylogeny (in preparation).

4. Liuling Yan, Bob Fairclough and Mrinal Bhave (1998) Molecular evidence supporting the origin of B genome of *Triticum turgidum* from *T. speltoides*. Proceeding of the 9th International Wheat Genetics Symposium, Saskatoon, Canada. 2: 119-121

5. Yan L-L., Fairclough R. and Bhave M. (1999) Characterisation of individual *waxy* genes in wheat and their application in food industry. 10th World Congress of Food and Technology, Sydney, Australia, October 1999. p56

6. Liuling Yan, Robert J. Fairclough and Mrinal Bhave (1999) Development of primer sequences and restriction enzymes representing the individual *waxy* locus of common wheat. 48th Cereal Chemistry Conference, Melbourne, September 1999.

7. Yan L. L., Fairclough R. and Bhave M. (1998) Molecular basis of electrophoretic differences in granule-bound starch synthase (GBSS) in wheat. Proceedings of 42nd Annual ASBMB and 38th Annual ASPP Conference, Adelaide, Australia. ISSN 13284924. Pos-Mon-67

8. Liuling Yan, R. J. Fairclough and Mrinal Bhave (1997) Characterisation of the waxy proteins of *T. monococcum.* The 19th Annual Conference on the Organisation and Expression of the Genome, Lorne, Victoria, Australia. Poster 2-41.

9. L-L. Yan, R. Fairclough and M. Bhave (1997) Structure of the *waxy* genes from *T.monococcum* and *T. speltoides.* 47th Cereal Chemistry Conference, Perth, Australia. pp 475-478

ABBREVIATIONS

ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
CTAB	hexadecyltrimethylammonium bromide
dATP	adenosine deoxyribonucleoside triphosphate
dCTP	cytidine deoxyribonucleoside triphosphate
dGTP	guanosine deoxyribonucleoside triphosphate
dNTPs	equimolar mixture of dATP, dCTP, dGTP, dTTP
dTTP	thymine deoxyribonucleoside triphosphate
DDT	dithiothreitol
DNase	deoxyribonuclease
EB	ethidium bromide
EDTA	ethylenediaminotetra-acetic acid
g	times gravitational force
GBSS	Granule-bound starch synthase
HMW	Higher molecular weight
IgG	immunoglobulin G
kb	kilobase pairs
kDa	kilodaltons
LacZ	*E.coli* gene for β-galactosidase
LMW	Lower molecular weight
mRNA	messenger RNA
N-terminus	amino-terminus
OD	optical density
PBS	phosphate buffered saline
PCR	Polymerase Chain Reaction
PEG	polyethylene glycerol
pI	isoelectric points
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute

SBE	starch branching enzymes
SDE	Starch debranching enzymes
SDS	sodium dodecyl sulphate
SGP	Starch granule proteins
SSS	soluble starch synthases
TE	Tris-EDTA buffer
UV	ultraviolet

TABLE OF CONTENTS

CHAPTER 4 CLONING AND CHARACTERIZATION OF THE *WAXY* GENE OF *T. monococcum* AND *T. speltoides*

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

Starch consists of two major components, amylose and amylopectin. The appropriate end-use of starch is dependent on an important property of starch, the ratio of amylose:amylopectin. Wheat with reduced amylose concentration of starch is important for food industries such as those involved in manufacture of Japanese Udon noodles and the white salted noodles. Development of amylose-free starch would open up new end-uses of wheat in the food industry and non-food industry. Amylose is synthesized by the waxy protein, a predominant component of proteins embedded within starch granules in endosperm of wheat. Electrophoretic properties of three different waxy proteins in common wheat have been reported extensively. However, the molecular properties and interaction of the different *waxy* genes and their individual contributions to starch quality are largely unknown, due to the complex nature of the genomes of common wheat. One strategy to resolve these problems is to identify, clone and sequence individual *waxy* genes and thus obtain information on the proteins encoded thereby. This project has the objective of cloning and characterising the *waxy* genes in the various diploid progenitors of wheat and to clone the individual *waxy* genes in polyploid wheat. In this Chapter, the biochemical pathway of starch synthesis and the functions of enzymes involved are discussed, with particular emphasis on the waxy proteins and *waxy* genes. The current status of our knowledge of wheat evolution is also reviewed.

1.1 Production of wheat and importance of wheat starch

1.1.1 The demand for wheat starch.

Wheat is the most widely grown and consumed food crop and the main source of starch. The global annual demand for wheat has been forecast to increase from 560 million tonnes in 1998 to 1050 million tonnes in 2020 (Kronstad, 1998). The annual yield has been predicted to increase from the current 2.5 tonnes per hectare to 3.8 tonnes per hectare (Braun *et al.*, 1998). With the annual 1.6% increase in yield required to fulfil the demand for wheat of the world for the next 20 years, there is a challenge to meet the demand for end-use of wheat grain in different catalogues of food industry and non-food industry. End-use of wheat grain is dependent on several major determinants including protein quantity and quality, starch composition and quality, grain hardness, content of lipid and lipoprotein, and pigments of endosperm and pericarp (Morris, 1998). Wheat grain consists of the embryo, endosperm and pericarp (Fig. 1.1). The endosperm is the major part of wheat grain and accounts for >80% of the grain weight. Starch comprising approximately three-fourths of the endosperm plays an important role in determining the end-use of wheat grain.

1.1.2 Uses of starch and starch quality

Nearly 35% of the world's population uses wheat starch as the staple food (Braun *et al.*, 1998). Wheat starch is a major material processed for food such as biscuit, cake and bread, and is also an important raw material for gelling and pasting in the non-food industry (Morell *et al.*, 1995; Ball *et al.*, 1996; Smith *et al.*, 1997; Zeng *et al.*, 1997; Graybosch, 1998). Starch with high amylose concentration is useful in non-food industry; for example, such starch may improve the texture and quality of plastics (Watanabe *et al.*, 1998). Breeding for high-amylose varieties has already been achieved successfully in barley (Schondelmaier *et al.*, 1992) and maize (Ferguson, 1994). Selection has also been carried out using the hybrids of *T. turgidum* and *T. tauschii* to obtain hexaploid common wheat with a high amylose concentration (Watanabe *et al.*, 1998). However, studies on amylose-free starch are discussed here in greater details, due to its high potential application in the food and non-food industries.

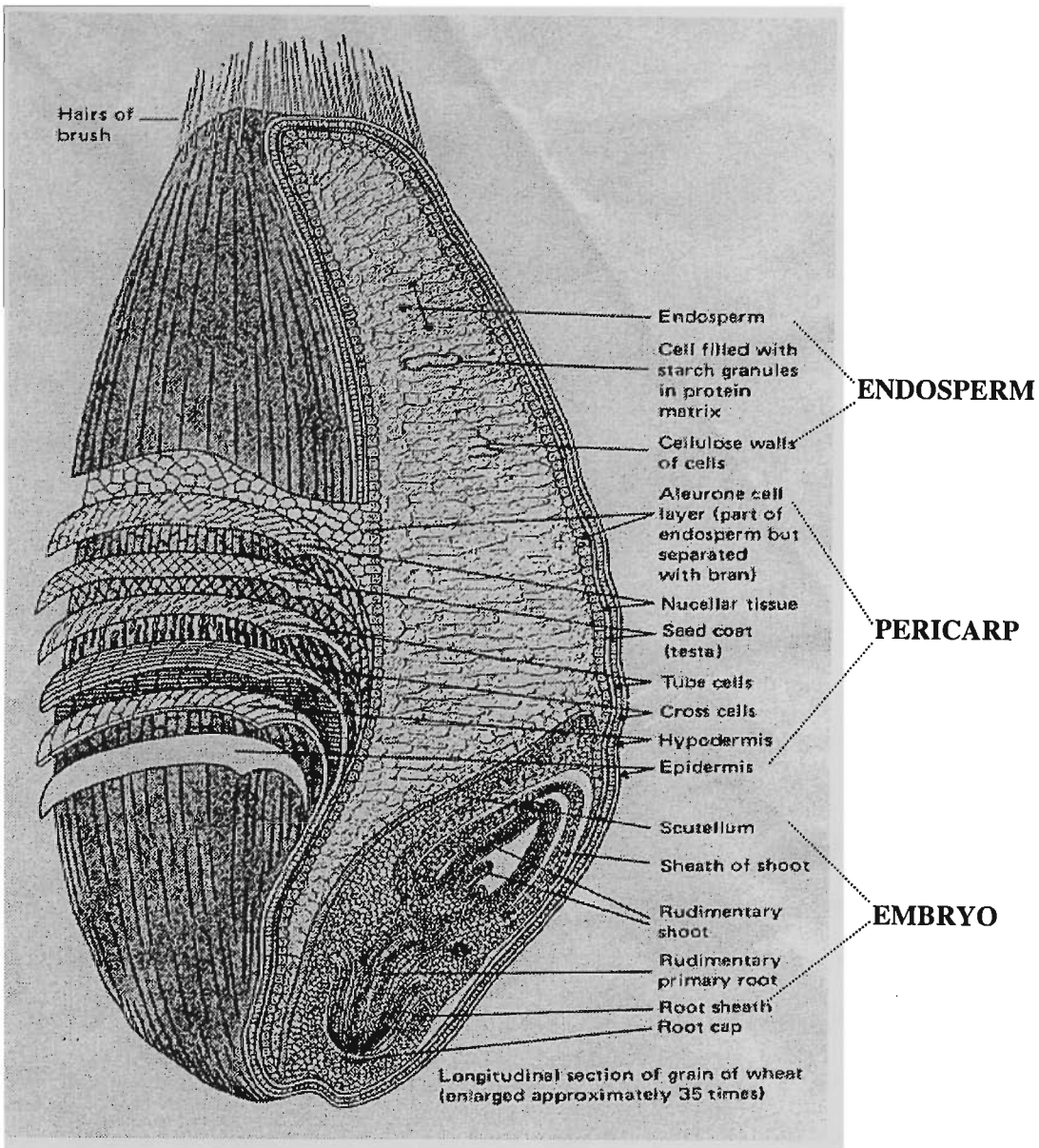


Fig. 1.1 Structure of the wheat grain. Longitudinal section of the wheat grain (Cornell and Hoveling, 1998).

1.2 Chemical properties of starch

1.2.1 Chemical structure and composition of starch

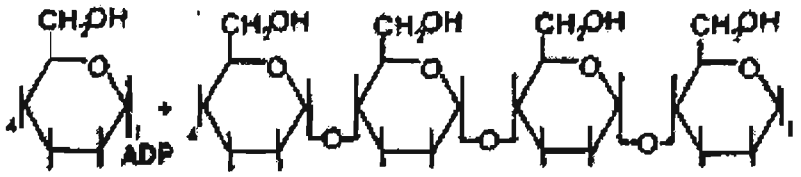
Starch consists of two major components; amylose, the essentially linear chain component of α-(1-4) linked D-glucose residues, and amylopectin, the branched-chain component wherein the chains of α-(1-4) linked D-glucose branch every 20-25 residues through the α-(1-6) linkage (Fig. 1.2). The relative composition of amylose and amylopectin in starch is variable in plants. For example, the amylose concentration of starch is 0-30 % in maize and rice, but is relatively high, 25-30 %, in wheat.

1.2.2 Amylose concentration of starch and its relationship to starch quality

As discussed above, the ratio of amylose:amylopectin is an important criterion in determining the end-use of starch. The amylose-free starch of maize has been noted for its paste clarity, high water binding capacity and resistance to gel formation and retrogradation of paste (Crow and Dove, 1987). It is thus a preferred substrate for food, papermaking and adhesive industries (Zeng *et al.*, 1997; Graybosch, 1998). The amylose-free starch of potatoes also yields a clear paste that does not retrograde and is useful in preparation of foods (Visser *et al.*, 1993). The amylose-free starch of rice is valuable in the production of certain food products such as cakes and porridge (McKenzie and Rutger, 1983; Sano *et al.*, 1986).

In some countries such as Australia, wheat is a major crop due to the climate conditions being better for growing wheat rather than maize. A large quantity of wheat is used currently in industrial applications such as manufacture of gluten, which produces starch as by-product (Graybosch, 1998). Commercial applications of wheat starch could be more extensive if a significant reduction in amylose concentration is achieved. Graybosch (1998) has recently summarised that the utilization of amylose-free starch of wheat could be in three areas: production of modified food starches, extension of the storage of baked goods without a concomitant dilution of gluten, and increased profitability of gluten manufacturers.

A



B

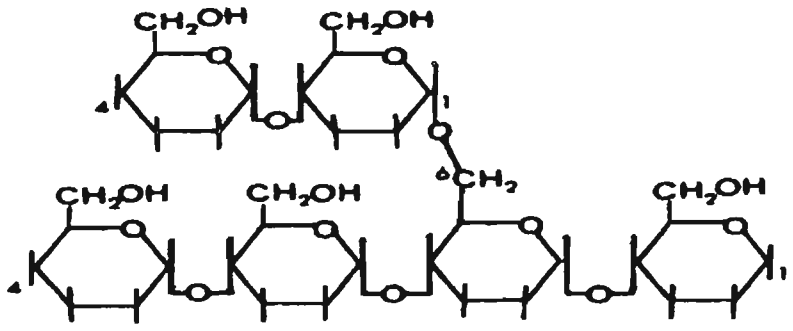


Fig. 1.2 Diagram of the chemical structure of starch. Starch consists of amylose and amylopectin. Amylose is the essentially linear chain component of α-(1-4) linked D-glucose residues (**A**). Amylopectin is the branched-chain component wherein the chains of α-(1-4) linked D-glucose branch every 20-25 residues through the α-(1-6) linkage (**B**) (Martin and Smith 1995).

1.2.3 Reduced-amylose starch of wheat

It has been reported that some wheat varieties produce a relatively low proportion of amylose in starch and this type of wheat flour, produced from the low-amylose wheat, is considered to provide improved physical properties of noodles and better storage and keeping qualities (Nelson, 1987). The major reason why Australian wheat has a great marketability in the food industry in Asian countries is that wheat starch of some Australian varieties contains a slight reduction in the level of amylose concentration. It is, therefore, considered to be of superior quality for the manufacture of Japanese Udon and white salted noodles (Oda *et al.*, 1980; Yamamori *et al.*, 1992; Miura and Tanii, 1994). Due to interests in reduced amylose starches, attempts are underway to identify mutant wheat varieties with reduced-amylose starches. Such mutants and their flour quality will be discussed in section 1.6.4.

1.2.4 Development of amylose-free starch of wheat

The ratio of amylose:amylopectin in starch of common wheat does not show a wide degree of variation, compared to that found in the diploid cereals, maize and rice. Investigations so far have failed to report any naturally occurring amylose-free wheat varieties (Yamamori *et al.*, 1994; Miura *et al.*, 1994; Hoshino *et al.*, 1996; Graybosch, 1996, 1998; Zeng *et al.*, 1997; Demeke *et al.*, 1997; Yasui *et al.*, 1997). Wheat starch with the only slightly reduced amylose concentration cannot meet the requirements of the market for amylose-free starch for certain food and non-food industries. Therefore, the amylose-free starches of diploid maize, rice and potato plants command a larger portion of the world's starch markets. In theory, wheat with amylose-free starch could be utilized in the same way as amylose-free starch of maize, rice and potato, as described previously (section 1.2.2). The lack of significant natural variation in the amylose concentration of wheat starch could be related to the complex genomic makeup of wheat; it is a hexaploid, thus the expression of genes on one genome may be masked by that on another genome (Ainsworth *et al.*, 1993). More and more attention is being paid to the production of amylose-free starch of wheat using biotechnology or by crossing of different null lines with amylose-reduced starch in breeding programs (section 1.6.6.2).

1.3 Biochemical pathway of amylose and amylopectin synthesis in cereals

Starch synthesis in plants is confined to chloroplast in green photosynthetic tissues and amyloplasts in non-green storage tissues such as pollen and endosperm tissue of developing grains. Intricate details for the biochemical pathways of starch synthesis in cereals will not be presented here, since these have been reviewed in recent years (Keeling *et al.*, 1988; Tetlow *et al.*, 1994; Morell *et al.*, 1995; Martin and Smith 1995; Ball *et al.*, 1996; Koβmann 1997; Wang *et al.*, 1998). Based on this literature, the process of starch biosynthesis can be summarized as follows (Fig. 1.3):

1. Hexose phosphates such as glucose-1-phosphate, glucose-6-phosphate, or fructose-6-phosphate enter into the amyloplast from cytosol;
2. Glucose-1-phosphate is converted into ADP-glucose, the substrate for starch synthesis, by the enzyme ADP-glucose pyrophosphorylase (ADPase);
3. ADP-glucose is converted to an α-(1-4) linked D-glucopyronosyl residues to form amylose, by starch synthases (SS);
4. ADP-glucose is also converted to α-(1-6) linked D-glucopyronosyl residues to form amylopectin, by starch synthases and starch branching enzymes (SBE).

Amylose and amylopectin are synthesized in two independent pathways; however, these two polymers can inter-convert due to actions of the enzymes such as SBE and starch debranching enzymes (SDE) (Koβmann 1997).

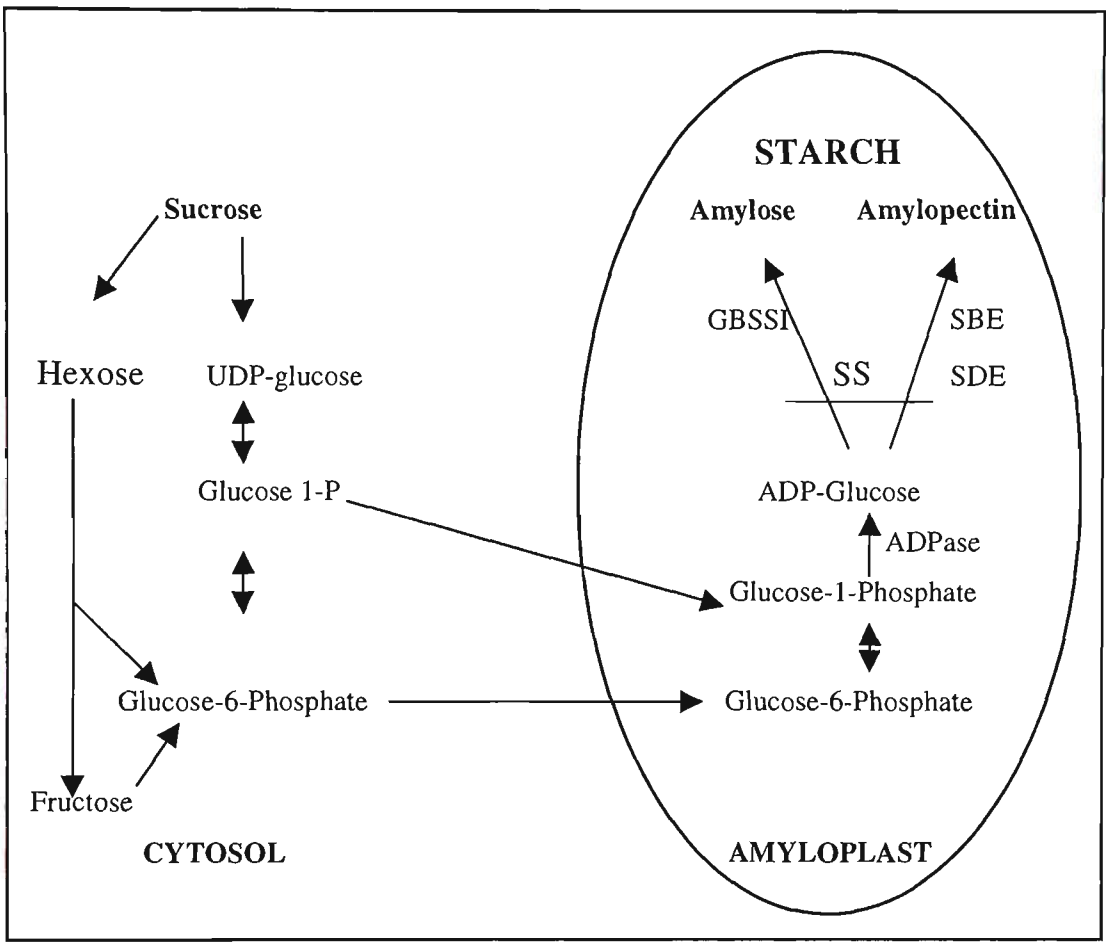


Fig 1.3 Pathway of starch biosynthesis in amyloplast of cereals. Hexose phosphates enter into amyloplast from cytosol and are converted to ADP-glucose. ADP-glucose is the substrate of starch synthesis and is polymerized into amylose and amylopectin by two independent pathways (Wang *et al.*, 1998). GBSSI: granule-bound starch synthases; SS: starch synthases; SBE: starch branching enzymes; SDE: starch debranching enzymes.

1.4 Functions of enzymes involved in synthesis of amylose and amylopectin from ADP-glucose

The enzymes in the starch biosynthetic pathway have been characterised based on their roles in the formation of various substrates and in the amylose and amylopectin biosyntheses. Each enzyme involved in amylose and amylopectin synthesis is also related to properties of starch structures. A model microbial system in *Chlamydomonas* has been constructed to help understand the biogenesis of the plant starch granule. The enzymes involved in the synthesis of an ordered crystalline structure of starch have been described (Ball *et al.*, 1996; Buleon *et al.*, 1997). However, the functions of enzymes in determining the structures of starch are not known. The functions of the three key enzymes, SS, SBE and SDE, which are directly involved in the polymerisation of ADP-glucose to amylose and amylopectin, are discussed below.

1.4.1 Starch synthase (SS)

Starch synthases (SS, EC 2.4.1.21) catalyse the conversion of ADP-glucose to amylose and amylopectin by the two independent pathways (Fig. 1.3). On the basis of the solubility properties of enzymes when extracted from starch granules, starch synthases are divided into two classes: soluble starch synthases (SSS) which can be separated partly from the starch granules and granule-bound starch synthases (GBSS) which are exclusively granule-bound.

1.4.1.1 Soluble starch synthases (SSS)

Preiss (1991) described two classes of soluble starch synthases (SSS), defined as a type I and type II (in the order of their elution from DEAE-cellulose). These SSS have been purified from rice (55 kDa and 57 kDa, Baba *et al.*, 1993), pea (77 kDa and 60 kDa, Denyer and Smith, 1992), maize (76 kDa, Mu *et al.*, 1994), and potato (70 kDa, 85 kDa, 135 kDa, Koβmann, 1997). All starches in plants probably contain at least one isoform of soluble starch synthases (Smith *et al.*, 1995). A number of genes for soluble starch synthases have been characterized recently, and they include SSSI in rice (Baba *et al.*, 1993); SSSI, SSSII and SSSIII in potato (Koβmann, 1997); SSSIa

and SSSIb in maize (Mu *et al.*, 1994); and SSSI and SSSII in pea (Dry *et al.*, 1992). Three types of SSS cDNA have been cloned recently from common wheat (Gao and Chibbar, 1998); the details of the work are discussed in section 1.5.2. A SSS genomic gene has been cloned and characterized (Li *et al.*, 1999).

The enzymatic activity of SSS can be reduced either by mutation in SSS genes, e.g., in barley or pea (Tynela and Schulman, 1993), or by an increase in growth temperature during endosperm development of wheat that affects the enzyme stability of SSS (Jenner *et al.*, 1995). However, the precise effect of reduction in SSS activity on starch synthesis is unknown. Recent studies using transgenic potatoes (Koβmann, 1997) have shown that specific inhibition of SSSI gene, encoding the 70 kDa SSS, did not change any significant structural properties of starch. The repression of the SSSII gene, encoding the 85 kDa SSS, led only to slight changes in the starch structure. However, repression of the SSSIII gene, encoding the 135 kDa SSS, reduced the amylose and phosphate contents and altered the gelation characteristics of the starch. These observations indicate that the various SSS enzymes might have different roles in starch synthesis.

1.4.1.2 Granule-bound starch synthase (GBSSI, waxy protein)

The first report of an unusual phenotype of maize, where the kernel had the soft wax-like sheen, occurred in the early of this century; this type of maize was thus described to possess the *waxy* locus (Nelson, 1987). In 1923, the starch in waxy endosperm of maize was reported to produce a reddish-brown colour instead of the blue-black colour produced by ordinary starch, when reacted with an I/KI stain (Nelson, 1987). Nelson and Rhines (1962) discovered that mutants of the maize *waxy* locus lacked an approximately 59 kDa protein in the endosperm. Since then, the 59 kDa protein in endosperm has been called waxy protein and plants which are free of waxy proteins have been called waxy plants. The hexaploid common wheat that lacks one or two of the three waxy proteins is called partial waxy wheat and is discussed further in section 1.6.3. Because the waxy protein is the major component of the proteins bound within the starch granules when these are extracted from endosperm of cereals, it is called the granule-bound starch synthase (GBSSI) (Preiss and Levi, 1980). The *waxy* locus of cereals is designed as the *amf* locus of potato and *lam* locus of pea (Smith *et al.*, 1997).

Sprague was the first to report in 1943 that the starch in the haploid pollen and the triploid endosperm tissues of waxy maize lacked amylose and was composed entirely of amylopectin (Nelson, 1987, and references within). Since then, the role of the waxy proteins has been confirmed through a number of experiments; the waxy protein is thought to play a major role in the regulation of amylose concentration. In maize, the waxy protein activity, as well as amylose concentration, were found to decrease with the addition of inactive or mutant *waxy* genes, in the series WX/WX/WX, wx/WX/WX, wx/wx/WX and wx/wx/wx in the triploid endosperm tissues of maize (Tsai, 1974). The starch of the *waxy* mutant of maize was reported to be amylose free (Shure *et al.*, 1983; MacDonald and Preiss 1985). The waxy protein activity and amylose concentration were reported to have an apparently positive relationship in different non-*waxy* mutant cultivars, i.e., the waxy protein activity decreased with increase in the dosage of mutant *waxy* genes in the endosperm of rice, as in maize (Sano *et al.*, 1986). Shimada *et al.* (1993) reported that the introduction and expression of the antisense RNA of the *waxy* gene into rice affected the level of expression of the target gene and reduced the amylose concentration in endosperm starch. In potato, Visser *et al.* (1991) and Koβmann (1997) reported that the expression of antisense RNA of the *waxy* gene inhibited the waxy protein activity in starch of potato tuber and gave rise to tubers containing amylose-free starch.

1.4.2 Starch branching enzymes (SBE)

Starch branching enzymes (EC 2.4.1.18) introduce the branch points, α-1,4-glucans (Fig. 1.1), into the amylopectin and thus have been shown to play an important role in amylopectin synthesis. Two forms of branching enzymes have been found in most of the higher plants studied to date. The nomenclature of these forms, SBEI and SBEII, is based on their order of elution from maize endosperm during the extraction of enzymes (Morell *et al.*, 1995). The genes or cDNA for SBE have been isolated from maize SBEI (Baba *et al.*, 1991) and SBEII (Fisher *et al.*, 1993), rice SBEI (Kawasaki *et al.*, 1993) and SBEIII (Mizuno *et al.*, 1993), pea SBEI and SBEII (Smith 1988; Burton *et al.* 1995), potato SBE (Poulsen and Keiberg, 1993). The cDNA for wheat SBEI (Repellin *et al.*, 1997) and SBEII (Nair *et al.*, 1997) and a genomic gene for wheat

SBEI (Rahman *et al.*, 1997) have been cloned recently; these clones are discussed further in section 1.5.3.

The roles of SBE in pea have been revealed by investigations on mutants of the SBE locus which exhibited significantly lower amylopectin concentration relative to the wild type (Martin and Smith, 1995; Wang *et al.*, 1998). The assignment of roles of SBE in other plants is based mainly on the high homology of the deduced amino acid sequences of these SBEs to the *E. coli* branching enzyme (1,4-glucoan-6-glycosyltransferase, *glg*B). However, not all SBEs appear to have the same roles or make equal contributions to amylopectin synthesis. In transgenic potato plants created by introduction of the antisense RNA of the SBE gene, the repression of the native SBE gene altered the structure of some soluble glucans but, surprisingly, did not result in any increase of amylose concentration of the starch granules (Koβmann, 1997). These observations suggest that additional SBE genes might exist which could compensate for the repressed SBE gene. In maize, it has been revealed that SBEII preferentially branched amylose and SBEI preferentially branched amylopectin (Smith *et al.*, 1997). It is unknown whether variations in functions of different SBEs exist also in other plants. It should be noted that, based on the analyses of homology of their deduced amino acid sequences, the homologues of maize SBEI are pea SBEII, rice SBEI and potato SBEI and that the homologues of maize SBEII are pea SBEI, rice SBEIII and potato SBEI. (Morell *et al.*, 1995; Smith *et al.*, 1997).

1.4.3 Starch debranching enzymes (SDE)

Starch debranching enzymes (EC 2.4.1.18) play a role in the determination of structure and composition of amylopectin and thus the end-uses of starch (Ball *et al.*, 1996; Repellin *et al.*, 1998). The enzymatic properties of SDE have been described (Pan and Nelson, 1984; Doehlert and Knutson, 1991; Nakamura *et al.*, 1996a and 1996b; Smith *et al.*, 1997). Two types of enzymes, isoamylase in maize (EC 3.2.1.68) and limit dextrinase or pullulanase or R-enzyme in rice (EC 3.2.1.41), have been found to have starch debranching activity. However, relative to SS and SBE, information on the biochemical properties and roles of SDE in starch biosynthesis is rather limited. Approximately twenty SDE genes or their cDNA have been registered in GenBank

(updated 16/4/1999). These include two cDNA for starch debranching enzymes (or R-enzyme) in rice, cDNAs for pullullanse-type starch debranching enzyme (*zpu*1) and SU1 isoamylose (*sugaryl*) in maize, and one SDE genomic gene for limit dextrinase (HvLD99) in barley. A partial cDNA of common wheat has been cloned too (Repellin *et al.*, 1998). However, direct evidence for the precise roles of the individual SDE genes in plants is still lacking.

1.4.4 Interactions of various enzymes

Although the unique role of GBSSI in amylose synthesis has been described by a number of groups (see section 1.4.1.2), GBSSI has also been found to be involved in amylopectin synthesis *in vivo*. Evidence for this comes from two reports. Firstly, mutants of cereals, potatoes and *Chlamydomonos* that lacked the *waxy* gene differed in amylopectin concentration from their wide-type lines (Delrue *et al.*, 1992; Doehlert *et al.*, 1993; Maddelein *et al.*, 1994; Takeda *et al.*, 1997). Secondly, the waxy protein was found to elongate chains within amylopectin in isolated starch granules (Baba *et al.*, 1987; Denyer *et al.*, 1996). Our analyses indicate that the homology of amino acid sequences is high not only between different GBSSI but also between GBSSI and SSS. For example, there is 85% identity between GBSSI of wheat and rice, but GBSSI in wheat exhibits approximately 60% identity to SSSI in rice as well (data not shown). It thus seems likely that the GBSSI protein also has a function that is similar to that of SSS.

Starch synthesis is a complex biochemical pathway involving many enzymes and the role of each individual enzyme has still not been determined unequivocally. The enzymes in starch biosynthetic pathways in plants are related to those involved in glycogen synthesis in cyanobacteria and *E. coli*, wherein the functions of each enzyme are easier to investigate by the creation of a number of mutants. In summary, there are four enzymes in *E. coli* glycogen synthesis, glycogen synthase (*glg*A), 1,4-glucoan-6-glycosyltransferase (*glg*B, branching enzymes), ADP-glucose synthetase (*glg*C) and glycogen debranching enzyme (*glg*X). The functions of these can be analyzed individually by creation of specific mutants; this information can be then utilized to identify and predict the roles of enzymes involved in starch biosynthesis in plants.

1.5 Enzymes embedded within the starch granules in endosperm of wheat

Starch granules isolated from the developing or mature endosperm of wheat contain two types of proteins, those that are embedded within the starch granule and those that have come in contact with the starch granule and are exclusively associated with its surface (Schofield and Greenwell, 1987; Rahman *et al.*, 1995). Only the proteins embedded within the starch granules in wheat have been studied extensively, as these enzymes are believed to have important functions in starch synthesis due to their direct association with starch granule (Rahman *et al.*, 1995). Electrophoretic analyses of the enzymes embedded within the granules indicate that these enzymes fall into four major classes based on their molecular weights. These include a 59 kDa granule-bound starch synthase (GBSSI), a 75 kDa soluble starch synthase (SSS), a 85 kDa starch branching enzyme (SBE) and a 100-105 kDa starch granule protein (SGP) (Rahman *et al.*, 1995; Denyer *et al.*, 1995; Takaoka *et al.*, 1997; Yamamori *et al.*, 1998) (Fig. 1.4). There are no reports that starch debranching enzymes (SDE) are bound within the starch granules during the endosperm development. The enzymes GBSSI, SSS, SBE and SGP embedded within the starch granules are reviewed in this Chapter.

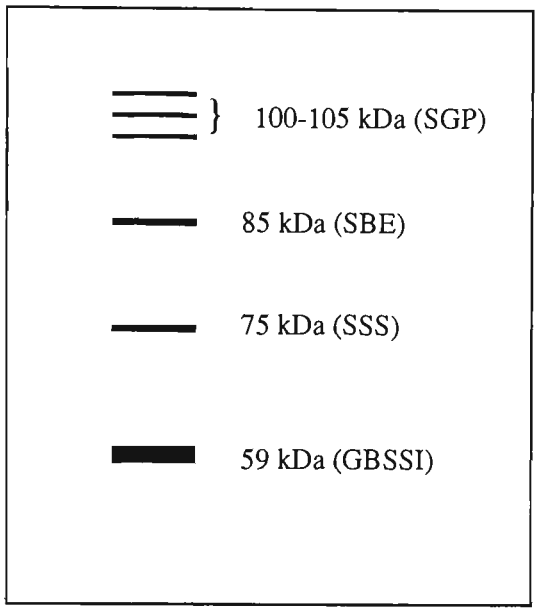


Fig. 1.4 A diagrammatic representation of electrophoretic profile of the enzymes embedded within the starch granules in developing endosperm of wheat on the SDS-PAGE. The 59 kDa, 75 kDa, 85 kDa and 100-105 kDa proteins represent the granule-bound starch synthases (GBSSI), soluble starch synthases (SSS), starch branching enzymes (SBE) and starch granule proteins (SGP), respectively. The four classes of proteins are shown to differ in band thickness, to indicate their unequal amounts.

1.5.1 The 59 kDa proteins: GBSSI, granule-bound starch synthase

The 59 kDa protein, GBSSI, is the predominant protein embedded within starch granules (Fig. 1.4), is exclusively granule-bound in endosperm of wheat and is thus designated as the granule-bound starch synthase (GBSSI) (Fig. 1.5). Proteins of three other size classes, 75 kDa, 85 kDa and 100-105 kDa, are also present in the granule-bound fraction (Schofield and Greenwell, 1987; Mu *et al.*, 1994; Rahman *et al.*, 1995; Denyer *et al.*, 1993, 1995) (Fig. 1.5) and discussed in section 1.5.2-1.5.4. However, these three size classes of enzymes exhibit significantly lower amounts than GBSSI when determined using SDS-PAGE (Fig. 1.4). On the basis of the results described using Western blotting experiments (Taira *et al.*, 1995; Rahman *et al.*, 1995) and the sequences of amino acids at the N-terminus (Denyer *et al.*, 1995; Taira *et al.*, 1995; Rahman *et al.*, 1995; Fujita *et al.*, 1996), the 59 kDa protein has been confirmed to be starch synthase in different types of wheat. The role of the 59 kDa GBSSI in wheat is similar to that of the waxy protein in other plants, as discussed in section 1.4.1.2. Further details of properties of the three different GBSSI or waxy proteins in common wheat are presented in section 1.6.

A cDNA clone for GBSSI of common wheat was reported by Clark *et al.* (1991). The amino acid sequence deduced from this cDNA had high homology with that of the waxy proteins in other plants (Ainsworth *et al.*, 1993). Three other cDNA clones for GBSSI in common wheat have been isolated recently (sequence unpublished, Matus *et al.*, 1998). These three cDNAs were suggested to be originated from the A, B or D genomes, based on comparison of the waxy proteins deduced from these cDNAs with the N-terminal amino acid sequences reported by Nakamura *et al.* (1995b). However, the sequence and organization of genomic copies of *waxy* genes in wheat are unknown. These have been characterized in this project.

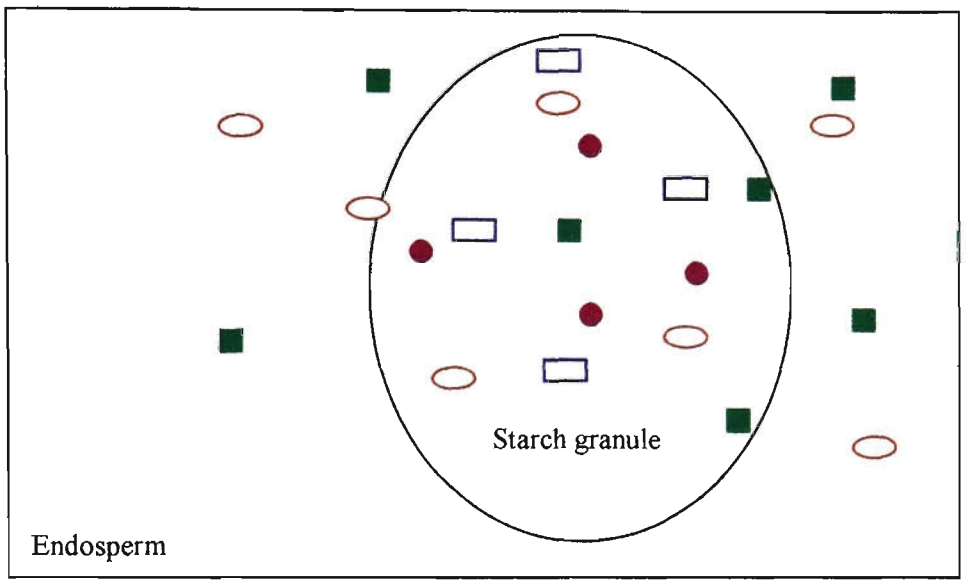


Fig. 1.5 Diagrammatic representation of the locations of enzymes involved in starch synthesis in wheat. GBSSI ( ), SSS ( ), SBE ( ) and SGP ( ) represent the 59 kDa, 75 kDa, 85 kDa and 100-105 kDa proteins embedded within starch granules. SSS and SBE are present in both granule-bound and soluble fractions; GBSSI and SGP are present only within the granules.

1.5.2 The 75 kDa proteins: SSS, soluble starch synthase

The 75 kDa protein was found to be a starch synthase and was antigenically related to the 77 kDa starch synthase of the pea embryo (Denyer *et al.*, 1995). It is called the soluble starch synthase (SSS) as it can be separated from the starch granules (Denyer *et al*, 1995; Rahman *et al.*, 1995, Yamamori *et al.*, 1998). The 75 kDa protein has been detected in both soluble and granule-bound fractions in endosperm of wheat (Fig. 1.5), however, it makes a minor contribution to the starch synthase activity in the soluble fraction (Denyer *et al.*, 1995). In order to investigate the role of this SSS in starch synthesis, three cDNA clones for SSS in wheat have been isolated (Gao and Chibbar, 1998). These cDNA sequences exhibited 55-70% similarity to maize SSSIa and SSSIb, and thus were considered to be counterparts of maize SSS. Two of these three different clones were deduced to encode 88 kDa and 80 kDa polypeptides (it was not reported whether these polypeptides contain the transit peptides of SSS or not). Studies on functions and identification of the phylogenetic relationship of these cDNAs are being carried out (Gao and Chibbar, 1998). A genomic clone for SSS of *T. tauschii* has been isolated and work on transgenic plants using the promoter of this gene to drive the SSS cDNA is underway (Li *et al.*, 1999).

1.5.3 The 85 kDa protein: SBE, starch branching enzymes

The 85 kDa protein was found to be the starch branching enzyme (SBE), based on the immunoprecipitation experiments and amino acid sequences at the N-terminal, and it has been detected in both soluble and granule-bound fractions (Rahman *et al.*, 1995, Yamamori *et al.*, 1998). In wheat, two full-length cDNAs encoding SBEI (85.4 kDa) and SBEII (85.5 kDa) in the endosperm of common wheat have been cloned (Repellin *et al.*, 1997; Nair *et al.*, 1997). The genomic gene for SBEI has been isolated from *T. tauschii* (Rahman *et al.*, 1997) and common wheat (Baga *et al.*, 1998). Two separate promoter regions have been identified for directing expression of the SBEI gene (Baga *et al.*, 1998). The transgenic wheat with the SBEI gene has been shown to result in a reduction in starch branching enzyme activity level and the modified starch has a reduced crystalline structure and a lower temperature demand for starch gelatinisation has been observed (Chibbar *et al.*, 1998).

1.5.4 The 100-105 kDa proteins: SGP, starch granule proteins

In the higher molecular weight region of proteins extracted from starch granules in endosperm of common wheat, only one band was shown in the papers by Yamamori *et al.* (1992), Nakamaru *et al.* (1993) and Ainsworth *et al.* (1993); these authors investigated only GBSS proteins. Two bands, 100 kDa and 105 kDa, were then reported by Rahman *et al.* (1995), followed by the report that the isoforms of 100-105 kDa comprised three bands (Denyer *et al.*, 1995). Yamamori *et al.* (1998) reported recently that the sizes of the proteins in this region are 100 kDa, 108 kDa and 115 kDa. This group of proteins is also exclusively bound to the starch granule (Rahman *et al.*, 1995; Denyer *et al.*, 1995). They are referred to as starch granule proteins (SGP) here, in order to differentiate these from the three other groups of proteins embedded within the starch granules, GBSS, SSS and SBE, discussed above.

The three SGP polypeptides have similar N-terminal amino acid sequences (Rahman *et al.*, 1995; Denyer *et al.*, 1995). The genes encoding these three proteins are homoeologous and have been located on chromosomes 7A, 7B and 7D of common wheat, by analysis of variation in these three proteins in the nullisomic-tetrasomic lines (Denyer *et al.*, 1995). Although these SGP are suggested to have starch synthase activity (Rahman *et al.*, 1995; Denyer *et al.*, 1995; Yamamori *et al.*, 1998) based on hybridization with other starch synthases, their precise roles in starch biosynthesis are unclear (Denyer *et al.*, 1995; Rahman *et al.*, 1995). Among a range of cereals analysed, these proteins have been found only in wheat (Rahman *et al.*, 1995). Mutants of all three loci for these SGP have been reported recently in common wheat (Yamamori, 1998). Understanding the biochemical functions of the SGP in wheat would be dependent on further analysis of such mutants and cloning and characterization of the genes encoding these proteins.

1.5.5 Comparisons of properties of different enzymes in starch synthesis

The above overview indicates that four major types of proteins occur within starch granules in endosperm of wheat (Fig. 1.5). Except for the 85 kDa protein which is the starch branching enzyme, the three other groups of proteins, 59 kDa, 75 kDa and 100-105 kDa, have starch synthase activity (Table 1.1). However, the role of only the 59 kDa waxy protein has been defined clearly, by comparison with its counterparts in other cereal plants, as discussed in section 1.4.1.2. It is uncertain whether the 75 kD SSS and 100-105 kDa SGP have a role in determining the composition or structures of amylose or amylopectin or both. Complete understanding of the starch biosynthesis process will depend on isolation and characterisation of all the relevant enzymes and their genes.

1.5.6 Genes encoding proteins embedded within starch granules

Search of Genbank for expressed sequence tags (ESTs) indicates 1256 entries (as at 21/4/1998); they include starch synthase, starch branching enzymes, starch debranching enzymes and other enzymes involved in starch synthesis. However, only a limited number of complete genes encoding enzymes involved in starch synthesis in wheat have been cloned (Table 1.2). A great number of DNA and protein sequences for starch synthesis in diploid plants such as *Arabidopsis* and rice have been revealed and would allow rapid progress in cloning and characterisation of individual genes from different genomes of wheat. Therefore, it is anticipated that in the next 5-10 years ESTs for all genes expressed in the developing endosperm of wheat will be available (Clarke *et al.*, 1998).

Table 1.1 Summary of properties of enzymes embedded within starch granules of wheat

Protein size	Name	Location	Activity	Function
59 kDa	GBSSI	Granule-bound fraction	Starch synthases	Amylose synthesis
75 kDa	SSS	Soluble and granule-bound fraction	Starch synthases	(?)
85 kDa	SBE	Soluble and granule-bound fractions	Starch branching Enzymes	Amylopectin synthasis
100-105 kDa	SGP	Granule-bound fraction	Starch synthases	(?)

The known four classes of proteins embedded within starch granules are GBSSI (granule-bound starch synthase), SSS (soluble starch synthses), SBE (starch branching enzymes) and SGP (starch granule proteins). (?) indicates that the precise functions of SSS and SGP are not known.

Table 1.2 Inventory of clones of genes encoding enzymes embedded within starch granules in endosperm of wheat

Starch synthetic/degrading enzymes (cDNAs and genomic clones)	Accession No	Sources
GBSSI clones		
1. cDNA from *T. aestivum*	X57233	Clark *et al.* (1991)
2. genomic clone from *T. monococcum*	AF110973	Yan *et al.* (this project)
3. genomic clone from *T. speltoides*	AF110974	Yan *et al.* (this project)
4. genomic clone from *T. tauschii*	AF110975	R. Appels (pers. commun.)
5. cDNA (GBSSI-DI) from *T. aestivum*	N/A	Matus *et al.* (1998)
SSS clones		
1. SA10 cDNA from *T. aestivum*	N/A	Gao and Chibbar (1998)
2. A3a I cDNA from *T. aestivum*	N/A	Gao and Chibbar (1998)
3. A3X I cDNA from *T. aestivum*	N/A	Gao and Chibbar (1998)
4. SSS genomic clone from *T. tauschii*	N/A	Li *et al.* (1999)
SBE clones		
1. wSBEI-D2 genomic clone from *T. tauschii*	AF00282	Rahman *et al.* (1997)
2. SBEI from *T. aestivum*	Y12320	Repellin *et al.* (1997)
3. SBEII cDNA from *T. aestivum*	Y11282	Nair *et al.* (1997)
4. SBEII genomic clone	N/A	Rahman *et al.* (1998)
SGP clones		
Two cDNAs from *T. aestivum*	N/A	M. Morell (pers. commun.)
SDE clones		
Partial cDNA from *T. aestivum*	N/A	Repellin *et al.* (1998)

N/A: not available, indicates that the sequence data have not been entered into GenBank or a published journal.

1.6 Waxy proteins of common wheat

1.6.1 Identification of three waxy proteins

The three waxy proteins of common wheat, WX-A1, WX-B1, and WX-D1, were first reported to exhibit slightly different molecular weights and isoelectric points by the use of a modified ratio of acrylamide:bisacrylamide (30:0.135) in two-dimensional SDS-PAGE analysis (Nakamura *et al.*, 1993a). The WX-A protein was reported to have a slightly higher molecular weight and a more basic isoelectric point (pI) than the WX-B1 and WX-D1 proteins, while WX-B1 protein was shown to have the same molecular weight but a slightly acidic pI relative to WX-D1 protein. The three waxy proteins have been distinguished also by a modified pH system in a one-dimensional SDS-PAGE (Zhao and Sharp, 1996). Such methods that can distinguish between the three waxy proteins of common wheat are useful for study of genetics and expression of the triplicated set of genes in the allohexaploid common wheat (discussed in following sections).

1.6.2 The three *waxy* loci in common wheat

In common wheat (*T. aestivum*, AABBDD), structural genes are found as a triplicate set on homoeologous chromosomes, due to its being a hexaploid. For instance, the genes which control high molecular weight (HMW) subunits of the storage protein glutenin are located on the long arms of chromosomes 1A, 1B, and 1D (Payne *et al.*, 1983) and those encoding grain esterase (*Est*-5) are on the long arms of chromosomes 3A, 3B and 3D (Ainsworth *et al.*, 1984). The nullisomic-tetrasomic lines of common wheat var. Chinese Spring have been utilised extensively to identify locations of the triplicate set of genes. In such lines, one specific homologous pair of chromosomes has been deleted and is replaced by additional copies of a homologous pair from one of the two remaining genomes. The three *waxy* loci have been shown to be located on different chromosomes of common wheat by the use this plant material in the following experiments.

The analysis of waxy proteins in nullisomic-tetrasomic lines of common wheat var. Chinese Spring have shown that lines with deletions of chromosomes 7A, 4A or 7D

contain only two waxy proteins, while lines with other deletions contain all three waxy proteins (Nakamura *et al.*, 1993a). Therefore, chromosomes 7A, 4A and 7D of common wheat each can be assumed to contain one *waxy* locus on chromosomes 7A, 4A and 7D respectively (Fig. 1.6). The analyses of isoenzyme markers and restriction fragment length polymorphism (RFLP) have identified the three *waxy* loci to be located on the short arm of chromosome 7A (7AS), the long arm of chromosome 4A (4AL) and the short arm of chromosome 7D (7DS) (Chao *et al.*, 1989). Moreover, Chao *et al.* (1989) showed that part of the short arm of the "original" chromosome 7B, which should contain a *waxy* locus homoeologous to that of 7AS and 7DS, had been translocated to the long arm of 4A (Fig. 1.6). Ainsworth *et al.* (1993) reported similar results for locations of the *waxy* loci in common wheat by analyses of Southern blots of nullisomic-tetrasomic lines. Therefore, the three *waxy* loci of common wheat are designated as the *wx-A1* (*wx-A* or *wx-7A1*), *wx-B1* (*wx-B* or *wx-4A1*) and *wx-D1* (*wx-D or wx-7D1*) genes, respectively (McIntosh et al., 1998; Yamamori et al., 1995). These studies have provided important information for cloning of individual *waxy* genes from the corresponding genomes of different diploid wheat, in the present study.

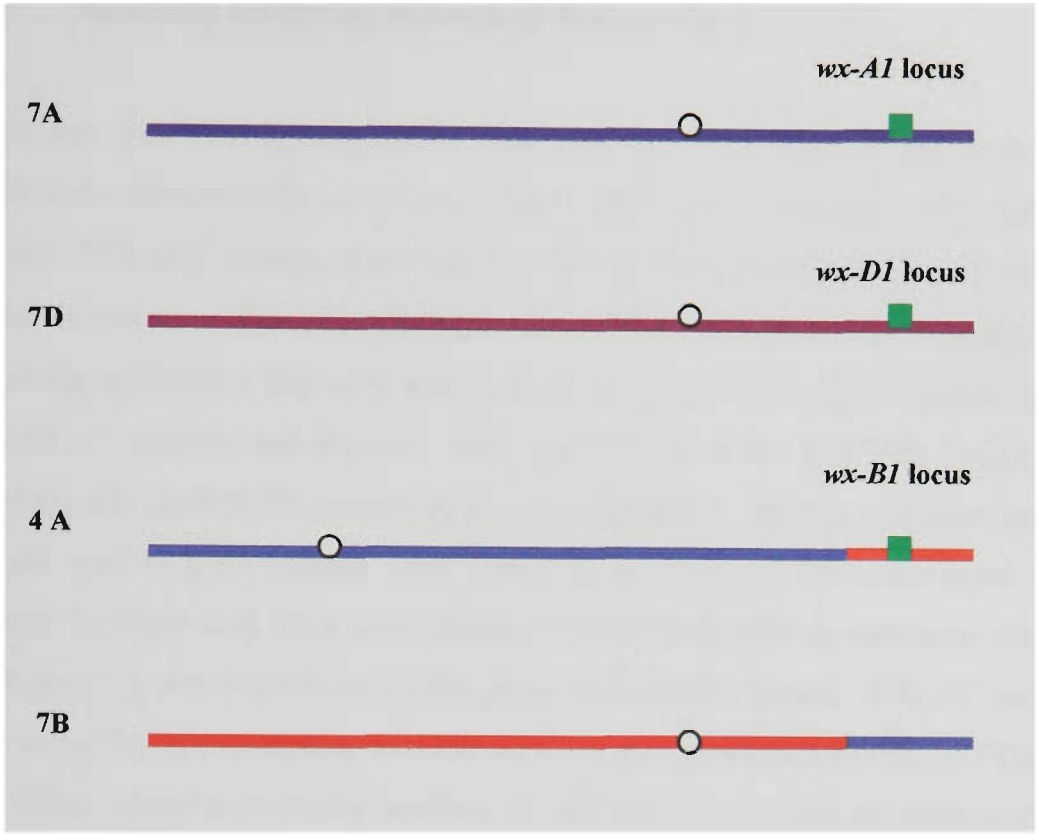


Fig. 1.6 A diagrammatic representation of the locations of the three *waxy* loci of common wheat. The *wx-A1*, *wx-B1* and *wx-D1* loci are located on chromosomes 7AS, 4AL and 7DS. "O" represents the centromere. Part of the "original" chromosome 7BS containing the *wx-B1* locus has been translocated to chromosome 4BL (Naranjo *et al.*, 1987), but 4B was redesignated as 4A (Chao *et al.*, 1989; Dvorak *et al.*, 1990). This translocation occurred in tetraploid wheat (Naranjo *et al.*, 1987; Devos *et al.*, 1995).

1.6.3 Naturally occurring mutants of the *waxy* loci

Yamamori *et al.* (1994) investigated 1560 varieties of common wheat from around the world and reported different types of waxy proteins in null lines. The null lines for WX-B1 (WX-4A) protein were found to be common (about 48%) in Australian and Indian varieties. The null allele for the WX-A1 protein was reported to occur frequently in Korean, Japanese and Turkish wheat but was relatively rare in cultivars from other countries and regions. Only one variety so far, Bai Huo from China, was found to lack the WX-D1 protein in all tested varieties. Double null lines lacking both WX-A1 and WX-B1 protein were found to be rare and no other types of double mutants or triple null lines were found to occur naturally in common wheat. Two corresponding waxy proteins in the tetraploid emmer wheat, WX-A1 and WX-B1, were identified but no double null line of these two proteins were found (Yamamori *et al.*, 1995a). Despite extensive analysis of a number of varieties, no mutant of the waxy protein in diploid *T. monococcum* was found (M. Bhave, 1997, pers. commun.). Extensive efforts have been made to identify mutants of *waxy* genes in Australia (Zhao and Sharp, 1996), US (Graybosch, 1996; Zeng *et al.*, 1997; Graybosch *et al.*, 1998), Canada (Demeke *et al.*, 1997) and Japan (Miura *et al.*, 1994; Hoshino *et al.*, 1996; Yasui *et al.*, 1997). However, the types of null lines found are similar to those reported byYamamori *et al.* (1995).

1.6.4 Effects of waxy proteins on starch quality

As described in section 1.4.1.2, the waxy protein has been shown to regulate the amylose concentration of starch in plants. This theory is also supported by the observations on mutants of various *waxy* loci in wheat. For example, a null allele for the *waxy* gene on chromosome 4A (null *wx-B1*) has been associated with lower amylose concentration (Yamamori *et al.*, 1992; Nakamura *et al.*, 1993b; Miura and Tanii, 1994; Zhao *et al.*, 1996; Zeng *et al.*, 1997). It has been reported that starch from wheat lines containing the null *wx-B* allele confers high peak paste viscosity (Crosbie, 1991; Konik *et al.*, 1992) or high swelling property of the flour (Toyokawa *et al.*, 1989; Crosbie, 1991; McCormic *et al.*, 1991; Crosbie *et al.*, 1992; Konik *et al.*, 1993; Yun *et al.*, 1996). Such wheat starches have been associated with the desirable soft and elastic eating quality of Japanese Udon noodles and high quality Korean white

salted noodles (Yamamori *et al.*, 1992; Miura and Tanii, 1994; Ross *et al.*, 1997; Zhao *et al.*, 1998). However, it has been shown that the effect of the *wx-B1* null mutation was not simply due to a decrease in amylose concentration; instead, it most likely caused a subtle change in the starch structure, resulting in high starch viscosity and high flour swelling volume (FSV) (Zhao *et al.*, 1998).

1.6.5 Contributions of different waxy proteins to amylose concentration

Nakamura *et al.* (1993a) reported from an analysis of nullisomic lines of the experimental variety Chinese Spring that the relative amounts of the three waxy proteins in wheat endosperm are unequal; the amount of the WX-B1 is the largest, followed by the WX-D1 and WX-A1 proteins (WX-B1>WX-D1>WX-A1). Moreover, WX-B1 protein plays a greater role in amylose synthesis than the other two (Yamamori *et al.*, 1994). Studies on the effects of chromosomal locations of different *waxy* genes on amylose concentration in a monosomic series of Chinese Spring have shown that the amount of the WX-B1 protein in the monosomic 4A line is reduced together with a decrease in the amylose concentration. However, the WX-A1 and WX-D1 proteins are decreased in the monosomic 7A and 7D lines respectively, without a concomitant decrease in their amylose concentration (Miura *et al.*, 1994). Miura and Sugawara (1996) measured amylose concentration in nullisomic-tetrasomic lines of the experimental wheat "Chinese Spring". Removal of chromosomes 4A, the home of the *wx-B1* locus, reduced the amylose concentration from 25.5% to 22.5%; when chromosomes carrying either *wx-A1* and *wx-D1* were removed, the amylose concentration declined, but the reduction was only 50% of that observed when *wx-B1* was removed. These observations indicate that the WX-B1 protein makes more contribution to amylose synthesis than the WX-A1 and WX-D1 proteins.

Several points are of particular interest regarding the differential effects of the three *waxy* genes on amylose concentration (Miura *et al.*, 1994). Firstly, only a decrease in dosage of the most potent gene, *wx-B1*, gives a sufficient reduction in the WX-B1 protein to affect the level of amylose. Secondly, a reduction in either of the WX-A1 or WX-D1 proteins appears to be compensated for by the WX-B1 protein. In addition, a regulatory gene on chromosome 7B, separate from the *waxy* genes, possibly suppresses

the activities of the *wx-B1* genes. However, the molecular mechanism of these interactions and compensations between various waxy proteins of wheat are unclear.

1.6.6 Production of waxy wheat

Due to the interactions mentioned above, lack of one or two *waxy* genes is insufficient for creation of wheat varieties with considerably reduced amylose concentration. Therefore, a number of techniques are being investigated in order to produce totally or almost completely waxy wheat.

1.6.6.1 Mutagenic treatments

Mutagenic treatments such as X-rays or ethylmethane-sulfonate (EMS) have been used to produce waxy maize (Briggs *et al.*, 1965), waxy rice (Amano 1981), waxy potato (Hovenkamp-Hermelink *et al.*, 1987), and waxy *T. monococcum* (Kanzaki and Noda, 1988). However, these methods have proven unsuitable for producing waxy tetraploid and hexaploid wheat because polyploid wheat have been shown to have a high resistance to mutagenes, as revealed by a low frequency of mutants (Sears, 1972; Nakamura *et al.*, 1995a).

1.6.6.2 Crossing of the partial waxy mutants

The crossing of different null lines to get triple null mutants can result in the production of completely waxy or amylose-free starch. Nakamura and colleagues (1995a) produced the world's first completely waxy wheat by this method; traditional crossing between the *wx-D1* single null line Bai Huo and the *wx-A1* and *wx-B1* double null line Kanto 107 resulted in progeny that lacked all isoforms of waxy proteins and had no starch amylose. Due to the significance of amylose-free starch discussed previously (section 1.6.4), numerous efforts have been made to develop waxy wheat varieties, in laboratories in USA (Graybosch, 1998), Australia (Zhao and Sharp, 1996), Canada (Demeke *et al.*, 1997) and Japan (Yamamori *et al.*, 1995; Hoshino *et al.*, 1996; Yasui *et al.*, 1997; Kiribuchi-Otobe *et al.*, 1998; Miura *et al.*, 1998). However, only one naturally occurring line of null *wx-D1* gene from China has been discovered so far

and is being applied in over world for crossing purposes; it is thus difficult to produce waxy wheat in different locally suitable genetic backgrounds using this single mutant line in wheat breeding programs. Therefore, two additional sources have been described to solve this problem. Yasui *et al.* (1997) have recovered two waxy wheat lines after treating seed of the double null line Kanto 107 (lacking *wx-A1* and *wx-B1* genes) with EMS to inactivate *wx-D1* gene in this double null line. Kiribuchi-Ocobe (1997) has also identified five waxy wheat lines from a doubled haploid breeding program designed to remove rapidly a low amylose characteristic from a mutant line, in order to adapt these to different genetic backgrounds.

1.6.6.3 Transgenic wheat

The advances in gene manipulation and creation of transgenic plants in the past decade have made it possible to increase yield and create novel agricultural products (Wasserman *et al.*, 1995). One strategy to produce waxy wheat in world-wide varieties would be to develop transgenic lines through antisense RNA technology (Wasserman *et al.*, 1995). These transgenic potato and rice plants, expressing the antisense RNA of the *waxy* gene, have been produced successfully in producing the amylose-reduced or free starch as described in section 1.4.1.2. *Agrobacterium tumefaciens*-mediated transformation has significant advantage over naked DNA delivery techniques such as microprojetcile bombarment. The technique in *Agrobacterium tumefaciens*-mediated transformation in monotyledenous wheat has been difficult, although some success has been achieved (Cheng *et al.*, 1997). Chibbar *et al.* (1998) have successfully obtained the transgenic wheat with the SBE gene (section 1.5.3), by the microprojetcile bombarment techniques. This technique has raised the possibility of generation of transgenic wheat with other genes such as *waxy* gene. It is anticipated that these methods would allow production of waxy wheat in the near future. For this and other genetic engineering approaches, it is a prerequisite to clone the respective *waxy* genes in wheat.

1.7 Taxonomic status and genomes of wheat

1.7.1 Taxonomic background of wheat

Wheat (*Triticum*) is a member of the tribe *Triticeae*, a subdivision of the grass family *Gramineae*. *Triticeae* is divided into two subtribes, *Triticinae* (consisting of about 35 genera, such as *Triticum*, *Aegilops*, *Secale*), and *Hordeinae* (consisting of *Hordeum* and related genera) (Breiman and Graur, 1995). The genus *Triticum* contains wheat species with cultivated forms, such as the hexaploid common wheat, *T. aestivum*, while *Aegilops* (goat grasses) included the wild relatives of common wheat such as *Aegilops speltoides* (Breiman and Graur, 1995). Several proposals have been made to merge the genus *Aegilops* within the genus *Triticum* (Bowden, 1959; Morris and Sears, 1967); this designation has been used by different groups (Dvorak *et al.*, 1993; Appels *et al.*, 1989; Talbert *et al.*, 1995; Friebe and Gill, 1996), though the genus *Aegilops* is retained for use as an independent genus (Kerby and Kuspira, 1987; Breiman and Graur, 1995; Mori *et al.*, 1997). In this project, *Aegilops* is desinated as *Triticum*, i.e., *Ae. bicornis*, *Ae. longissima*, *Ae. searsii*, *Ae. sharonensis* and *Ae. speltoides* are designated as *T. bicornis*, *T. longissima*, *T. searsii*, *T. sharonensis*, and *T. speltoides*, respectively, and *Ae. squarrosa* is designated as *T. tauschii*.

1.7.2 Types of cultivated wheat

On the basis of the morphology and response to fungal pathogens, the cultivated species in *Triticum* has been divided into three groups: einkorn, emmer, and dinkle (Breiman and Graur, 1995). The einkorn is the cultivated diploid wheat, *T. monococcum*. *T. monococcum* is separated into the two subspecies, *T. monococcum* ssp. *boeoticum* and *T. monococcum* ssp. *urartu* (Johnson *et al.*, 1976). The wild equivalent of *T. monococcum* is *T. boeoticum* (Takumi *et al.*, 1993; Heun *et al.*, 1997). The emmer group of wheat is the cultivated tetraploid *T. turgidum* that includes three subspecies, *T. turgidum* ssp. *turgidum*, *T. turgidum* ssp. *dicoccum*, and *T. turgidum* ssp. *durum*, and their wild relative is *T. turgidum* ssp. *dicoccoides* (Kerby and Kuspira, 1987). The dinkle or bread wheat, i.e., common wheat, is hexaploid *T. aestivum*. In addition, the tetraploid *T. timopheevi* and the hexaploid *T. zhukovskyi* are also cultivated in the Caucasus region (Kimber and Sears, 1987).

1.7.3 Genomes of different types of wheat

The genetic sources of different types of wheat have been designated by different genomes, for examples, genomes AABB for *T. turgidum*, genomes AABBDD for *T. aestivum*, genomes AAGG for *T. timopheevi* and genomes AAAAGG for *T. zhukovskyi*. These genomes in polyploid wheat are proposed to have originated from those in diploid wheat. It has been suggested that *T. monococcum* (AA), *T. urartu* (AA and/or BB), *T. tauschii* (DD), and the five diploid species belonging to section *Sitopsis* (genomes BB or SS), *T. bicornis*, *T. longissima*, *T. searsii*, *T. sharonensis*, and *T. speltoides*, are the candidates for being the diploid progenitors of the A, B, D or G genomes of the polyploid wheat (Kerby and Kuspira, 1987, Breiman and Graur, 1995).

1.8 Origins of the genome of common wheat.

1.8.1 Conventional theories for origin of common wheat

Common wheat, *T. aestivum* (AABBDD, 2n=6x=42) is an amphidiploid, in that it is a polyploid species containing three distinct diploid genomes, A, B and D, each with its own independent evolutionary history, but the genetic behaviour of *T. aestivum* resembles that of a diploid. The origins of each genome in this polyploid wheat have been studied extensively during this century. On the bases of geographical distributions, morphological characteristics, meiotic chromosome behaviour of hybrids and amphiploids, the concentration and the restriction patterns of nuclear DNA and the protein constitution, the evolution of common wheat has been reviewed (Kerby and Kuspira, 1987; Kimber and Sears, 1987) and summarised in Fig. 1.7.

1. **The tetraploid progenitor of common wheat was *T. turgidum*.** Common wheat evolved from the hybridisation of a diploid species and tetraploid species. This progenitor contributed A and B genomes to common wheat. Three cultivated types of *T. turgidum* and their wild species mentioned in section 1.7.2 are all possible tetraploid progenitors of common wheat.

2. ***T. tauschii* was the most likely diploid progenitor of D genomes of common wheat.** *T. tauschii* is unequivocally identified to contribute D

genome to common wheat (Kihara, 1944; Mcfadden and Sears, 1946; Kimber, 1974).

3. ***T. monococcum* and/or *T. urartu* might have contributed the A genome to common wheat**. The diploid *T. monococcum* is suggested to be the donor of A genome to common wheat (Sax, 1922, Lilienfeld and Kihara, 1934, Kerber, 1964; Morris and Sears, 1967; Johnson, 1975; Gill and Kimber, 1974; Jaaska, 1980; Jones *et al.*, 1983). However, *T. urartu* might also be the donor of A genome to common wheat, as it contains the same genome as *T. monococcum* (Johnson, 1975; Konarev, 1983).

4. **The origin of B genomes is unclear**. The results on the origin of B genome are controversial from different groups. Five diploid species in the *Sitopsis* section, *T. bicornis*, *T. longissima*, *T. sharonensis* *T. speltoides* and *T. searsii* (Kerby and Kuspira, 1987, and references therein), and another diploid species, *T. urartu* (Johnson, 1975), have been suggested as candidates for being the donors of the B genome to common wheat.

Diploid ancestor

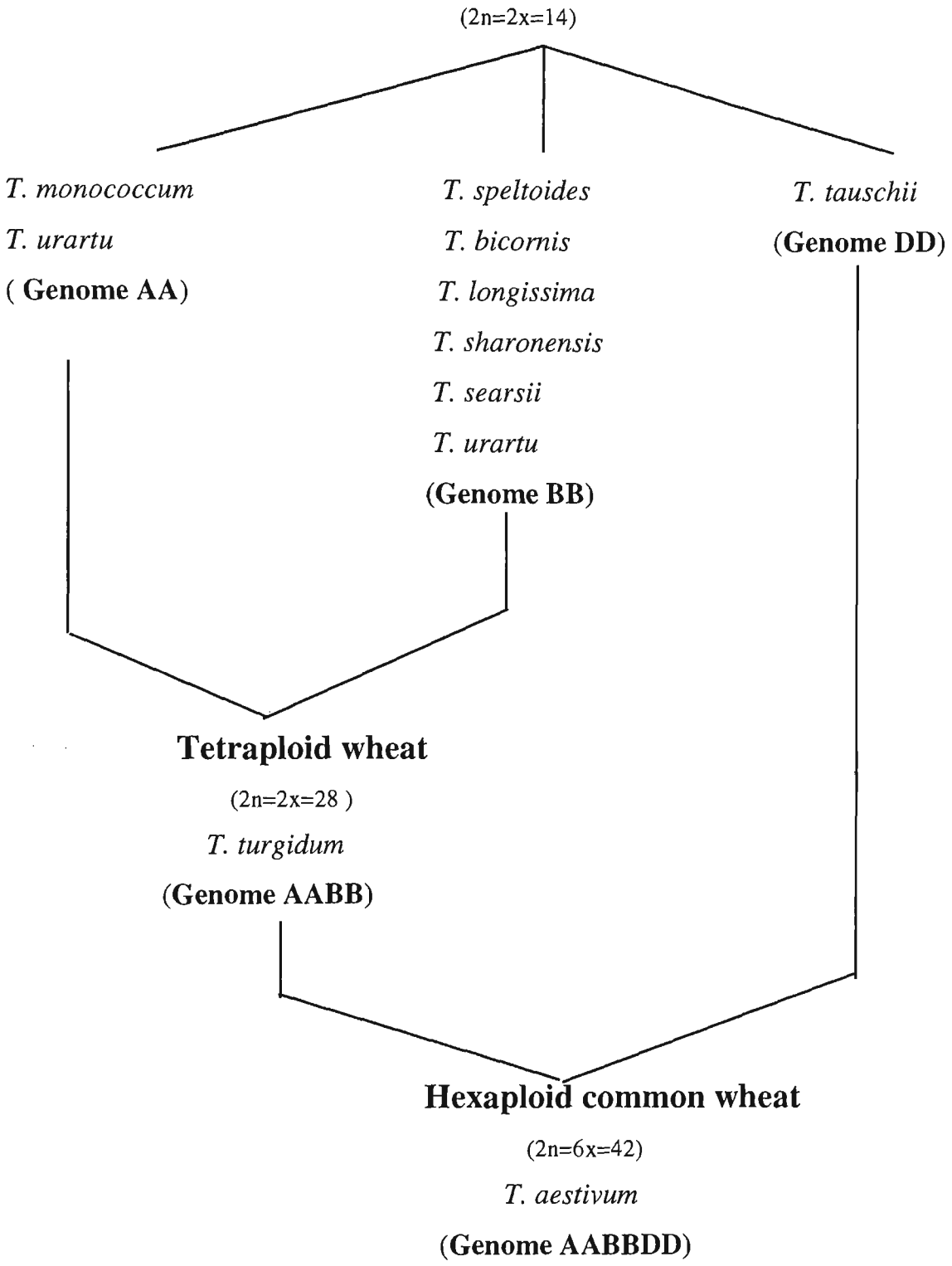


Fig. 1.7 Origins of genomes A, B and D of common wheat. The diagram is based on the review on the evolution of wheat by Kerby and Kuspira (1987).

1.8.2 Current theory for the origins of common wheat

A large number of molecular techniques has been employed recently to obtain more data in order to resolve the ambiguity as to the origins of A and B genomes in common wheat.

1.8.2.1 Origin of D genome

T. tauschii has been identified to contribute the D genome to *T. aestivum* by through recent investigations (Lagudah *et al.*, 1991; Hohmann *et al.*, 1993; Gill *et al.*, 1991; Gill *et al.*, 1993; Friebe and Gill, 1996).

1.8.2.2 Origin of A genome

On the basis of an analysis of the *NOR* locus in different types of wheat, *T. monococcum* was suggested to contain A genome (Gill and Appels, 1988). Kerby *et al.* (1988) reported that both *T. monococcum* and *T. urartu* had a purothionin identical to that from in *T. turgidum* and *T. aesivum*, thus suggesting that these two species possessed the A genome and were possible donors of it to the polyploid wheat. Analysis of restriction fragment length polymorphism (RFLP) of several repeated DNA families and unique sequences identified *T. urartu* as the donor species of the A genome in *T. turgidum*, *T. timopheevi* and *T. aestivum* (Dvorak *et al.*, 1988; Dvorak and Zhang, 1990, Dvorak *et al.*, 1993; Takumi *et al.*, 1994); C-banding analysis also strengthened this suggestion (Jiang and Gill, 1994). Further evidence for *T. urartu* being the donor of the A genome has been provided by the comparison of isoenzymes (acid phosphatase, esterase and superoxide dismutase) between diploid and tetraploid wheat (Jaaska, 1997).

1.8.2.3 Origin of B genome

On the bases of different methods, three different types of theories have been proposed for the origin of the B genome.

1. ***T. speltoides* being the donor of B genome.** Gill and Appels (1988) reported that only *T. speltoides* among the diploid wheat shares close homology with the polyploid *T. aestivum*, on the basis of the rDNA spacer of the major *NOR* locus. Talbert *et al.* (1991, 1995) reported that *T. speltoides* was more closely related than the other four *Sitopsis* diploids (section 1.7.3) to the B genome of *T. aestivum*, based on the comparison of two repetitive sequences and the low-copy DNA sequences. Dau and Gustafson (1996) isolated a DNA fragment which contained *T. speltoides*-specific short repeated sequences, which could not be found in the other species of the *Sitopsis* section, and found that it was present in the B genomes of *T. turgidum* and *T. aestivum*. Badaeva *et al.* (1996) reported that *T. speltoides* was most similar to the B genome of polyploid wheat but different from the other species of the *Sitopsis* section in C-banding patterns. Several studies have indicated that *T. speltoides* shows an intra-specific variation in the *coxII* intron, which is consistent with the reported DNA polymorphism in *T. aestivum* in both organelle and nuclear DNAs (Miyashita *et al.*, 1994; Ohsako *et al.*, 1996; Sasanuma *et al.*, 1996). In addition, Blake *et al.* (1998) also supported the idea that the B genome of common wheat had a monophyletic origin and that *T. speltoides* was the only possible donor of the B genome, based on comparison of single-copy DNA sequences of fourteen loci between different types of wheat.

2. **More than one species being the likely donors of B genome.** While the above results all support the view that *T. speltoides* might have contributed B genome to polyploid wheat, other reports indicate that more than one species could be the likely B donors, i.e., the origin of B genome of common wheat might be polyphyletic. Three of the five putative B genome donors from the *Sitopsis* section had amino acid sequences very similar to that of the purothionin form specified by the *Pur-B1* locus of *T. turgidum* and *T. aestivum* (Kerby *et al.*, 1990). More than one of the several candidates for the B genome in the *Sitopsis* section also had the same RFLP patterns as *T. aestivum* (Takumi, 1994).

3. ***T. speltoides* not being the donor of B genomes.** Miyashita *et al.* (1994) found that *T. speltoides* was not identical to *T. turgidum-T. aestivum*, on analyses of variations in restriction maps. Mori *et al.* (1997), on the basis of the variation in the *coxII* intron, did not agree that *T. speltoides* was the B genome donor to *T. turgidum* and *T. aestivum*. Randhawa *et al.* (1997) reported that some species other than *T. speltoides* in the *Sitopsis* section contributed the B genome to polyploid wheat, based on the variations in the glutenin subunit composition in wheat.

Conclusion: When consideration is given to the above evidence and the various theories are taken together, the information on the possible progenitor of B genome of common wheat remains still unclear and controversial, although the majority of reports support that *T. speltoides* might be the donor of B genome. Three further possibilities regarding the identity of the ancestor of the B genome have been suggested (Sears, 1977; Breiman and Graur, 1995; Dau and Gustafson, 1996):

1. The donor is now extinct;
2. The donor of B genome has not been discovered yet;
3. The B genome originates from an introgression of two or more parental species.

For the third model of B genome evolution, however, Dvorak and Zhang (1990) and Talbert *et al.* (1991) were unable to find any evidence of such introgression into *T. aestivum*, based on the studies of repeated DNA sequences in wheat. The origin of B genome thus remains to be identified unequivocally.

1.9 Relationship of genomes between common wheat and its relatives

1.9.1 Genomes of *T. timopheevi-T. zhukovskyi* and their origins

Among the relatives of common wheat, the origins of the different genomes of the tetraploid *T. timopheevi* (AAGG) and the hexaploid *T. zhukovskyi* (AAAAGG) have been studied extensively, due to their possible application in wheat production in the Caucasus region (Kimber and Sears, 1987). The two types of tetraploid wheat, *T. turgidum* (AABB) and *T. timopheevi* (AAGG), offer an attractive model for the study

of genome evolution and speciation of wheat (Gill and Appels, 1988). This model is also important in studies of cytoplasm, because studies so far have shown that *T. turgidum* and *T. timopheevi* have the same nucleus but different sources of cytoplasm (Friebe and Gill, 1996).

1.9.1.1 A genome

T. zhukovskyi (AAAAGG) has been considered to originate from the hybridization of *T. timopheevi* (AAGG) and *T. monococcum*, the latter contributing one A genome to of *T. timopheevi* and the second A genome to *T. zhukovskyi* (Jakubzizer, 1958; Upadhya and Swaminathan, 1963; Johnson 1968). However, based on repeated DNA sequence profiles, the origin of A genome in *T. timopheevi* was found to be *T. urartu* instead of *T. monococcum* (Dvorak *et al.*, 1993; Breiman and Graur, 1995). Therefore, *T. zhukovskyi* does not really possess two identical A genomic sets, but two distinct A-like genomes derived from two different diploid species, *T. monococcum* and *T. urartu* (Dvorak *et al.*, 1993). The donors of A genomes in *T. timopheevi* and *T. zhukovskyi* are still in debate. In order to distinguish between the different A genomes between *T. timopheevi* and *T. zhukovskyi*, the A genome *T. timopheevi* (AAGG) is designated as A^t genome and the second A genome of *T. zhukovskyi* (AAAAGG) is designed as A^z genome in this project. Thus the constitutions of genomes are A^tA^tGG for *T. timopheevi* and $A^tA^tA^zA^zGG$ for *T. zhukovskyi*.

1.9.1.2 G genome

Cytoplasmic analyses and studies of repeated DNA sequences indicate that *T. timopheevi* (A^tA^tGG) and *T. turgidum* (AABB) could have originated from a common tetraploid progenitor. *T. speltoides* was suggested to be the donor of G genome to *T. timopheevi* (Johnson, 1975; Dvorak and Appels, 1982; Chen and Gill, 1983; Ogihara and Tsunewaki, 1988; Dvorak and Zhang, 1990; Jiang and Gill, 1994; Shands and Kimber, 1973; Jaaska, 1978; Tsunewaki and Ogihara, 1983; Talbert *et al.*, 1991). *T. sharonensis* was also suggested to be the possible contributor of G genome (Kushnir *et al.*, 1983).

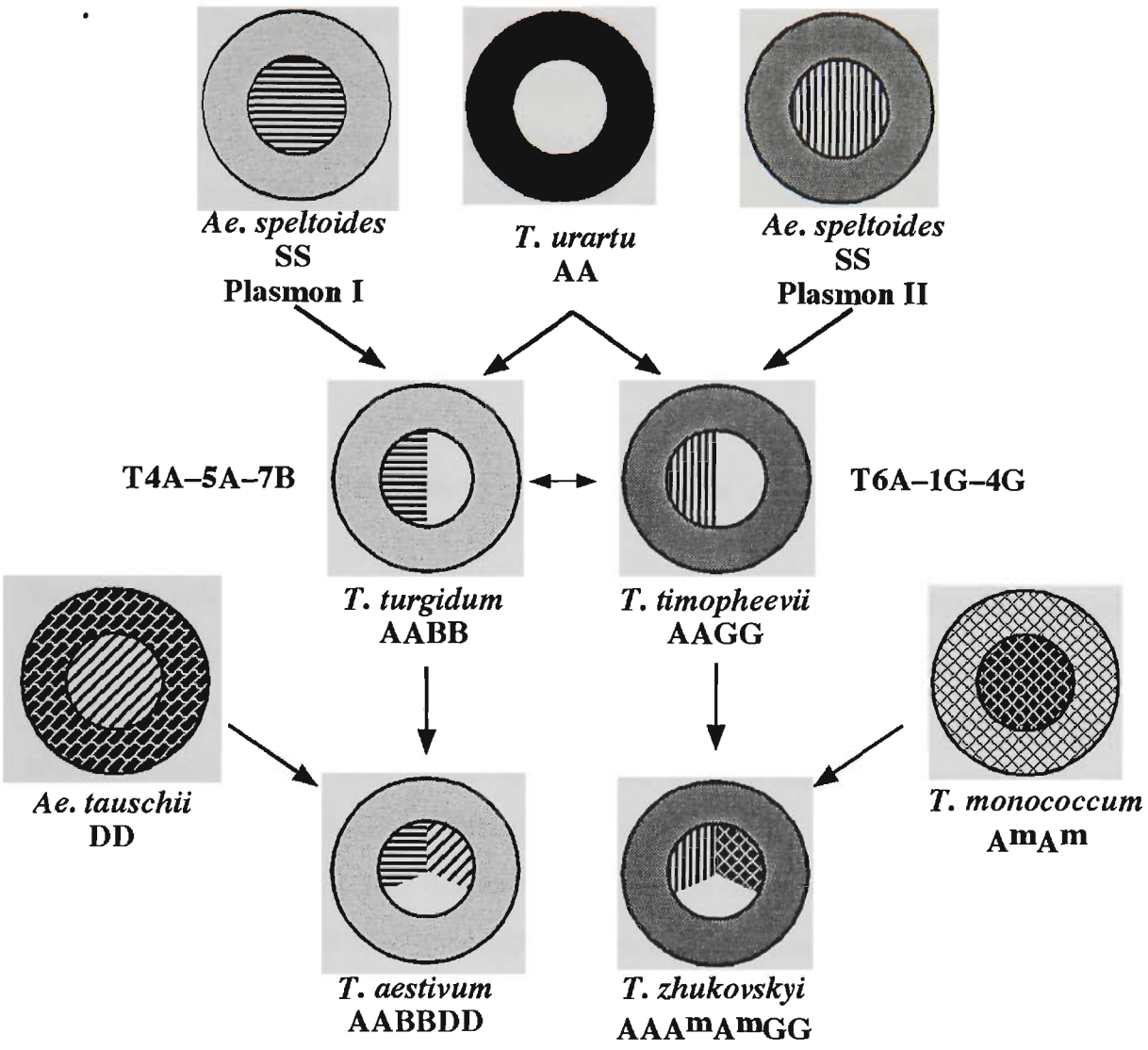


Fig. 1.8 Current theory of the evolution of wheat (Updated homepage of Wheat Genetics Resource Centre, Kansas State University, USA, as at 26, April, 1999). The theory for evolution of wheat is described in section 1.9.1.

Summary: A current theory of the phylogenetic and evolutionary relationships of different genomes of common wheat and its relatives has been summarised in Fig. 1.8, by Gill and his colleagues. In this theory, the donors of A and B genomes of *T. turgidum* are *T. urartu* and *T. speltoides*, respectively; *T. aestivum* is a hybrid of *T. turgidum* and *T. tauschii* which contribute D genome to *T. aestivum.* The donors of A and G of *T. timopheevi* are *T. urartu* and *T. speltoides*, respectively; *T. zhukovskyi* is a hybrid of *T. timopheevi* and *T. monococcum* which contributed the A genome to *T. zhukovskyi*. *T. turgidum* and *T. timopheevi* have the same nucleus sources but different cytoplasm sources.

1.9.2 Complexity of origins of various genomes of wheat

Extensive efforts have been made to understand the origins of different genomes of wheat. The reason for the intensive and continuous research work on wheat phylogeny has been its economic importance and evolutionary peculiarities. Both reasons are closely linked, as the wild wheat species can furnish important material for genetic resources such as genes for disease resistance or cold hardiness. However, it is difficult to identify the actual origins of genomes in different types of polyploid wheat, due to lack of unequivocal evidence, appropriate existing material and the ancient record of breeding carried out by humans. Moreover, a number of translocations have been reported to exist between different genomes. For example, a translocation between sections of long arm of 7B chromosome and short arm of 4A chromosome in common wheat has been reported (Chao *et al.*, 1989). Chromosomes 4A, 5A and 7B in *T. turgidum* and *T. aestivum* and chromosomes 6A, 1G and 4G in *T. timopheevi* are thought to be rearranged due to the presence of a cyclic translocation (Friebe and Gill, 1996). In addition, polymorphism within species has also been found in diploid and polyploid wheat. For example, *T. tauschii*, a putative donor of D genome to common wheat was found to be polymorphic for RFLP patterns (Talbert *et al.*, 1998). Studies have shown that RFLPs and polymorphic protein profiles in different varieties of common wheat (Chao *et al.*, 1989; Kam-Morgan and Gill, 1989; Liu *et al.*, 1990; Lubbers *et al.*, 1991, Frederiksen and Seberg, 1992; Monte *et al.*, 1993). It has also been suggested recently that common wheat had polyphyletic origins (Talbert *et al.*, 1998). These observations make it even more complicated to identify accurately the

origins of the different genomes in wheat. More studies at molecular level are thus required to reveal the origins of different genomes in different types of wheat unequivocally.

1.10 The aims of the project

1.10.1 The need to identify and investigate the individual *waxy* genes

As explained in section 1.4.1.2, the waxy protein plays a critical role in regulating the amylose concentration of wheat starch, a criterion highly significant for certain food and non-food industries. However, the following questions are raised:

- Do the three waxy proteins in common wheat, which exhibit distinguishable properties upon SDS-PAGE, have differences in their biochemical structures and/or functions, or make unequal contribution to the overall amylose concentration?

- Is it possible to develop molecular markers for the three *waxy* loci of common wheat, for identification of specific waxy varieties?

- Is it possible to obtain complete molecular information for the individual *waxy* genes of different genomes of wheat, in order to further understand the phylogenetic relationship between these and hence evolution of wheat?

- Are the three *waxy* genes in common wheat sufficiently identical so that a single antisense cDNA approach can be utilised to inhibit the expression of all *waxy* genes to obtain a transgenic wheat without amylose?

It is evident that in order to answer these questions it would be necessary to clone and sequence the individual *waxy* genes. Diploid wheat offers the opportunity to clone single *waxy* genes representing the appropriate counterparts of those in polyploid wheat; furthermore, the sequences identified in diploid wheat would provide an essential tool for the cloning of the respective genes in polyploid wheat.

1.10.2 Selection of plant materials

The information provided in section 1.9 has indicated that except for diploid *T. tauschii*, which is accepted as the donor of D genome, the donors of A and B genomes of common wheat are still unclear. On the basis of the evidence available at the time of commencement of this project (early 1996) which suggested that *T. monococcum* and/or *T. urartu* were the likely donors of A genome of common wheat, *T. monococcum* was chosen to clone the *waxy* A gene. Similarly, *T. speltoides* was chosen from several candidates of B genome to clone the *waxy* B gene.

It was essential also to clone and compare the *waxy* genes from other proposed progenitors of the A and B genomes of common wheat, other polyploid wheat, and the wild relatives of wheat. This information would provide a clearer understanding of the evolutionary relationships among the different genomes of wheat.

1.10.3 Specific aims

The main aim of this project was to clone and characterise the individual *waxy* genes from different diploid, tetraploid and hexaploid wheat and utilize the sequence data to develop molecular markers for these *waxy* loci and to extend our understanding of the evolution of wheat. The internal relationships between different sections of this project are shown in Fig. 1.8. These include the following specific aims.

- To investigate the properties of proteins embedded within the starch granules of endosperm in different diploid, tetraploid and hexaploid wheat (Chapter 3).

- To clone the genomic copy of the *waxy* genes, encompassing the entire coding region, of *T. monococcum* as a proposed donor of A genome and *T. speltoides* as a proposed donor of B genome (Chapter 4).

- To characterize the waxy proteins deduced from the sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* and understand their biochemical and physical properties predicted from their amino acid sequences (Chapter 5).

- To isolate the individual partial *waxy* genes from the two *waxy* loci each of the tetraploid *T. turgidum* (AABB) and *T. timopheevi* (A^tA^tGG), three *waxy* loci each of the hexaploid *T. aestivum* (AABBDD) and *T. zhukovskyi* ($A^tA^tA^zA^zGG$), and a single *waxy* locus of each of the other diploid progenitors. To find the *waxy* locus-specific sequences representing the A, B, D, G, A^t and A^z genomes and further understand the phylogenetic relationship of these genomes between different polyploid wheat (Chapter 6).

- To develop the molecular markers for the individual *waxy* loci of common wheat and to investigate the molecular basis of the waxy null lines (Chapter 7).

Taken together, the great volume of sequence data proposed for this project will provide valuable information for the *waxy* genes and the waxy protein encoded thereby in wheat. This information will be valuable for understanding wheat evolution, analysing mutants of the *waxy* genes and applications in transgenic technology.

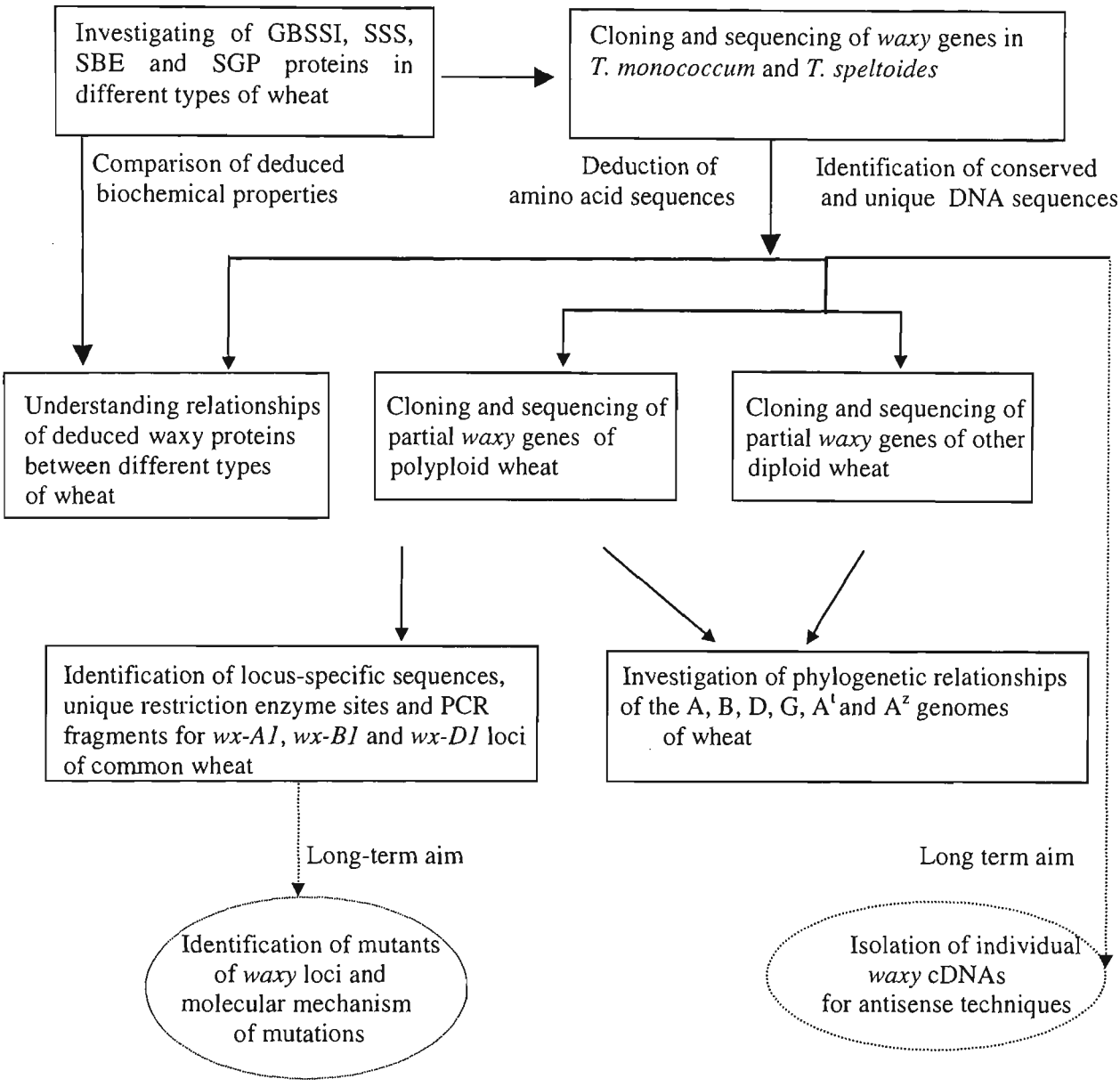


Fig. 1.9 Specific aims of this project. The main objective of this project was to clone and characterise the *waxy* genes in *T. monococcum* and *T. speltoides*. After this was achieved, this was extended to achieve some additional aims that have further applications in wheat breeding and biotechnology programs.

CHAPTER 2

MATERIALS AND METHODS

List of Figures and Tables in Chapter 2

Chapter 2 Materials and Methods

2.1 Plant material

2.1.1 Plant material for electrophoretic analysis of proteins

Different types of diploid, tetraploid and hexaploid wheat were investigated for the proteins embedded within the starch granules in the endosperm of mature seeds. The sources of the plant material are shown in Table 2.1. The material includes:

1. *T. monococcum*, a proposed contributor of A genome to polyploid wheat;
2. *T. tauschii*, the proposed donor of D genome to common wheat;
3. *T. speltoides T. bicornis*, *T. longissima*, *T. searsii* and *T. sharonensis*, the five species in the *Sitopsis* section, proposed by various investigators to have donated B genome to polyploid wheat; *T. speltoides* or/and *T. sharonensis* are also proposed to have donated the G genome to polyploid wheat;
4. *T. urartu*, a proposed donor of both A genome and B genome to polyploid wheat;
5. Three forms of emmer wheat (AABB), the putative tetraploid progenitor of *T. aestivum*, i.e., *T. turgidum* ssp. *turgidum*, *T. turgidum* ssp. *dicoccum* and *T. turgidum* ssp. *durum*, and their wild relative, *T. dicoccoides*;
6. Common wheat, *T. aestivum* (AABBDD). Three types of the hexaploid common wheat were utilized: one type with high paste viscosity, considered appropriate for making noodles and bread (varieties Rosella and Machete); the second type with low paste viscosity and commonly used to make biscuits (varieties Wyuna and Vectis); and the third type, experimental wheat (varieties Sturdy and Chinese Spring);
7. *T. timopheevi* (A^tA^tGG);
8. *T. zhukovskyi* ($A^tA^tA^zA^zGG$); proposed to be a hybrid of *T. timopheevi* and *T. monococcum.*

Table 2.1 Plant materials of diploid, tetraploid and hexaploid wheat used in the protein analyses

Species	Accessions tested	Ploidy of Genomes	Genome	Source
T. moncoccum	15*	2x	AA	MU and AWCC
T. urartu	Aus 17649, Aus 90459	2x	AA or/and BB	AWCC
T. speltoides	Aus 18995, Aus 21638 Aus 18942	2x	BB or/and GG	AWCC
T. bicornis	Aus 18795, Aus 18930	2x	BB	AWCC
T. longissima	Aus 18797, Aus 18798	2x	BB	AWCC
T. searsii	CPI 115475	2x	BB	CSIRO
T. sharonensis	Aus 18802, Aus 18747	2x	BB or/and GG	AWCC
T. tauschii	Aus 18967, Aus 18916	2x	DD	MU
	CPI 799			CSIRO
T. turgidum	Aus 19785, Aus 19592 Aus 764	4x	AABB	AWCC
T. aestivum	Chinese Spring, Rosella. Wyuna	6x	AABBDD	VIDA
T. timopheevi	Aus 19794, Aus 19796	4x	A^tA^tGG*	AWCC
T .zhukovskyi	Aus 18689, Aus 18690	6x	$A^tA^tA^zA^zGG$*	AWCC

* C68-73; C68-18; C68-19; C68 24G 16; Sile 500 Sp 3750; Sile 506 Sp 3376, Aus. 16280; Aus 3701; Aus. 16278; Aus. 15361; Aus. 22983; Aus. 22986; Aus. 90442; Aus. 90624 and Aus. 95051.

MU: The University of Melbourne, Department of Agriculture. AWCC: Australia Winter Cereals Collection, Tamworth, NSW. CSIRO: Commonwealth Scientific and Industrial Research Organization, Canberra, ACT, Australia; VIDA: Victoria Institute for Dryland Agriculture, Horsham, Victoria, Australia. * As discussed in section 1.9.1, in order to between distinguish between different A genomes of the polyploids *T. timopheevi* and *T. zhukovskyi*, A genome of *T. timopheevi* (AAGG) is designated as A^t genome and the second A genome of in *T. zhukovskyi* (AAGGAA) is designated as A^z genome, in this project.

2.1.2 Plant materials for cloning the *waxy* genes

On the basis of the results of analyses of the waxy proteins in SDS-PAGE, a typical accession of each type of wheat was used to clone its *waxy* gene (Table 2.2). *T. monococcum* (AA) and *T. speltoides* (BB) were used to clone their respective, almost complete genomic *waxy* gene, from the first translated exon to the last one (Chapter 4). The corresponding section of the *T. tauschii* (DD) *waxy* gene was investigated simultaneously at CSIRO in a collaborative effort (R. Appels, pers. commun.). Other species in Table 2.2 were used to clone their respective partial *waxy* genes (Chapter 6).

2.2 Electrophoretic analysis of proteins embedded within starch granules

2.2.1 Preparation of starch granules

Starch granules from endosperm of mature seeds were prepared by the method of Echt and Schwartz (1981). The pericarp and embryo were cut out and 200 mg endosperm homogenised in 3 mL of the precooled extraction buffer (0.055 M Tris-Cl pH 6.8, 2.6% SDS, 5% (V/V) β-mercaptoethanol, 10% (V/V) glycerol and 0.05 M dithiothreitol) with a precooled mortar and pestle held on ice. The suspension was filtered through a layer of Miracloth and centrifuged at 14000 rpm in an Eppendorf 5415C microcentrifuge, for 1 minute, at 4^0C. This centrifuge was used for centrifugations of all small-volume samples, unless mentioned otherwise. The pellet of starch granules was washed three times with extraction buffer, followed by double distilled water and then methanol, in 3 mL of each elution buffer, in this order, and then dried under vacuum. During the procedure, all elution buffers were precooled to 4^0C to avoid swelling of starch granules and the release of proteins.

2.2.2 Extraction of proteins embedded within starch granules

The starch granules extracted and purified as above should contain only those proteins that are embedded within them, as those proteins that are on the surface of starch granules would have been removed. The predominant protein embedded within the starch granules is the waxy protein or GBSSI, as discussed in section 1.5.

Table 2.2 Plant material used for cloning *waxy* genes

Species	Name	Origin	Usage
T. monococcum	Aus 22986	AUS	The *waxy* gene (Chapter 4)
T. speltoides	Aus 21638	IRQ	The *waxy* gene (Chapter 4)
T. turgidum ssp. *durum*	Aus 764	CYP	One partial *waxy* gene (Fragment 1) (Chapter 6). Two partial *waxy* genes (Fragment 2) (Chapter 6).
T. aestivum	C.S.*	CHI	Three partial *waxy* genes (Chapter 6).
T. bicornis	Aus 18795	UTK	Partial *waxy* gene (Chapter 6).
T. longissima	Aus 18799	UTK	Partial *waxy* gene (Chapter 6).
T. sharonensis	Aus 18802	USA	Partial *waxy* gene (Chapter 6).
T. urartu	Aus 17649	USR	Partial *waxy* gene (Chapter 6).
T. timopheevi	Aus 19794	KEN	Two partial *waxy* genes (Chapter 6).
T. zhukovskyi	Aus 18689	USR	Three partial *waxy* genes (Chapter 6).

* C.S., Chinese Spring.

All wheat weeds were kindly provided by Australia Winter Cereals Collection (AWCC), Tamworth, NSW.

2.2.2.1 Preparation of protein samples for the one-dimensional SDS-PAGE

For analysis of proteins by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), the proteins were extracted from the starch granules by the method of Echt and Schwartz (1981). 150 µL of extraction buffer (section 2.2.1) was added to an Eppendorf tube containing 10 mg of starch granules. The contents of the tube were mixed and heated in boiling water for 10 minutes. The gelled starch solutions were cooled on ice immediately, then centrifuged at 14000 rpm for 10 minutes at 4^0C. The supernatants were used for further analysis.

2.2.2.2 Preparation of protein samples for two-dimensional electrophoretic analysis

The proteins from the starch granules were extracted using an extraction buffer [8 M urea, 2% (V/V) Nonidet-P40, 2% (V/V) ampholine (pH 4-7), 5% polyvinylpyrrolidone (PVP, M.W. 360000, Sigma), 5% β-mercaptoethanol] (Nakamura *et al.*, 1993). 10 mg of starch granules were mixed with 150 µL of this buffer and heated for 5 minutes in boiling water, then centrifuged at 14000 rpm for 10 minutes at 4^0C. The supernatants were used for further analysis.

2.2.3 One dimensional SDS-PAGE

2.2.3.1 Preparation of gels

SDS-PAGE was used to determine the molecular weights of proteins by the discontinuous electrophoresis system (Laemmli, 1970; Hames, 1990). Briefly, the gels were prepared and electrophoresed using the following gel solutions and buffers.

1. Resolving or separating gel: 0.375 M Tris-Cl (pH 8.8), 0.1% SDS. The electrophoresis system had a 8% separating gel, but the ratio of acrylamide:bisacrylamide used was 30:0.135; this ratio has been reported to allow high resolution of the waxy proteins (Nakamura *et al.*, 1992).
2. Stacking gel: 0.125 M Tris-Cl (pH 6.8), 0.1% SDS. The electrophoretic systym had a 4% stacking gel, but the ratio of acrylamide:bisacrylamide was

traditional 29:1 rather than 30.0:0.135 ratio, as the latter was found to result in gels that were too soft to allow proper loading of the protein samples.

3. Electrode buffer (pH 8.3): 0.025 M Tris, 0.192 M glycine, 0.1% SDS.
4. Protein Sample Buffer (loading buffer): 0.125 M Tris-Cl (pH 6.8), 10% glycerol, 5% β–mercaptoethanol, 2% SDS, 0.013% Bromophenol blue.

All gels were polymerised by the addition of 0.05% of TEMED and 1% of freshly-made 10% ammonium persulphate.

2.2.3.2 Electrophoretic conditions

Protein samples were prepared for electrophoresis by addition of an equal volume of the protein loading buffer to the supernatants in section 2.2.2.1, mixing these together, and incubating for 3 minutes in boiling water, followed by cooling on ice. Protein samples were loaded alongside the high-range of molecular weight markers (45 kDa to 200 kDa, Bio-Rad). Electrophoresis was conducted at a constant voltage of 100 V until the dye had run out of the Mini-Protean gels, then a further 15 minutes after this point, for best separation of the waxy proteins, or until the dye had reached bottom of the gels in case of the larger Protean-I system (Bio-Rad).

2.2.4 Two dimensional SDS-PAGE

2.2.4.1 Preparations of gels

The first dimension: The Immobiline DryStrip pH 4-7 (Pharmacia) was applied for isoelectric focussing of proteins in the first dimension. The two-dimensional electrophoretic unit (217 Multiphor II, Pharmacia) was used following the instruction manual provided by the supplier.

The second dimension: Preparation of the second dimensional SDS-PAGE was the same as in section 2.2.3.1, using 30:0.135 acrylamide:bisacrylamide. A wide, single-well comb was used for loading the Immobiline DryStrip from the first dimensional gel.

2.2.4.2 Electrophoretic conditions

25 μL of protein samples and 25 μL of loading buffer (supernatants in section 2.2.2.2) (24 g urea, 1.0 mL β-mercaptoethanol, 1.0 mL Pharmalyte pH 3-10, 0.25 mL Triton X-100, and 0.01% Bromphenol blue in 50 mL solution) were mixed, then loaded into the sample cup of the first dimensional gel. The first dimensional gel was run at 300 V, 1 mA for 1 hour; 2000 V, 1 mA for 5 hours; then 3000 V, 1 mA for 18 hours. All procedures for the first dimensional gel were followed the instruction manual of the supplier (Pharmacia). The second dimensional gel was run at a constant voltage of 100 V until the dye had reached the bottom of the gel in the Protean-I system (Bio-Rad).

2.2.5 Processing of gels following electrophoresis

Upon completion of electrophoresis, the gel plates were separated and the stacking gels discarded. The resolving gels were stained by silver stain using the methods given in Table 2.3.

The gels were photographed using a MP-4 Land Camera (Polaroid) and Polaroid 667 positive film, with 5.6 aperture and 1/125 exposure time. The gels were then placed in distilled water for at least 2 hours before drying them using the GelAir Dryer (Bio-Rad) for 3 hours, for permanent storage.

Table 2.3 Procedures for staining gels

Solution	Composition*		Time of Treatment
Fixing solution	100 mL	Ethanol	30 minutes
	25 mL	Acetic acid	
Incubation solution	75 mL	Ethanol	
	17.0 g	Sodium acetate 3 H_2O	30 minutes
	0.5 g	Sodiumthiosulfate	
	1.3 mL	Glutardialdehyde (25% W/V)	
Washing	H_2O		3×10 minutes
Silver solution	0.25 g	Silver nitrate	40 minutes
	50 µL	Formaldehyde	
Developing solution	6.25g	Sodium carbonate	15-45 minutes**
	25 mL	Formaldehyde	
Stop solution	3.65 g	EDTA	5-10 minutes
Preserving solution	25 mL	Glycerol	20 minutes

* All recipes are for making up to 250 mL.

** Time for development of dye was decided depending on the intensity of the protein bands. It was necessary to keep some of the gels in the developing solution up to 45 minutes, in order to develop faint bands of the proteins in the high molecular weights.

2.3 Cloning of the *waxy* genes from genomic DNA

2.3.1 Selection of methods for cloning the *waxy* genes

A number of techniques are applied in cloning genes from plants for different purposes, for example, constructing a genomic DNA library to clone a gene containing the promoter region, constructing a cDNA library to investigate only the coding sequences without the intervening introns and to develop antisense approaches for transgenic plants. The exons often serve to indicate the degree of conservation of the coding sequences, due to the functional constraints and thus serve as "biological clocks" for determining evolutionary history and rate of divergence of species. Since this project was to analyse the sequence and organization of a number of *waxy* genes from different genomes of wheat, it was essential to clone and investigate the sequences of both introns and exons. It was thus decided to use the Polymerase Chain Reaction (PCR), a rapid and efficient approach for amplifying, cloning and/or sequencing multiple waxy genes, in a relatively short period of time, compared to constructing and screening individual genomic libraries.

Comparison of the amino acid sequences deduced from the six cDNA sequences of the *waxy* genes in plants indicated that the four monocotyledonous plant species have 83.4% and 93.3% identity while those from the two dicotyledonous plants share a 75.4% identity (Ainsworth *et al.*, 1993). These results indicate that the exon sequences of the *waxy* genes are highly conserved in plants, and thus these might be highly conserved also between different species of wheat. The PCR approaches could then be developed utilising these, to amplify, clone and/or sequence various overlapping fragments of the *waxy* genes, including the interrupting and possibly variable introns.

2.3.2 Growth of seedlings

Seedlings of different types of wheat (Table 2.2) were grown in a dark room. Young yellow leaves were collected from these seedlings, quick-frozen in liquid nitrogen and stored at -80^{0}C for future use.

2.3.3 Preparation of genomic DNA

The protocol for the extraction of genomic DNA from plant material was based on that of Bernatzky and Tankley (1986), with some modifications. 3 g of the frozen leaf samples were placed in a pre-cooled mortar and pestle and ground under liquid nitrogen by adding 5 g of glass beads. The leaf powder was left at room temperature for about 10 minutes to allow the liquid N_2 to evaporate, ensuring that the pulverized sample remained dry and powdery and did not thaw. 10 mL of pre-cooled extraction buffer [100 mM Tris-Cl pH 8.0, 20 mM EDTA, 1.4 M NaCl, 2% CTAB (hexadecyltrimethylammonium bromide), and 2% β-mercaptoethanol] was added to the leaf powder, and the suspension was incubated at 65°C for 90 minutes with occasional gentle mixing. The suspension was cooled to room temperature, and then extracted with an equal volume of 24:1 chloroform/octanol by centrifuging 2400x g at 4°C for 5 minutes. The upper aqueous phase was removed and 1/10 its volume of 10% CTAB added to it; the mixture was then extracted with the chloroform/octanol at 20°C (10 minutes, 2400x g). The aqueous phase was collected and 3 volumes of CTAB precipitation solution (50 mM Tris-Cl pH 8.0, 10 mM EDTA and 1% CTAB) was added to it. The mixture was then centrifuged at 8000x g in the JA-20 rotor at 4°C for 10 minutes. The DNA/CTAB pellet was resuspended in 1 mL of high-salt TE buffer (TE pH 8.0, 1 M NaCl). The DNA was precipitated by addition of precisely 0.6 volume of pre-cooled isopropanol and centrifugation at 8000x g for 10 minutes at 4^0C. The DNA pellet was dissolved in 200 µL water; the purity of the DNA samples was then assayed by reading their absorbance of DNA samples at 260nm/280nm using a UV spectrophotometer (LKB,. Model 4045). All samples were checked on agarose gels for the presence of RNA. In case RNA was detected on the gels or the ratio of A_{260}/A_{280} readings was >1.80, RNA in the DNA sample was removed by treating with DNase-free RNase A at 37^0C for 30 minutes. The DNA sample was then recovered by re-extraction of the sample with phenol:chloroform (24:1 in volume), and precipitation with 1/10 volume of ammonia acetate (3 M, pH 5.2) and 2 volumes of 95% chilled ethanol on ice for 30 minutes, followed by centrifugation at 8000x g in the JA-20 rotor at 4°C for 10 minutes to collect the DNA pellet. The purified DNA sample with a 1.75-1.80 ratio of A_{260}/A_{280} absorbance readings was

dissolved in distilled water and its concentration was determined according to the formula: 1 A_{260} unit = 50 μg DNA/mL (Sambrook *et al.*, 1989).

2.3.4 Primers for cloning the *waxy* genes of *T. monococcum* and *T. speltoides*

2.3.4.1 Criteria for primer designation

The following criterion were applied for designing all primers (Sambrook *et al.*, 1989):

1. There was no self-complementary or self-looping sequence within or at the ends of each primer.
2. The annealing temperature of a pair of primers was similar.
3. The length of primers was >17 bp, which is statistically essential to obtain a unique annealing site in an eukauyotic genome.
4. Proportions of A, T, G and C in the primers followed the principle of 40 %-55% G+C, and there was higher proportion of G and C at 3' ends (G/C clamps) to enable more stable binding between the primers and their target sequences.

2.3.4.2 Primer sequences

Based on the analyses of amino acid sequences of the waxy proteins from barley (Rohde *et al.*, 1988), rice (Wang *et al.*, 1990), maize (Klösgen *et al.*, 1986), potato (van der Leij *et al.*, 1991) and wheat (Ainsworth *et al.*, 1993) (Fig. 5.4 in Chapter 5), the conserved domains of these were determined. The DNA sequences corresponding to these conserved domains were used to design four pairs of primers, in order to amplify the single *waxy* gene present in *T. monococcum* and *T. speltoides* each (Table 2.4). The *waxy* genes were amplified as four overlapping DNA fragments, which together covered the area from upstream of the start codon to the stop codon, as per the structure of the barley *waxy* gene (Rohde *et al.*, 1988). The primer sequences are shown in Table 2.4.

Table 2.4 Sequences of oligonucleotide primers used for the PCR amplifications

Primer	Direction	Sequences	Length (bases)	Melting Temperature (^{0}C)
		-31 -13		
F1	Forward	5'-TTGCTGCAGGTAGCCACAC-3'	19	59.7
		603 581		
R1	Reverse	5'-CTCAAGTGCTGCCTGGCAGAGAA-3'	23	63.7
		513 533		
F2	Forward	5'-GACCAAGGAGAAGATCTATGG-3'	21	62.0
		1080 1063		
R2	Reverse	5'-CATGCCGTTGACGATGCC-3'	18	59.5
		908 891		
F3	Forward	5'-ATCAACTGGATGAAGGCC-3'	18	57.2
		1573 1556		
R3	Reverse	5'-ATAGTGTCGACGAGCCCG-3'	18	59.5
		1490 1471		
F4	Forward	5'-AGCTCCAGGGAATGCGCTAC-3'	20	61.9
		1845 1826		
R4	Reverse	5'-TCAGGGAGCGGCGACGTTCT-3'	20	63.9

All primer sequences were based on the sequence of a *waxy* cDNA of common wheat (Clark *et al.*, 1991). F1 was directly based on the sequence of the 5' untranslated leader region of this cDNA. The other seven primer sequences were based on this cDNA sequence, which encoded amino acids conserved in the waxy proteins in barley (Rohde *et al.*, 1988), rice (Wang *et al.*, 1990), maize (Klösgen *et al.*, 1986) and potato (van der Leij *et al.*, 1991).

2.3.4.3 Primer locations

The locations of the above primers used to amplify the *waxy* genes of *T. monococcum* and *T. speltoides* are shown in Fig. 2.1. The locations of primers were chosen such that they were unlikely to be interrupted by introns, on the basis of alignment of the *waxy* cDNA of common wheat (Clark *et al.*, 1991) and a genomic copy of the barley *waxy* gene (Rohde *et al.*, 1988) (as there was no published information for a wheat genomic gene at that time). There are 11 exons and 10 introns the barley *waxy* gene, excluding the first untranslated exon and the first intron. The *wx-TmA* gene of *T. monococcum* and *wx-TsB* gene of *T. speltoides* cloned and sequenced in Chapter 4 cover the entire translated section, i.e., the 11 exons and the 10 introns interrupting these. The designation of Exon 1 (E1) has been used in this project for the first translated exon of the *waxy* gene of wheat and barley. This exon is equivalent to the exon 2 of Rohde *et al.* (1988). Thus, the exons 2 to 12 of Rohde *et al.* have been designated as exons 1 to 11 throughout this project. The predicted fragment lengths and sections amplified by each pair of primers are shown in Fig. 2.1 and Table 2.5.

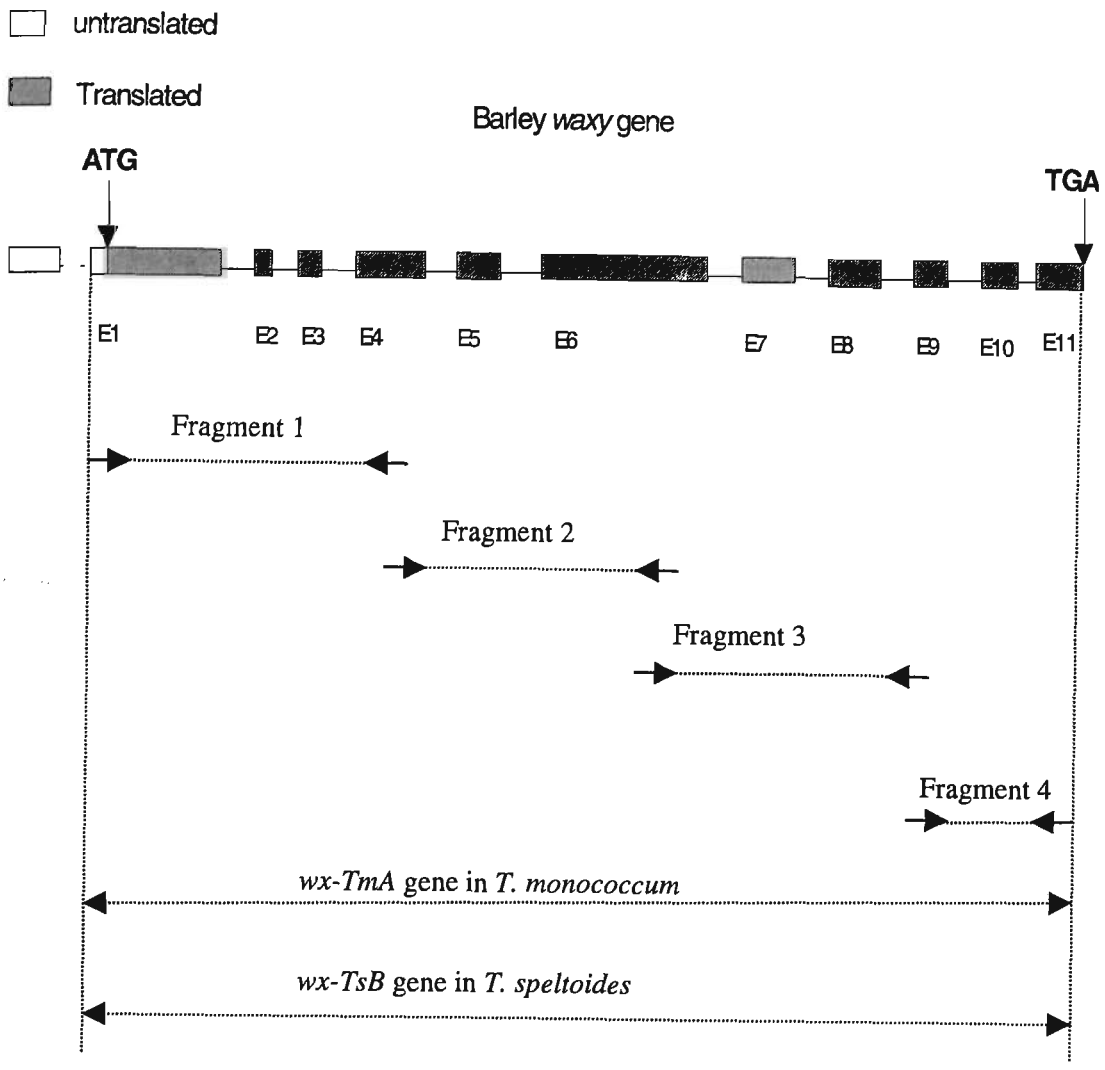


Fig. 2.1 Locations of the four pairs of primers (Table 2.4) used for cloning overlapping fragments of the *waxy* genes of *T. monococcum and T. speltoides*. The 12 boxes represent exon 1 in the untranslated region and the 11 translated exons of the *waxy* gene in barley (Rohde *et al.*, 1988). The lines between boxes indicate introns. The *waxy* genes of *T. monococcum* and *T. speltoides* are cloned from the partial untranslated region through the translation initiation codon ATG to termination codon TGA. There is a 41 bp untranslated leader at the 5' end of exon 1 of the barley *waxy* gene, 34 bp of which, upstream of the translation initiation codon, was included in Fragment 1 in this study.

Table 2.5 The fragments of the *waxy* genes obtained by PCR and the expected lengths of PCR products

Fragment (bp)	Primer		Regions covered In fragments	Expected sizes (bp)
	Upstream	Downstream		
1	F1	R1	Exons 1, 2, 3 and 4 Intron 1, 2 and 3	933
2	F2	R2	Exon 4, 5 and 6 Introns 4 and 5	777
3	F3	R3	Exons 6, 7, 8 and 9 Introns 6, 7 and 8	896
4	F4	R4	Exons 8, 9, 10 and 11 Introns 8, 9 and 10	614

The eleven grey boxes represent the eleven translated exons (E1 to E11) of the *waxy* gene of barley (Rohde *et al.*, 1988); E1 encodes the start codon and E11 encodes the stop codon. The open area of E1 indicates the untranslated region of this exon, preceding the start codon. The additional open box upstream of E1 represents the first untranslated codon of the barley *waxy* gene (Rohde *et al.*, 1988); this section has not been cloned in the present study. The broken lines represent introns. The *waxy* genes of *T. monococcum* and *T. speltoides* have been cloned from a 31 bp section of E1 upstream of the start codon to stop codon in E11 (Chapter 4).

2.3.5 Primers for cloning different regions of the *waxy* genes in *T. turgidum*

Based on the results that all four fragments were successfully cloned from *T. monococcum* and *T. speltoides* (Chapter 4) by using the four pairs of primers shown in Table 2.4, primers F1 and R1 were used initially to isolate Fragment 1 of the two *waxy* genes in *T. turgidum* ssp. *durum*. Comparison of the expected two genes of *T. turgidum*, a putative tetraploid progenitor of A and B genomes of common wheat (section 1.8), with those of *T. monococcum* and *T. speltoides*, would provide information regarding the origin of the *wx-A* and *wx-B* genomes. However, the sequences of the fragments obtained from several clones of Fragment 1 of *T. turgidum* were identical, representing either only the *wx-A* gene or only the *wx-B* gene. The primers F2 and R2 were then used in an attempt to isolate two different types of Fragments 2, and the attempt was successful. The two different types of Fragment 2 of *T. turgidum* were designated as *wx-TdA* and *wx-TdB*. The internal sequences of *wx-TdA* and *wx-TdB* were later redesignated respectively as *wx-Td2* and *wx-Td1*, respectively, in order to align the latter two with the partial, smaller *waxy* genes cloned from other diploid and polyploid wheat (section 2.3.6, results in Chapter 6). The three different Fragments of the *waxy* genes cloned from *T. turgidum* are shown in Fig. 2.2.

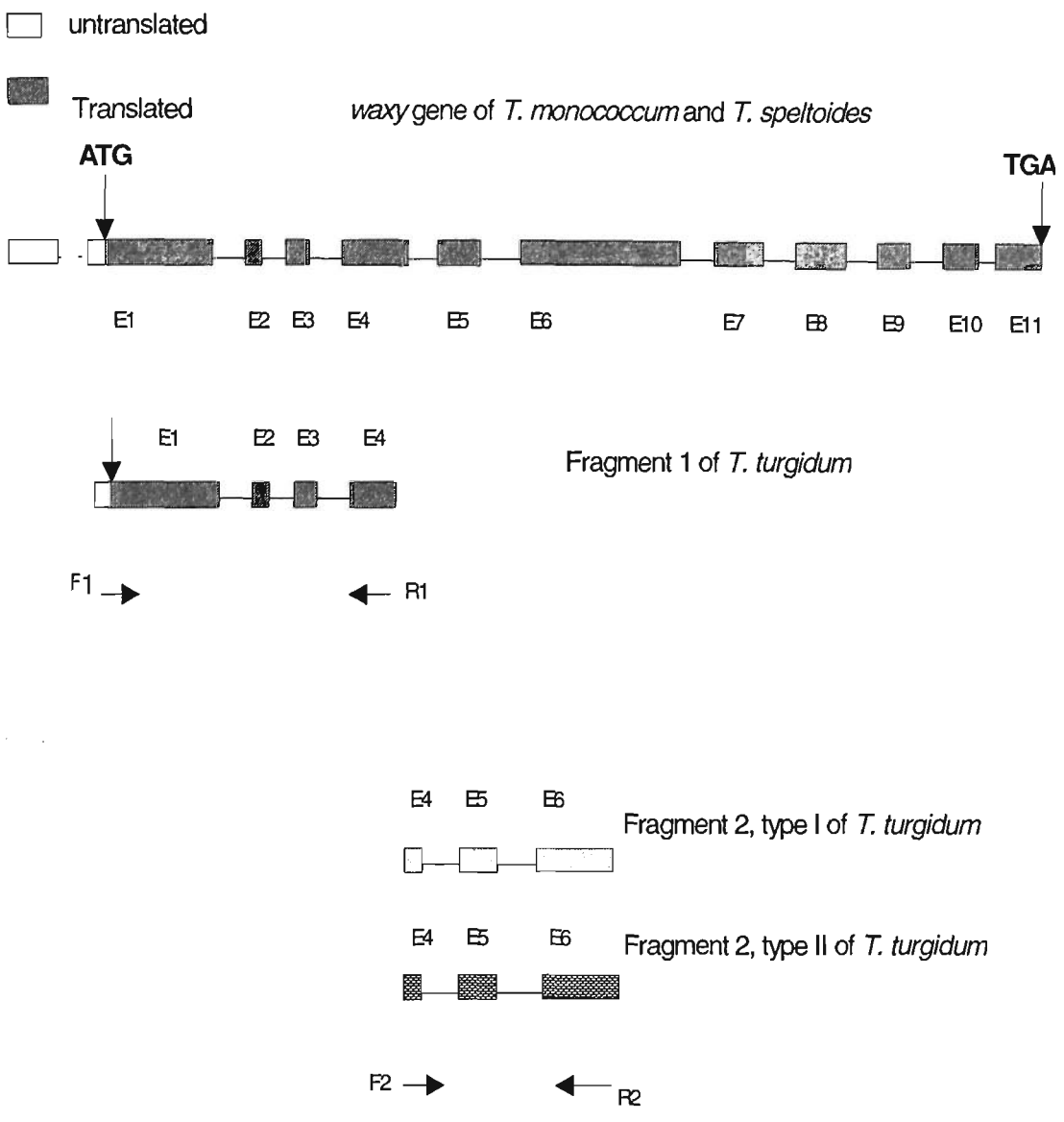


Fig. 2.2 The cloned regions of the *waxy* genes of *T. turgidum*. Boxes indicate exons and lines between boxes indicate introns. One type of Fragment 1 and two different types of Fragment 2 (*wx-TdA* and *wx-TdB*) were obtained from *T. turgidum*.

2.3.6 Primers for cloning partial *waxy* genes in different types of wheat

When primers F2 and R2 (Table 2.4) were used to amplify Fragment 2 from different genomes of polyploid wheat and diploid wheat, no expected PCR products were obtained, except for *T. monococcum* and *T. speltoides* (section 2.3.4) and *T. turgidum* (section 2.3.5). To obtain information for comparison of the *waxy* genes between different genomes of wheat (Chapter 6), a new reverse primer, PR2, and the same forward primer F2, were used together to amplify sections of the *waxy* genes covering the intron 4. This intron had displayed variable lengths in the *waxy* genes of *T. monococcum*, *T. speltoides* and *T. tauschii*. The reverse primer PR2 is located on exon 5 (Fig. 2.3) and its sequence was 5'-GCCGTCCTATAGATGCCATTG-3'. The predicted length of the PCR products amplified by F2 and PR2 is approximately 360 bp-390 bp, the variation occurring due to the variation in length of intron 4; the amplified fragment covers exon 4-intron 4-exon 5 (Fig. 2.3). The partial *waxy* gene was amplified and cloned from the individual genomes of polyploid *T. aestivum* (AABBDD) and *T. timopheevi* (A^tA^tGG)-*T. zhukovskyi* ($A^tA^tA^zA^zGG$) lineages and from other diploid wheat (Table 2.2) proposed to be the progenitors of A, B, D, G, A^t or A^z genomes in wheat (sections 1.8 and 1.9).

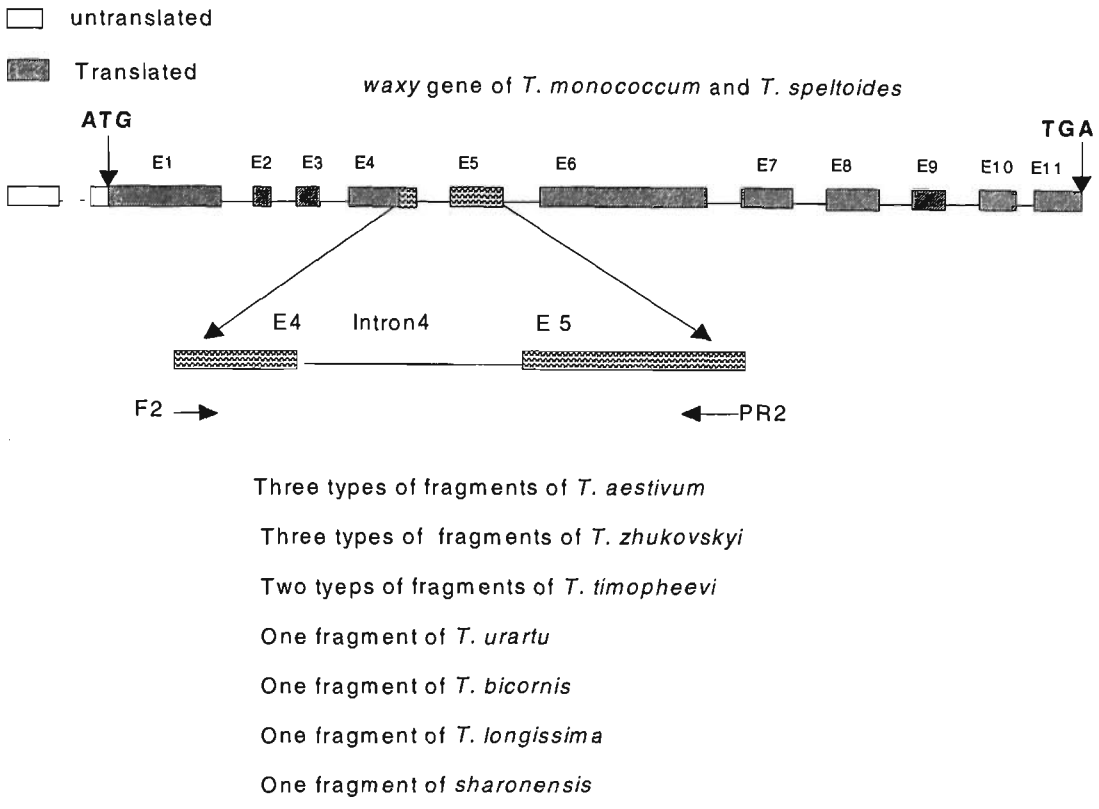


Fig. 2.3 The cloned regions of the *waxy* genes from different types of wheat. Boxes indicate exons and lines between boxes indicate introns. These partial *waxy* genes in wheat covered the major section of exons 4 and 5 and complete intron 4.

2.3.7 Primers for cloning partial *waxy* genes as molecular markers for the A, B and D genomes of common wheat

In order to develop an AFLP (Amplified Fragment Length Polymorphism)-based technique for identification of the individual *waxy* loci and their mutants in common wheat, one pair of primers was designed to attempt to amplify the three different fragments from the same region of the three *waxy* genes. The sequence of the forward primer (GF) on exon 4 was 5'-CCTGCTACCTCAAGAGCAAC-5' and the sequence of the reverse primer (GR) located in exon 5 was 5'-GGCTTGTCGTAGCCGTCGAT-3', and their locations are shown in Fig. 7.2. Based on the comparison of the two types of Fragment 2 of the *waxy* genes of *T. turgidum* (2.3.5) and the corresponding region of the *waxy* gene *T. tauschii* (R. Appels., per, commun.), it was expected that this pair of primers would lead to amplification of three fragments significantly different in length, due to the variable length of intron 5. The three fragments originated from *wx-A*, *wx-B* and *wx-D* loci of common wheat in a single PCR should be separated clearly on a 3% agarose gel (Chapter 7).

2.3.8 Synthesis and preparation of primers

Except for primers GF and GR in section 2.3.7 which were synthesized commercially by Bresatec, the oligonucleotide primers in sections 2.3.4-2.3.6 were synthesized in house, using DNA synthesizer 391 (Applied Biosystems). After synthesis, the cleavage from column, deprotection, dehydration and precipitation of the oligonucleotides were carried out as per the instruction manual provided by the supplier (Applied Biosystems).

Cleavage of oligonucleotides from synthesis column:

1. 5-7 mL concentrated ammonia solution (30-35%) was dispensed into a clean glass bottle (McCartney). The tubes were capped tightly.
2. The tip of a sterile 1 mL syringe was inserted into one end of the column. The syringe was fitted snugly with the plunger at the needle end.
3. 0.75 mL concentrated ammonia was drawn into another 1 mL syringe. Any air bubbles at the tip were removed.

4. The second syringe was inserted into the other end of the column. The syringe-column-syringe unit was held horizontally.
5. The plunger of the syringe containing ammonia was pushed gently and the ammonia was allowed to pass thorough the column into the opposite syringe. Pushing the ammonia back and forth between the syringes was repeated 4-5 times.
6. The syringe-column-syringe unit was laid down on a flat surface and incubated for 15 minutes at room temperature, making sure that there was ammonia solution inside the column.
7. Step 5 was repeated.
8. The ammonia solution was drawn into one of the syringes, which was then removed from the column. The ammonia was expelled into a pre-cleaned 4 mL vial, and the vial was tapped tightly with a teflon-lined screw cap.
9. Steps 2-8 were repeated four times. The ammonia solution containing the oligonucleotides eluted at each stage was pooled (total volume, 3 mL) into a vial, capped tightly and stored at -20^{0}C or deprotected.

Deprotection of oligonucleotides: The tightly closed collection vial containing the oligonucleotides was placed at 55^{0}C overnight or for 8-12 hours

Dehydration of oligonucleotides: The deprotected oligonucleotide was incubated in a waterbath at 40^{0}C for 16 hours, with the vial being capped loosely so that the ammonia evaporated, then frozen at -70^{0}C before dehydration. The frozen oligonucleotides were dehydrated overnight under vacuum in a freeze-drier.

Precipitation of oligonucleotides: The freeze-dried oligonucleotides were dissolved in 600 µL of water and 100 µL of 3 M sodium acetate (pH 5.2); these volumes were calculated from the expected 20 optical density unit (ODU) of oligonucleotides (30 µL of water and 5 µL of 3 M sodium acetate per 1 ODU of oligonucleotide products, according to the supplier's instructions). 2000 µL of ethanol was added to the oligonucleotides solutions, which was vortexed briefly. The mixture was stored at –20^{0}C for 30 minutes, then centrifuged at 14000 rpm in a microcentrifuge for 5 minutes. The supernatant was removed and the oligonucleotides pellet was washed using 70%

ethanol. Finally, the oligonucleotide pellet was dried under vacuum and dissolved in 300 µL of distilled water. The oligonucleotide concentrations were determined according to the formula: 1 A_{260} unit =32 µg DNA /mL (Sambrook *et al.*, 1989).

2.3.9 PCR amplifications

The Polymorase Chain Reactions (PCR) to isolate fragments of the *waxy* genes (sections 2.3.4-2.3.6) were carried out, using the GeneAmp kit and AmpliTaq DNA polymerase (Perkin-Elmer) and as per the supplier's instructions. *Taq* polymerase (5 U/µL) was added to the reactions, after the initial denaturation of samples at 95°C for 10 minutes. The amplifications were carried out for 35 cycles; each cycle consisting of denaturation (94°C, 1 minute), annealing (58°C, 45 seconds) and extension (72°C, 2 minutes). The annealing temperature was decreased to 55°C in some reactions such as amplification of Fragment 1 of the *wx-TsB* gene, when no PCR products were detected at 58°C annealing temperature. A typical 100 µL amplification reaction consisted of 10 µL PCR buffer (10X), 6 µL $MgCl_2$ (25 mM), 20 µL dNTP mixture (0.25 mM each of dATP, dTTP, dGTP, and dCTP), 1 µL each of the two primers (30-50 pmol/µL), 1µL genomic DNA template (1µg/µL) and 61 µL of sterile water. PCR amplifications were carried out on the DNA Thermal Cycler-480 (Perkin-Elmer) or Peltcer Thermal Cycler (PTC-200) (Bresatec).

2.3.10 Gel electrophoresis of DNA

To determine the size of PCR products, aliquots (9 µL) of the amplification reactions were mixed with 1 µL of 10X loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF and 40% sucrose) and electrophoresed on 1% agarose gels. The PCR products from primers GF and GR were electrophoresed on 3% agarose gels as the fragments were expected to have a difference of only 20-60 bp in their lengths. All gels were prepared using 1X TAE (appendix III) and 1/500 volume of the ethidium bromide (EB) stock (10 mg/mL) (Sambrook *et al.*, 1989). Plasmid DNA and the DNA products of digestion reaction thereof (section 2.3.15) were also electrophoresed using 1% agarose gels.

The gels were eletrophoresed in 1X TAE buffer at typically 80-100 V for 1-2 hours. Appropriate marker, e.g., λDNA-*Eco*RI/*Hind* III digestion (0.564 kb-21.22 kb, Progen) or PCR marker (100 bp to 2kb, Progen, Appendix III) were used for estimating the size of the tested DNA samples, as per Sambrook *et al.* (1989). The gels were viewed on a UV transilluminator and photographed using MP4 Land Camera (Polaroid) and 667 positive film (Polaroid) using a 5.6 aperture and 1/8 second exposure time.

2.3.11 Purification of PCR products

The PCR products of expected sizes were purified using two methods:

2.3.11.1 Wizard™ Minicolumns (Promega)

Wizard™ PCR Preps DNA Purification System, for rapid purification of DNA fragments (400-1000 bp), was used as per the supplier's instructions. Briefly, the following protocol was used:

1. The DNA band was excised from the agarose gels and placed into an Eppendorf tube.
2. 1 mL of resin was added into the tube and the agarose slices were incubated at 65^0C for 5 minutes.
3. The melted agarose slices were loaded into one column (provided with the System), then the column was centrifuged at 13000 rpm for 20 seconds to dry the resin.
4. The column was washed using 80% isopropanol, centrifuged and dried.
5. 30-50 μL of water was added into column and DNA was eluted by centrifuging the water containing the DNA into a sterile tube.

The purification efficiency of this method for <300 bp and >1000 bp PCR products was found to be low and the BANDPURE DNA purification kit (Progen) was utilized for these PCR products.

2.3.11.2 BANDPURE^{TM} DNA Purification kit (Progen)

The kit was used as procedures of the supplier's protocol. Briefly, the steps were as follows:

1. At least 2.5 volumes of the NaI binding buffer was added to the DNA slice, excised from the TAE agarose gel and the agarose slice was melted at 55^{0}C for 5 minutes.
2. The BANDPURE silica matrix suspension was vortexed until it was resuspended thoroughly and then 5 µL of it was added to the DNA solution. For amounts of DNA greater than 5 µg, an additional 1 µL silica matrix suspension was added for each 1 µg of DNA.
3. The melted gel-silica matrix suspension was incubated at room temperature for 5 minutes with frequent gentle mixing, then centrifuged for about 30 seconds to pellet the silica matrix with the DNA bound to it. The supernatant was discarded.
4. The pellet was washed by adding 500 µL of ice-cold ethanol wash solution to the pellet. The pellet was resuspended and centrifuged for about 30 seconds to pellet the matrix, and the ethanol wash solution was discarded. This step was repeated 3 times, retaining the pellet each time.
5. The DNA was eluted from the pellet by adding 30-50 µL of water, incubation at 55^{0}C for 5 minutes, then centrifugation for 30 seconds to collect the supernatant containing DNA.

The approximate DNA concentration was determined on agarose gels, using the intensity of one band of the 50 base or 100 base PCR markers for comparison, according to instructions of the supplier (Promega). The PCR products purified by either method were cloned into vectors (section 2.3.12) or used directly for DNA sequencing (section 2.5).

2.3.12 Ligation of PCR products into the vectors

PCR products were cloned into either of the two types of vectors, pCR-Script Amp SK(+) (Stratagene) or pGEM-T (Promega), for sequencing purposes.

2.3.12.1 pCR-Script Amp SK(+) cloning vectors

The pCR-Script cloning vectors (Stratagene) were initially used to clone the PCR products of *waxy* genes of *T. monococcum* and *T. speltoides* (section 2.3.4). The vector map is provided in Appendix I. All procedures of cloning followed the instruction manual of the supplier. The main procedures were as follows:

1. **Polishing the ends of the purified PCR products:** The ends of PCR products were polished with *Pfu* DNA polymerase to create blunt-ended DNA fragments, which facilitated cloning into pCR-Script Amp SK(+) cloning vector.

2. **Ligation into the pCR-script vectors:** PCR products were cloned into the pCR-script vectors that were predigested with the *Srf* I restriction enzyme provided with this kit. A typical 10 µL ligation reaction included 1µL pCR-Script vector (10 ng/µL), 1 µL pCR-Script 10X reaction buffer, 0.5 µL 10 mM rATP, 5.5 µL blunt-ended PCR product (10-20 ng/µL), 1 µL *Srf* I restriction enzyme (5 U/µL), 1 µL T_4 DNA ligase. The ligation reaction was performed at 4^0C for 24 hours.

3. **Transformation of the ligation products into the competent cells:** The competent cells provided with this kit were *E. coli* XL1-blue MRF′ Kan^R. The procedures for transformation of the ligation products are described in section 2.3.13.

The Fragments 1 and 4 of both the *wx-TmA* gene of *T. monococcum* and the *wx-TsB* gene of *T. speltoides* (Fig. 2.1) were cloned into this vector. However, it was found

later that this kit was not as efficient as some other vectors such as pGEM-T vectors, for the following reasons:

1. The cloning procedure required additional steps (such as the polishing reaction) and reagents (such as rATP and β-mercaptoethanol) compared to the pGEM-T vector system (Promega).

2. The kit was not cost-effective.

All further cloning work was then carried out with the pGEM-T vector system.

2.3.12.2 pGEM-T cloning vectors

The pGEM-T vectors (Promega) were prepared for cloning by adding a 3' terminal thymidine to both ends. These single 3'-T overhangs at the insertion site improve the efficiency of ligation of a PCR product into these plasmids greatly, by preventing the recircularization of the vector and providing a compatible overhang for PCR products generated by certain polymerases, such as AmpliTaq. The AmpliTaq often adds a single deoxyadenosine, in a template-independent fashion, to the 3′ ends of amplified fragments (information provided by Promega instruction manual). The pGEM-T vector map is provided in Appendix II.

The pGEM-T vectors were used to clone the partial *waxy* genes amplified by F2 and PR2 primers from the polyploid *T. turgidum-T. aestivum*, *T. timopheevi-T. zhukovskyi* lineages of wheat and some diploid wheat (sections 2.3.5 and 2.3.6, Fig. 2.2 and 2.3). The procedures for cloning were as per the technical manual provided with the kit. A typical 10 μL ligation reaction contained 1 μL pGEM-T vector (50 ng/μL), 4 μL PCR products (10-20 ng/μL), 1 μL T_4 ligase 10X buffer, 1 μL T_4 DNA ligase (5U/μL) and 3 μL deionized water. Competent JM109 cells provided with the kit were used to transform the ligation products; the procedures for transformation were as described in 2.3.14.

2.3.13 Preparation of competent *E. coli* cells

E. coli XL1-blue MRF' Kan^R strains and JM109 provided with the respective kits were also used for preparation of competent cells, by the $CaCl_2$ treatment method, as described in Sambrook *et al.* (1989). Briefly, a single colony of the *E. coli* strain of interest was inoculated into 6 mL Luria Broth (LB) medium (composition in Appendix III) and incubated at 37^0C overnight. 5 mL of this culture was transferred into 100 mL of fresh LB medium in a 500 mL flask and grown to A_{600} of 1.0-1.5. An appropriate volume of this culture was transferred into another flask and then mixed with a further 100 mL LB medium till the A_{600} reached 0.1. This diluted culture was then incubated at 37^0C under vigorously shaking (250 cycles/minute) to A_{600} of 0.6-0.8 (2-3 hours). The cells were placed on ice for 10 minutes, then harvested by centrifugation at 1500 g, 4^0C for 5 minutes. The cells were washed with 20 mL of sterile 15% glycerol and then precooled 0.1 M $CaCl_2$. The cells were then spun at 1500 g, 4^0C for 5 minutes and resuspended in 10 mL of 15% glycerol and 0.1 M $CaCl_2$ solution. The cell suspension was aliquoted immediately into 100 µL aliquots immediately and stored at -70^0C until required.

2.3.14 Transformation of ligation reaction into competent cells

The transformation of competent cells of *E. coli* XL1-blue MRF' Kan^R and JM109 was carried out as per the protocols provided with the respective kits. The competent cells stored at -70^0C were thawed on ice, mixed with the appropriate ligation reactions described in sections 2.3.12.1 and 2.3.12.2 and held for 60 minutes on ice. The cells were then heat shocked at 42^0C for 45 seconds, then allowed to recover for 5 minutes on ice before the addition of 250 µL SOC medium (Appendix III). After incubation of these cells at 37^0C for 1 hour, 100 µL of these were plated onto LB agar plates (composition in Appendix III) and incubated overnight at 37^0C. These plates contained 50 µg/mL ampicillin and 20 µL each of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (10%) and isopropylthio-β-D-galactoside (IGPT) (0.2 M) and incubated at 37^0C overnight.

2.3.15 Identification of positive clones

Both pCR-Script Amp SK(+) and pGEM-T vectors contain the *LacZ* gene, which provides α-complementation of the ß-galactosidase subunit encoded by the plasmid, with that encoded by the appropriate host cells. This allows colour selection of recombinant phagemids (colourless colonies) and nonrecombinant phagemids (blue colonies) on plates containing X-gal and IPTG. The 'white' colonies were identified as 'positives' for successful insertion of PCR products. These "positive" clones were further identified so by two methods. The plasmid DNAs from the white colonies were prepared (section 2.4), cut with restriction enzymes with cutting sites on either side of the insertion site, and the size of the insertions was determined by agarose gels as described in section 2.3.10. Some of the clones containing the correct insert size were identified further by nested PCR using the appropriate internal primers. For example, primers F2 and R1 were used as internal primers to identify Fragment 1 and primers F4 and R3 were used as internal primers to identify the inserts of Fragment 4 (Fig. 2.1, Table 2.4). The conditions for the nested PCR were the same as those described in section 2.3.4. The clones which produced the appropriate fragment sizes upon nested PCR were sequenced.

2.4 Isolation of plasmid DNA

2.4.1 Preparation of plasmid DNA

2.4.1.1 Small scale preparation

Small scale plasmid DNA preparation from the recombinant *E. coli* colonies (section 2.3.15) were carried out by the miniprep procedure (Sambrook *et al.*, 1989). Single recombinant colonies of *E. coli* were inoculated into 5 mL LB + 5 µL ampicillin (50 mg/mL) and grown at 37^0C overnight in a shaker (250 cycles/minute). Cells were then harvested and resuspended in 350 µL of STET solution (Appendix III). 25 µL of a freshly prepared solution of lysozyme (10 mg/mL in 10 mM Tris·Cl pH 8.0) was added and mixed by vortexing for 3 seconds. The tube was placed in a boiling-water bath for exactly 40 seconds and cooled on ice for 5 minutes, followed by centrifugation of the bacterial lysate at 12000 rpm for 10 minutes at room temperature. The pellet of bacterial debris was removed with a sterile toothpick and 40 µL of 2.5 M sodium

acetate (pH 5.2 or 4.8) and 420 µL of isopropanol were added to the supernatant. After mixing gently, the tube was stored for 5 minutes at room temperature. The pellet of plasmid DNA was recovered by centrifugation at 12000 rpm for 5 minutes at 4^0C. The supernatant was removed by gentle aspiration, then the DNA pellet was washed with 70% chilled ethanol and dried under vacuum. The DNA was resuspended in 50 µL dH_2O for further use.

2.4.1.2 Large scale preparation

The large scale preparations of plasmid DNA were carried out according to the following method (Sambrook *et al.*, 1989).

Single recombinant colonies of *E. coli* were grown overnight in 100 mL LB + 100 µL ampicillin (50 mg/mL) at 37^0C, in a shaker (250 cycles/minute). The cells were harvested and resuspended in 6 mL ice-cold STE buffer (Appendix III) and lysozyme was added to a final concentration of 2 mg/mL. The cells were resuspended gently and incubated for 20 minutes on ice, before the addition of 12 mL of freshly prepared solution of 0.2 M NaCl/1% SDS. After a further incubation for 10 minutes on ice, 7.5 mL of 3 M sodium acetate (pH 4.8 or 5.2) was added to the cell lysate which was then incubated on ice for 20 minutes and then centrifuged in a Beckman JA-20 rotor at 13000 rpm at 4^0C for 15 minutes. The supernatant was collected (avoiding the white precipitate of genomic DNA and cell debris) and 10 µL of RNase A (1mg/mL) was added to it, followed by incubation at 37^0C for 30 minutes. The mixture was extracted twice with phenol:chloroform:isoamyloalchohol (preparation in 25:24:1 in volume). The DNA pellet was collected by centrifugation at 10000 rpm in Beckman JA20 rotor and dissolved in 100 µL sterile H_2O for further use. The DNA sample was used for determination of DNA concentration by UV spectrometer (section 2.3.3) and further agarose gels (section 2.3.10), restriction enzyme digestion (2.3.15) or further purification by agarose gel electrophoresis (2.3.10).

2.4.1.3 Quantum Prep Plasmid Miniprep Kit

Plasmid DNA was also extracted by an alternative protocol, Quantum Prep Plasmid Miniprep Kit (Bio-Rad). 5-10 μg of plasmid DNA was obtained from 1.5 mL *E. coli* culture and this plasmid DNA was used directly for sequencing reactions. The purification solutions provided with the kit consisted of Cell Resuspension Solution, Cell Lysis Solution, Neutralization Solution and Quantum Prep Matrix. The protocol followed was as per the instruction manual provided by the supplier.

2.4.2 Recovery of the plasmid DNA from agarose gels

The DNA samples extracted on the small and large scale (sections 2.4.1.1 and 2.4.1.2) could be used for digestion with restriction enzymes, but needed to be purified further for sequencing purposes. The BANDPURE™ DNA Purification kit (Progen) was used to purify the plasmid DNAs after electrophoresing these on agarose gels; the procedures were as described in section 2.3.11.2.

2.5 Determination of DNA sequences

2.5.1 Primers for sequencing reactions

The purified PCR products and plasmid DNAs described in sections 2.3.10, 2.4.1.3 and 2.4.2 were subject to sequencing. In addition to the primers used for cloning fragments of the *waxy* gene (Table 2.4), some primers, based on the sequence of the pCR Script vector (Appendix 1) and provided in the kit, were also used in sequencing reactions. As the sequence data accumulated, additional internal primers were designed based on this data, to confirm the sequences and obtain overlaps, to allow the reconstruction of the contiguous sequences.

2.5.2 Sequencing reactions

Two kits (Perkin Elmer) were used to carry out the DNA sequencing reactions, the ABI Prism™ Dye Terminator Cycle Sequencing Ready Reaction Kit (Standard FS) and ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit. The composition of reaction mixes is given in Table 2.6. Cycle sequencing reactions were

performed in 20 μL reaction volumes for 25 cycles, each cycle consisting of step 1 (96°C, 30 seconds), step 2 (50°C, 15 seconds) and step 3 (60°C, 4 minutes), on the DNA Thermal Cycler-480 (Perkin-Elmer) and Peltcer Thermal Cycler (PTC-200) (Bresatec). All sequences were determined on an ABI 373A DNA Sequencer at DNA Sequencing Service, Department of Microbiology, Monash University, Clayton 3168, Australia.

2.5.3 Determination of sequences

Nucleotide at each site of all *waxy* genes in Chapters 4, 6 and 7 were determined by two sequence reactions. 400-600 bp sequences were obtained usually from sequencing reactions. When PCR products were used as template for sequencing reactions, sequences of each *waxy* gene fragment were determined by sequencing two independent PCR products. When plasmid DNA was used as templates, sequences of each *waxy* gene fragment were determined by sequencing DNA from two separate colonies. On the few occasions when the two sequence reactions produced two different results for a nucleotide, or where the nucleotide could not be identified clearly ("N" nucleotide), sequence data was obtained by conducting additional sequencing reactions.

Table 2.6 Composition of sequencing reactions

Composition	Standard FS kit	BigDye kit
DNA sample	30-90 ng PCR products or 300-500 ng plasmid DNA	30-90 ng PCR products or 300-500 ng plasmid DNA
Primer	1 μL (3-5 pmol)	1 μL (3-5 pmol)
*Premixed reagents	8 μL	6 μL
Total volume	20 μL	16 μL

* Premixed reagents were provided by Perkin-Elmer.

2.6 Analysis of sequences of the *waxy* genes and proteins

Sequence editing, assembly, translation and alignments were carried out by using the Australian National Genomic Information Service (ANGIS) on WebANGIS interface (Greta 1997), as described briefly below.

2.6.1 Editing and analysis of DNA sequences

The Ereverted program was used to reverse the sequence data determined by the reverse primers into the forward direction for further analysis. The Ecomposition program was used to analyse A, T, G and C composition of the DNA sequences of the *waxy* genes.

2.6.2 Comparison of two sequences

The Bestfit program was used to determine the extent of identity between two DNA sequences and the extent of identity or similarity in amino acid sequences of any two proteins. The conventional conditions for this comparison were used for gap weight 5.00, length weight 0.30, average match 1.00, and average mismatch -0.90. Any two sequences were compared by the Prettybox program to exhibit the locations of the altered nucleotides or amino acids.

2.6.3 Alignment of multiple sequences

Simultaneous alignment of DNA sequences of more than two genes (Chapter 6) or amino acid sequences of more than two proteins (Chapter 4) was performed by the Pileup program of ANGIS. The conventional conditions were gap weight 3.00 and length weight 0.10. The multiple sequence alignments were also analyzed further with the Readseq program and the Prettybox program to identify the location of variable sequences.

2.6.4 Analysis of restriction enzyme sites

DNA sequences of the *wx-TmA* gene of *T. monococcum* and *wx-TsB* gene of *T. speltoides* (Chapter 4) and the three individual *waxy* genes, *wx-W1*, *wx-W2* and *wx-W3*, of common wheat (Chapter 7), were analysed by the Mapping program to search for the sites of 6-base cutter restriction enzymes within these genes. The maps for any unique restriction enzyme sites present on these *waxy* genes were also developed.

2.6.5 Analysis of short repeated sequences

In order to analyze the role of variation of DNA sequences in different *waxy* genes during the evolution of wheat, the DNA sequences of the *waxy* genes in *T. monococcum* and *T. speltoides* (Chapter 4) were analysed by the Erepeat program. The short repeated sequences in these genes were identified by searching for >7 bp repeated sequences within every 50 bp range in the DNA sequences.

2.6.6 Translation of DNA into protein sequences

The intron/exon structure of the *wx-TmA* gene of *T. monococcum* and *wx-TsB* gene of *T. speltoides* was identified based on the AG/GT splice junctions of the plant *waxy* genes (Chapter 4). The exon sequences of the *wx-TmA* and *wx-TsB* genes were translated into putative amino acid sequences of their respective proteins, WX-TmA and WX-TsB, using the Etranslation program (Chapter 4 and Chapter 5).

2.6.7 Analyses of the deduced physical properties and structures of proteins

The putative amino acid sequences of the two waxy proteins, WX-TmA of *T. monococcum* and WX-TsB of *T. speltoides* (section 2.6.6), and those of other plant waxy proteins (published data, Chapter 5) were analysed by Pepstats program for the physical properties including molecular weight, isoelectric point and composition of amino acids (Chapter 5). The sequences of the WX-TmA and WX-TsB proteins and a waxy protein (WX-W1) deduced from a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991) were analyzed by the Peptidestructure program to predict their secondary

structures (Chapter 5). The Peptidestructure program has been constructed based on the original method by Chou and Fasman in 1978 (Greta, 1997).

2.6.8 Construction of phylogenetic trees

A family of amino acid sequences with various degrees of identity of the waxy proteins among 11 grass species was obtained (Chapter 5). The waxy protein sequences were deduced from the cDNA sequences and/or the coding sections of genomic sequences of the following: wheat (Clark *et al.*, 1991), barley (Rohde *et al.*, 1988), rice (Wang *et al.*, 1990), maize (Klösgen *et al.*, 1986), potato (van der Leij *et al.*, 1991), pea (Dry *et al.*, 1992), cassava (Salehuzzaman *et al.*, 1993), sorghum (Rosenfield *et al.*, 1996), *T. tauschii* (R. Appels, pers. commun.), and *T. monococcum* and *T. speltoides* (this project). These sequences were initially aligned using the Pileup program of the ANGIS and submitted to the Eprotdist program to calculate the extent of identity between every two sequences. Based on the analysis of the Eprotdist program, the Eneighbor program was then used to construct their phylogenetic tree. The phylogenetic relationships among the 17 partial *waxy* genes from the different genomes of wheat were constructed by a phylogenetic tree following the above procedures (Chapter 6).

2.7 Chemicals and reagents

The lists of chemicals and reagents used in this project are provided in Appendix III.

CHAPTER 3

VARIATIONS IN PROTEINS EMBEDDED WITHIN THE STARCH GRANULES IN DIFFERENT WHEAT

List of Figures and Tables in Chapter 3

Chapter 3 Variations in Proteins Embedded within the Starch Granules in Different Wheat

ABSTRACT

The investigation of proteins embedded within the starch granules in endosperm of mature seeds of various genome types of wheat showed the existence of four types of proteins. These included the 59 kDa granule-bound starch synthase (GBSSI, also called the waxy protein), the 75 kDa soluble starch synthase (SSS), the 85 kDa starch branching enzymes (SBE) and the >85 kDa starch granule proteins (SGP). The single band of the waxy protein appeared slightly variable in its molecular weight among various diploid species. The waxy proteins of the tetraploid *T. turgidum* and the hexaploid *T. aestivum* were detected as a doublet of bands, while those of the tetraploid *T. timopheevi* and the hexaploid *T. zhukovskyi* could not be separated. SSS and SBE protein profiles appeared invariable; variation, however, was found in the electrophoretic pattern of SGP in wheats. Two bands (110 kDa SGP1 and 100 kDa SGP3) were detected in the tetraploid *T. turgidum* and three bands (110 kDa SGP1, 105 kDa SGP2 and 100 kDa SGP3) were detected in the hexaploid *T. aestivum*. A 95 kDa SGP4 was observed in both tetraploid *T. timopheevi* and hexaploid *T. zhukovskyi*, in addition to the SGP1. This is the first report of a 95 kDa SGP protein in starch granules of wheat. A single band of either SGP1, SGP2, SGP3 or SGP4 was detected in the proposed diploid progenitors of polyploid wheat. Based on the electrophoretically detectable variation of both waxy proteins and SGP in diploid wheat, the origin of different genomes of wheat is discussed.

3.1 Introduction

Investigations of proteins embedded within the starch granules in the cereal endosperm have been of particular interest to wheat physiologists and geneticists, as these proteins are believed to be the enzymes involved in starch synthesis. There are four major types of proteins embedded within starch granules in endosperm of common wheat (AABBDD): the predominant 59 kDa GBSSI (or waxy protein), the 75 kDa SSS (soluble starch synthase), the 85 kDa SBE (starch branching enzymes) and the 100-105 kDa SGP (starch granule proteins) (section 1.5). These proteins have been analyzed in common wheat (Rahman *et al.*, 1995; Denyer *et al.*, 1995; Takaoka *et al.*, 1997; Yamamori *et al.*, 1998). In particular, the waxy proteins have been studied extensively in common wheat (Nakamura *et al.*, 1993; Ainsworth *et al.*, 1993; Miura *et al.*, 1994; Miura and Sugawara, 1996; Grayborsch, 1996; Hoshino *et al.*, 1996; Zhao and Sharp, 1996; Demeke *et al.*, 1997; Yasui *et al.*, 1997; Zeng *et al.*, 1997; Graybosch *et al.*, 1998; Miura *et al.*, 1998) (section 1.6). In common wheat, the WX-A protein has a slightly higher molecular weight and a more basic isoelectric point than WX-B1 and WX-D1 proteins; the WX-B1 and WX-D1 proteins have approximately the same molecular weight but a slightly different isoelectric point (Nakamura *et al.*, 1993). In the other types of wheat, Yamamori *et al.* (1995) reported that the two waxy proteins of *T. turgidum* (AABB, 2n=4x=28) corresponded to the WX-A1 and WX-B1 proteins of common wheat as was previously described by Nakamura *et al.* (1993). Taira *et al.* (1995) and Fujita *et al.* (1996) reported that only one waxy protein existed in several proposed diploid progenitors, except for *T. bicornis* and *T. sharonensis*, of the A, B and D genomes of common wheat.

All the proteins embedded within the starch granules in different genome types and ploidy levels of wheat were analyzed in this project for the following reasons:

1. The three waxy proteins in common wheat exhibit variations in electrophoretic properties, as explained above. The three waxy proteins also show variation in band intensity in SDS-PAGE, indicating quantitative differences (Nakamura *et al.*, 1993; Yamamori *et al.*, 1994) and possibly different contributions to amylose content (Miura *et al.*, 1994; Miura and Sugawara, 1996; Zhao *et al.*, 1998) (reviewed in section 1.6). It is important

that the waxy proteins could be used as protein markers for detecting variations in different genomes of wheat. Further analysis of the waxy proteins in different types of wheat will also provide clues for determining the phylogenetic relationship of wheat.

2. Except for the SGP which has been studied in common wheat, these proteins have not been reported in diploid or tetraploid wheat. Because the genes encoding these SGP have been located on chromosomes 7A, 7B and 7D of common wheat (Denyer *et al.*, 1995), it is likely that these genes could come from different diploid progenitors of wheat. Therefore, an understanding of the SGP in wheat of different ploidy levels will provide more information on their phylogenetic relationships.

3. The properties of GBSSI, SSS, SBE and SGP in the *T. timopheevi* (A^tA^tGG) -*T. zhukovskyi* ($A^tA^tA^zA^zGG$) lineage of wheat are little-known. The two types of tetraploid wheat, *T. turgidum* (AABB) and *T. timopheevi* (A^tA^tGG), offer an attractive model for the study of genome evolution and speciation in various polyploid wheat (Gill and Appels, 1988). Although a great number of efforts have been devoted to understanding the phylogenetic relationship between B and G genomes (Gill and Appels, 1988; Breiman, 1987; Dvorak and Zhang, 1990; Terachi and Tsunewaki, 1992; Tsunewaki, 1993; Miyashita *et al.*, 1994; Tsunewaki and Nakamura, 1995; Mori *et al.*, 1995; Sasanuma *et al.*, 1996; Ohsako *et al.*, 1996; Badaeva *et al.*, 1996; Mori *et al.*, 1997), the origins of B and G genomes have been in debate (as reviewed in section 1.9). The analysis of proteins embedded within the starch granules in *T. timopheevi* and *T. zhukovskyi* would thus provide new information on the identification of the different genomes.

4. Finally, investigations of waxy proteins in different varieties of wheat will help identify the 'typical' accession of different genomic types of wheat with functional *waxy* genes. Such representative of each type, exhibiting the respective typical waxy protein patterns, will then be used to clone and analyze their respective *waxy* genes for this project (Chapter 4 and Chapter 6).

3.2 Variations of waxy proteins in different types of wheat in SDS-PAGE

3.2.1 Variation of waxy proteins in diploid wheat

Fig. 3.1 shows the molecular weights of the granule-bound proteins of *T. monococcum*, *T. speltoides* and *T. tauschii* upon SDS-PAGE analysis, using the method described in section 2.2. All three diploid wheat showed a single band of size approximately 59 kDa. This band has been identified previously as the waxy protein (Ainsworth *et al.*, 1993; Nakamura *et al.*, 1993); therefore, it has been designated as the waxy protein in the present work. The molecular weight of the waxy protein of *T. monococcum* was slightly higher than that of *T. speltoides* in the SDS-PAGE. The molecular weight of the waxy protein of *T. tauschii* appeared to be very similar to that of *T. monococcum*. No variations in molecular weight were detected within the various accessions of each of the three diploid species tested.

In *T. urartu* and five proposed donors of B genome in the *Sitopsis* section, the waxy proteins exhibited one band also; however, subtle differences were detected among the various species (Fig. 3.2). The molecular weights of the waxy proteins of *T. speltoides* (lane 1), *T. longissima* (lane 3) and *T. sharonensis* (lane 5) were similar to each other, but slightly higher than those of *T. bicornis* (lane 2), and *T. searsii* (lane 4). The molecular weight of the waxy protein of *T. urartu* (lane 6) was very similar to that of *T. sharonensis*, *T. speltoides* and *T. longissima*.

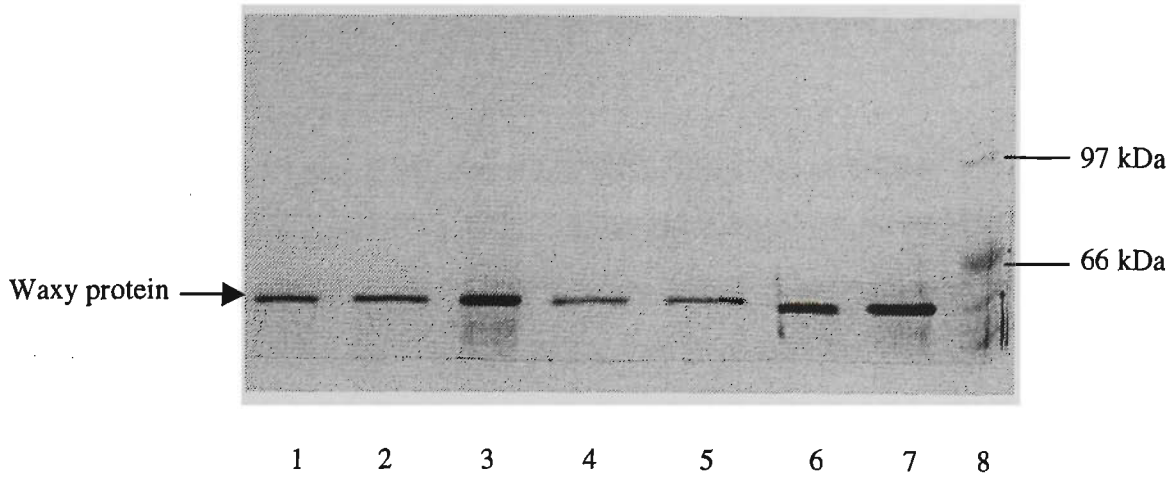


Fig. 3.1 Waxy proteins in *T. monococcum*, *T. speltoides* and *T. tauschii*. 1. *T. monococcum. Aegilopoides* var. Aus 90624; 2. *T. monococcum* var. Aus 22986; 3. *T. monococcum* var. Sile 506 Sp 3376; 4. *T. tauschii* var. Aus 18967; 5. *T. tauschii* var. Aus 18916; 6. *T. speltoides* var. Aus 21638; 7. *T. speltoides* var. Aus 18995; 8. Protein markers. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

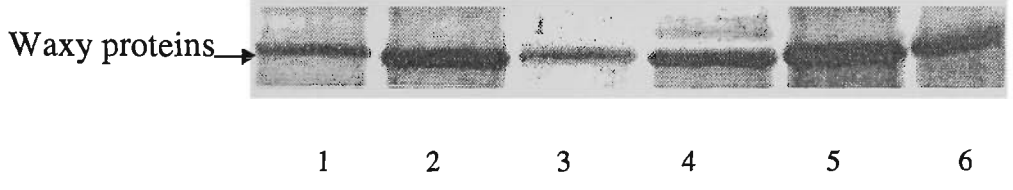


Fig. 3.2 Waxy proteins in several proposed diploid progenitors of B genome to common wheat. 1. *T. speltoides* var. Aus 21638; 2. *T. bicornis* var. Aus 18930; 3. *T. longissima* var. Aus 18798; 4. *T. searsii* var. CPI 115475; 5. *T. sharonensis* var. Aus 18747; 6. *T. urartu* var. Aus 17649. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

3.2.2 Variation of waxy proteins in the *T. turgidum-T. aestivum* lineage of wheat

T. turgidum (AABB) and *T. aestivum* (AABBDD) wheat exhibited a doublet of bands of the waxy proteins, consisting of the higher molecular weight (HMW) and the lower molecular weight (LMW) bands (Fig. 3.3), as reported in the literature (Nakamura *et al.*, 1993a). The molecular weights of the two bands in the tetraploid wheat species (lanes 1 and 2) corresponded to those observed in the hexaploid wheat species (lanes 3-6).

In *T. aestivum*, two types of varieties were evident as indicated by differences in the protein amount of protein of two bands. Varieties Chinese Spring (lane 3), Wyuna (lane 4), and Vectis (lane 5) had much more intense LMW bands, while Rosella (lane 6) had a much fainter LMW band. The HMW protein is proposed to be the WX-A1 and the LMW proteins are proposed to be a mixture of WX-B1 and WX-D1 in common wheat var. Chinese Spring (Nakamura *et al.* 1993a). Rosella has been reported to lack the WX-B1 protein (Yamamori *et al.*, 1994; Zhao and Sharp, 1996). In Rosella, the lack of the WX-B protein resulted in significant decrease in the LMW component of the protein, relative to normal wheat (Fig. 3.3). However, the LMW band of Rosella, representing the unaltered WX-D1 protein, was slightly more intense than the HMW band representing the WX-A1 protein (Fig. 3.3).

It was also noted that the two bands of the waxy proteins in the tetraploid wheat species showed different intensities, the LMW band being much more intense than the HMW band. This observation was true for both *T. turgidum* ssp. *dicoccum* (Fig. 3.3, lane 1) and *T. turgidum* ssp. *durum* (Fig. 3.3, lane 2), even though the two exhibited variations in the total amount of waxy proteins loaded on the gel.

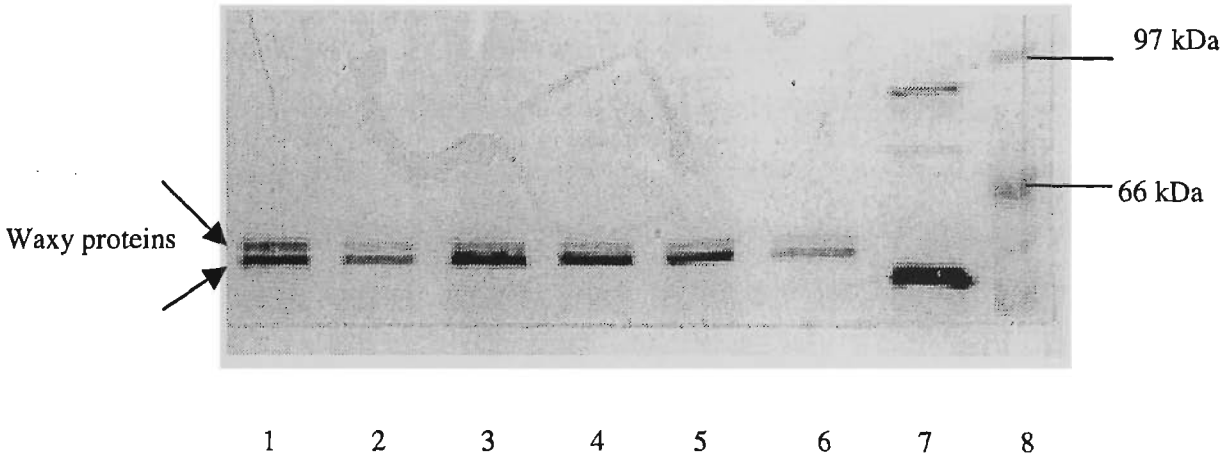


Fig. 3.3 Waxy proteins in polyploid wheat species. The doublet of bands of waxy proteins in polyploid wheat species is indicated by two arrows. 1. *T. turgidum* ssp. *dicoccum* var. Aus 21048; 2. *T. turgidum* ssp. *durum* var. Aus 19415; 3. *T. aestivum* var. Chinese Spring; 4. *T. aestivum* var. Wyuna; 5. *T. aestivum* var. Vectis; 6. *T. aestivum* var. Rosella; 7. Rice; 8. Protein markers. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

3.2.3 Variations in waxy proteins in the *T. timopheevi-T. zhukovskyi* lineage of wheat

In another group of polyploid wheat, the tetraploid *T. timopheevi* (A^tA^tGG) (Fig. 3.4, lane 4) and the hexaploid *T. zhukovskyi* ($A^tA^tA^zA^zGG$) (Fig. 3.4, lane 5), the waxy proteins were not separated into different bands while they were run on the same gel as those of *T. aestivum* (Fig. 3.4, lanes 1 and 2). Therefore, the molecular weights of the waxy proteins encoded by the genes on A^t and G genomes of *T. timopheevi* and on A^t, A^z and G genomes of *T. zhukovskyi* appear to be almost identical.

3.2.4 Simultaneous comparison of the waxy proteins in wheat of different ploidy levels

Fig. 3.5 shows the simultaneous comparison of the waxy proteins of *T. monococcum* (lane 2), *T. speltoides* (lane 3) and *T. tauschii* (lane 4) with those of *T. aestivum* (lanes 5 and 6). The waxy protein of *T. speltoides* was lower in molecular weight than those of *T. monococcum* and *T. tauschii*. The waxy protein of *T. speltoides* was more similar in its molecular weight to the LMW band of *T. aestivum* (lanes 4 and 5), while the waxy protein of *T. monococcum* was more similar in its molecular weight to the HMW band of *T. aestivum*. The molecular weight of the waxy protein of *T. tauschii* appeared similar to that of *T. monococcum*.

3.3 Analysis of the waxy proteins in diploid wheat in two dimensional SDS-PAGE

To further resolve any differences in biochemical properties of the single waxy protein bands in the various diploid wheat, the waxy proteins of *T. monococcum* and *T. tauschii* were chosen to be run in a 2D-PAGE system as described by Nakamura *et al.* (1993a) (section 2.2.4). The waxy proteins of *T. monococcum* appeared to be a complex with 5-6 subunits, 2-3 subunits more than the WX-A1 protein in *T. aestivum* reported by Nakamura *et al.* (1993a). There were 3-4 subunits in the waxy protein of *T. tauschii*, similar to the WX-D1 protein in *T. aestivum* (Nakamura *et al.*, 1993). The isoelectric points (pI) of the waxy protein subunits in *T. monococcum* were very similar to those of *T. tauschii*, in the range of pH 5.5-6.5.

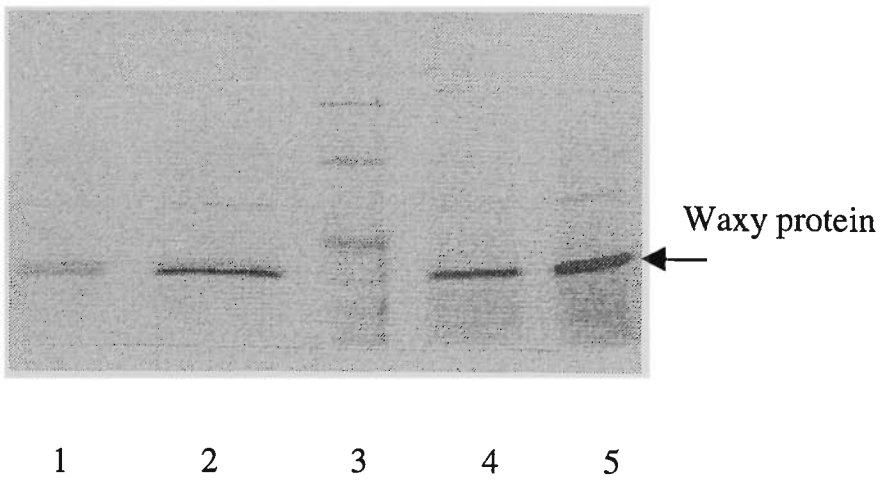


Fig. 3.4 Waxy proteins in the *T. timopheevi* and *T. zhukovskyi* lineage of wheat. 1. *T. aestivum* var. Rosella; 2. *T. aestivum* var. Chinese Spring; 3. Protein markers; 4. *T. timopheevi* var. Aus 19794; 5. *T. zhukovskyi* var. Aus 18689. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

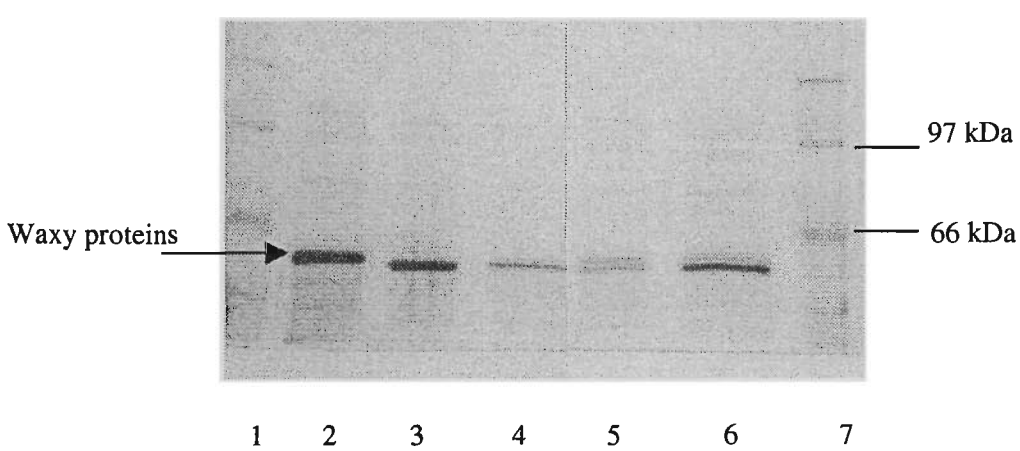


Fig. 3.5 Waxy proteins in wheat of different ploidy levels. 1. Protein markers; 2. *T. monococcum* var. Sile 506 Sp 3376; 2. *T. speltoides* var. Aus. 21638; 4. *T. tauschii* var.18916; 5. *T. aestivum* var. Rosella; 6. *T. aestivum* var. Chinese Spring; 7. Protein markers. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

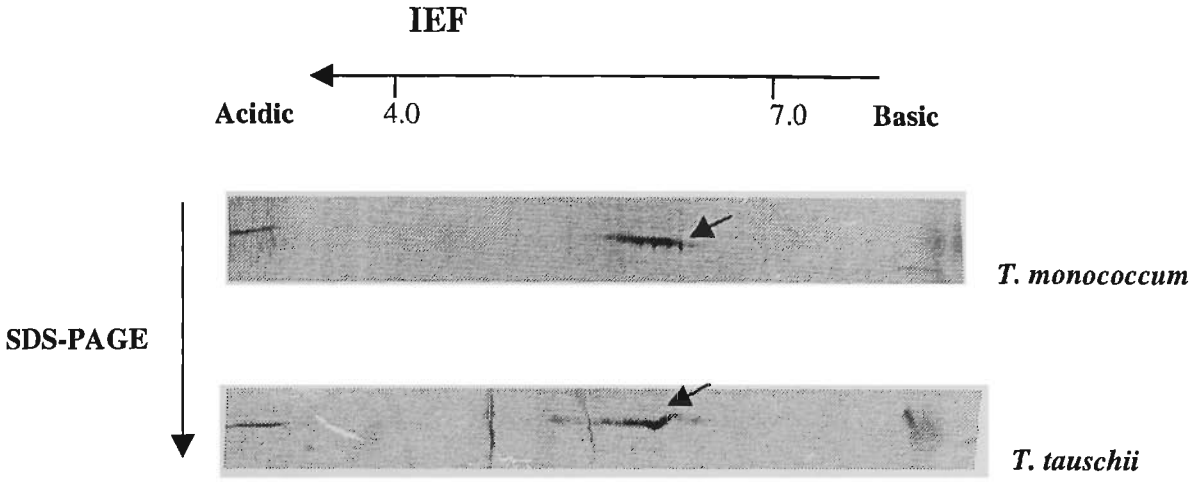


Fig. 3.6 Waxy proteins of *T. monococcum* and *T. tauschii* in 2D-SDS-PAGE. The waxy proteins of *T. monococcum* var. Aus 22986 and *T. tauschii* var. Aus 18916, denoted by arrows, were electrophoresed as described in section 2.2.4.

3.4 Identical patterns of SSS and SBE proteins in different wheat

The investigation of proteins embedded within the starch granules in the endosperm of mature seeds showed that all tested types of wheat had an approximately 75 kDa band and an approximately 85 kDa band (Fig. 3.7-3.11). These two bands have been identified as soluble starch synthase (SSS) and starch branch enzymes (SBE), respectively (Rahman *et al.*, 1995; Denyer *et al.*, 1995); the results thus indicated that these proteins are invariant in their molecular weights in different types of wheat.

3.5 Variation in SGP in different types of wheat

3.5.1 Variation in SGP in polyploid wheat

Electrophoretically detectable differences were observed in the patterns of the >85 kDa molecular weights of the starch granule proteins (SGP) among the different types of wheat (Fig. 3.7).

In the hexaploid common wheat, *T. aestivum*, the SGP were separated into three bands, corresponding to approximately 110kDa SGP1, 105 kDa SGP2 and 100 kDa SGP3, respectively (lane 1). This result is consistent with the previous reports that this group of proteins was 100-105 kDa (Denyer *et al.*, 1995) or 115 kDa, 108 kDa and 100 kDa (Yamamori, 1998). Only two bands, i.e., SGP1 and SGP3, were detected in the tetraploid *T. turgidum* ssp. *durum* (lane 2), as well as *T. turgidum* ssp. *dicoccides* and *T. turgidum* ssp. *dicoccum* (data not shown). Both the tetraploid *T. timopheevi* (lane 3) and the hexaploid *T. zhukovskyi* (lane 4) had the SGP1 seen in *T. turgidum-T. aestivum* and an additional approximately 95 kDa SGP4 protein, but did not show the SGP3 of *T. turgidum* and *T. aestivum* or SGP2 of *T. aestivum*. The 95 kDa SGP in *T. timopheevi* (lane 3) and *T. zhukovskyi* has not been reported in the previous investigations of the proteins embedded the starch granules of common wheat.

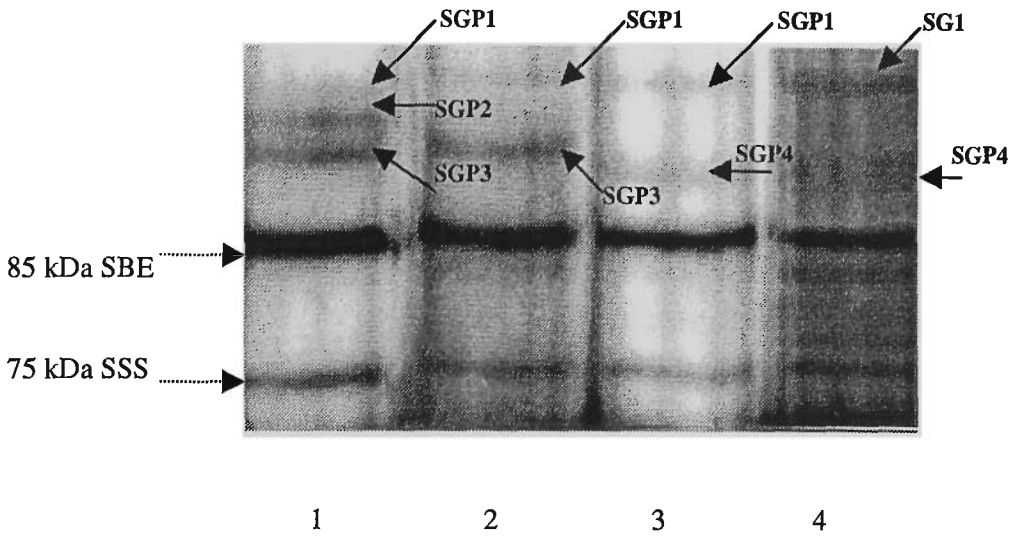


Fig. 3.7 SGP, SSS and SBE proteins in polyploid wheat. The electrophoretic patterns of higher molecular weights of (>85 kDa) granule-bound proteins (SGP) are shown: 1. *T. aestivum* var. Chinese Spring; 2. *T. turgidum* ssp. *durum* var. Aus. 764; 3. *T. timopheevi* var. Aus. 19794; 4. *T. zhukovskyi* var. Aus. 18689. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5. The SGP are denoted by arrows with solid lines. The two SGP in *T. turgidum* ssp. *durum* (110 kDa SGP1 and 100 kDa SGP3) corresponded to two of the three SGP in *T. aestivum* (110 kDa SGP1, 105 kDa SGP2 and 100 kDa SGP3) reported by Denyer *et al.* (1995). In addition to the 110 kDa SGP1, an approximately 95 kDa SGP4 was detected in *T. timopheevi* and *T. zhukovskyi.*

3.5.2 Variation in SGP in diploid wheat

A single protein band of 110 kDa SGP1, 105 kDa SGP2, 100 kDa SGP3 or 95 kDa SGP4 was detected in all proposed diploid progenitors of A, B, D or G genomes (Fig. 3.8-3.11).

3.5.2.1 *T. monococcum*

By screening fifteen varieties of *T. monococcum*, it was found that there were two distinct groups of *T. monococcum*. The majority of accessions belonged to Type I, which contained a single SGP4 (Fig. 3.8, lane 2). A single accession formed Type II, which contained a SGP1 (Fig. 3.8, lane 1).

3.5.2.2 *T. speltoides*

The five accessions of *T. speltoides* investigated had three different size classes of SGP (Fig. 3.9). Type I of *T. speltoides* had a SGP4 (lane 2). This band was not very clear in the photo due to loss of resolution during reproduction of the Figure, but it could be detected in the original SDS-PAGE gel. Type II of *T. speltoides* had a SGP3 (lane 1) and Type III of *T. speltoides* had a SGP1 (lane 3).

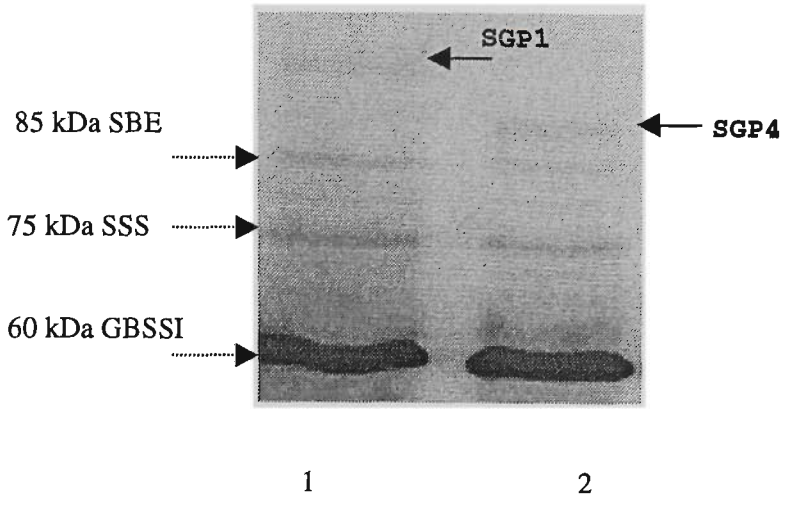


Fig. 3.8 SGP in *T. monococcum*. 1. *T. monococcum* var. Sile 506 Sp 3376; 2. *T. monococcum* var. Aus 22986. The starch granules and proteins were extracted as described in section 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

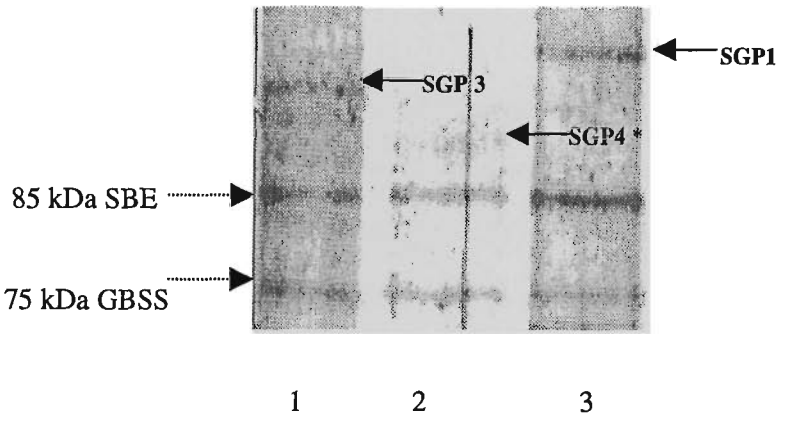


Fig. 3.9 SGP in *T. speltoides*. 1. *T. speltoides* var. Aus 21638; 2. *T. speltoides* var. Aus 18995; 3. *T. speltoides* var. Aus. 18942. *In lane 2, a faint of SGP4 band was detected in the gel but lost during reproduction of the photo. The starch granules and proteins were extracted as described in sections 2.2.1 and 2.2.2. The electrophoretic conditions were as in section 2.2.3.2 and staining procedures were as in section 2.2.5.

3.5.2.3 *T. tauschii*

T. tauschii had a unique protein, the SGP2 (Fig. 3.10, lane 1), among all of the tested diploid wheat. The molecular weight of this protein was intermediate to that of the 110 kDa SGP1 and 100 kDa SGP3 of *T. turgidum* and it has been designated as the 105 kDa SGP2. SGP1 of *T. turgidum* (lane 2) could not be detected clearly in this photo due to loss of resolution of the fainter band during reproduction of the Figure. The three SGP of *T. aestivum* (lane 3) could not be detected in the Figure for the same reason (these can be seen in Fig. 3.7). Lane 4 and lane 5 represent SGP1 and SGP4 of type II and type I of *T. monococcum* respectively, as mentioned previously (section 3.6.2.1).

3.5.2.4 Other diploid species

Fig. 3.11 shows the SGP pattern in *T. urartu* and five diploid wheat species in the *Sitopsis* section. All of these had only one band in the >85 kDa region, with 110 kDa SGP1 or 100 kDa SGP3 (Fig. 3.11). *T. searsii* exhibited an approximately 110 kDa thick doublet (lane 4).

The variation in SGP detected in the different accessions and species of wheat is summarized in Table 3.1. Overall, there were four major size classes of SGP, i.e., 95 kDa, 100 kDa, 105 kDa or 110 kDa, in different diploid wheat. Variation in the molecular weights of SGP was detected within species of the three diploid wheat, *T. monococcum*, *T. urartu* and *T. speltoides*. Furthermore, variation in sizes and number of bands of SGP was detected in different genomes of polyploid wheat.

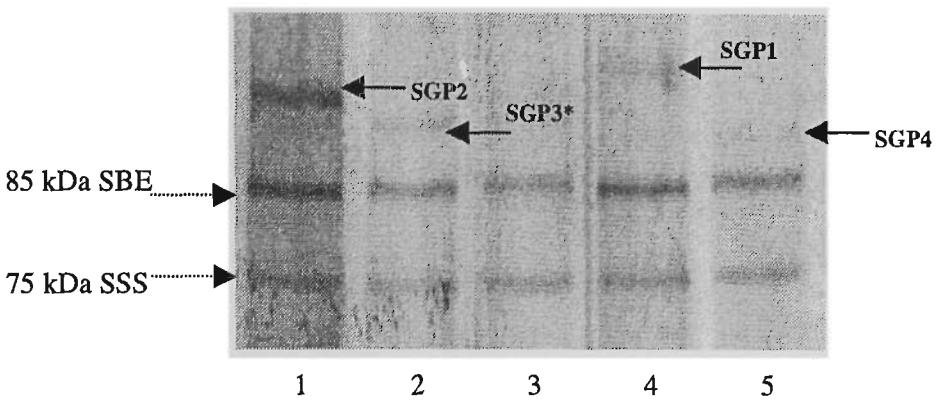


Fig. 3.10 SGP in *T. tauschii*. 1. *T. tauschii* var. CPI 799; 2. *T. turgidum* ssp. *durum* var. Aus. 19415; 3. *T. aestivum* var. Chinese Spring; 4. *T. monococcum* ar. Aus. Sile 506 Sp 3376; 5. *T. monococcum* var. 22986.

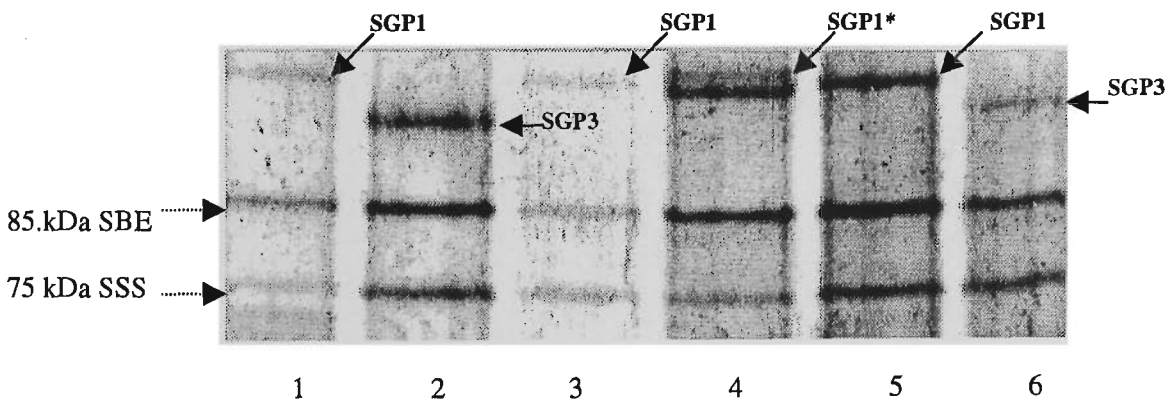


Fig. 3.11 SGP in diploid wheat. 1. *T. speltoides* var. Aus 21638); 2. *T. bicornis* var. Aus. 18930; 3. *T. longissima* var. Aus. 18798; 4. *T. searsii* var. CPI 115475; 5. *T. sharonensis* var. Aus. 18747; 6. *T. urartu* var. 90459. All of these diploid species had only one 110 kDa SGP1 or 100 kDa SGP3. * *T. searsii* exhibited an approximately 110 kDa thick doublet; the origin of an additional band is uncertain.

Table 3.1 Profiles of the SGP in diploid, tetraploid and hexaploid wheat

Species	Size of SGP				Accessions tested
	95 kDa SGP4	100kDa SGP3	105 kDa SGP2	110 kDa SGP1	
T. moncoccum, type I	+	-	-	-	14*
T. moncoccum, type II	-	-	-	+	Sile 506 Sp 3376
T. urartu, type I	-	+	-	-	Aus.17649
T. urartu, type II	-	-	-	+	Aus.90459
T. bicornis	-	+	-	-	Aus.18795, Aus.18930
T. longissima	-	+	-	+	Aus.18797, Aus.18798
T. searsii	-	+	-	+	CPI 799
T. sharonensis	-	-	-	+	Aus.18802, Aus.18747
T. speltoides, Type I	+	-	-	-	Aus.18995
T. speltoides, Type II	-	+	-	-	Aus.21638
T. speltoides, Type III	-	-	-	+	Aus.18942
T. tauschii	-	-	+	-	Aus. 18967, Aus. 18916, CPI 799
T. turgidum	-	+	-	+	Aus. 19785, Aus. 19592 Aus. 764
T. aestivum	-	+	+	+	Chinese Spring, Rosella, Wyuna
T. timopheevi	+	-	-	+	Aus. 19794, Aus. 19796
T .zhukovskyi	+	-	-	+	Aus. 18689, Aus. 18690

"+" and "-" indicate the presence or absence of that particular SGP, respectively. All protein samples were prepared from mature seeds. Sources of seeds are shown in Table 2.1 of Chapter 2. * The fourteen varieties of *T. monococcum* are: C68-73; C68-18; C68-19; C68 24G 16; Sile 500 Sp 3750; Aus. 16280; Aus 3701; Aus. 16278; Aus. 15361; Aus. 22983; Aus. 22986; Aus. 90442; Aus. 90624 and Aus. 95051.

3.6 Discussion

3.6.1 The electrophoretic profiles of proteins embedded within starch granules of wheat

In the papers published so far (Rahman *et al.*, 1995; Denyer *et al.*, 1995; Takaoka *et al.*, 1997; Yamamori *et al.*, 1998), the proteins embedded within the starch granules in the endosperm of common wheat have been divided into four major classes: 59 kDa GBSSI (waxy protein), 75 kDa SSS, 85 kDa SBE and >85 kDa SGP. Our observation in different types of wheat (e.g., Fig. 3.8) is consistent with these previous reports. However, the profiles of these proteins appeared variable in wheat of different ploidy levels. The 75 kDa SSS and 85 kDa SBE appeared to have identical molecular weights among all tested wheat of different ploidy levels; however, the 59 kDa GBSSI and >85 kDa SGP exhibited variations in their molecular weights. These variations could be useful for understanding of roles of different GBSSI and SGP and origins of different genomes of wheat.

3.6.2 Variations in waxy proteins among wheat of different ploidy levels

There was only one band of the waxy protein, of molecular weight approximately 59 kDa, in all tested diploid wheat. The tetraploid group of *T. turgidum* including the four subspecies (ssp. *turgidum*, ssp. *dicoccoides*, ssp. *durum* and ssp. *dicoccum*), and the hexaploid group of *T. aestivum*, contained two waxy protein bands with slightly different molecular weights. These results are consistent with a number of previous reports (Nakamura *et al.*, 1993; Yamamori *et al.*, 1994; Taira *et al.*, 1995). However, other interesting observations on waxy proteins in different wheat were also made in this study.

3.6.2.1 Variations in waxy proteins in diploid progenitors

On the basis of the band size of the waxy proteins, the five species in the *Sitopsis* section and *T. urartu*, which are all proposed to be donors of B genomes to common polyploid wheat (section 1.8), could be divided into two types (Fig. 3.2). One type included *T. speltoides*, *T. longissima*, *T. sharonensis* and *T. urartu*, and had slightly higher molecular weights for the waxy proteins, compared to the other type which

included *T. bicornis* and *T. searsii*. The waxy protein of *T. speltoides* had a molecular weight similar to that of the WX-B protein of *T. turgidum-T. aestivum* (Fig. 3.5). Taira *et al.* (1995) reported that the molecular weights of waxy proteins in *T. speltoides*, *T. longissima*, *T. searsii and T. urartu* were 59.0 kDa, 59.2 kDa, 59.1 kDa and 59.4 kDa, respectively in SDS-PAGE. In their report the molecular weight of the waxy protein in *T. speltoides* was the lowest among the four proposed donors of B genomes tested. However, our results showed that the protein size of *T. speltoides* was slightly higher than that of *T. searsii* and similar to that of *T. urartu*. Our results also showed the size of waxy proteins of *T. monococcum* and *T. tauschii* was very similar (Fig. 3.1). These comparisons indicate that interspecies variation might exist in the molecular weights of the waxy proteins in diploid species of wheat.

3.6.2.2 Possible mutations in the *waxy* gene during evolution from diploid to polyploid wheat

Our observations have indicated that the waxy protein in *T. monococcum* was not detectably different in size from that of *T. tauschii* on the one-dimensional SDS-PAGE and that the mobility of the waxy protein of *T. monococcum* or *T. tauschii* did not correspond to that of the WX-A1 or WX-D1 protein in *T. aestivum*. In the two-dimensional gel, the waxy protein in *T. monococcum* (Fig. 3.6) also had more subunits than those reported for the WX-A1 protein of *T. aestivum* reported by Nakamura *et al.* (1993). In contrast, the number of subunits of the waxy proteins observed for *T. tauschii* was very similar to those of the WX-D1 protein in *T. aestivum* (Nakamura *et al.*, 1993). These observations indicate that there might be some differences in the electrophoretic properties of the waxy protein of *T. monococcum* and the WX-A1 protein of *T. aestivum*. These differences could be explained by two possibilities. Firstly, mutations could have occurred in the *waxy* gene on A genome during the evolution of wheat from a diploid through tetraploid to hexaploid genome, if *T. monococcum* was the donor of A genome to *T. turgidum*. The other possibility is that another diploid, *T. urartu*, rather than *T. monococcum*, might be the donor of A genome to the polyploid *T. turgidum-T. aestivum* lineage of wheat. Sequence comparison of the genes of various diploid wheat with the individual *waxy* genes in polyploid wheat will be carried out to address these possibilities in Chapter 6.

3.6.2.3 A possible diphyletic origin of *T. turgidum* and *T. timopheevi*

The two waxy protein bands in the tetraploid *T. turgidum* (AABB) could be distinguished readily (Fig. 3.3), however, only one band of the waxy protein was detected in *T. timopheevi* (A^tA^tGG) in one-dimentional SDS-PAGE (Fig. 3.5). This difference in molecular weight of waxy proteins between *T. turgidum* and *T. timopheevi* might be due to their independent lineages from different diploid progenitors and thus slight biochemical differences. This observation agrees with the previous reports that there was a diphyletic origin of *T. turgidum* and *T. timopheevi* (Kihara, 1963; Tsunewaki *et al.*, 1980, 1993; Tsunewaki and Nakamaru, 1995; Mori *et al.*, 1995, 1997). However, a number of reports have indicated that *T. turgidum* and *T. timopheevi* originated from a common tetraploid wheat (Johnson, 1975; Tanaka *et al.*, 1978; Dvorak and Appels, 1982; Chen and Gill, 1983; Dvorak and Zhang, 1990, Sasanuma *et al.*, 1996; Breiman, 1987; Terachi and Tsunewaki, 1992; Tsunewaki, 1993; Miyashita *et al.*, 1994; Ohsako *et al.*, 1996; Badaeva *et al.*, 1996). The electrophoretic difference in the waxy proteins between *T. turgidum* and *T. timopheevi* observed here is insufficient by itself to unequivocally support one theory or the other, as the differences could be due to mutation in their *waxy* genes and/or different origins of their genomes. In order to provide more information for distinguish between the genomes of these tetraploid wheat and extend our understanding on evolution of these wheat species, DNA sequences of the partial *waxy* genes from their individual genomes will be analysed in Chapter 6.

3.6.2.4 Variation of expression in different *waxy* proteins in the *T. turgidum-T. aestivum* lineage

The two waxy proteins of tetraploid *T. turgidum* ssp. *dicoccum* exhibited significantly different densities in SDS-PAGE, the LMW band of WX-B1 showing a much more intense band than the HMW band of WX-A1 (Fig. 3.3). This was also observed in *T. turgidum* ssp. *durum* (Fig. 3.3) and *T. turgidum* ssp. *turgidum* (Fig. 3.4). This result is consistent with the previous report of Yamamori *et al.* (1995a); however, in our results, the difference in the intensity of the two waxy proteins in these different tetraploid wheat appeared more marked than that in the report of Yamamori *et al.*(1995). In common wheat, the WX-B1 is reported to have stronger intensity than

the WX-D1 and WX-A1 on SDS-PAGE, as described in section 1.5.2. Two possible reasons have been proposed for this observation; the WX-B1 protein could be more extractable than the WX-A1 and WX-D1 proteins, or it might be produced more that the WX-A1 protein (Yamamori *et al.*, 1995). However, other studies have shown that the WX-B1 protein had a higher expression than WX-A1 and WX-D1 proteins, and thus makes more contribution to amylose content in common wheat (Miura *et al.*, 1994; Miura and Sugawara, 1996). The molecular basis of such differences in expression of the different waxy proteins in wheat is largely unknown. In rice, the expression of the *waxy* gene controlling amylose content is regulated by temperature during grain development (Sano *et al.*, 1985). Unique microsatellite alleles containing a single-nucleotide polymorphism in the untranslated region of the *waxy* gene resulted in variations in the apparent amylose content in rice (Ayres *et al.*, 1997). In maize, it was reported that the waxy protein was not produced or its amount was reduced, in the event of insertion of a controlling element in the *waxy* gene (Echt and Schwartz, 1981). The molecular mechanisms of the unequal expression of the WX-A1, WX-B1 and WX-D1 proteins in common wheat need to be investigated further, to test these possibilities as well as others, such as unequal turnover rates of these proteins, unequal stabilities of their mRNA, or unequal promoter activities. Some of these will be discussed further in section 5.7.5.

3.6.3 Properties of SGP in different ploidy wheat

In view of the potential importance of the SGP in starch production, as reviewed in section 1.4, the SGP in different types of wheat also have been investigated extensively in the present study.

3.6.3.1 A novel 95 kDa protein embedded within starch granules

The electrophoretic patterns of the proteins embedded within starch granules of common wheat observed in this study were identical to those described Denyer *et al.* (1995) and Yamamori *et al.* (1998). The sizes of the three SGP in common wheat are reported to be 100-105 kDa (Rahman *et al.*, 1995; Denyer *et al.*, 1995), or 100 kDa, 108 kDa and 115 kDa (Yamamori *et al.*, 1998). The sizes of the three SGP in common wheat were estimated to be 110 kDa SGP1, 105 kDa SGP2 and 100 kDa

SGP3, in the present studies, although the precise sizes cannot be determined by electrophoretic gels.

In the >85 kDa region, only a 95 kDa SGP4 was detected in type I of *T. monococcum* (Fig. 3.8) and type I of *T. speltoides* (Fig. 3.9), while all other SGP of *T. monococcum* and *T. speltoides* and other diploid species contained a single band 110 kDa SGP1, 105 kDa SGP2 or 100 kDa SGP3. The tetraploid *T. timopheevi* and hexaploid *T. zhukovskyi* also contained the SGP4. The 95 kDa SGP4 has not been reported so far; thus it is a novel protein in wheat. It is possible that the SGP4 belongs to the same family as the SGP1, SGP2 and SGP3, based on their similar locations in the starch granules. This needs to be investigated further by Western blotting using an anti-SGP antibody.

3.6.3.2 The phylogenetic relationship of genomes of common wheat based on analysis of SGP

Denyer *et al.* (1995) analysed SGP in 20 nullisomic-tetrasomic (NT) lines of Chinese Spring (*T. aestivum*). The results showed that the110 kDa SGP1, 105 kDa SGP2 and 100 kDa SGP3 were encoded by genes located on chromosomes 7A, 7D and 7B, respectively, of common wheat. When this data and our own observations are combined, it would suggest that information on the origins of A, B and D genomes of common wheat may be obtained by comparing the patterns of SGP in wheat of different ploidy levels.

The tetraploid *T. turgidum* (AABB) exhibited two bands of SGP, 110 kDa SGP1 and 100 kDa SGP3 (Fig. 3.7). *T. tauschii* (DD) had a unique 105 kDa SGP2 among all the tested diploid wheat (Fig. 3.10). A hypothetical hybrid of SGP between *T. turgidum* and *T. tauschii* would correspond precisely to the three SGP (110 kDa, 105 kDa and 100 kDa) in common wheat (AABBDD). This observation strongly suggests that the SGP2 would be coded by the D genome of common wheat, and that *T. turgidum* and *T. tauschii* could be the tetraploid and diploid donors, respectively, to common wheat; this suggestion thus supports previous studies by Kerby and Kuspira (1987). By extending this line of reasoning further, it appears that the SGP1 and SGP3 should be encoded by the genes on A genome and genome B respectively.

However, the size of SGP appeared to have further intraspecies and interspecies variation. SGP3 appeared not only in *T. bicornis* and one accession of *T. speltoides*, the two proposed donors of the B genome, but also in one accession of *T. urartu* which has been considered recently to be a donor of the A genome (Breiman and Graur, 1995; Friebe and Gill, 1996; Devorak, 1998). SGP1 appeared not only in the proposed donor of A genomes, one accession of both *T. monococcum* and *T. urartu*, the proposed donors of the genome, but also in the proposed donors of B genome, *T. longissima*, *T. searsii*, *T. sharonensis* and one accession of *T. speltoides*. These observations suggest that the phylogenetic relationship of A and B genomes in polyploid wheat is much more complicated than suggested above.

3.6.3.3 Profiles of SGP also support a diphyletic origin of *T. turgidum* and *T. timopheevi*

The tetraploid *T. timopheevi* (A^tA^tGG) exhibited two bands: SGP1 and SGP4 (Fig. 3.7, lane 2), different from *T. turgidum* (AABB) which exhibited SGP1 and SGP3 (Fig. 3.7, lane 4). This observation is consistent with our results on the waxy proteins in the two tetraploid wheat (Fig. 3.4); that is, *T. turgidum* and *T. timopheevi* might be two independent lineages. However, the locations of the two genes encoding SGP1 and SGP4 on specific genomes in *T. timopheevi* are unclear. It cannot be determined which diploid species contributed the genes encoding SGP1 and SGP4 to *T. timopheevi*, for the above reasons given above that the variations of the SGP occurred within and between its possible diploid progenitor species.

3.6.4 Methods for investigation of proteins embedded within starch granules

Although the waxy proteins and the three SGP of common wheat have been analysed here, slightly variable results have been obtained by different groups of researchers, as reviewed in section 1.5. A number of reasons can explain these variations. Firstly, the gels were prepared by using different ratios of acrylamide to bisacrylamide, i.e., 29:1 (Schofield and Greenwell, 1987; Yamamori *et al.*, 1992; Ainsworth *et al.*, 1993; Rahman *et al.*, 1995), 30:0.135 (Nakamura *et al.*, 1993a; Yamamori *et al.*, 1994), 37:1 (Denyer *et al.*, 1995) or 30:0.1 (Zhao and Sharp, 1996); or the gels were made using a different pH buffer system (Zhao *et al.*, 1996). Secondly, the gels could have been run

under different conditions such as voltages and time and stained by different methods such Coomassie Brilliant Blue or silver stains. Different staining times were also used. Finally, different plant materials, such as endosperm from mature or developing seeds, were used for different studies. The present study provides a direct, thorough and valid comparison of all waxy proteins and SGP from mature seeds of common wheat and all types of its relatives, by using the same ratio of acrylamide:bisacrylamide (30:0.135) for all gel, electrophoresing for similar length of time and staining by the same methods.

In most experiments conducted by other researchers, the starch granules were extracted according to the method described by Echt and Schwartz (1981). The protein samples in this study were prepared following the same method. However, the protocol for the preparation of proteins in this project (section 2.2.2) was based on the method of Nakamura *et al.* (1993a), in which proteinase K was not used to remove any possible proteins left on the surface of the granules after preparation of the starch granules. This possibly resulted in a very faint protein band of approximately 80 kDa in some of the samples (Fig. 3.8) and some other proteins besides GBSSI, SSS, SBE and SGP, which were detected occasionally (Fig. 3.7, lane 4).

3.6.5 Cloning of genes encoding waxy proteins.

Characterization of waxy proteins and SGP in the different types of wheat allowed us to choose appropriate plant material to clone the genes encoding these proteins from diploid species. The fact that SGP is likely to play an important role in starch biosynthesis in wheat led to further work on the genes encoding these proteins, at the collaborating laboratory at CSIRO (R. Appels, pers. commun.).

Characterization of the waxy proteins of the different diploid progenitors allowed us to clone the single *waxy* A gene from *T. monococcum* (Aus. 22986) and the single *waxy* B gene from *T. speltoides* (Aus. 21638) for molecular studies in this project. Because investigation of SGP was carried out after the cloning of the *waxy* genes had been initiated, the *T. monococcum* and *T. speltoides* used for cloning the *waxy* genes were later found to contain 95 kDa SGP4 and 100 kDa SGP3 respectively. As a part of a collaborative effort to clone the *waxy* genes in the three possible diploid

progenitors to common wheat, the *waxy* gene of the proposed progenitor of the D genome of common wheat, *T. tauschii* (DD), was cloned in Professor R. Appels's laboratory, Plant Industry, CSIRO. Based on the sequence data of the *waxy* genes obtained for *T. monococcum* and *T. speltoides* (Chapter 4), a set of primers was designed to clone and characterize partial *waxy* genes from all other proposed diploid progenitors (except for *T. searsii*) and the polyploid wheat species further (Chapter 6), to analyze the phylogenetic relationships of different genomes of wheat.

CHAPTER 4

CLONING AND CHARACTERISATION OF THE *WAXY* GENES OF *T. monococcum* AND *T. speltoides*

List of Figures and Tables in Chapter 4

Chapter 4 Cloning and Characterisation of the *waxy* Gene of *T. monococcum* and *T. speltoides*

ABSTRACT

The genes encoding the *waxy* proteins have been isolated from two diploid wheat species, *T. monococcum* (AA) and *T. speltoides* (BB), by the Polymerase Chain Reaction (PCR) approach. The length of the *wx-TmA* and *wx-TsB* genes cloned was 2834 bp and 2826 bp, respectively, which covered the region from initiation codon to termination codon, including the 11 translated exons and 10 intervening introns, and the partial untranslated leader. Multiple sequence alignment indicated that the *wx-TmA* gene of *T. monococcum* and *wx-TsB* of *T. speltoides* had the same exon/intron structure as the barley *waxy* gene (Rohde *et al.*, 1988). The two diploid wheat *waxy* genes had identical lengths for the respective exons and shared approximately 95% identity to each other in exon sequences. The majority of the introns, however, appeared significantly variable in length and sequences, which mainly resulted from different lengths (1-18 bp) of insertions/deletions and from accumulation of 7-9 bp repeated DNA sequences. The striking observation was that no exon of either of these two diploid genes had a particular 33 bp sequence coding for an additional 11 amino acids reported from a cDNA of common wheat. Instead, there was a high identity between this sequence and that at the 3' end of intron 4 of the *waxy* genes of the two diploid wheat. The sequence information of the *wx-TmA* and *wx-TsB* identified conserved sequences as well as variable sections and was used to assist in the design of primers for cloning the individual *waxy* genes in polyploid wheat (Chapter 6).

4.1 Introduction

The results of SDS-PAGE analysis of waxy proteins in Chapter 3 (section 3.2.1.1) indicated that there was only one approximately 59 kDa waxy protein in all proposed diploid progenitors of wheat, in comparison to the two waxy proteins in tetraploid wheat and three waxy proteins in hexaploid wheat. These results suggested that there was only one *waxy* locus in diploid wheat, as expected from data on other diploid cereals, maize (Shure *et al.*, 1983), rice (Wang *et al.*, 1990) and barley (Rohde *et al.*, 1988).

The waxy proteins encoded by the different genomes of wheat showed some variations in molecular weight on SDS-PAGE gels (Chapter 3). However, the common wheat (*T. aestivum*, AABBDD) consists of a triplicate set of the *waxy* genes which are homoeologous and probably very similar to each other in their sequences; it is thus almost impossible to clone and characterise easily and efficiently the three individual *waxy* genes using the published data from other plant *waxy* genes. It is thus necessary to clone the individual genes from the diploid progenitors containing different genomes to obtain preliminary results on their sequences identities and differences. It is incorrect to assume that the diploid genomes would remain unchanged during polyploidisation and evolution; however, one would expect to see some fundamental, albeit minor, differences in the various genomes of different diploid and polyploid wheat.

On the basis of the results obtained on the comparison of the waxy proteins in different wheat, no definite conclusion for the origins of A and B genomes of common wheat could be drawn as discussed in chapter 3 (sections 3.3.2.2 and 3.3.3.2). One accession each from *T. monococcum* (AA) and *T. speltoides* (BB) was, therefore, chosen for cloning their respective *waxy* genes. As a part of a collaborative effort to clone the *waxy* genes in the three possible diploid progenitors to common wheat, the *waxy* gene of the proposed progenitor of the D genome of common wheat, *T. tauschii* (DD), was cloned in Professor R. Appels's laboratory, Plant Industry, CSIRO, Canberra, Australia.

The *waxy* genes have been isolated from diploid plant species such as barley (Rohde *et al.*, 1988), rice (Wang *et al.*, 1990; Okagaki, 1992), maize (Shure *et al.*, 1983; Klösgen *et al.*, 1986), potato (Hovenkamp-Hermelink *et al.*, 1987; Visser *et al.*, 1989; van der Leij *et al.*, 1991) and pea (Dry *et al.*, 1992). A cDNA of the *waxy* gene in common wheat was reported (Clark *et al.*, 1991) and two other clones have been described recently in Dr R. Chibbar's laboratory, Plant Biotechnology Institute, Sasktchewan, Canada (sequence unpublished, Matus *et al.*, 1998). Analysis of the published data shows that the amino acid sequences of plant waxy proteins are highly conserved (section 2.3.1). On the basis of the conserved domains in these sequences, four pairs of primers were designed to amplify and clone the complete translated region of the *waxy* genes in *T. monococcum* (*wx-TmA*) and *T. speltoides* (*wx-TsB*) by the Polymerase Chain Reaction (PCR) approach in this study (section 2.3.3). The amino acid sequences of the respectively, complete waxy proteins deduced from these two waxy genes have been obtained, their predicted biochemical and physical properties will be described and analyzed in Chapter 5.

4.2 Isolation of the *waxy* gene of *T. monococcum* (*wx-TmA*) and *T. speltoides* (*wx-TsB*)

4.2.1 PCR products of the *wx-TmA* and *wx-TsB* genes

As described in section 2.3.4, the *waxy* genes of *T. monococcum* and *T. speltoides* were amplified by designing four pairs of primers (section 2.3.4.2, Table 2.4; section 2.3.4.3, Fig. 2.1), cloned and/or sequenced. These fragments are designated as Fragment 1 (amplification product of primers F1 and R1), Fragment 2 (amplification product of primers F2 and R2), Fragment 3 (amplification product of primers F3 and R3) and Fragment 4 (amplification product of primers F4 and R4), as described in section 2.3.4.2 (Table 2.5). These four fragments overlapped and covered the partial untranslated region and all translated exons and the intervening introns, as shown in Fig 2.1 (section 2.3.4.3).

4.2.2 Cloning of PCR products

The expected sizes of PCR products for the four fragments of the *waxy* genes of *T. monococcum* and *T. speltoides* were obtained from different reactions (Fig. 4.1A). Fragments 1 and 4 of the *waxy* genes of the two diploid wheat were cloned into pCR-Script Amp SK(+) vectors (section 2.3.12.1) (Fig. 4.1B). Both PCR products for Fragments 2 and 3 of both genes and cloned inserts for Fragments 1 and 4 of both genes were sequenced (section 2.5).

4.2.3 Determination for sequences of the *wx-TmA* and *wx-TsB* genes

The sequencing results (raw data not shown) indicated that Fragments 1, 2, 3, and 4 of the *T. monococcum waxy* gene (*wx-TmA*) were 931 bp, 759 bp, 881 bp and 656 bp in length respectively and that the lengths of these four fragments of the *T. speltoides waxy* gene (*wx-TsB*) were 938 bp, 759 bp, 877 bp and 645 bp respectively. Fragments 1, 2 and 3 of *T. monococcum* and *T. speltoides* were very similar in size to those predicted for the barley *waxy* gene (Table 4.1). However, Fragment 4 of *T. monococcum* and *T. speltoides* both exhibited significantly different sizes in comparison to the corresponding region of the barley *waxy* gene (Rohde *et al.*, 1988).

4.2.4 Differences in sequences of the original primers and the corresponding regions amplified from the *wx-TmA* and *wx-TsB* genes

Some mismatches were observed between sequences of some of the original primers and sequences of the amplified regions (Table 4.2). No such difference was found in sequences in the F1, R2 or R4 primer region. In the case of the *wx-TmA* gene, one nucleotide mismatch was detected in R3 and F4 primer regions. In case of the *wx-TsB* gene, one nucleotide was different between the amplified regions and their primer regions, F2, F3 and F4. Surprisingly, four out of the 23 nucleotides of the original primer R1 were different from the amplified sequences of *wx-TsB*.

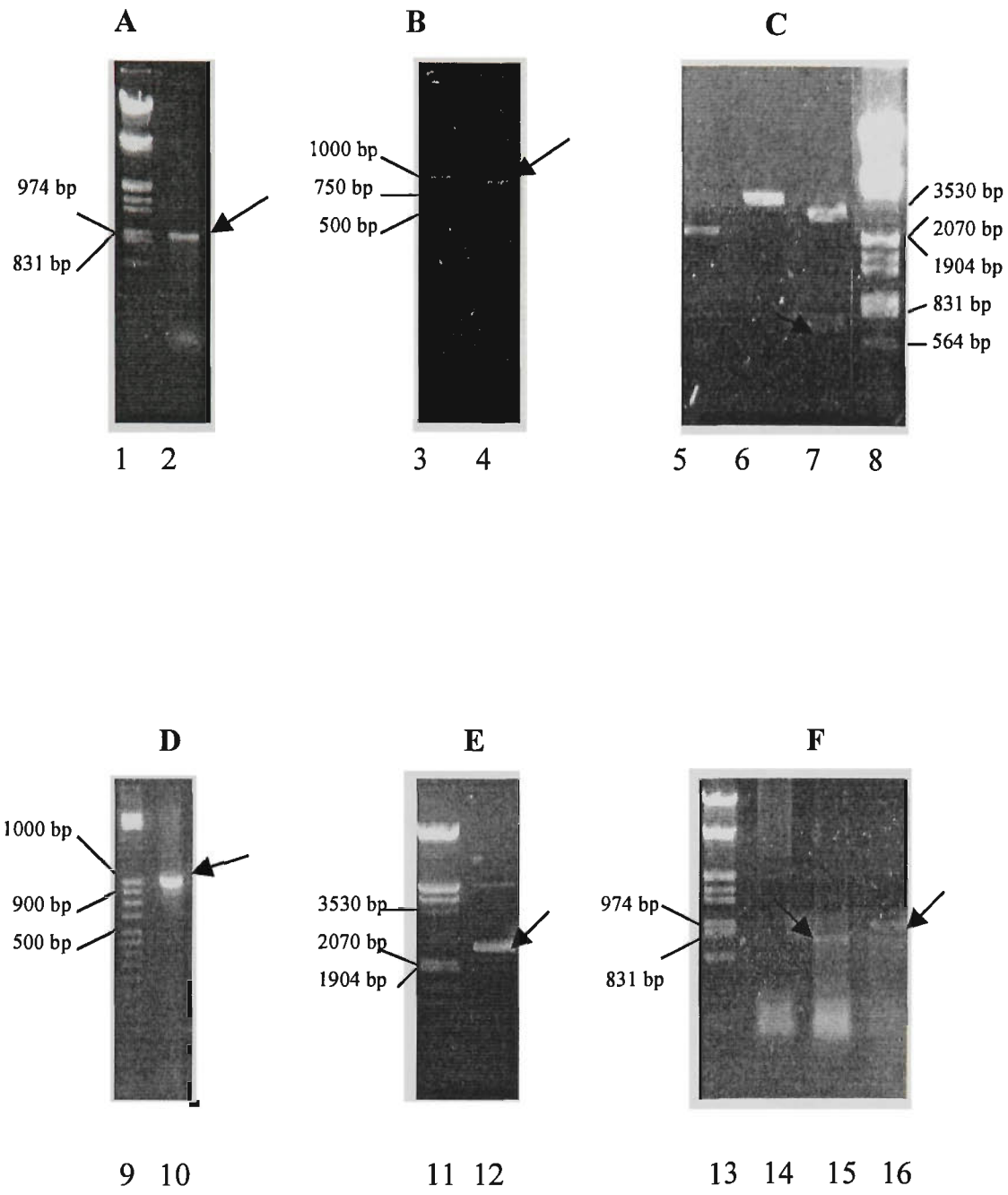


Fig. 4.1 Representative gels of the PCR products and clones of the fragments of the *waxy* genes of *T. monococcum* and *T. speltoides*. **A:** 1. DNA marker (λ-DNA-EcoRI/Hind III, Progen); 2. Fragment 1 of an approximate 930 PCR product from *T.*

Fig. 4.1 (*continued*)

monococcum; **B:** 3. PCR marker (Promega); 4. The purified PCR product from lane 2 of **A,** using Wizard minicolumn (Promega); **C:** 5. Uncut recombinant pCR-Script Amp SK(+) vectors containing an insert of Fragment 4 of the *waxy* gene of *T. monococcum*; 6. The recombinant plasmid DNA (approximately 3.6 kb) in lane 5, linearised with *Sac*I; 7. Double digestion of the recombinant plasmid in lane 5 with *Sac*I and *Cla*I, showing the 2961 bp vector and an approximate 650 bp insert of fragment 4 of the *waxy* gene of *T. monococcum* (shown by the arrow); 8. DNA markers (λ-DNA-EcoRI/Hind III, Progen); **D:** 9. DNA markers (50 bp ladder, MBI); 10. An approximate 940 bp of PCR product of Fragment 1 of the *waxy* gene from *T. speltoides*; **E:** 11. DNA markers (λ-DNA-EcoRI/Hind III, Progen); 12. Uncut recombinant pCR-Script Amp SK(+) vector containing an insert of approximate 940 bp of Fragment 1 of the *waxy* gene from *T. speltoides*; **F:** 13. DNA markers (λ-DNA-EcoRI/Hind III, Progen); 14. Negative control for PCR. 15. PCR product of Fragment 2 (approximate 760 bp) of the *waxy* gene of *T. speltoides*; 16. PCR product of Fragment 3 (approximate 880 bp) of the *waxy* gene of *T. speltoides*.

4.3 Sequences of the *wx-TmA* and *wx-TsB* genes

The sequence data of the four overlapping fragments of the *T. monococcum* and *T. speltoides waxy* genes were used to construct contigs of the *wx-TmA* sequence (Fig. 4.2) and *wx-TsB* sequence (Fig. 4.3), respectively. The lengths of these *wx-TmA* and *wx-TsB* genes were 2834 bp and 2826 bp, respectively. By comparison with the nucleotide sequence of the *waxy* barley gene (Rohde *et al.*, 1988), the translation initiation codon ATG and the termination codon TGA of the *wx-TmA* and *wx-TsB* genes were found to be at sites homologous to that of the barley *waxy* gene, as expected. Therefore, the *wx-TmA* gene cloned here consisted of a 2803 bp from initiation codon to termination codon and a 31 bp untranslated leader (Fig. 4.2), and the *wx-TsB* gene consisted of 2795 bp from initiation codon to termination codon and 31 bp of untranslated leader (Fig. 4.3). The barley *waxy* gene included 34 bp of the 5' untranslated leader and 2794 bp of the corresponding region to the *wx-TmA* and *wx-TsB*.

Table 4.1 Comparison of lengths (bp) of fragments from the *waxy* genes of *T. monococcum* and *T. speltoides*

Fragments	Primer pairs	*wx-TmA* (bp)	*wx-TsB* (bp)	Expected length for *waxy* gene of barley (bp)
1	F1-R1	931	938	933
2	F2-R2	759	759	777
3	F3-R3	881	877	896
4	F4-R4	656	645	614

The four amplified fragments of the *waxy* genes *T. monococcum* and *T. speltoides* were very similar in length to the corresponding sections of the barley *waxy* gene (Rohde *et al.*, 1988).

Table 4.2 Comparison of sequences of the original primers and the corresponding regions amplified from *wx-TmA* and *wx-TsB* genes

Primer sequences	*wx-TmA* sequences (5'→3')
F1 5'-TTGCTGCAGGTAGCCACAC-3'	TTGCTGCAGGTAGCCACAC
*R1 5'-TTCTCTGCCAGGCAGCACTTGAG-3'	TTCTCTGCCAGGCAGCACTTGAG
F2 5'-GACCAAGGAGAAGATCTATGG-3'	GACCAAGGAGAAGATCTATGG
*R2 5'-GGCATCGTCAACGGCATG-3'	GGCATCGTCAACGGCATG
F3 5'-ATCAACTGGATGAAGGCC-3'	ATCAACTGGATGAAGGCC
*R3 5'-CGGGCTCGTCGACACTAT-3'	CGGGCTCGTCGACAC**C**AT
F4 5'-AGCTCCAGGGAATGCGCTAC-3'	AGCTCCAGGG**G**ATGCGCTAC
*R4 5'-AGAACGTCGCCGCTCCCTGA-3'	AGAACGTCGCCGCTCCCTGA

Primer sequences	*wx-TsB* sequences (5'→3')
F1 5'-TTGCTGCAGGTAGCCACAC-3'	TTGCTGCAGGTAGCCACAC
*R1 5'-TTCTCTGCCAGGCAGCACTTGAG-3'	T**C**CTCTGCCAGGCAGC**G**CT**G**GA**A**
F2 5'-GACCAAGGAGAAGATCTATGG-3'	GACCAAGGAGAAGATCTA**C**GG
*R2 5-GGCATCGTCAACGGCATG-3'	GGCATCGTCAACGGCATG
F3 5'-ATCAACTGGATGAAGGCC-3'	ATCAACTGGATGAAGGC**T**
*R3 5-CGGGCTCGTCGACACTAT-3'	CGGGCTCGTCGACACTAT
F4 5'-AGCTCCAGGGAATGCGCTAC-3'	AGCTCCAGGG**G**ATGCGCTAC
*R4 5'-AGAACGTCGCCGCTCCCTGA-3'	AGAACGTCGCCGCTCCCTGA

Primer sequences were designed based on the cDNA of common wheat (Clark *et al.*, 1991). The nucleotides in *wx-TmA* and *wx-TsB* genes that were observed to be different from the original primer sequences are shown in bold print and underlined. * Indicates that the sequences of the reverse primers (Table 2.4 in Chapter 2) have been converted into the forward direction here, to allow easy comparison with the *waxy* gene sequences.

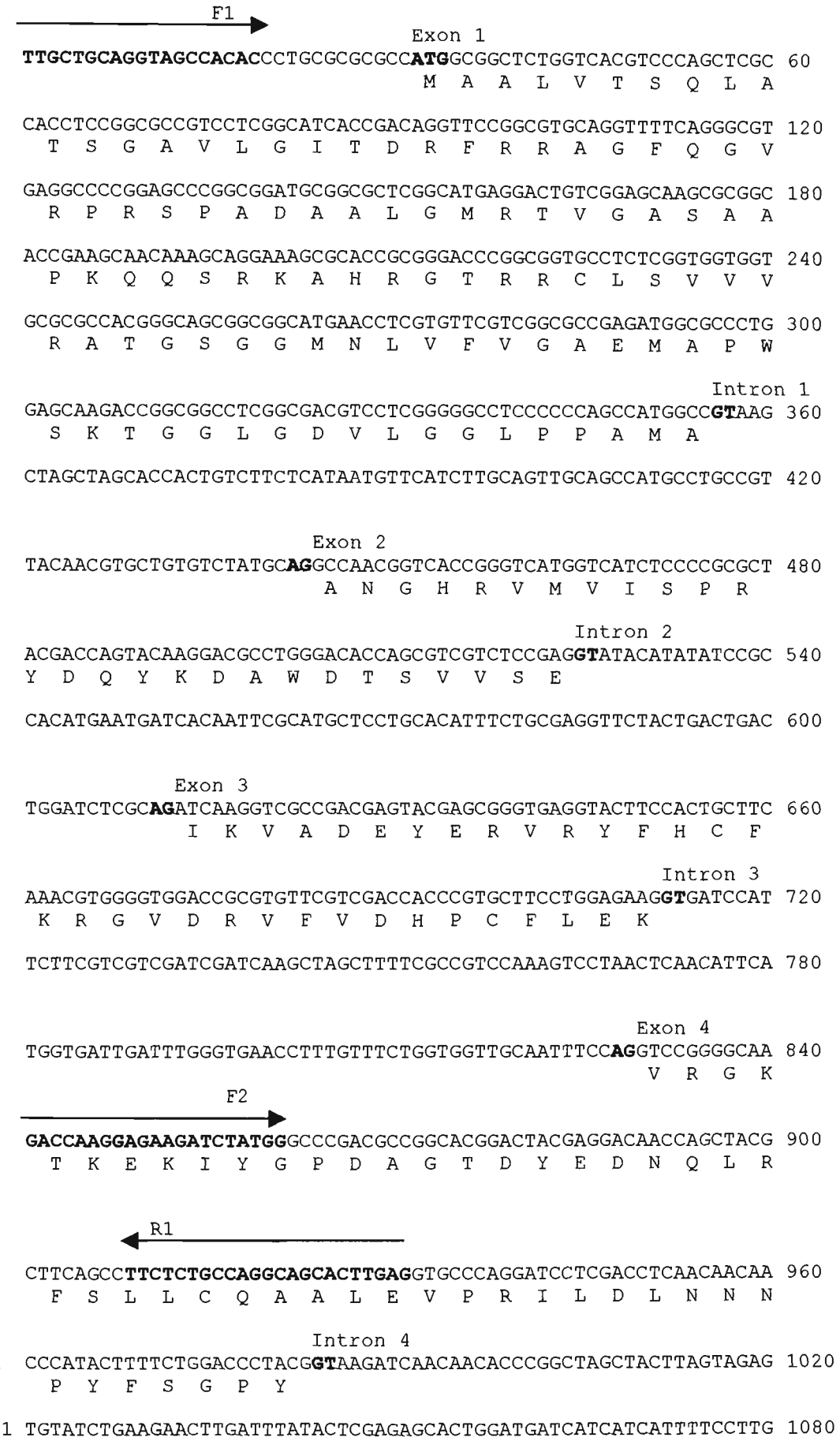


Fig. 4.2 *to be continued*

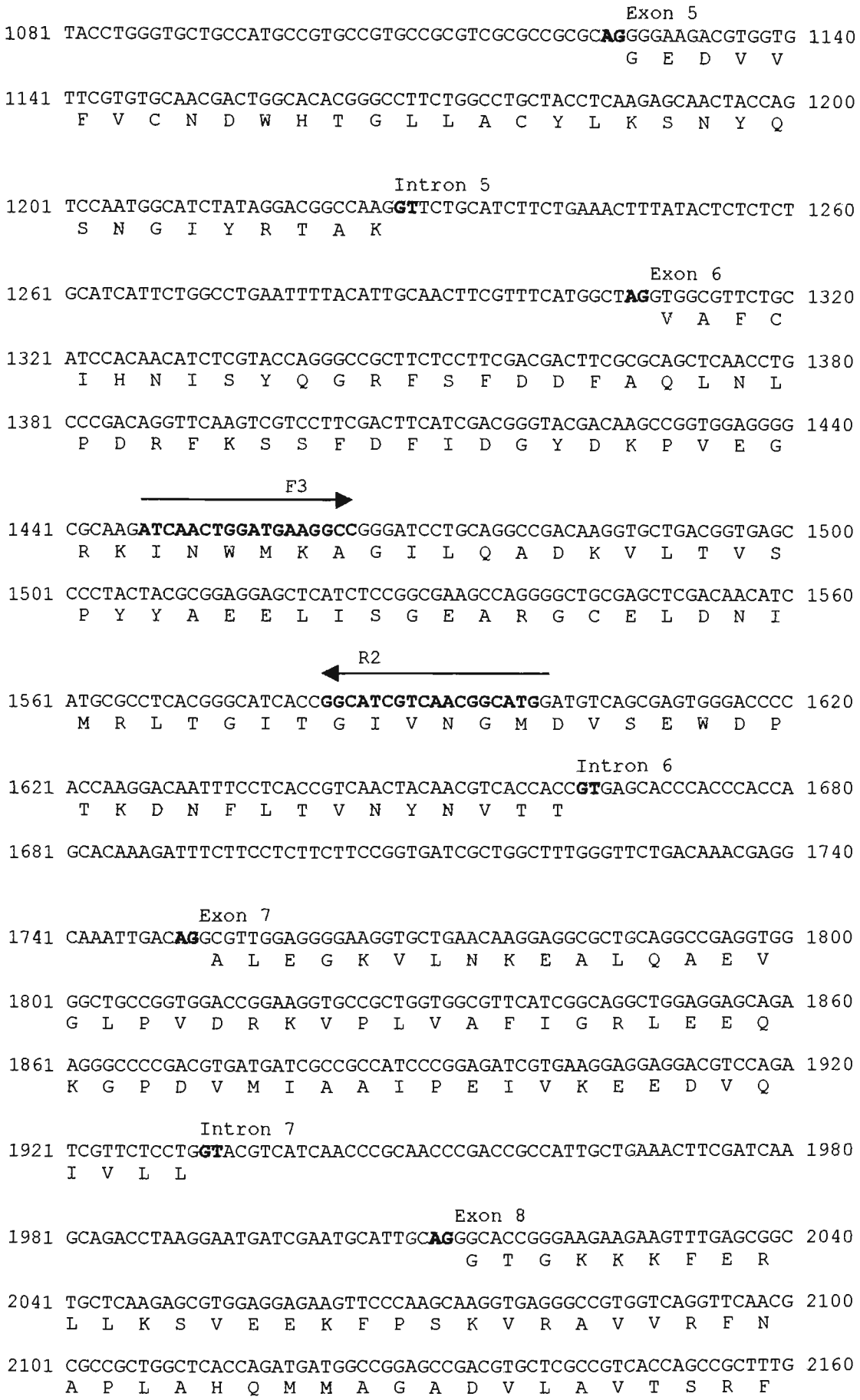


Fig. 4.2 *to be continued*

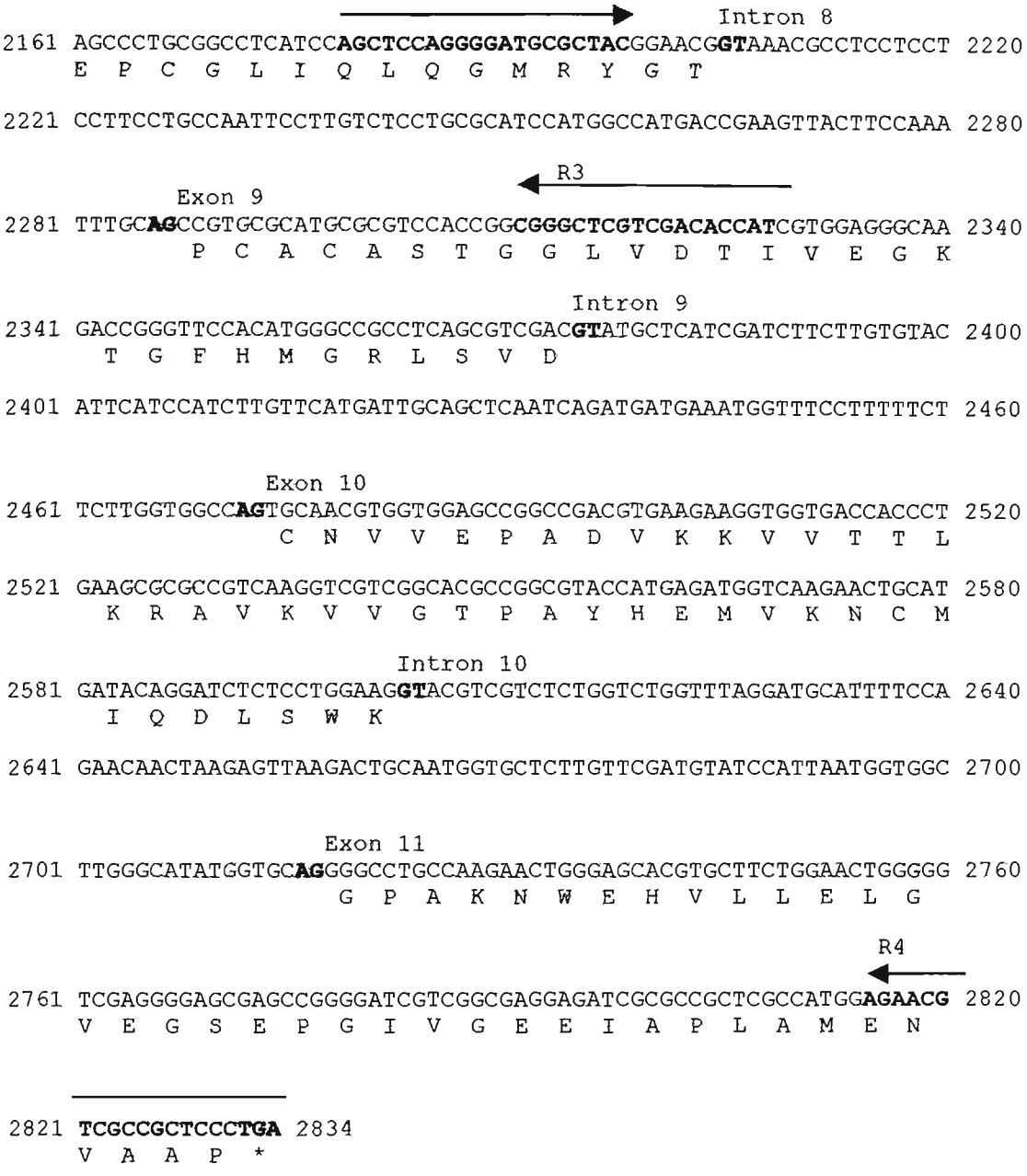


Fig. 4.2 Sequence of the *wx-TmA* gene of *T. monococcum*. The *waxy* gene of *T. monococcum* was amplified by using four pairs of primers (Fig. 2.3.4.2). Locations of these primers are indicated by arrows and sequences in the *wx-TmA* gene corresponding to these primer sequences are shown in bold. The amplified DNA fragments were sequenced and used for the construction of the configure of the *wx-TmA* gene. The initiation and termination codons and the GT/AG spice junctions of introns are shown in bold and shaded; these were deduced from the sequence of the *waxy* gene in barley (Rohde *et al.*, 1988). The amino acids deduced from the exon sequences are shown below these.

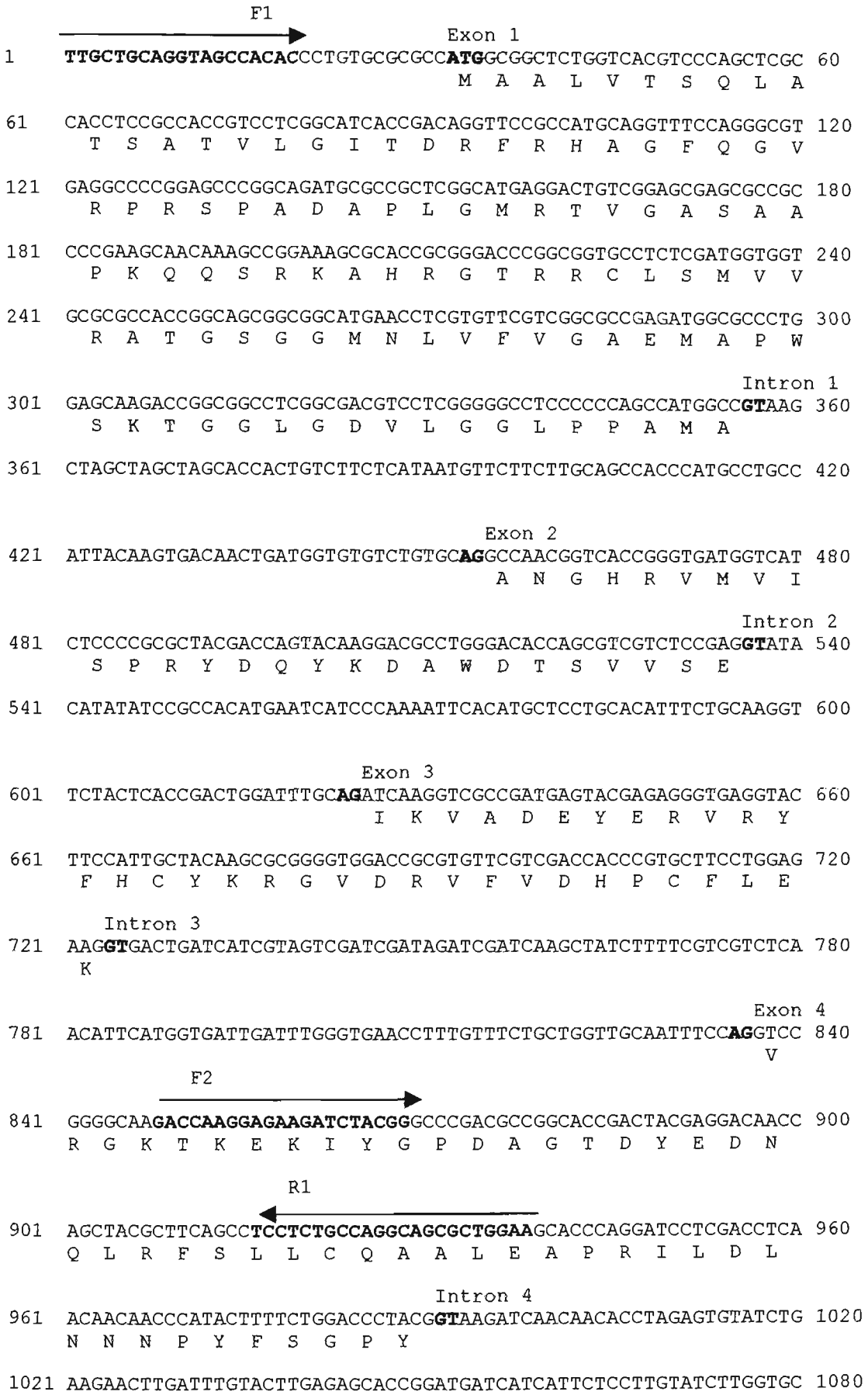


Fig. 4.3 *to be continued*

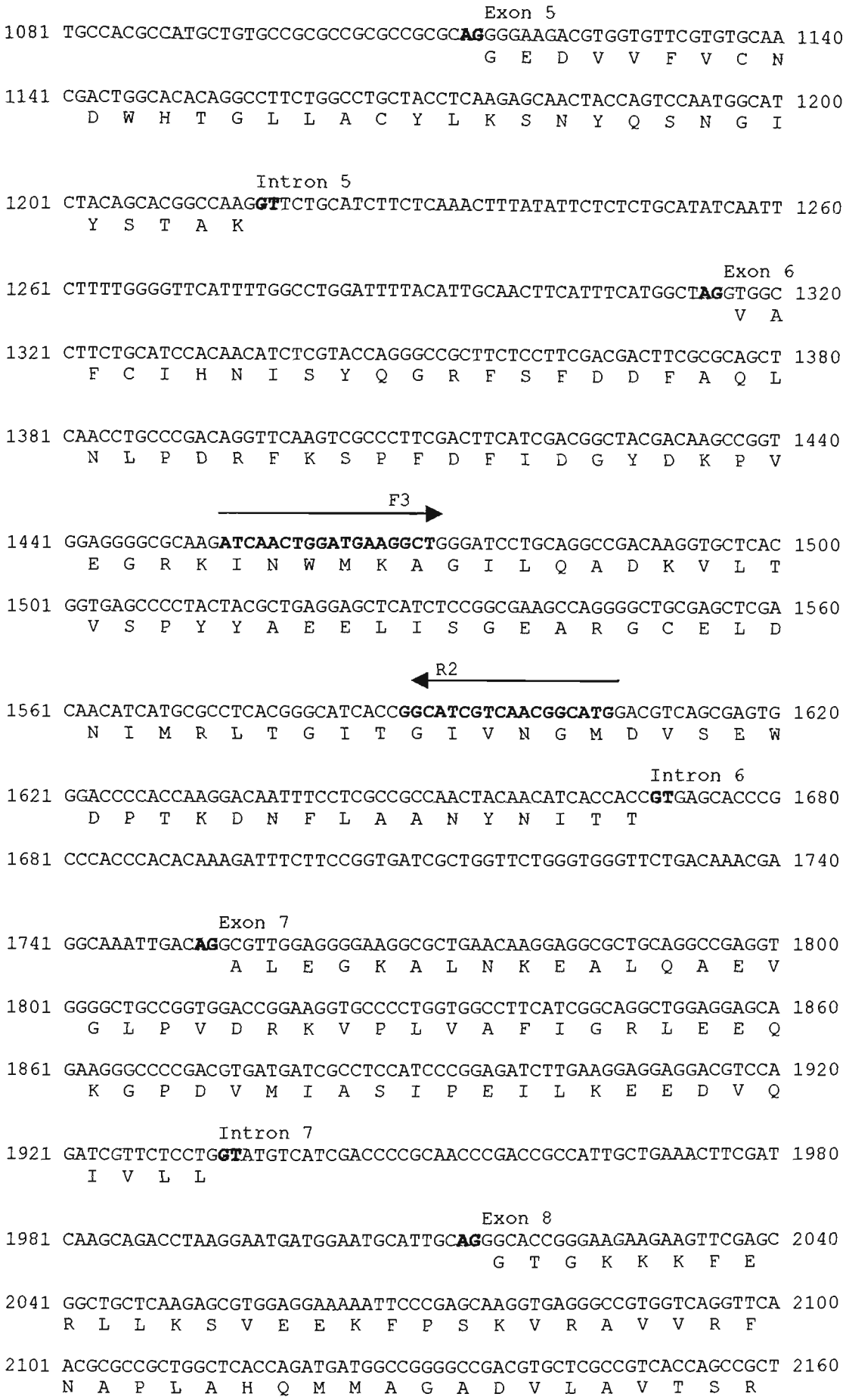


Fig. 4.3 *to be continued*

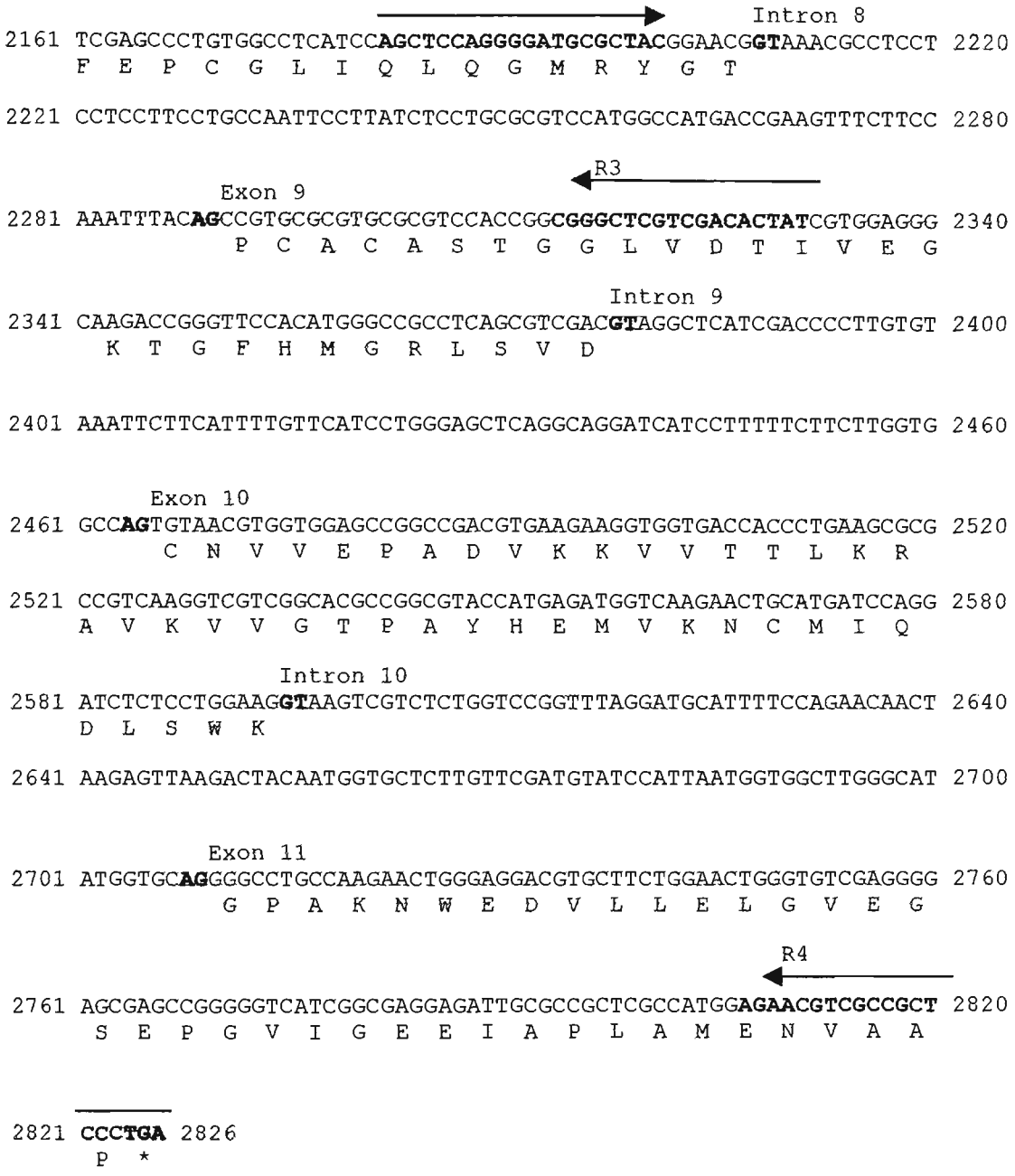


Fig. 4.3 Sequence of the *wx-TsB* gene of *T. speltoides*. The *waxy* gene of *T. speltoides* was amplified by using four pairs of primers (Fig. 2.3.4.2). Locations of these primers are indicated by arrows and sequences in the *wx-TsB* gene corresponding to these primer sequences are shown in bold. The amplified DNA fragments were sequenced and used for the construction of the configure of the *wx-TsB* gene. The initiation and termination codons and the GT/AG spice junctions of introns are shown in bold and shaded; these were deduced from the sequences of the *waxy* gene in barley (Rohde *et al.*, 1988). The amino acids deduced from the exon sequences are shown below these.

4.4 Properties of the *wx-TmA* and *wx-TsB* genes

4.4.1 Structures of the *wx-TmA* and *wx-TsB* genes

As described in section 2.3.4 (Fig. 2.1), the work described in this Chapter was aimed at cloning all 11 translated exons and thus the intervening 10 introns of the *wx-TmA* and *wx-TsB* genes. Alignment of the sequence data obtained for these two genes with the barley waxy gene (Rohde *et al.*, 1988) indicated that the *wx-TmA* and *wx-TsB* gene sequences aligned perfectly with 'exon 1' onwards section, as expected from the experimental design (section 2.3.4.3). As explained in section 2.3.4.3, this first translated exon of the barley gene (exon 2 in Rohde *et al.*' sequence, 1988) has been designed as Exon 1 (E1) throughout this project for convenience. The alignment indicated that the translation initiation codon (ATG), termination codon (TGA) and the GT/AG splice junctions of each exon and intron of the barley *waxy* gene, exon 1 onwards, were highly conserved and present at the corresponding positions in the *wx-TmA* and *-TsB* genes. Therefore, the *wx-TmA* and *wx-TsB* genes consisted of 11 exons and 10 introns, from the first translated exon containing the translation initiation codon (ATG) to the termination codon (TGA); this structure was identical to that in the corresponding section of the barley *waxy* gene (Fig. 4.4). However, multiple sequence alignment indicated that the *waxy* genes in maize (Klösgen *et al.*, 1986), rice (Wang *et al.* 1990) and potato (van der Leij *et al.*, 1991) consisted of 13 exons and 12 introns in the corresponding region (Fig. 4.4). This resulted from the fact that exon 4 (154 bp) is split into two 90 bp and 64 bp sections, while exon 6 (354 bp) is split into 110 bp and 254 bp sections by two additional introns in these genes.

It should be noted that on the basis of the alignment of the genomic *waxy* gene of barley (Rohde *et al.*, 1988) with the barley cDNA sequence in the GenBank, the AG of intron 2 appears to be positioned incorrectly by Rohde *et al.* (1988), at positions 1442-1443 instead of 1410-1411. The position of the AG of intron 2 was conserved in all published *waxy* genes in maize (Klösgen *et al.*, 1986), rice (Wang *et al.* 1990) and potato (van der Leij *et al.*, 1991).

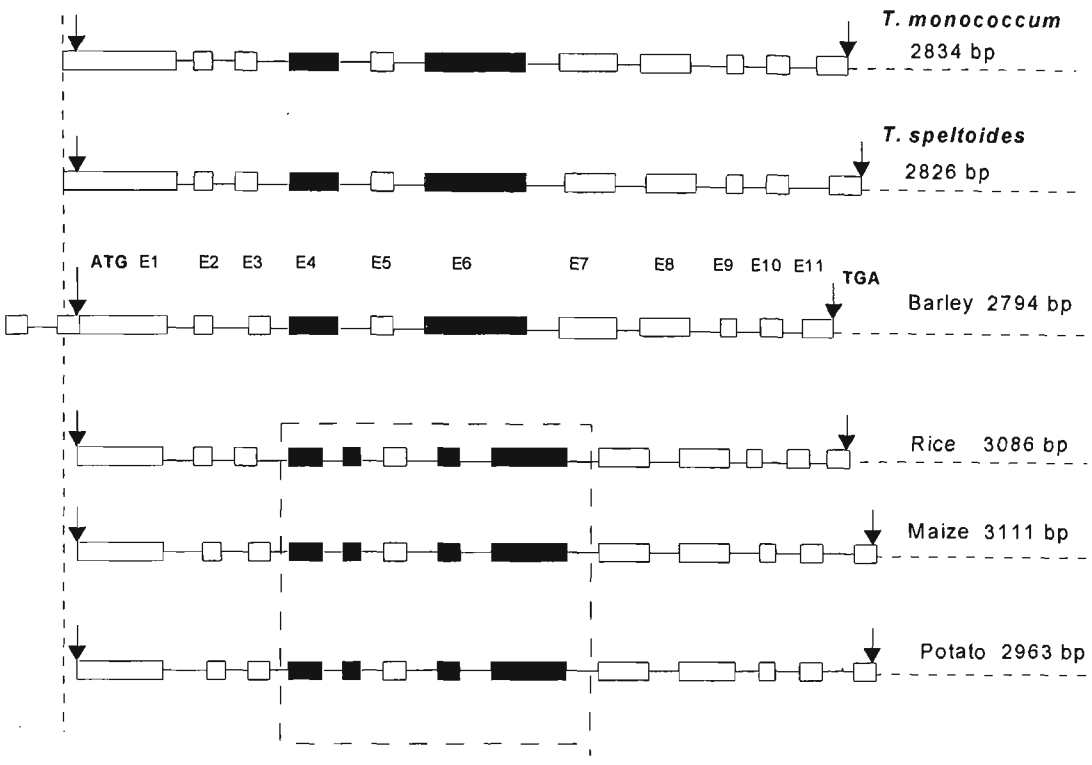


Fig. 4.4 Structure of the *waxy* genes in plants. Boxes indicate exons which are numbered in the barley gene. Lines between the boxes indicate introns. Arrows represent the sites for initiation codon (ATG) or termination codon (TGA). Box of exon 4 and box of exon 6 in the *wx-TmA*, *wx-TsB* and barley genes were spliced into two parts in the rice, maize and potato genes by additional introns. The length of all waxy genes were from the translation initiation codon to termination codon. Sources of gene sequences are as follows: barley (Rohde *et al.*, 1988), maize (Klösgen *et al.*, 1986), rice (Wang *et al.*, 1990) and potato (van der Leij *et al.*, 1991).

4.4.2 Base composition of the *wx-TmA* and *wx-TsB* genes

By comparison of the base compositions of the gene sequences between the initiation codon (ATG) and termination codon (TGA), it was found that there was a marked GC bias in the *wx-TmA* and *wx-TsB* genes. GC content of the two genes was about 58%, similar to that of the corresponding regions of barley and maize genes but slightly higher than that of the rice gene (52%). In contrast, the potato gene had a significant AT bias in base composition, with only 40% GC.

4.4.3 T-rich rather than AT-rich regions in introns

The T, C, A and G content of the 10 introns of *wx-TmA* and *wx-TsB* was 30-31%, 25-26%, 22-23% and 21-22% respectively (Table 4.4). This result indicated that the deduced introns of the *wx-TmA* and *wx-TsB* genes contained T-rich regions rather than conventional AT-rich regions of other plant genes (Simpson and Filipowi, 1996). However, the T-rich sequences were not distributed throughout all the introns, but significantly noticeable in intron 3, 5 and 9 of the two *waxy* genes (Table 4.6). In exons, A and T content made up about 35% of the total nucleotides in *wx-TmA* and *wx-TsB* genes. The difference in A and T content between intron and exon sequences was more significant in the two diploid wheat *waxy* genes than in other plant genes reported by Simpson and Filipowicz (1996).

Table 4.3 Base composition (%) of the *waxy* genes

Content %	*wx-TmA*	*wx-TsB*	Barley	Rice	Maize	Potato
A	20.48	20.64	21.19	23.36	19.58	27.51
C	29.29	29.80	29.21	25.47	30.25	19.17
G	29.33	28.84	28.49	26.51	29.83	21.16
T	20.91	20.72	21.12	24.66	20.35	32.16

The base composition was determined for the sequences between initiation codon (ATG), termination codon (TGA) of the *wx-TmA* (Fig. 4.2) and *wx-TsB* genes (Fig. 4.3) using the Ecomposition program of ANGIS (section 2.6.1) and then converted to % of each base out of the total. The DNA sequence of the *wx-TmA* and *wx-TsB* genes is given in Fig 4.2 and Fig. 4.3 respectively. Sources of published genes are as follows: barley (Rohde *et al.*, 1988), maize (Klösgen *et al.*, 1986), rice (Wang *et al.*, 1990) and potato (van der Leij *et al.*, 1991).

Table 4.4 Base composition (%) of introns of *wx-TmA* and *wx-TsB* genes

Intron	*wx-TmA*				*wx-TsB*			
	A	C	G	T	A	C	G	T
1	20.45	27.27	21.59	30.68	22.22	27.27	21.21	29.29
2	25.00	27.27	19.32	28.47	28.09	29.21	14.61	28.09
3	19.49	22.88	22.03	**35.59**	20.35	19.50	23.01	**37.17**
4	22.38	26.57	25.17	25.87	22.40	27.20	24.80	25.60
5	19.75	24.69	14.81	**40.74**	19.19	21.21	15.15	**44.44**
6	23.60	29.21	22.47	24.72	23.81	27.39	27.39	21.43
7	30.86	29.63	20.99	18.52	29.27	29.27	23.17	18.29
8	19.28	36.14	15.66	28.92	19.28	36.14	14.46	30.12
9	19.19	21.21	19.19	**40.40**	15.91	25.00	22.73	**36.36**
10	21.74	17.39	27.83	33.04	23.48	17.39	26.96	32.17
Average	22.17	26.23	20.91	30.70	22.40	25.96	21.35	30.30

Sequences of different introns of *wx-TmA* and *wx-TsB* genes were deduced from their sequences described in Fig. 4.2 and Fig. 4.3, respectively. Base composition of individual introns was determined using the Ecomposition program of ANGIS and then converted to % of each base out of total. The three T-rich introns are shown in bold.

4.5 Variation in exons of the *wx-TmA* and *wx-TsB* genes

4.5.1 Variation in exon 1

The 31 bp untranslated leader sequence in the exon 1 of the *wx-TmA* and *wx-TsB* genes lacked a GCG that occurred at -3 to -5 positions of the start codon AGT in the barley *waxy* gene (data not shown).

The alignment of the *wx-TmA* and *wx-TsB* genes with other plant genes showed four gap regions only in exon 1 (Fig. 4.5); all other exons had the same length (data not shown). In these four gap regions, a number of variations were observed in short nucleotide repeats (Fig. 4.5). For example, in the twenty nucleotides of region 3, the *wx-TmA* and *wx-TsB* genes shared three AAGC, and the rice gene had three tandem AGC or two tandem CGGT repeats. In the nineteen nucleotides of region 4, *the wx-TmA* and *wx-TsB* genes had $(GCN)_5$, while other genes had $(GCN)_{2\text{-}4}$. In region 1, the rice gene had a long insertion relative to other cereal genes, which apparently resulted from five tandem CGN. Such variations in copy numbers of short repeated DNA sequences could reflect instability of this region of the *waxy* gene.

4.5.2 Extent of identity of the exons of the *wx-TmA* and *wx-TsB* genes

Both *wx-TmA* and *wx-TsB* genes had a total length of 1818 nucleotides in exons from the initiation codon to the termination codon, and the length of the individual exons was identical between the two genes. The identity between the individual exons ranged from 94%-99%, with the overall identity being 96.59% (Table 4.5).

Table 4.5 Comparison of exons and introns of the *wx-TmA* and *wx-TsB* genes

Number	Exons			Introns		
	Length (bp)		Identity	Length (bp)		Identity
	wx-TmA	*wx-TsB*	(%)	*wx-TmA*	*wx-TsB*	(%)
1	324	324	95.75	88	99	90.36
2	81	81	98.77	88	89	91.57
3	99	99	93.94	118	113	95.46
4	154	154	94.81	143	125	92.00
5	101	101	96.04	81	99	93.83
6	354	354	97.18	89	84	93.75
7	180	180	96.67	81	82	97.53
8	192	192	96.88	83	83	95.18
9	87	87	97.70	99	88	85.23
10	129	129	98.45	115	115	97.39
11	117	117	95.73	-	-	-
Total	**1818**	**1818**	**96.64**	**985**	**977**	**93.23**

Number of nucleotides in the exons and introns were determined from their sequences given in Fig. 4.2 and Fig. 4.3, respectively. Sequence identities between the corresponding introns and exons of the *wx-TmA* and *wx-TsB* genes were obtained using the Bestfit program of ANGIS (section 2.6.2).

Fig. 4.5 Variation in sequence of exon 1 in different plant *waxy* genes. The exon 1 sequences from initiation codon ATG of the *waxy* genes were aligned using Pileup program of ANGIS (section 2.6.3). The number on top of the lines indicates the nucleotide number from 'A' of the initiation codon ATG. Only the four variable regions are shown, separated by double slashes that indicate locations of no insertions/deletions variable sequences. A: *wx-TmA*, B: *wx-TsB*, H: Barley (Rohde *et al.*, 1988), M: maize (Klösgen *et al.*, 1986), R: rice (Wang *et al.* 1990), and P: potato (van der Leij *et al.*, 1991).

4.6 Variation in introns of the *wx-TmA* and *wx-TsB* genes

4.6.1 Variable lengths and sequences of introns

Alignment of the *wx-TmA* and *wx-TsB* genes showed the presence of 18 gaps between the two genes (Fig. 4.6). These gaps mainly resulted from insertions/deletions of 1-18 bp sequences. Table 4.5 shows the length and the extent of identity of each intron between the *wx-TmA* and *wx-TsB* genes. Except for introns 8 and 10 that had the same length, the other 8 introns differed in their length, a fact especially observed for introns 4 and 5. The intron 4 of *wx-TmA* gene was longer by 18 bp than that of *wx-TsB* (Table 4.5), the extra length resulting from two gaps in the aligned sequences (Fig. 4.6). In contrast, the intron 5 of the *wx-TsB* was longer than that of *wx-TmA*, resulting in one large gap in the aligned sequences. The smallest difference in length, one nucleotide, was found in introns 2 and 7. These various insertion/deletions together accounted for the variation in the total lengths of introns of the two genes, 985 bp for *wx-TmA* and 977 bp for *wx-TsB*.

The corresponding intron sequences of the *wx-TmA* and *wx-TsB* genes shared a high identity, nine of the ten introns had >92% identity, only intron 8 had a more limited identity (85%). However, the identity of introns exhibited a greater variation, compared to the identity of the exon sequences (Table 4.5).

4.6.2 Short repeated sequences involved in variation of intron sequences of the *waxy* genes

The variation in intron sequences of the *wx-TmA* and *wx-TsB* genes was observed to be related also to the accumulation of repeated sequences. Repeated sequences of a size >7 bp occurring within a 50 bp range are shown in Fig. 4.6. For example, at the 5' end of intron 1, the *wx-TsB* gene contained three tandem repeats of AGCT, one more than the *wx-TmA* gene, resulting in a 4 bp gap between the two genes. In intron 3 of *wx-TsB* gene, an extra copy of GATCGAT mostly accounted for an 8 bp gap. Repetitive sequences were also found to occur in the vicinity of the gaps in the variable introns 4 and 5.

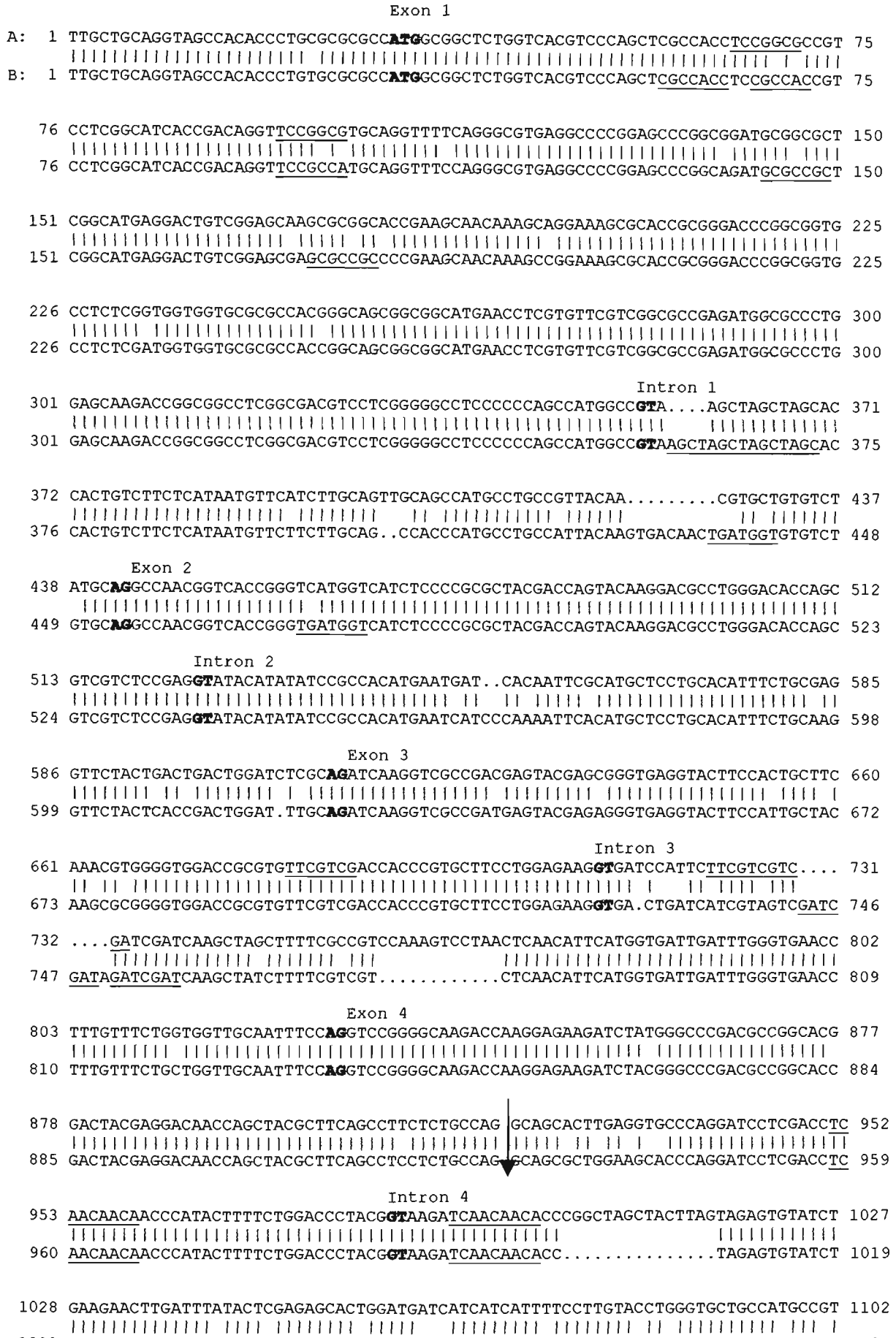


Fig. 4.6 *to be continued*

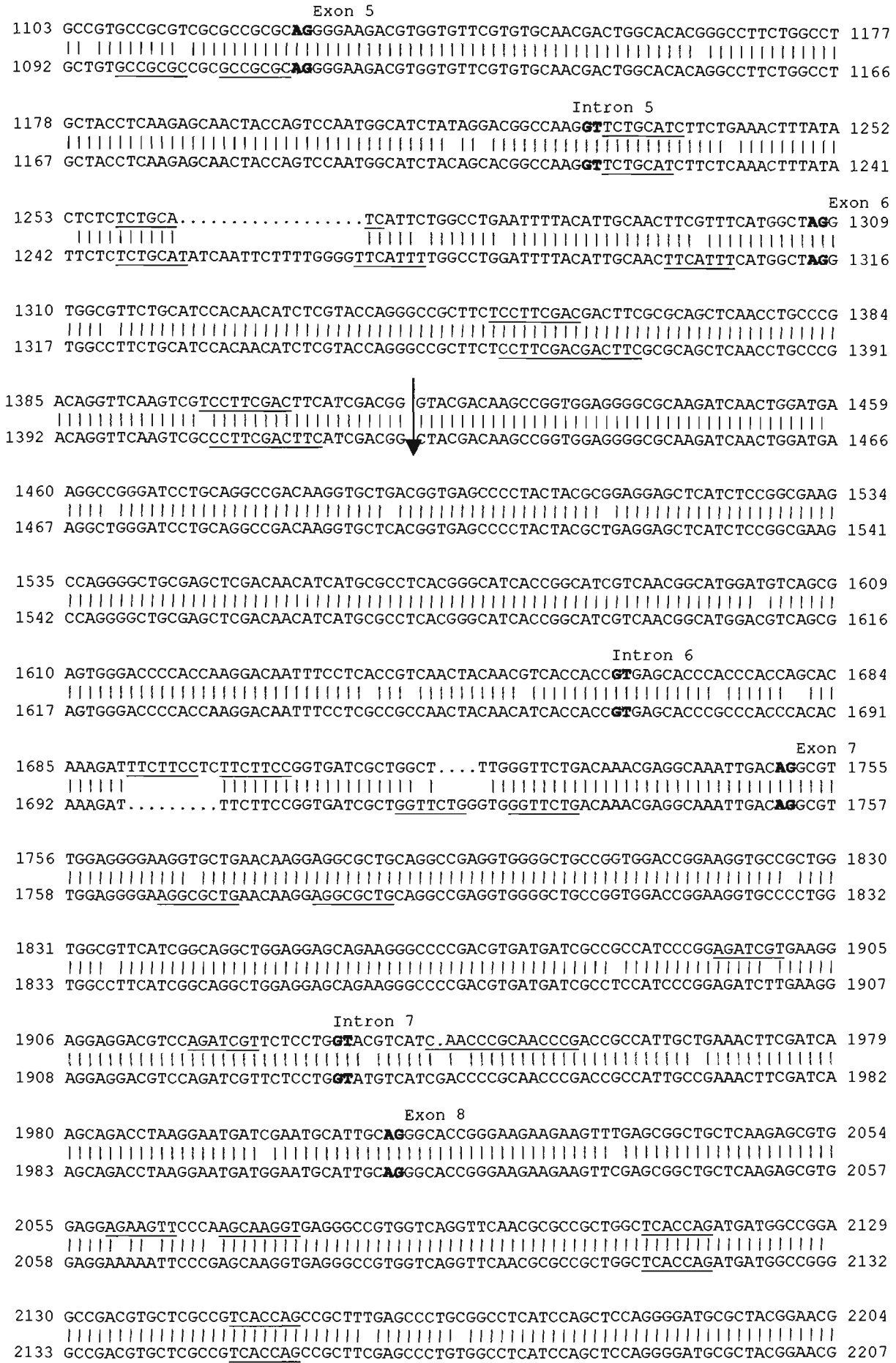


Fig. 4.6 *to be continued*

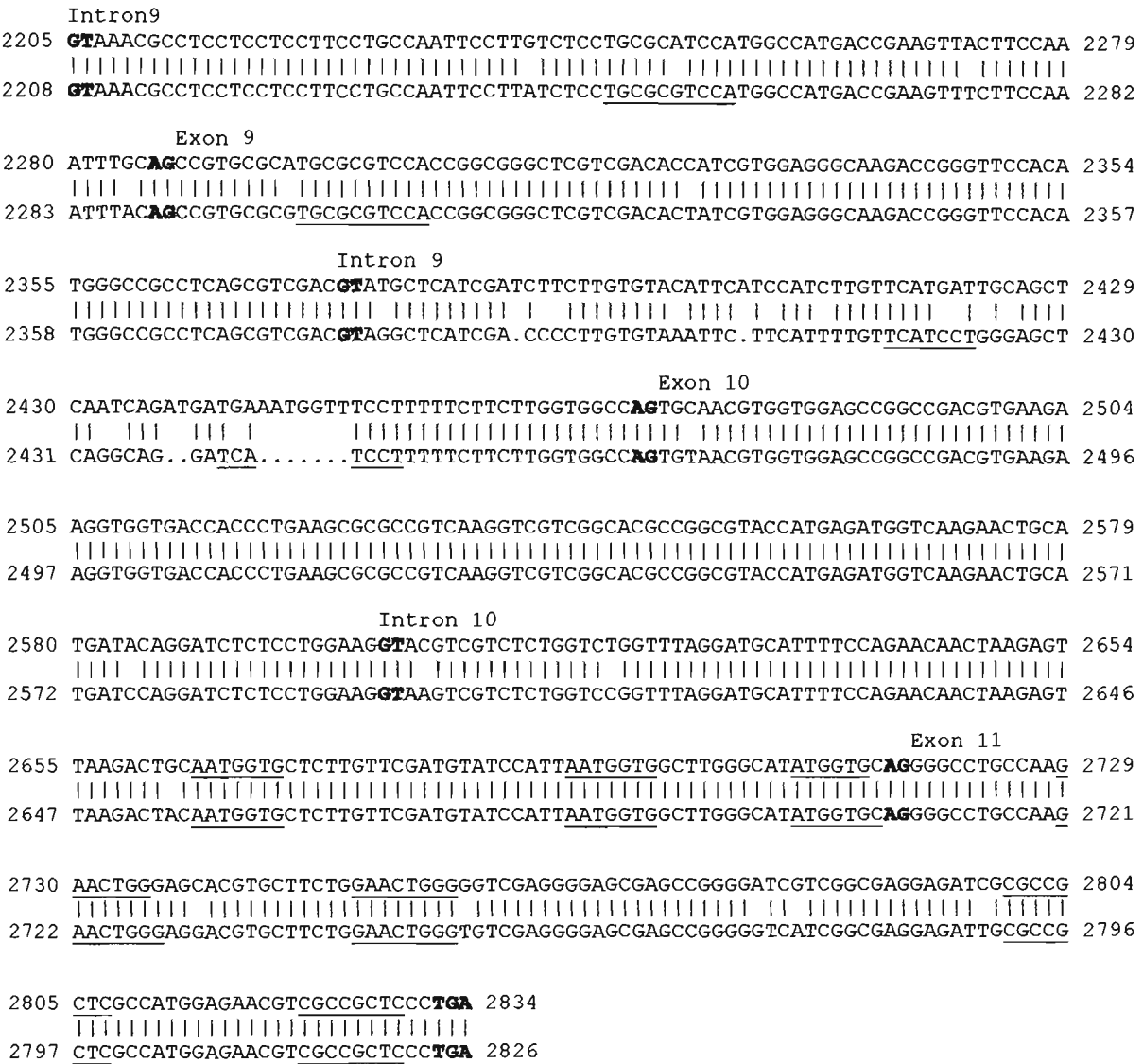


Fig. 4.6 Alignment of the *wx-TmA* and *wx-TsB* gene sequences. Sequence comparison was performed by Bestfit program of ANGIS, using the placement of gaps to get the maximal identity (section 2.6.2). The translation initiation and termination codons of translation and GT/AG spice junctions of introns are shown in bold and shaded. Vertical lines indicate nucleotides that are identical between the *wx-TmA* and *wx-TsB* genes; the identity of *wx-TmA* and *wx-TsB* is 95.31%. Dots denote gaps between the genes; there are 18 gaps. The underlined sequences indicate the >7 bp repeats found within a 50 bp range, by using the Erepeat program of ANGIS (section 2.6.5). Vertical arrows indicate the positions of the additional introns found in rice (Wang *et al.*, 1990), maize (Klösgen *et al.*, 1986) and potato (van der Leij *et al.*, 1991) *waxy* genes.

4.7 Sequence comparison of the *wx-TmA* and *wx-TsB* genes with other plant *waxy* genes

Table 4.6 indicates the extent of identity between DNA sequences of the plant *waxy* genes reported so far; all sequences were compared from the translation initiation codon to termination codon and include all exon and intron sequences in this section. The *wx-TmA* and *wx-TsB* genes exhibited the best identity to each other, relative to the other *waxy* genes. The range of identity of these two diploid wheat waxy genes to different plant genes varied greatly, from 89% to barley to 61-63% to potato.

Further comparison of sequences of the exons of *wx-TmA* and *wx-TsB* with a *waxy* cDNA of common wheat reported by Clarke *et al.* (1991) showed that neither *wx-TmA* nor *wx-TsB* was completely identical to this cDNA, although *wx-TmA* and *wx-TsB* had 95.98% and 94.82% identity, respectively, to this cDNA.

4.8 Deduced physical maps of the *wx-TmA* and *wx-TsB* genes

Although the *wx-TmA* and *wx-TsB* genes shared >95% identity in their DNA sequences, the slight difference in their DNA sequences created different restriction enzyme sites by 6 base cutter once only in the two regions (Fig. 4.7). These different restriction enzymes sites, particularly those found in exons, would be useful for identifying specific genome types of wheat.

Table 4.6 Identity (%) of nucleotides in the coding region of the *waxy* genes in plants

	wx-TmA	***wx-TsB***	Barley	Rice	Maize	Potato
wx-TmA	100	95.31	89.35	77.03	76.95	62.94
wx-TsB		100	89.65	77.15	74.76	61.30
Barley			100	76.00	76.52	63.10
Rice				100	75.77	62.00
Maize					100	61.39
Potato						100

Sequences of the different *waxy* genes were compared only from the translation initiation codon to stop codon. Sources of the published gene sequences are as follows: barley (Rohde *et al.*, 1988), maize (Klösgen *et al.*, 1986), rice (Wang *et al.*, 1990) and potato (van der Leij *et al.*, 1991).

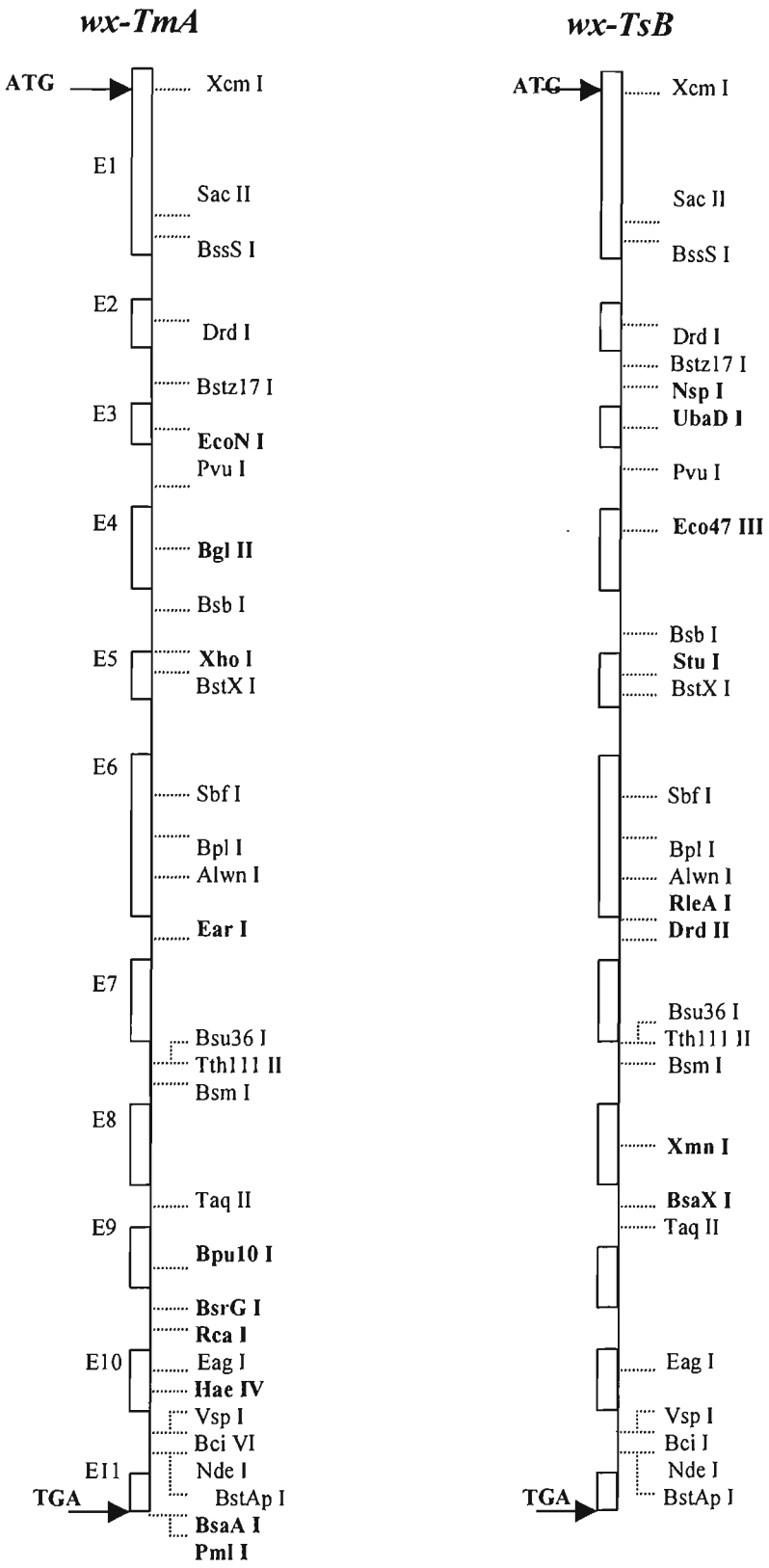


Fig. 4.7 Physical maps of the *waxy* gene of *T. monococcum* and *T. speltoides*. The sites of restriction enzymes were obtained to screen sequences of *wx-TmA* of *T. monococcum* and *wx-TsB* of *T. speltoides* by using the Mapping program of ANGIS (section 2.6.4) under the cutter with 6 bp and only once in the two genes. Specific restriction enzyme sites to *wx-TmA* or *wx-TsB* are shown in bold. 'E' represents the locations of exons. Lines between boxes represent introns. Arrows indicate the locations of initiation codon and termination codon.

4.9 Discussion

4.9.1 Application of the sequences of the *wx-TmA* and *wx-TsB* genes

This Chapter describes the sequences and exon/intron structures of the two homoeologous genes, *wx-TmA* and *wx-TsB*, encoding granule-bound starch synthase (*waxy* protein) in *T. monococcum* and *T. speltoides*, respectively. The *wx-TmA* and *wx-TsB* genes were found to share the highest homology (in exons and introns both) with each other, compared with other known plant *waxy* genes. The alignment of the two genes indicated that the lengths of exons were the same for the two genes, while 18 gaps were found in the intron region. These results led us to design "universal" primers based on the conserved sequences of exons in these two diploid wheat, to identify and clone the individual, partial *waxy* genes in other diploid and polyploid species of wheat (Chapter 6). The variable lengths and the insertions/deletions of sequences in the introns identified in this Chapter will also play a key role in allowing identification of the different genomes of wheat and of mutants of the *waxy* genes by cloning appropriate variable regions of different waxy genes in polyploid wheat. These aspects are discussed in Chapters 6 and 7. A number of other interesting observations have also been made from the results obtained in this Chapter, which are discussed below.

4.9.2 Introns evolve faster than exons

The nucleotide sequence data obtained here revealed that variability in the coding region was much smaller than that in the noncoding region (Table 4.5). This observation supports previous reports by van Campenhout *et al.* (1998) and Mason-Gamer *et al.* (1998). The data also indicated that insertion/deletions of short nucleotide sequences have occurred frequently during the diversification of the genomes in *Triticum* and *Aegilops*. The variations in the *waxy* genes of *T. monococcum* and *T. speltoides* observed here are consistent with the report on *rbcL* genes on the chloroplast genome of wheat and several *Aegilops* species (Terachi *et al.*, 1996). The gaps found in the alignment of the introns of the *wx-TmA* and *wx-TsB* genes (Fig. 4.6) are attributed to the different length and variable sequences of introns.

This is not surprising, as introns are suggested to undergo/retain more mutations compared to exons, due to lack of functional constraints on the introns. The comparisons of a particular gene between members within a given family or in different species have led to the conclusion that the protein-coding sequences evolve slowly, mainly by point mutations, while intron sequences evolve much more rapidly by deletion and insertion events (Breathnach and Chambon, 1981; van Campenhout *et al.*, 1998). The present data also indicated that the deletion/insertion events were related to the repeated sequences in introns, though these events are not explained simply by repeat sequences alone.

4.9.3 Variation in structures of *waxy* genes in plants

The *waxy* genes of *T. monococcum* and *T. speltoides* cloned here consist of 11 exons and 10 introns. The translated sections of the genomic copies of the *waxy* genes encode the entire waxy proteins and their transit peptides, which are the primary interest for distinguishing the different waxy proteins of wheat. Comparison of the intron/exon structures of *wx-TmA* and *wx-TsB* genes with corresponding sections of other *waxy* genes indicated that the *waxy* gene of barley retained the same exon/intron structure as those of *T. monococcum* and *T. speltoides* (Fig. 4.4); however, the *waxy* genes of rice (Wang *et al.*, 1990), maize (Shure *et al.*, 1983; KlÖsgen *et al.*, 1986) and potato (Hovenkamp-Hermelink *et al.*, 1987; van der Leij *et al.*, 1991) consist of 13 exons and 12 introns. The two additional exons in these plants result from splitting of two exons, exon 4 and exon 6 corresponding to the *wx-TmA*, *wx-TsB* and barley *waxy* genes, respectively (Fig. 4.5). The fact that the *waxy* genes in wheat and barley have lost two introns during plant evolution suggests a selective loss of introns of *waxy* genes after the divergence of wheat and rice. Although one would predict that there is a specific machinery for intron excision, the exact removal of an intron would be a very rare event. There are now several examples that show that precise intron deletion can occur in the course of evolution, such as globin pseudogenes (Nishioka *et al.*, 1980; Vanin, 1980) and a rat insulin gene (Lomedico *et al.*, 1979; Cordell *et al.*, 1979). The intron loss observed in the *waxy* gene of plants could be utilised as a molecular tool for further studies into plant evolution and taxonomy.

4.9.4 Relationships of the diploid wheat *waxy* genes to a cDNA of common wheat

Comparison of exons of the genomic genes of *T. monococcum* and *T. speltoides* with a cDNA of the *waxy* gene of common wheat (Clark *et al.*, 1991; Ainsworth *et al.*, 1993) showed that the exon sequences of neither of the two diploid wheat were completely identical to this cDNA. The striking difference was that a 33 bp sequence encoding an additional 11 amino acids in this cDNA (Ainsworth *et al.*, 1993) could not be found in the *wx-TmA* or *wx-TsB*. However, 28 bp of the 33 bp in this cDNA was identical to sequences at the 3' end of intron 4 of both *wx-TmA* and *wx-TsB* genes (Fig. 4.8). Further analysis showed that sequences of these sections of intron 4 of *wx-TmA* and *wx-TsB* could code for 7 out of 11 additional amino acids deduced from this cDNA. The intron 4 of the barley *waxy* gene also contained 22 bp identical to this 33 bp sequences.

Intron 4 occurred between the G and GG of the codon GGG for glycine (Fig. 4.2 and Fig. 4.3), which was highly conserved in *T. monococcum* and *T. speltoides* as well as the *waxy* genes of barley (Rohde *et al.*, 1988), rice (Wang *et al.*, 1990), maize (Klösgen *et al.*, 1986) and potato (van der Leij *et al.*, 1991). The following possibilities could explain the unusual match between the partial intron 4 of *wx-TmA* and *wx-TsB* and the cDNA of common wheat. Firstly, it is possible that this cDNA corresponded to the *waxy* gene encoded by the D genome to common wheat rather than the A and B genomes which would be related to *wx-TmA* or *wx-TsB*. However, this possibility has been ruled out by comparing this cDNA with the *waxy* gene of *T. tauschii* (*wx-TtD*) (R. Appels, pers. commun.); the sequence alignment of the section of intron 4 of *wx-TtD* to this cDNA was very similar to that of the *wx-TmA* and *wx-TsB*. Secondly, it is possible that a point mutation at the splice site allowed a part of the intron 4 sequence to contribute to the *waxy* cDNA characterized by Clark *et al.* (1991). This possibility has been investigated further by cloning and comparing the three individual *waxy* genes of common wheat with this cDNA in Chapter 7.

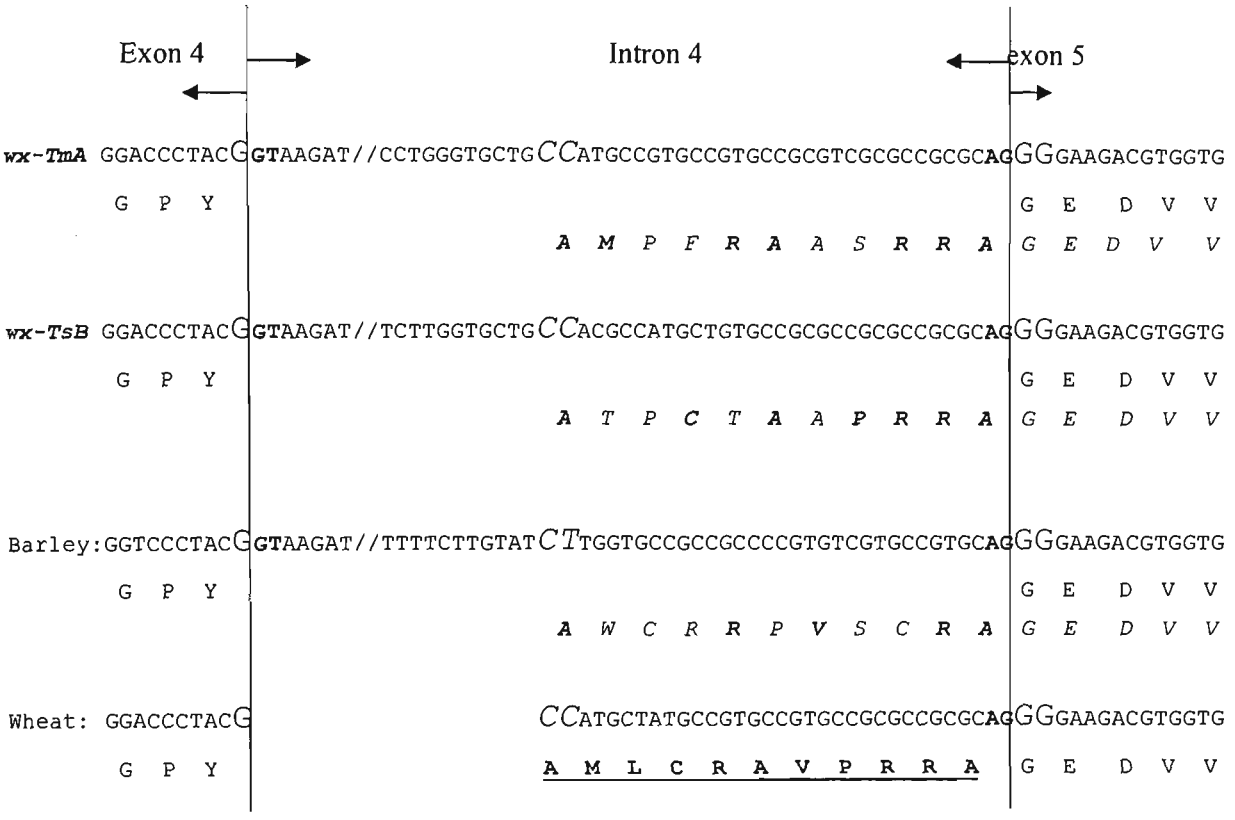


Fig. 4.8 Comparison of sequences of a section of intron 4 of *wx-TmA* and *wx-TsB* genes and a cDNA of common wheat. The sequence between G and GG in the wheat cDNA (Clark *et al.*, 1991) denotes the additional 33 bp sequence; the amino acid proposed to be encoded by this extra sequence are underlined. The sequence corresponding to this stretch occurs in intron 4 of the *waxy* genes of *T. monococcum* (*wx-TmA*), *T. speltoides* (*wx-TsB*) and barley gene (Rohde *et al.*, 1988); the amino acids hypothetically encoded by this section of intron 4 are italicized and amino acids identical to those encoded by the extra sequence of the cDNA are shown in bold. The GT/AG splice junction of intron 4 of *wx-TmA* and *wx-TsB* was deduced from that of the barley *waxy* gene (Rohde *et al.*, 1988).

4.9.5 Application of the PCR in gene cloning and identification of the progenitors of common wheat

The *waxy* gene of *T. monococcum* and *T. speltoides* reported here was cloned successfully using the PCR approach, basing the primers on a cDNA sequence of common wheat, whose source, i.e., the genome (A, B or D) it had originated from, was not known. This indicated that at least the coding sequences of the *waxy* genes of the two diploid wheat were very similar to some of the three *waxy* genes in common wheat. Talbert *et al.* (1994) and Nieto-Lopez and Blake (1994) have successfully used primers from gene sequences of *T. tauschii* and barley to amplify homologous sequences from wheat. Although there is a possibility of mismatch between the primer sequences and the DNA template, in those cases, such a mismatch could be tolerated if the amplification is carried out at lower annealing temperature. Occasionally, inconsistency does occur in sequences from different PCR reactions (<0.1% in this study), but does not seriously impair the identification of various genes or chromosomes or genomes. The PCR method should thus be applicable for amplification and cloning of the *waxy* genes for the purposes of identification of different genomes in *Triticeae*, if they are sufficiently similar in their sequences. With this hypothesis in mind and based on the conserved sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* discussed in this Chapter, two primers (section 2.3.4) were designed to clone the partial *waxy* genes covering the interesting intron 4 from the individual *waxy* loci of different types of wheat. Fourteen such partial *waxy* genes, representing different genomes of wheat, will be reported and analyzed in Chapter 6.

CHAPTER 5

ANALYSIS OF THE DEDUCED AMINO ACID SEQUENCES OF THE WAXY PROTEINS

List of Figures and Tables in Chapter 5

Chapter 5 Analysis of the deduced Amino Acid Sequences of the Waxy Proteins

ABSTRACT

Comparison of the amino acid sequences of the waxy proteins, WX-TmA and WX-TsB, deduced from the coding sequences of the *wx-TmA* gene of *T. monococcum* and the *wx-TsB* gene of *T. speltoides* respectively (Chapter 4), indicated that the deduced transit peptides of the two proteins had the same number of amino acids and shared highly homologous sequences. Both transit peptides had highly basic isoelectric points and were enriched in both polar and small amino acids, which accounted for certain physical properties of these transit peptides. The two mature proteins, WX-TmA and WX-TsB, had the same number of amino acids but slightly different molecular weights and isoelectric points, which were 58. 90 kDa and pI 5.63 for WX-TmA and 58.76 kDa and pI 5.29 for WX-TsB. There were thirteen pairs of amino acid substitutions between the mature WX-TmA and WX-TsB proteins, a key substitution accounting for a significant variation in their molecular weights being arginine in WX-TmA substituted by serine in WX-TsB. The variation in the amino acid sequences of the WX-TmA and WX-TsB proteins accounted for the differences in their mobility following SDS-PAGE (Chapter 3). Both WX-TmA and WX-TsB had the sections of amino acid sequences that are conserved in other plant waxy proteins. Approximately 5% difference in amino acid sequences was observed between WX-TmA, WX-TsB, the WX-TtD deduced from the *waxy* gene of *T. tauschii* (R. Appels, pers. commun.) and WX-W1 deduced from a cDNA of common wheat (Clark *et al.*, 1993); the difference resulted in variation in their predicted secondary structures. The deduced physical properties of the WX-TmA, WX-TsB or WX-TtD did not match perfectly with those of WX-W1; the possible explanations for this are discussed.

5.1 Introduction

The sequences of the *waxy* genes of *T. monococcum and T. speltoides* have been given in Chapter 4; the biochemical and physical properties of the waxy proteins deduced from these DNA sequences are analysed in this Chapter.

In plants, the product of the *waxy* gene, the waxy preprotein, is about 67 kDa and consists of two distinct functional parts, the approximate 59 kDa mature waxy protein that is involved in amylose synthesis and detected in SDS-PAGE and the 7 kDa transit peptide that is responsible for importing the waxy protein into the amyloplasts of endosperm (Shure *et al.*, 1983). The function of the waxy protein has been studied extensively and a number of null *waxy* locus mutants have also been found or produced and applied in the production of waxy maize and rice (MacDonald and Preiss, 1985; Sano, 1986; Visser *et al.*, 1989; van der Leij *et al.*, 1991). The significance of waxy varieties of cereals in food industry has been discussed in section 1.1.3. Slight differences were detected in the molecular weights of the three mature waxy proteins in common wheat (*T. aestivum*, AABBDD), as explained in Chapter 3; this observation supports the results obtained previously by Nakamura *et al.* (1993). The properties of the three waxy proteins in common wheat, WX-A1, WX-B1 and WX-D1, as known from literature and our results, can be summarized as follows: the WX-A1 protein has a slightly higher molecular weight and a more basic isoelectric point than WX-B1 and WX-D1 proteins, and the WX-B1 and WX-D1 proteins have the same molecular weight but a slightly different isoelectric point (Nakamura *et al.*, 1993).

However, the three waxy proteins of common wheat have also been separated into the three bands with apparently different molecular weights in an improved one-dimensional SDS-PAGE method which utilizes a lower pH (7.8) for the separating gel compared to the conventional one (pH 8.8) (Zhao *et al.*, 1996). The molecular weights of the three waxy proteins, WX-A1, WX-B1 and WX-D1, have been calculated as 62.8 kDa, 58.7 kDa and 56.7 kDa by this method. These differences in molecular weights could be suggestive of insertions or deletions of as many as 10-50 amino acids, and/or

amino acid substitutions, between them. However, experimental data is lacking on whether such variations do, in fact, exist, and if so, how these variations as to affect the proportions and functions of the different waxy proteins in the endosperm of wheat.

The functions of the three waxy proteins and their contributions of amylose content in common wheat are known to vary in monosomic lines (Miura *et al.*, 1994) and nullisomic lines (Miura and Sugawara, 1996), as reviewed in section 1.6.5. The WX-B1 protein is found to be expressed in higher amounts, followed by WX-D1 and WX-A1 proteins (section 3.2.2). However, the relationship of these functional variations in different waxy proteins to their sequences and structures is little-known.

In this chapter, the amino acid sequences of the two waxy proteins, WX-TmA of *T. monococcum* and WX-TsB of *T. speltoides*, have been deduced from their respective gene sequences (*wx-TmA* and *wx-TsB*) discussed in Chapter 4. These protein sequences have been aligned with other plant waxy proteins and the glycogen synthase of *E. coli* using the Pileup program of ANGIS (section 2.6.3) and the conserved amino acids among them have been analyzed. In order to reveal the differences in various waxy proteins of wheat, the WX-TmA and WX-TsB were compared with WX-TtD deduced from sequence of the *T. tauschii waxy* gene (R. Apples, pers. commun.) and the WX-W1 deduced from the sequence of a *waxy* cDNA of common wheat (Clark *et al.*, 1991). The deduced biochemical and physical properties, the molecular basis of electrophoretic differences observed and differences in the predicted secondary structures of these waxy proteins are analysed and discussed.

5.2 Properties of waxy preproteins

5.2.1 Sizes of waxy preproteins

As explained in Chapter 4, the *wx-TmA* gene of *T. monococcum* and *wx-TsB* gene of *T. speltoides* had the same total length of exons (1818 bp) from the translation initiation codon (ATG) to the termination codon (TGA) and encoded 605 amino acid of precursor proteins (Fig. 4.2 and Fig. 4.3). The WX-TmA preprotein deduced from

the sequence of the *wx-TmA* gene and WX-TsB preprotein deduced from the sequence of *wx-TsB* gene had a calculated molecular weight of 66.36 kDa and 66.30 kDa, respectively. This result was in complete agreement with other waxy preproteins deduced from the sequences of their respective genes in plants (Shure *et al.*, 1983; Ainsworth *et al.*, 1993).

5.2.2 Identification of the putative cleavage site of the transit peptides of the WX-TmA and WX-TsB preproteins

As explained in the section 5.1, the waxy preprotein consists of two distinct functional components, the mature waxy protein and its transit peptide. The bands normally observed on SDS-PAGE are the mature waxy proteins; the transit peptide is not seen on the gels as it is cleaved off during the importation of the *waxy* protein into the specific amyloplasts of endosperm cells. The cleavage site for transit peptides can be identified by determining the sequence of amino acids at the N-terminus of the mature waxy protein. This sequence has been reported to be ATGSGGMNLVFV in *T. aestivum* (Ainsworth *et al.*, 1993; Rahman *et al.*, 1995; Fujita *et al.*, 1996), *T. monococcum*, *T. speltoides*, as well as *T. turgidum* and *T. tauschii* (Taira *et al.*, 1995; Fujita *et al.*, 1996). This sequence was found to be at positions 72 to 83 of the deduced sequences of WX-TmA and WX-TsB preproteins (Fig. 5.1). The deduced cleavage site of the transit peptide of these preproteins is thus between arginine at position 71 and alanine at position 72; this site appears to be highly conserved in the genus *Triticum*. This arginine residue was not highly conserved in other plants, but the cleavage site was immediately before either alanine or glycine (Fig. 5.1).

5.3 Properties of transit peptides of the waxy proteins

5.3.1 Biochemical properties of the transit peptides

The transit peptides of WX-TmA and WX-TsB proteins were composed of 71 amino acids, very similar in length to those of other plants (Table 5.1). Interestingly, all transit peptides had a much more basic deduced isoelectric point (pI 11-12.75) and net positive charges, compared to their respective mature proteins which had pI in the range of pI 5.3-6.7 and net negative charges.

```
Maize:  MAALATSQLV ATRAGLGVPD ...ASTFRRG AAQG..... ...LR.G ARA 37
Sorghum:MSTLATSQLV ATHAGLGVPD ...ASMFRRG GVQG..... ...LRAA ARA 38
WX-TmA: MAALVTSQLA TSGAVLGITD R.....FRRA GFQG..... ...VRPR SPA 36
WX-TsB: MAALVTSQLA TSATVLGITD R.....FRHA GFQG..... ...VRPR SPA 36
WX-TtD: MAALVTSQLA TSGTVLGITD R.....FRRA GFQG..... ...VRPR SPA 36
WX-W1:  MAALVTSQLA TSGTVLSVTD R.....FRRP GFQG..... ...LRPR NPA 36
Barley: MAALATSQLA TSGTVLGVTD R.....FRRP GFQG..... ...LRPR NPA 36
Rice:   MSALTTSQLA TSATGFGIAD RSAPSSLLRH GFQG..... ...LKPR SPA 41
Potato: ........MA SITASHHFVS RSQ....TSL DTKS.TLSQI GLRNHTL THN 37
Cassava:........MA TVIAA.HFVS RSSHLSIHAL ETKANNLSHT GPWTQTI TPN 41
Pea:    MATITGSSMP TRTACFNYQG RSAESKLNLP QIHFNN.... ...NQAF PVL 43

                                        Region 2
Maize:  SAAA.DTLSM RTSARAAPRH QQQARR...RGG.RFPSLVVC 72 ASAGMNVVFV  82
Sorghum:SAAAGDALSM RTSACPAPRQ QPAARRGGRGG.RFPSLVVC 77 ATAGMNVVFV  87
WX-TmA: DAA....LGM RTVGASA.AP KQQSRKAHRGTRRQLSVVVR 71 ATGSGGMNLVFV 83
WX-TsB: DAP....LGM RTVGASA.AP KQQSRKAHRGTRRQLSMVVR 71 ATGSGGMNLVFV 83
WX-TtD: DAA....LGM RTVGASA.AP T.QSRKAHRGTRRQLSMVVR 70 ATGSGGMNLVFV 82
WX-W1:  DAA....LGM RTVGASA.AP K.QSRKPHRFDRRQLSMVVR 70 ATGSGGMNLVFV 82
Barley: DAA....LGM RTIGASA.AP K.QSRKAHRGSRRQLSVVVS 70 ATGSGMNLVFV  81
Rice:   GGD....ASS LSVTTSARAT PKQQRSVQRGSRRFPSVVVY 77 ATGAGMNVVFV  88
Potato: GLRAVNKLDG LQSRTNTKVT PKMASRTETKRPGCSATIVC 77 GKGMNLIFV    86
Cassava:GLRSLNTMDK LQMKTQSKAV KKV.SATGNGRPA..AKIIC 78 GHGMNLIFV    87
Pea:    GLRSLNKLHV ........RT ARATSGSSDTSEKSLGKIVC 75 GMSLVFV     82
```

Fig. 5.1 Alignment of the transit peptides of waxy proteins in plants. The sequences of transit peptides and amino acids at the N-terminus of the mature waxy proteins were aligned by the Pileup program of ANGIS (section 2.6.3). The arrow indicates the sites for cleavage of the transit peptides from the mature waxy proteins. Arginine at position 71 and alanine at position 72, which appear to be highly conserved in the genus *Triticum*, are shown in red. Two regions of transit peptides are boxed and basic amino acids enriched in these regions are shown in bold. WX-TmA and WX-TsB, deduced from sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* (Chapter 4); WX-TtD, deduced from the sequence of the *waxy* gene of *T. tauschii* (R. Apples, pers. commun.); WX-W1, deduced from the sequence of a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991). Other amino acid sequences were deduced from their respective gene sequences: maize (Klösgen *et al.*, 1986), sorghum (Rosenfield *et al.*, 1996), potato (van der Leij *et al.*, 1991), pea (Dry *et al.*, 1992), cassava (Salehuzzaman *et al.*, 1993), barley (Rohde *et al.*, 1988) and rice (Wang *et al.*, 1990).

Table 5.1 Comparison of transit peptides and mature waxy proteins in different plants

Proteins	Mature proteins				Transit Peptides			
	Number	MW (kDa)	pI	Charges	Number	MW (kDa)	pI	Charges
WX-TmA	534	58.90	5.63	-10	71	7.48	12.75	+11
WX-TsB	534	58.76	5.29	-12	71	7.56	12.68	+10
WX-TtD	534	58.95	5.65	-9	70	7.38	12.74	+10
WX-W1	545	60.14	6.27	-6	70	7.62	12.58	+10
Barley	533	58.89	5.40	-10	70	7.33	12.68	+10
Rice	532	58.46	6.42	-4	77	8.01	12.51	+8
Maize	533	58.58	5.49	-11	72	7.41	12.74	+9
Sorghum	531	58.36	5.39	-11	77	7.74	12.50	+8
Potato	530	58.24	5.90	-8	77	8.35	11.60	+8
Pea	528	58.27	6.06	-9	75	8.11	10.95	+6
Cassava	530	58.36	6.70	-5	78	8.37	11.45	+8

The molecular weights (MW), isoelectric points (pI) and charges of the mature waxy proteins and their transit peptides were calculated using the Pepstats program of ANGIS (section 2.6.7). WX-TmA and WX-TsB, deduced from sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* (Chapter 4); WX-TtD, deduced from the sequence of the *waxy* gene of *T. tauschii* (R. Apples, pers. commun.); WX-W1, deduced from the sequence of a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991). Other amino acid sequences were deduced from their respective gene sequences: maize (Klösgen *et al.*, 1986); sorghum (Rosenfield *et al.*, 1996), potato (van der Leij *et al.*, 1991), pea (Dry *et al.*, 1992), cassava (Salehuzzaman *et al.*, 1993), barley (Rohde *et al.*, 1988) and rice (Wang *et al.*, 1990).

5.3.2 Identities of sequences of the transit peptides

Although the physical properties of the transit peptides, their actual amino acid sequences were varied among different plants (Fig. 5.1). Five amino acid replacements were detected between the transit peptides of WX-TmA and WX-TsB proteins. The identity of transit peptides of WX-TmA and WX-TsB was approximately 93%, much higher than most of the other plant waxy proteins (Table 5.2). The identities of amino acid sequences of the transit peptides of potato or pea to those of other plants were only 17-28%.

5.3.3 Unique features of sequences of the transit peptides

Analysis of sequences of transit peptides of WX-TmA and WX-TsB exhibited two Arginine-enriched regions, at position 21-34 in region 1 and positions 56-64 in region 2 (Fig. 5.1), and the content of arginine reached 12-16% in the two transit peptides. The transit peptides were significantly enriched in serine and threonine in case of potato, pea and cassava, respectively, with 28%, 22% and 21% of serine and threonine to the total amino acids, respectively. The rice transit peptide had a curious combination of arginine-enriched region and a serine/threonine enriched region, while the maize and sorghum transit peptides contained only the arginine-enriched region. In addition, a common property of transit peptides of all waxy proteins was the high content of small amino acids, alanine and glycine. The transit peptides of WX-TmA and WX-TsB exhibited 15-17% of alanine and 9-10% of glycine, and the alanine content accounted for 24%, and 21% of maize and sorghum waxy transit peptides, respectively.

Table 5.2 Identity of transit peptides of the waxy proteins in different plants

Proteins	**WX-TmA**	**WX-TsB**	WX-TtD	WX-W1	Barley	Rice	Maize	Sorg-hum	Potato	Pea	Cas-sava
WX-TmA	100										
WX-TsB	92.96	100									
WX-TtD	95.71	92.49	100								
WX-W1	85.74	84.29	87.14	100							
Barley	87.14	82.86	85.71	88.57	100						
Rice	54.29	54.29	57.97	55.07	60.87	100					
Maize	51.47	50.75	50.75	49.25	54.41	47.22	100				
Sorghum	54.41	50.00	51.47	50.00	50.07	42.30	81.94	100			
Potato	26.47	27.94	25.00	26.47	26.47	24.66	25.35	24.29	100		
Pea	18.57	17.14	18.18	24.61	20.00	21.92	20.00	12.00	33.33	100	
Cassava	20.59	15.71	20.29	16.67	15.25	15.07	22.86	20.83	47.95	28.77	100

The identities (%) of sequences between different transit peptides were determined by the Bestfit program (section 2.6.2). WX-TmA and WX-TsB, deduced from sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* (Chapter 4); WX-TtD, deduced from the sequence of the *waxy* gene of *T. tauschii* (R. Apples, pers. commun.); WX-W1, deduced from the sequence of a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991). Other amino acid sequences were deduced from their respective gene sequences: maize (Klösgen *et al.*, 1986), sorghum (Rosenfield *et al.*, 1996), potato (van der Leij *et al.*, 1991), pea (Dry *et al.*, 1992), cassava (Salehuzzaman *et al.*, 1993), barley (Rohde *et al.*, 1988) and rice (Wang *et al.*, 1990).

5.4 Properties of the mature WX-TmA and WX-TsB proteins

5.4.1 Biochemical properties of the mature WX-TmA and WX-TsB proteins

The mature waxy proteins, WX-TmA of *T. monococcum* and WX-TsB of *T. speltoides*, deduced from their respective genes (Chapter 4), possessed the same number of amino acids (Table 5.1, Fig. 5.2). The mature WX-TmA protein had the calculated molecular weight of 58.90 kDa, slightly higher than that of the mature WX-TsB protein, 58.76 kDa. The deduced isoelectric points of WX-TmA and WX-TsB were 5.63 and 5.29, respectively.

5.4.2 Identities of amino acid sequences between WX-TmA and WX-TsB proteins

The mature WX-TmA and WX-TsB proteins had 534 amino acids; their sequence identity was 97.57%. There were thirteen amino acid substitutions between the WX-TmA and WX-TsB proteins (Fig. 5.2). Table 5.3 shows the positions and properties of the thirteen pairs of amino acid substitutions. The WX-TmA protein possessed five Valine (MW 117 Da) more than the WX-TsB protein, while the WX-TsB protein had two Alanine (MW 89) more than the WX-TmA protein. A substantial (51%) difference in the molecular weights of the two proteins was accounted for by the amino acid at position 179, which was Arginine (174 Da) in the WX-TmA protein but Serine (105 Da) in the WX-TsB.

There were two pairs of amino acid substitutions involving significant changes of net charges. At position 179, the positively charged arginine (pI 10.76) in the WX-TmA protein was substituted by the polar and uncharged Serine (pI 5.98) in the WX-TsB protein. At position 504, the WX-TmA protein had the positively charged histidine (pI 7.59), whereas WX-TsB protein had the negatively charged aspartate (pI 2.77).

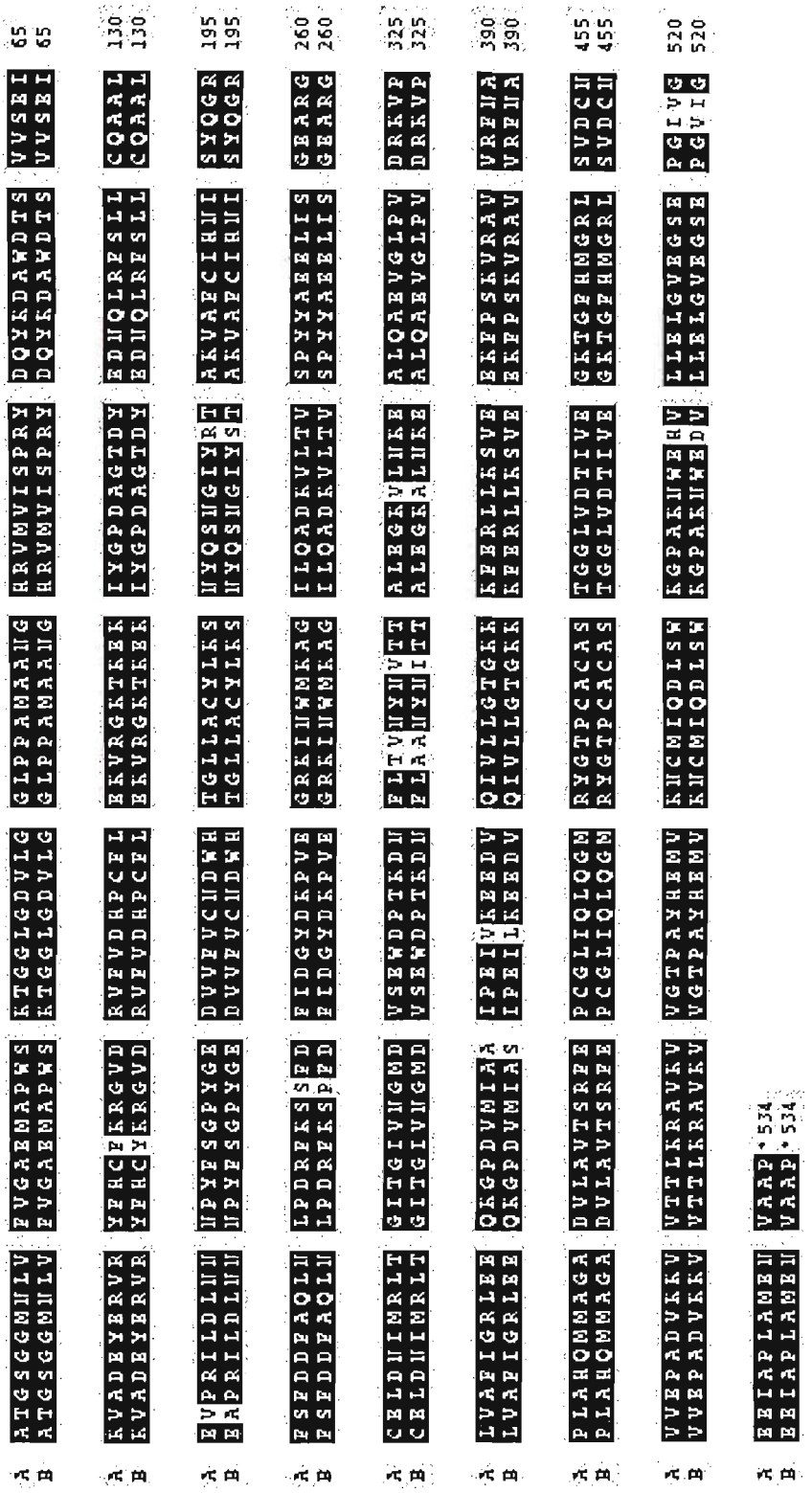


Fig. 5.2 Alignment of the deduced amino acid sequences of the mature waxy proteins of *T. monococcum* (WX-TmA) and *T. speltoides* (WX-TsB). The upper line (A) is the deduced amino acid sequence of the mature WX-TmA protein and the lower line (B) is the deduced amino acid sequence of the mature WX-TsB protein. Shared amino acids are indicated by shaded background and amino acids not shared are indicated by unshaded background. The alignment was performed using the Prettybox program of ANGIS (section 2.6.3).

Table 5.3 Differences in amino acid sequences of the mature WX-TmA and WX-TsB proteins

Sites	WX-TmA					WX-TsB			
	Amino acid		Codon	MW	pI	Amino acid		Codon	MW	pI
80	F	Phenylalanine	TTC	165	5.48	Y	Tyrosine	TAC	181	6.66
132	V	Valine	GTG	117	5.97	A	Alanine	GCA	89	6.01
179	**R**	**Arginine**	**AGG**	**174**	**10.76**	**S**	**Serine**	**AGC**	**105**	**5.68**
213	S	Serine	TCC	105	5.68	P	Proline	CCC	115	6.48
293	T	Threonine	ACC	119	5.87	A	Alanine	GCC	89	6.01
294	V	Valine	GTC	117	5.97	A	Alanine	GCC	89	6.01
298	V	Valine	GTC	117	5.97	I	Isoleucine	ATC	131	6.02
306	V	Valine	GTG	117	5.97	A	Alanine	GCG	89	6.01
345	A	Alanine	GTC	89	6.01	S	Serine	TCC	105	5.68
350	V	Valine	GTG	117	5.97	L	Leucine	TTG	131	5.98
504	**H**	**Histidine**	**CAC**	**155**	**7.59**	**D**	**Aspartate**	**GAC**	**133**	**2.77**
518	I	Isoleucine	ATC	131	6.02	V	Valine	GTC	117	5.97
519	V	Valine	GTC	117	5.97	I	Isoleucine	ATC	131	6.02

Amino acids differing between the mature WX-TmA and WX-TsB were obtained from Fig. 5.2. Codons for these variable amino acids were obtained from DNA sequences of the respective *waxy* genes, given in Chapter (Fig. 4.1 and Fig. 4.2). Molecular weight and isoelectric point of single amino acid were then calculated. Two pairs of amino acids which resulted in significant differences between WX-TmA and WX-TsB are shown in bold.

5.4.3 DNA sequence alterations responsible for the amino acid substitutions in WX-TmA and WX-TsB proteins

All amino acid substitutions between WX-TmA and WX-TsB proteins were caused by single nucleotide mutations, with the exception of the valine-alanine substitution at position 132 which required two base changes (Table 5.3). Seven, three, and one substitutions of amino acids resulted from the first, second, and third nucleotide change in codons, respectively, as would be expected from the degenerate nature of the genetic code.

5.4.4 Composition of amino acids of the mature WX-TmA and WX-TsB proteins

Although the biochemical properties of the mature WX-TmA and WX-TsB proteins were very similar (Table 5.1), there were subtle differences to be found in proportions of various amino acid groups (Table 5.4). For example, composition of charged amino acids in the D+E+H+K+R group in WX-TmA and WX-TsB was 25.66% and 25.47%, respectively. However, these apparently minor alterations might have a relatively large effect on the predicted secondary structures of these waxy proteins, as analyzed in section 5.5.

Table 5.4 Amino acid composition of WX-TmA and WX-TsB proteins

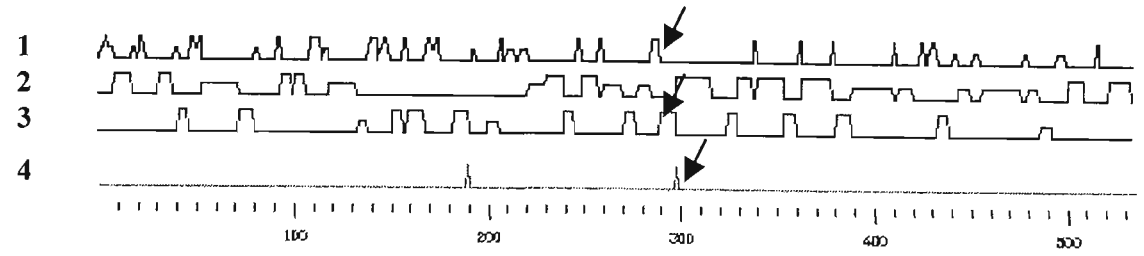
Types of amino acids		WX-TmA		WX-TsB	
		Number	%	Number	%
Small	A+G	89	16.67	92	17.23
Hydroxyl	S+T	46	8.61	46	8.61
Acidic	D+E	69	12.92	70	13.11
Acid/Amide	D+E+N+Q	106	19.85	107	20.04
Basic	H+K+R	68	12.73	66	12.36
Charged	D+E+H+K+R	137	25.66	136	25.47
Small hphob	I+L+M+V	139	26.03	136	25.47
Aromatic	F+W+Y	47	8.80	47	8.80

Amino acids of the WX-TmA and WX-TsB proteins were grouped using the Pepstats program of ANGIS (section 2.6.7). Number and % indicate the number of amino acids in each group and its percentage to the total number of amino acids of the waxy protein, respectively.

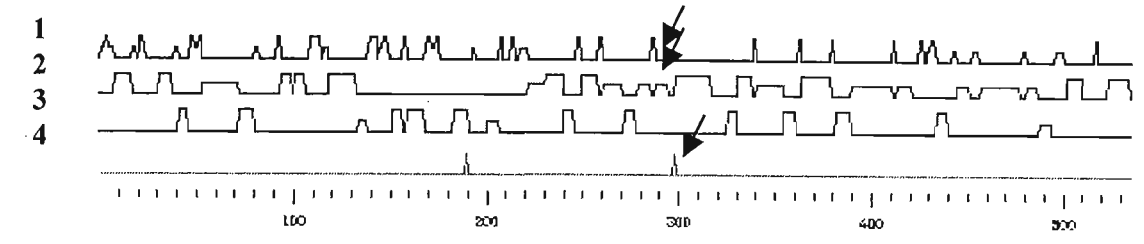
5.5 Subtle differences in the predicted secondary structures of waxy proteins of different types of wheat

The deduced amino acid sequences of the WX-TmA and WX-TsB proteins (this study), the WX-TtD protein of *T. tauschii* (R. Appels, pers. commun.) and the WX-W1 protein deduced from a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991) were used to predict their secondary structures (Fig. 5.3) using the PeptideStructure program of ANGIS (section 2.6.7). Overall, the WX-TsB protein had one more α-helix but one less β-sheet, at around amino acid 300, than the other three waxy proteins. The WX-TtD and WX-W1 proteins had one more α-helix than WX-TmA and WX-TsB proteins between amino acid 150-170. WX-TmA and WX-TsB had two potential glycosylation sites, while WX-TtD and WX-W1 had only one. WX-W1 protein lacked one turn at about amino acid 300, relative to the other three waxy proteins. Other subtle differences in heights of peaks could be detected.

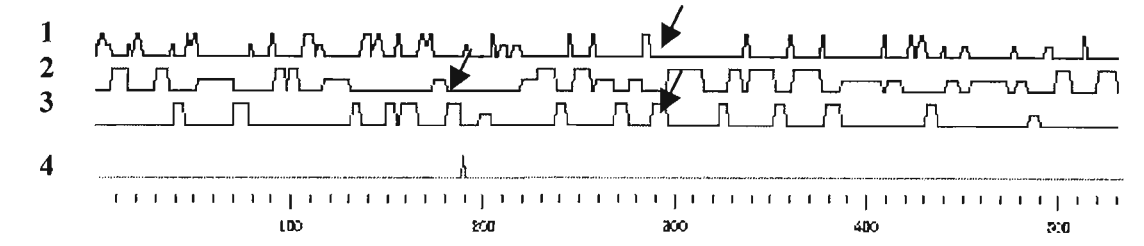
WX-TmA



WX-TsB



WX-TtD



WX-W1

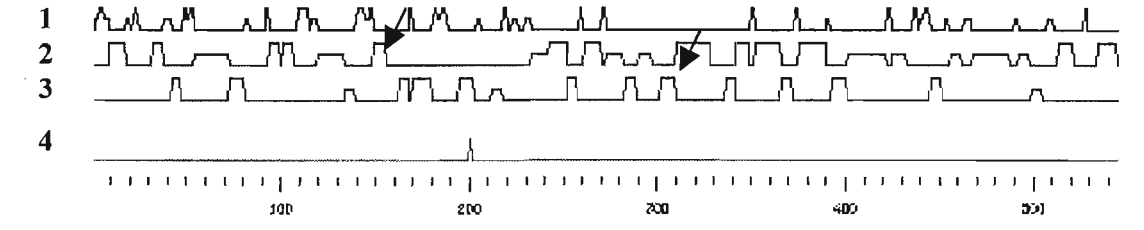


Fig. 5.3 Secondary structure of the different mature waxy proteins. The secondary structures of the waxy proteins were deduced using the Peptidestruture program of ANGIS (section 2.6.7). Lines 1, 2, 3 and 4 indicate turns, α-helices, β-sheets, and glycosylation sites in protein structures, respectively. The significant differences are indicated by arrows among the four waxy proteins, WX-TmA of *T. monococcum* (this study), WX-TsB of *T. speltoides* (this study), WX-TtD of *T. tauschii* (R. Appels, pers. commun.) and WX-W1 deduced from a cDNA of common wheat (Clark *et al.*, 1991).

5.6 Comparison of the WX-TmA and WX-TsB proteins with the other plant waxy proteins

5.6.1 Sequences specific for the different waxy proteins

When the Swiss-PROT databases were searched for text strings associated with WX-TmA and WX-TsB proteins, 54 sequences satisfied the expectation threshold parameter value 10. Glycogen synthase of *E. coli* (GLG) (Kumar *et al.*, 1986) and soluble starch synthase of rice (SSS, Baba *et al.* 1993) were chosen for alignment with all other known plant waxy proteins using the Pileup program of the ANGIS (section 2.6.3).

Fig. 5.4 shows that some conserved sequences occurred in all plant waxy proteins as well as in SSS and GLG. The putative motif KTGGL in Box 1, for the binding site for the substrates for starch synthesis, ADP-glucose and UDP-glucose (Furukawa *et al.*, 1990), was found in the WX-TmA and WX-TsB proteins at sites homologous to that in the other reported waxy proteins. In SSS and GLG, there is only one amino acid replacement (threonine → serine) at this binding site. In Box 4, the conserved sequence was STGGLVDT close to C-terminal of plant waxy proteins, and it was GTGGL in SSS, ETGGL in GLG. Other two conserved domains were NDWHT (Box 2) and SRFEPCGL (Box 3). The two conserved amino acid domains, PRYDQYKDAWDT, close to the N-terminal, and EIAPLA, close to the C-terminal, were found only in the waxy proteins but not in SSS and GLG, suggesting that they might be associated with some special property of the waxy proteins.

Multiple sequence alignment of the eleven different waxy proteins mentioned above (Fig. 5.4) showed that some amino acids were unique to certain wheat waxy proteins. WX-W1 of common wheat had 'an additional 11 amino acid insert' at positions 159-169, relative to all other plant waxy proteins, as reported by Ainsworth *et al.* (1993). These observations imply that these amino acids may represent the unique characteristics of the waxy proteins of these two diploid wheats.

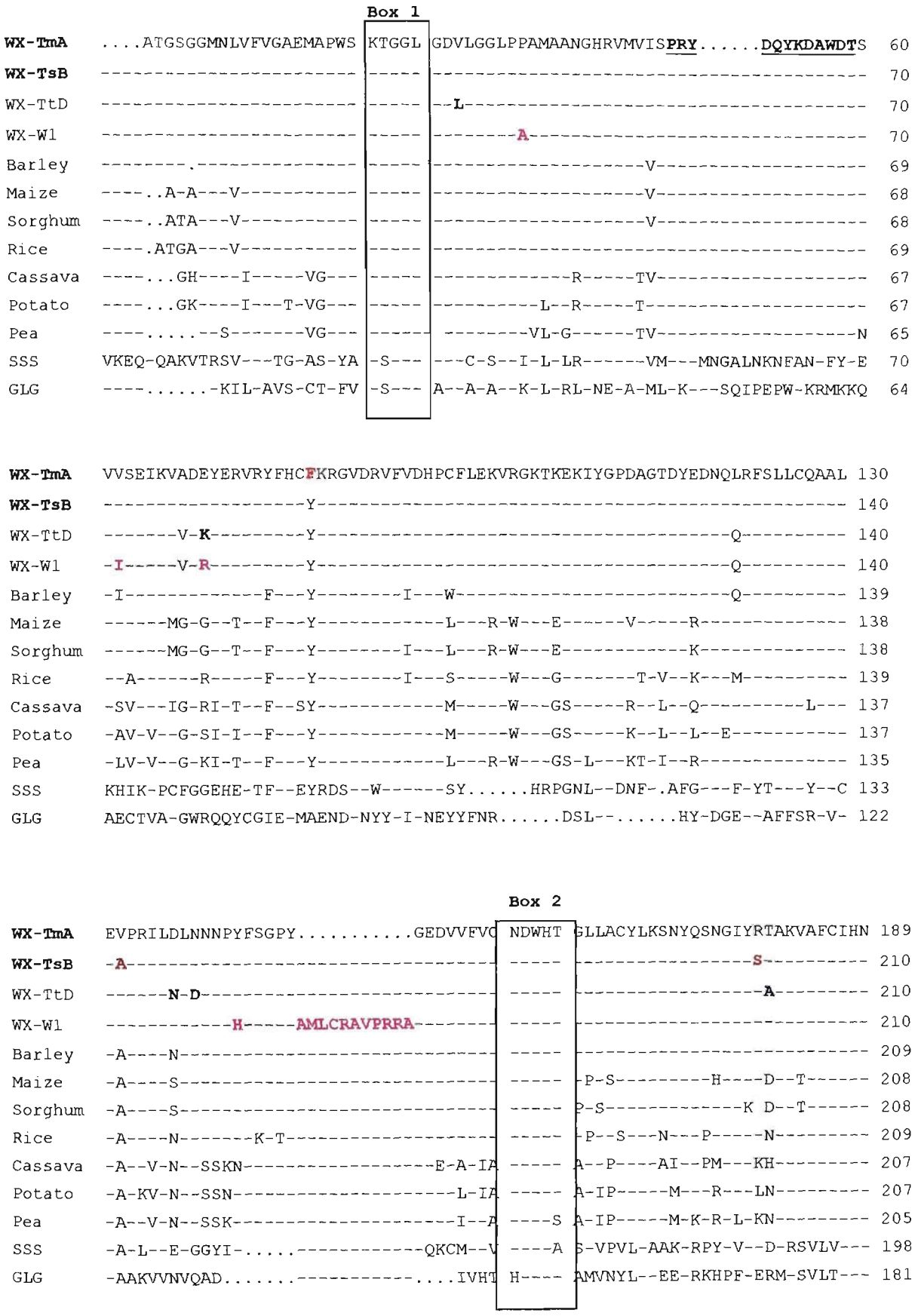


Fig. 5.4 *to be continued*

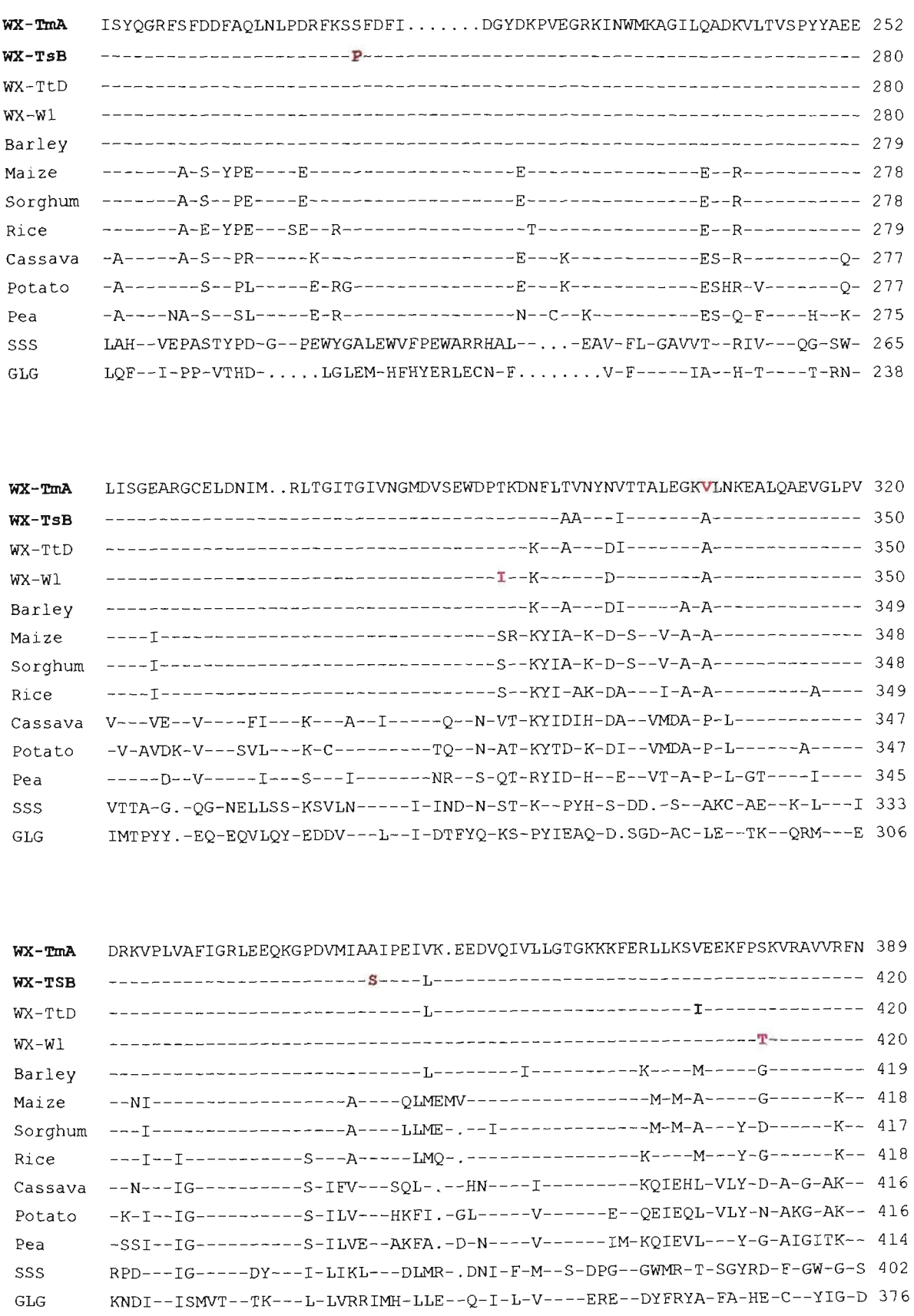


Fig. 5.4 *to be continued*

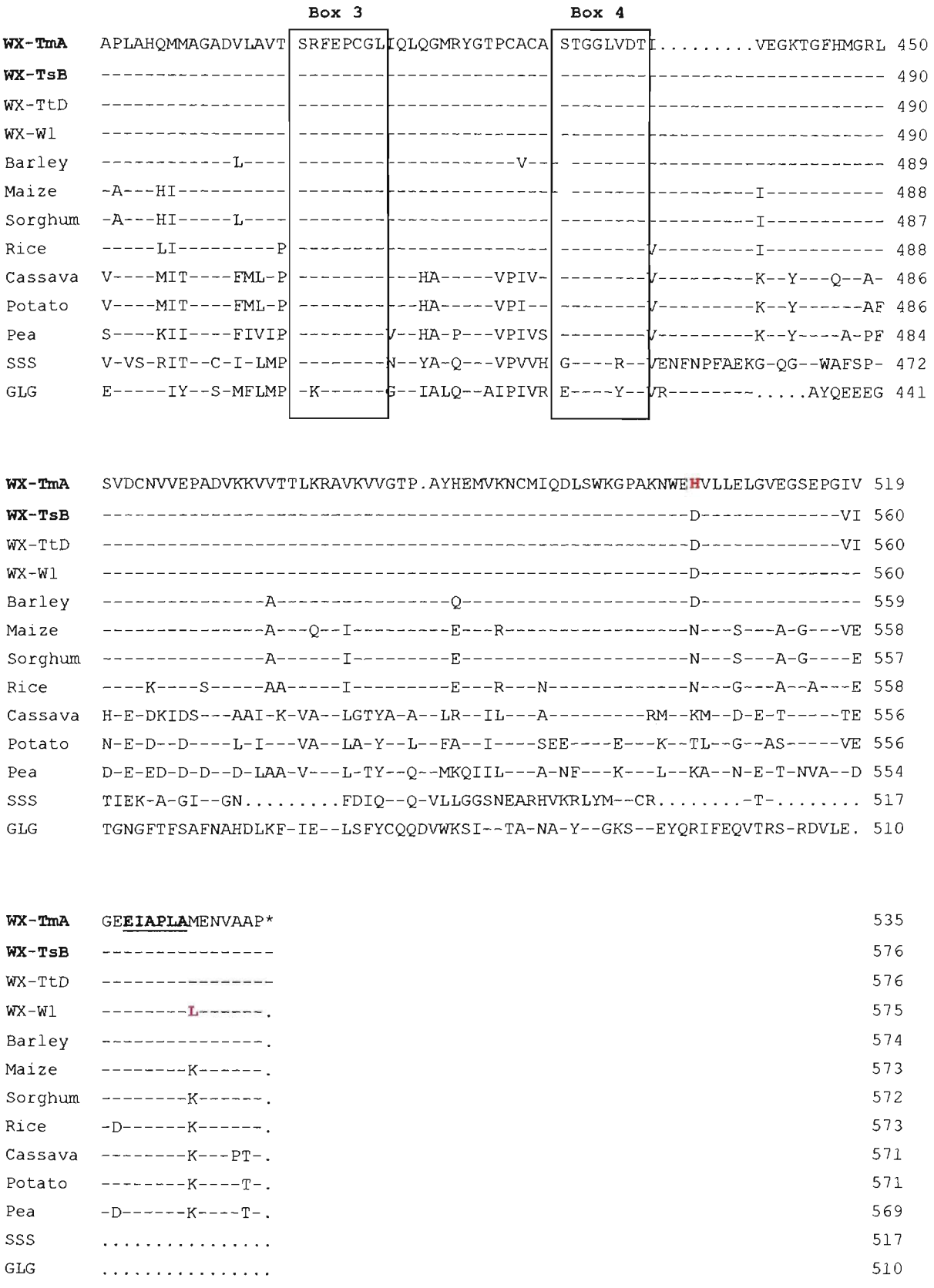


Fig. 5.4 Alignment of deduced sequences of mature waxy proteins in plants. The amino acid sequences of the mature waxy proteins were aligned using the Pileup program of ANGIS (section 2.6.3). "-" represents amino acids identical to the top

Fig. 5.4 (*continued*)

line and "." denotes gaps. The amino acids shared among all plant waxy proteins, SSS and GLG are boxed. Amino acid regions found only in plant waxy proteins are highlighted and underlined. Amino acids specific to a waxy protein in the genus *Triticum* are shown in colour print. A total of twenty-eight positions were found to contain different amino acids among WX-TmA, WX-TsB, WX-TtD and WX-W1. WX-TmA and WX-TsB, deduced from sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* (Chapter 4); WX-TtD, deduced from the sequence of the *waxy* gene of *T. tauschii* (R. Appels, pers. commun.); WX-W1, deduced from the sequence of a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991). Other amino acid sequences of the waxy proteins were deduced from their respective gene sequences: maize (Klösgen *et al.*, 1986), sorghum (Rosenfield *et al.*, 1996), potato (van der Leij *et al.*, 1991), pea (Dry *et al.*, 1992), cassava (Salehuzzaman *et al.*, 1993), barley (Rohde *et al.*, 1988) and rice (Wang *et al.*, 1990). SSS is the soluble starch synthase of rice (Baba *et al.*, 1993) and GLG is the glycogen synthase of *E. coli* (Kumar *et al.*, 1986).

5.6.2 Relationship of the four waxy proteins of wheat

The waxy proteins of none of the three diploid wheat studied were 100% identical to the WX-W1 protein deduced from the cDNA of common wheat (Clark *et al.*, 1991), in their sequences (Fig. 5.4) or in the deduced molecular weight and pI (Table 5.1). Even after disregarding the additional 11 amino acids in the WX-W1 protein reported by Ainsworth *et al.* (1993), its molecular weight and pI would be 58.92 kDa and 5.78, which was still different from that of the three diploid wheat proteins.

The calculated molecular weight of the waxy protein (WX-W1) deduced from a cDNA of common wheat (Clark *et al.*, 1991) was more than that of *T. monococcum*, *T. speltoides* and *T. tauschii* by 1.38 kDa, 1.25 kDa and 1.43 kDa, respectively. If the calculated molecular weight of the WX-W1 protein was correct, no matter which genome (A, B or D) of common wheat this protein originated from, the protein band of common wheat should have had a considerably reduced mobility on SDS-PAGE, compared of that of the diploid wheat. However, no such significant differences in the electrophoretic molecular weights could be detected between the diploid, hexaploid wheat and rice waxy proteins 3.2.2 (Fig. 3.3) and section 3.2.4 (Fig. 3.5). In contrast, the calculated molecular weight of the rice waxy protein (58.46 kDa) was smaller than that of the waxy proteins of *T. monococcum*, *T. speltoides* and *T. tauschii* by only 0.44 kDa, 0.30 kDa and 0.49 kDa, respectively. These differences were clearly visible on SDS-PAGE (Chapter 3, Fig. 3.3). It thus appears that the enhanced calculated molecular weight of the WX-W1 protein resulted from the additional 11 amino acids (Fig. 5.4) as reported by Ainsworth *et al.* (1993).

5.6.3 Extent of identity between the mature waxy proteins in plants

Analyses of sequence identities (Table 5.5) indicated that the WX-TmA and WX-TsB proteins had the best identity (97.57%) to each other, and had approximately 95-97% identity to the waxy proteins of *T. tauschii*, wheat and barley, 84-86% to rice and maize, and 65-69% to potato and pea waxy proteins.

Table 5.5 Extent of identity between the mature waxy proteins in different plants

Protein	**WX-TmA**	**WX-TsB**	WX-TtD	WX-W1	Barley	Rice	Maize	Sorg-hum	Potato	Pea	Cas-sava
WX-TmA	100										
WX-TsB	97.57	100									
WX-TtD	96.63	97.19	100								
WX-W1	97.38	96.07	96.82	100							
Barley	95.31	95.50	95.87	95.50	100						
Rice	84.59	84.40	85.59	84.77	86.47	100					
Maize	86.09	86.09	86.47	86.09	86.47	88.14	100				
Sorghum	86.82	86.82	87.01	86.82	88.14	88.70	71.13	100			
Potato	68.68	68.87	69.25	68.70	69.43	69.62	69.81	70.00	100		
Pea	65.53	65.15	65.15	64.96	66.48	67.80	67.42	74.81	75.38	100	
Cassava	68.68	68.49	68.68	69.06	69.62	70.38	71.13	71.13	81.13	74.81	100

The identity (%) of sequences between different mature waxy proteins was obtained by the Bestfit program (section 2.6.2). WX-TmA and WX-TsB, deduced from sequences of the *waxy* genes of *T. monococcum* and *T. speltoides* (Chapter 4); WX-TtD, deduced from the sequence of the *waxy* gene of *T. tauschii* (R. Apples, pers. commun.); WX-W1, deduced from the sequence of a reported *waxy* cDNA of common wheat (Clark *et al.*, 1993). Other amino acid sequences of the waxy proteins were deduced from their respective gene sequences: maize (Klösgen *et al.*, 1986), sorghum (Rosenfield *et al.*, 1996); potato (van der Leij *et al.*, 1991), pea (Dry *et al.*, 1992), cassava (Salehuzzaman *et al.*, 1993), barley (Rohde *et al.*, 1988) and rice (Wang *et al.*, 1990).

5.7 Discussion

5.7.1 The comparison of the deduced WX-TmA and WX-TsB proteins

The mature waxy proteins of *T. monococcum* and *T. speltoides*, deduced from the sequences of their genes (Chapter 4), had the same number of amino acids, but exhibited slight differences in their calculated molecular weights and isoelectric points (Table 5.1). There were thirteen amino acid substitutions in a total of 534 amino acids of the mature WX-TmA and WX-TsB proteins, probably resulting in different physical properties of the proteins. A key substitution occurred at position 179 of the mature *waxy* protein, where arginine (MW 174 Da, pI 10.76) in the WX-TmA protein was replaced with serine (MW 105 Da, pI 5.68) in the WX-TsB protein. The minor difference in the deduced molecular weights between these two waxy proteins is consistent with the difference observed in molecular weights of these proteins in SDS-PAGE (Chapter 3, Fig. 3.1). These results are also consistent with reports that the WX-A protein has a higher molecular weight and more basic pI compared to the WX-B protein, in the two-dimensional SDS-PAGE analysis of waxy proteins of common wheat (Nakamura *et al.*, 1993) and *T. turgidum* (Yamamori *et al.*, 1994).

5.7.2 The origins of A, B and D genomes based on the predicted physical properties of the waxy proteins

Comparison of WX-TsB (58.76 kDa), WX-TtD (58.95 kDa) and WX-TmA (58.90 kDa) showed that WX-TmA did not have the highest molecular weight, as was expected from the electrophoretic mobility of the three waxy proteins in common wheat (WX-A>WX-D>WX-B) (Nakamura *et al.*, 1993; Zhao and Sharp, 1996), based on the assumption that the proteins from the progenitors have remained relatively unchanged. This observation suggests the possibility that the A genome could have undergone subtle changes, i.e., some point mutation, during evolution from diploid to polyploid wheat. Fujita *et al.* (1996) reported the amino acid sequence of approximately 10% of waxy protein chains in *T. monococcm*, *T. urartu*, *T. turgidum* and different nullisomic/tetrasomic lines of *T. aestivum* var. Chinese Spring. They found that the sequence of the waxy protein in *T. monococcum* was identical to that of

a section of protein peptides from *T. turgidum* and *T. aestivum*, whereas there was an amino acid substitution between *T. urartu* and common wheat. They thus suggested that *T. monococcum* and not *T. urartu* might be the progenitor of the A genome of the *T. turgidum-T. aestivum* lineage of wheat. Their partial amino acid sequence of the *T. monococcum* waxy protein is consistent with our predicted sequence in the corresponding region. Based on the above comparisons, either *T. monococcum* or *T. urartu* could not be the donor of A genome to *T. turgidum-T. aestivum*. Two possibilities could explain the difference of amino acid sequences between diploid and polyploid wheat.

1. Firstly, a few point mutations in the *waxy* A gene in common wheat would have arisen; this possibility is supported by our results which indicated that a point mutation could result in distinguishable difference between WX-TmA and WX-TsB proteins (Table 5.3).

2. Secondly, it is possible that genetic polymorphism might exist within species of *T. monococcum* or *T. urartu* and that *T. monococcum* or/and *T. urartu* used for the studies here and for Fujita *et al.*' s work (1995) might not be correct accession for the progenitor of the A genome. Such genetic polymorphism might also exist between various varieties of diploid *T. speltoides*, based on the fact that there is a difference of one amino acid between the sequences of the *T. speltoides* used by Fujita *et al.* (1996) and that of the WX-TsB protein in this study.

These comparisons provided some information for the phylogenetic and evolutionary relationships of different genomes, however, the origins of A and B genomes could not be determined. Comparison of DNA sequences between the *waxy* genes of diploid wheat and the individual *waxy* genes from different genomes of polyploid wheat would reveal more precise information to further understand the evolution of wheat. This will be discussed in Chapter 6.

5.7.3 Properties of amino acid sequences of waxy proteins

Our results indicate that the deduced amino acid sequences of the WX-TmA and WX-TsB proteins have a high homology to the other plant waxy proteins. Interestingly, four regions of conserved sequences (Fig. 5.4) were detected, not only in glycogen synthase of *E. coli* but also in the soluble starch synthase of rice, suggesting that these sequences might play an important role in starch synthesis. The KTGGL sequence of Box 1 has been recognized as an ADP-glucose binding site of starch and glycogen synthases (Furukawa *et al.*, 1990), but the significance of the conserved sequences in the three other domains remains to be investigated.

5.7.4 Unique properties of the transit peptides of waxy proteins

The transit peptides of *waxy* proteins appeared much more variable than the mature *waxy* proteins and exhibited, for example, only 12-15% identity in their amino acid sequences between the three dicot plants, potato, pea and cassava. However, such divergent sequences do not appear to affect their role in transport of waxy proteins to amyloplast. In contrast to the variable amino acid sequences, the predicted physical properties of transit peptides were highly conserved. Firstly, all transit peptides had a net positive charge and were rich in polar amino acids, arginine, or threonine and serine. The presence of these amino acids results in a significant difference in polarity and net charges between the mature waxy proteins and their transit peptides. Secondly, the transit peptides shared a high content of small amino acids, alanine and glycine. It is suggested from our findings that the transporting property of the transit peptides of the waxy proteins would rely on their amino acid composition but not on amino acid sequences.

5.7.5 Possible explanations for the differential expression and function of the various waxy proteins

As described in section 3.7.2.4, the WX-B1 protein band was more intense than WX-D1 and WX-A1 protein bands in SDS-PAGE of the waxy proteins of common wheat.

There are a number of possible explanations for the differences in the expression of the WX-A1, WX-B1 and WX-D1 proteins of common wheat, on the basis the analysis of amino acid sequences of waxy proteins in possible diploid progenitors of common wheat.

1. Promoter. Waxy protein is dominated by a major gene which is highly conserved in its size and sequence in plants (Fig. 5.7, Table 5.1), however, elements regulating the expression of the *waxy* gene have been reported in rice (Wang *et al.*, 1995 and 1998). It is possible that positive regulating elements located in the promoter region of the *wx-B1* gene, would result in the WX-B1 protein being expressed at a higher level than WX-A1 and WX-D1. Alternatively, negative regulating elements may be located in the promoter region of the *wx-1A* and *wx-D1* genes so that WX-A1 and WX-D1 proteins would cause them to be expressed at a lower level relative to WX-B1 protein.

2. Post-transcriptional regulations. In rice, some varieties, that have been shown to have a lower amount of waxy proteins, showed inefficient excision of intron 1 from *waxy* mRNA (Wang *et al.* 1995). It is also possible that this phenomenon exists for the *wx-A* gene and *wx-D1* gene so that WX-A1 and WX-D1 proteins are expressed at the lower level than the WX-B1 protein.

3. Transit peptides. The transit peptides of the waxy proteins have a role in importing the waxy proteins into amyloplast, for amylose synthesis in endosperm cells. However, whether the efficiency of the transport is affected by the variability in the amino acid sequences of transit peptides is not known. It is possible that there are differences in the functions of the transit peptides of WX-A1, WX-B1 and WX-D1 proteins (Table 5.1).

4. Extractability of proteins. The WX-A1 and WX-D1 protein could be expressed to the same extent as WX-B1 proteins; however, it is possible that there were differences in the extraction of WX-A1 and WX-D1 proteins

from starch granules, perhaps due to their altered protein structure relative to WX-B1 protein. Therefore, WX-B1 protein exhibited more intensive band on SDS-PAGE.

In order to reveal the molecular mechanism of the differential expression of the three waxy proteins of common wheat, it is of interest to see if such differences in expression of the waxy proteins exist among the corresponding diploid wheat as well. However, in spite of keeping the various steps of starch granule and protein extraction protocols (section 2.2) constant, it is often difficult to prepare precisely identical quantities of protein extracts, with identical total protein content from the various diploid wheat, due to a number of reasons. These include differences in sizes of seeds of different species, some being extremely small, (e.g., *T. speltoides*), the difficulty of preparing starch, and the possible differences in the extractability of certain waxy proteins.

5.7.6 A possible phylogenetic tree of plants based on the waxy proteins

To determine the relationships between all known plant waxy proteins, a phylogenetic tree was constructed using method described in section 2.6.8. Glycogen synthase of *E. coli* (Kumar *et al.*, 1986) was used as an outgroup. The phylogenetic tree (Fig. 5.5) showed that the waxy proteins from the monocot plants, rice, maize, sorghum, barley and four types of wheat, formed a clearly distinct group relative to the dicot plants, pea, cassava and potato. Further, monocots fell into two clusters, one including rice, maize and sorghum, and the other including barley and the four types of wheat. *T. monococcum* and *T. speltoides* were classified into one cluster, while *T. tauschii* and common wheat belonged to another cluster. The result on monocot plants is consistent with a previous report indicating that the nucleotide sequence of the rRNA consisted of two clusters in the grass genera: one comprised of wheat and barley and the other comprised of maize and rice (Hamby and Zimmer, 1992). As the tree developed here, barley possesses the same progenitor and belongs to the same cluster as wheat, which is consistent with a previous report (Wolfe, 1989).

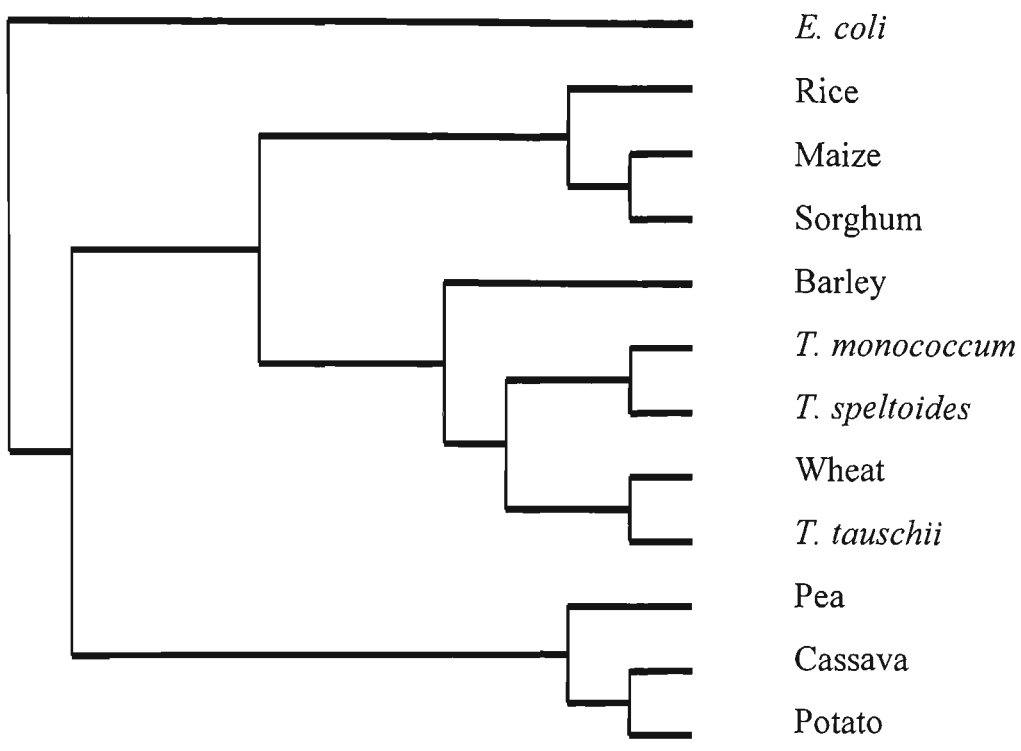


Fig. 5.5 The phylogenetic tree of the plant waxy proteins. Glycogen synthase of *E. coli* (Kumar *et al.*, 1990) was used an outgroup. The phylogenetic tree was constructed by the method described in section 2.6.8.

CHAPTER 6

GENOME-SPECIFIC SEQUENCES OF THE *WAXY* LOCI ANDTHEIR POTENTIAL APPLICATION IN FURTHER UNDERSTANDING THE EVOLUTION OF WHEAT

List of Figures and Tables in Chapter 6

Chapter 6 Genome-Specific Sequences of the *waxy* loci and their potential application in further understanding the evolution of wheat

ABSTRACT

Partial *waxy* genes have been cloned from the two *waxy* loci of tetraploid *T. turgidum* (AABB) and *T. timopheevi* (A^tA^tGG), the three *waxy* loci of hexaploid *T. aestivum* (AABBDD) and *T. zhukovskyi* ($A^tA^tA^zA^zGG$), and the single *waxy* locus each of the proposed diploid progenitors of different genomes of these polyploid wheat, respectively. These seventeen partial *waxy* genes contained exons 4 and 5 of the same lengths but intron 4 of varying lengths. The results have led to the separation of the two *waxy* loci of tetraploid wheat and the three *waxy* loci of hexaploid wheat by comparison of specific sequences of intron 4. Diploid *T. monococcum* and *T. tauschii* have been distinguished by the specific sequences of intron 4, but other five diploid wheat had the same length of intron 4 and could be distinguished only by minor replacement (1%) of sequences of the fragments cloned. Based on the sequence comparison among different diploid, tetraploid and hexaploid species, *T. urartu*, *T. spletoides* and *T. tauschii* were suggested to be the donors of A, B and D genomes of *T. turgidum* and *T. aestivum* and that *T. monococcum* was proposed to be the donor of A^t of *T. timopheevi* and *T. zhukovskyi*. However, the sequence of intron 4 of the cultivated *T. monococcum* tested in this study was different from that of the third set of genome (A^z) of *T. zhukovskyi*, a putative man-made hybrid of *T. monococcum* and *T. timopheevi*. This result suggests that intraspecies polymorphism existed in *T. monococcum* or that a wild *T. monococcum* (*T. monococcum* ssp *boeoticum*) would be the progenitor of the third set of genome *of T. zhukovskyi*. Owing to the fact that all the possible G genome donors had an identical length of intron 4, the progenitor of G genome needs to be investigated.

6.1 Introduction

The analyses of the *waxy* genes of *T. monococcum* and *T. speltoides* (Chapter 4) and *T. tauschii* (R. Appels, pers. commun.) indicated that the exons of none of these three *waxy* genes of diploid wheat were 100% identical to the sequence of a *waxy* cDNA of common wheat reported by Clarke *et al.* (1991). The result suggested that at least one of these three diploid wheat might not be the progenitor of common wheat (i.e., the cDNA originated from the genome that was *not* donated by anyone of these three species) or that mutations might have arisen in the DNA sequences of the *waxy* genes during evolution of wheat from diploid to hexaploid. The results also showed that the sequences of exons were highly conserved in the three diploid wheat; it would thus be very difficult to distinguish between the different genomes in polyploid wheat by finding unique exon sequences representing the different *waxy* loci. One solution to this problem could be to analyse the sequences of non-coding regions that have been proposed to evolve more rapidly than coding region (section 4.9.2). In particular, intron 4 was most variable in length between *wx-TmA* and *wx-TsB* and thus appeared to have the potential for being genome-specific (section 4.6.2, Fig. 4.6). The fact that partial sequences in a *waxy* cDNA of common wheat matched sequences at the 3' end of intron 4 of diploid wheat (section, 4.9.4, Fig. 4.8) led us to further investigate this interesting intron 4 of the *waxy* genes in polyploid species, in order to explain this uncommon match. Therefore, Interest was focussed on the intron 4 area of the *waxy* genes of different types of wheat in this work.

The origins of different genomes of common wheat, *T. aestivum* (AABBDD), have been studied extensively in this century, due to the interest in the evolutionary history of its complex allohexaploid genome and some investigations showing that useful genetic resources such as genes encoding disease resistance in the progenitors can be applied in genetic engineering and breeding programs (Gill *et al.*, 1983; Kerby and Kuspira, 1987; Rodriguez-Quijano *et al.*, 1997). Studies on the origins of genomes of *T. turgidum* (AABB)-*T. aestivum* (AABBDD) have been reviewed (Kerby and Kuspira, 1987), based on the studies of geographical distributions, morphological characteristics, meiotic chromosome behaviour of hybrids and amphiploids, the content and the restriction patterns of nuclear DNA and the protein constitution. The current theory for origin of different genomes in the *T. turgidum* (AABB)-*T. aestivum*

(AABBDD) lineage and the *T. timopheevi* (A^tA^tGG)-*T. zhukovskyi* ($A^tA^tA^zA^zGG$) lineage of wheat (conventional genome constituents are AAGG for *T. timopheevi* and AAAAGG for *T. zhukovskyi*; here the designations of A^t and A^z genomes are used to distinguish between the A genomes in different polyploid wheat) has been presented, based on the analyses of C-banding pattern, the repeated DNA sequences, restriction fragment length polymorphism (RFLP) and protein profiles (Breiman and Graur 1995; Friebe and Gill 1996) (reviewed in section 1.9). However, the diploid donors of the A, B, G and A^t genomes in these polyploid wheat are still in debate, except for showing that *T. tauschii* was the donor of D genome to *T. aestivum* and *T. monococcum* was the donor of A^z genome to *T. zhukovskyi* (sections 1.8 and 1.9).

A powerful method for determining evolutionary relationships is direct comparative DNA sequence analysis (Talbert *et al.*, 1991; Dvorak and Zhang, 1992). The *waxy* gene has been used recently to study the evolutionary nature of different genomes of wheat. Murai *et al.* (unpublished sequences, 1998) isolated and sequenced the major sections of respective *waxy* gene from *T. turgidum-T. aestivum* and the single *waxy* gene each from diploid *T. monococcum*, *T. searsii* and *T. tauschii*. They found that the sequences of the *waxy*-A and *waxy*-B genes were highly conserved from *T. turgidum* to *T. aestivum*, and that the *T. tauschii waxy* gene had 100% identity to the D gene in *T. aestivum*. However, the diploid donors of A and B genomes to polyploid wheat could not be deduced from these studies because only the three diploid species were investigated, but polyploid *T. timopheevi* and *T. zhukovskyi* were not investigated. Mason-Gamer *et al.* (1998) recently reported the sequences of 1.3 kb fragments covering translated exons 6 to 11 of the *waxy* genomic gene in several taxa. Eight diploid species, *T. boeoticum*, *T. monococcum*, *T. urartu*, *T. bicornis*, *T. longissima*, *T. searsii*, *T. speltoides* and *T. tauschii*, which are closely related to the above polyploid wheat, were included their report. However, neither polyploid wheat nor another diploid, *T. sharonensis*, was investigated in this report.

In the present project, in addition to investigate the *waxy* genes encoding the entire waxy proteins in *T. monococcum* and *T. speltoides* (Chapter 4), the partial *waxy* genes were cloned and sequenced from different diploid, tetraploid and hexaploid wheat. These partial *waxy* genes include those from the two *waxy* loci each of the tetraploid wheat, *T. turgidum* and *T. timopheevi*, from the three *waxy* loci each of the hexaploid

wheat, *T. aestivum* and *T. zhukovskyi*, and from the single *waxy* locus each of all the proposed diploid progenitors, except for *T. searsii*, of the different genomes of these polyploid wheat.

This Chapter addresses the sequences of the non-coding intron 4 representing the *waxy* loci of the A, B, D, G, A^t or A^z genomes, and has provided valuable phylogenetic and evolutionary information for different genomes of wheat.

6.2 Partial *waxy* genes from different types of wheat

6.2.1 Partial *waxy* genes from *T. turgidum*

The primers F2 and R2, used to amplify the Fragment 2 covering exon 4-intron 4-exon 5-intron 5-exon 6 of the *waxy* gene in *T. monococcum* and *T. speltoides* (section 2.3.5), were initially applied to clone the corresponding regions of the individual *waxy* genes in polyploid wheat. The two types of sequences of Fragments 2 were obtained from *T. turgidum* ssp. *durum* (called as *T. turgidum* here for convenience). The two fragments were designated as *wx-TdA* and *wx-TdB* and their complete sequences are shown in Chapter 7 (Fig. 7.2). In order to align with the exon 4-intron 4-exon 5 of the *waxy* genes cloned from other diploid and polyploid wheat (section 6.2.2), the two internal sections covering the exon 4-intron 4-exon 5 region of *wx-TdA* and *wx-TdB* of *T. turgidum* were redesignated as *wx-Td2* and *wx-Td1* in this Chapter, for the reason of comparison of different partial *waxy* genes showed in Fig. 6.1. The sequences of the other sections of *wx-TdA* and *wx-TdB* (i.e., exon 5-intron 5-exon 6) of *T. turgidum* will be analysed in Chapter 7.

The primers F1 and R1, used to amplify the Fragment 1 of the *waxy* genes in *T. monococcum* and *T. speltoides* (section 2.3.5), were also applied to clone the corresponding regions of the two individual *waxy* genes in *T. turgidum*. Only one type of sequence of Fragment 1 of the *waxy* gene in the *T. turgidum* has been isolated and will be described in section 6.9.

6.2.2 Partial *waxy* genes from other genomes of wheat

Except for the two types of fragments covering the exon 4-intron 4-exon 5-intron 5-exon 6 sections of the *waxy* genes in *T. turgidum* obtained by F2 and R2 (section 6.2.1), the corresponding fragments of the *waxy* genes in polyploid *T. aestivum*, *T. timopheevi* or *T. zhukovskyi* could not be amplified readily. A reverse primer PR2, which was designed from the conserved domain of exon 5 of the *waxy* gene in *T. monococcum* and *T. speltoides*, and the forward primer F2 were used to amplify the partial *waxy* genes in polyploid *T. aestivum*, *T. timopheevi* and *T. zhukovskyi* and diploid *T. bicornis*, *T. longissima*, *T. sharonensis* and *T. urartu* (section 2.3.6). These partial genes covered the exon 4-intron 4-exon 5 section.

6.3 Isolation of partial *waxy* genes

All fragments of the *waxy* genes from different genomes of wheat, as described in sections 6.2.1 and 6.2.2, were obtained. PCR products from the multiple *waxy* genes of polyploid wheat did not exhibit bands detectably different in size on 1% agarose gels (data not shown). For each species of wheat, the purified PCR products from two separated PCR reactions were mixed and cloned into pGEM-T vectors for sequencing the individual *waxy* genes (section 2.3.11.2). Two single clones were sequenced from each diploid wheat, and no differences were detected in the sequences of different clones. Five to eight clones were sequenced from polyploid wheat, until two different types of sequences from tetraploid wheat and three different types of sequences from hexaploid wheat were obtained (Table 6.1). Based on the sequences obtained, clones from each polyploid wheat were grouped into different classes and redesignated as shown, in Table 6.1.

Table 6.1 Designation of the *waxy* genes of different types of wheat and the clones obtained for these

Species	Ploidy level	Names of genes	No of clones	Names of clones
T. urartu	2	*wx-TuA*	2	U1 and U2
T. bicornis	2	*wx-TbB*	2	Bic1 and Bic2
T. longissima	2	*wx-TlB*	2	L1 and L2
T. sharonensis	2	*wx-TshB*	2	Sha1 and Sha2
T. turgidum	4	*wx-TdB*	2	Du1 and Du3
		wx-TdA	3	Du2, Du4 and Du6
T. aestivum	6	*wx-W1*	2	W1 and W5
		wx-W2	3	W2, W4 and W5
		wx-W3	3	W6, W7 and W8
T. timopheevi	4	*wx-Tt1*	2	V2 and V4
		wx-Tt2	2	V3 and V5
T. zhukovskyi	6	*wx-Tz1*	2	Z4 and Z6
		wx-Tz2	3	Z2, Z3 and Z5
		wx-Tz3	1	Z1

Fourteen *waxy* genes from different types of wheat were cloned. Except for *wx-TdA* and *wx-TdB* genes of *T. turgidum* which were amplified by primers F2 and R2 (section 2.3.5), the other twelve genes were amplified by primers F2 and PR2 (section 2.3.6). The PCR products were cloned into pGEM-T vectors (2.3.11.2).

6.4 Sequences of seventeen different partial *waxy* genes

All expected fragments of individual *waxy* genes were obtained from the different genomes of the diploid, tetraploid and hexaploid wheat. These partial *waxy* genes thus included a single gene from each diploid *T. urartu* (*wx-TuA*), *T. bicornis* (*wx-TbB*), *T. longissima* (*wx-TlB*) and *T. sharonensis* (*wx-TshB*); two different genes from each tetraploid *T. turgidum* (*wx-Td1* and *wx-Td2*) and *T. timopheevi* (*wx-Tt1* and *wx-Tt2*), and three different genes from each hexaploid *T. aestivum* (*wx-W1*, *wx-W2* and *wx-W3*) and *T. zhukovskyi* (*wx-Tz1*, *wx-Tz2* and *wx-Tz3*). The sequences of the above fourteen *waxy* genes were aligned with the sequences of the corresponding regions of the *waxy* genes of *T. monococcum* (*wx-TmA*) and *T. speltoides* (*wx-TsB*) (Chapter 4) and *T. tauschii* (*wx-TtD*) (R. Appels, pers. commun.), as shown in Fig. 6.1.

6.5 Variation in length of intron 4 in *waxy* genes

The partial *waxy* genes from the different types of wheat exhibited identical lengths of the sections of exon 4 (143 bp) and exon 5 (97 bp). The length of intron 4 appeared variable, ranging from 125 bp for most of the *waxy* genes to 152 bp for *wx-W3* of *T. aestivum* (Table 6.2). The GT/AG junction of intron 4 of the *waxy* genes in this study was deduced from that of the *waxy* gene of barley (Rohde *et al.* 1988). The variation in length of intron 4 allowed rapid identification of some *waxy* genes from the others. For example, the length of intron 4 of *T. monococcum* and *T. tauschii* was unique. The *wx-Td1* of *T. turgidum*, *wx-W1* of *T. aestivum* and *wx-Tz3* of *T. zhukovskyi* all had intron 4 of the same length (133 bp), different from other *waxy* genes. The sequences of the *wx-Tt1 of T. timopheevi* and *wx-Tz1* of *T. zhukovskyi* were clearly different from those of the *wx-TsB* of *T. speltoides* and the other eight *waxy* genes (Fig. 6.1), although they all had a 125 bp intron 4. The relationship of these different *waxy* genes was further analyzed at the DNA sequence level, as follows.

```
          1                                                          60
wx-TtD    GACCAAGGAG AAGATCTACG GGCCCGACGC CGGCACGGAC TACGAGGACA ACCAGCAGCG
 wx-W3    GACCAAGGAG AAGATCTACG GGCCCGACGC CGGCACGGAC TACGAGGACA ACCAGCAGCG
wx-Tt1    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACGGAC TACGAGGACA ACCAGCTACG
wx-Tz1    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACTGAC TACGAGGACA ACCAGCTACG
wx-TmA    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACGGAC TACGAGGACA ACCAGCTACG
wx-Td1    GACCAAGGAG AAGATCTATG GGCCCGATGC CGGCACGGAC TACGAGGACA ACCAGCTACG
 wx-W1    GACCAAGGAG AAGATCTATG GGCCCGATGC CGGCACGGAC TACGAGGACA ACCAGCTACG
wx-Tz3    GACCAAGGGG AAGATCTATG GGCCCGATGC CGGCACGGAC TACGAGGACA ACCAGCTACG
wx-TbB    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACTGAC TACGAGGACA ACCAGTAGCG
wx-TuA    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACTGAC TACGAGGACA ACCAGTAGCG
wx-Tz2    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACTGAC TACGAGGACA ACCAGTAGCG
 wx-W2    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACGGAC TACGAGGACA ACCAGTAGCG
wx-Tt2    GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACTGAC TACGAGGACA ACAAGTAGCG
wx-Td2    GACCAAGGAG AAGATCTATG GACCCGACGC CGGCACCGAC TACGAGGACA ACCAGCAGCG
wx-TlB    GACCAAGGAG AAGATCTATG GGCCCGATGC CGGCACGGAC TACGAGGACA ACCAGCTACG
wx-TsB    GACCAAGGAG AAGATCTACG GGCCCGACGC CGGCACCGAC TACGAGGACA ACCAGCTACG
wx-TshB   GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACCGAC TACGAGGACA GCCAGCTACG

          61                                                         120
wx-TtD    CTTCAGCCTT CTCTGCCAGG CGGCGCTGGA AGTGCCCAGG ATCCTGAACC TCGACAATAA
 wx-W3    CTTCAGCCTT CTCTGCCAGG CGGCGCTGGA AGTGCCCAGG ATCCTGAACC TCGACAATAA
wx-Tt1    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-Tz1    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-TmA    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-Td1    CTTCAGCCTG CTCTGCCAGG CAGCGCTTGA GGCACCCAGG ATCCTCGACC TCAACAACAA
 wx-W1    CTTCAGCCTG CTCTGCCAGG CAGCGCTTGA GGCACCCAGG ATCCTCGACC TCAACAACAA
wx-Tz3    CTTCAGCCTG CTCTGCCAGG CAGCGCTTGA GGCACCCAGG ATCCTCGACC TCAACAACAA
wx-TbB    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-TuA    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-Tz2    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
 wx-W2    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-Tt2    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC CCAACAACAA
wx-Td2    CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-TlB    CTTCAGCCTG CTCTGCCAGG CAGCGCTTGA GGTGCCCAGG ATCCTCGACC TCAACAACAA
wx-TsB    CTTCAGCCTC CTCTGCCAGG CAGCGCTGGA AGCACCCAGG ATCCTCGACC TCAACAACAA
wx-TshB   CTTCAGCCTG CTCTGCCAGG CAGCGCTTGA GGCACCCAGG ATCCTCGACC TCAACAACAA
```

Fig. 6.1 *to be continued*

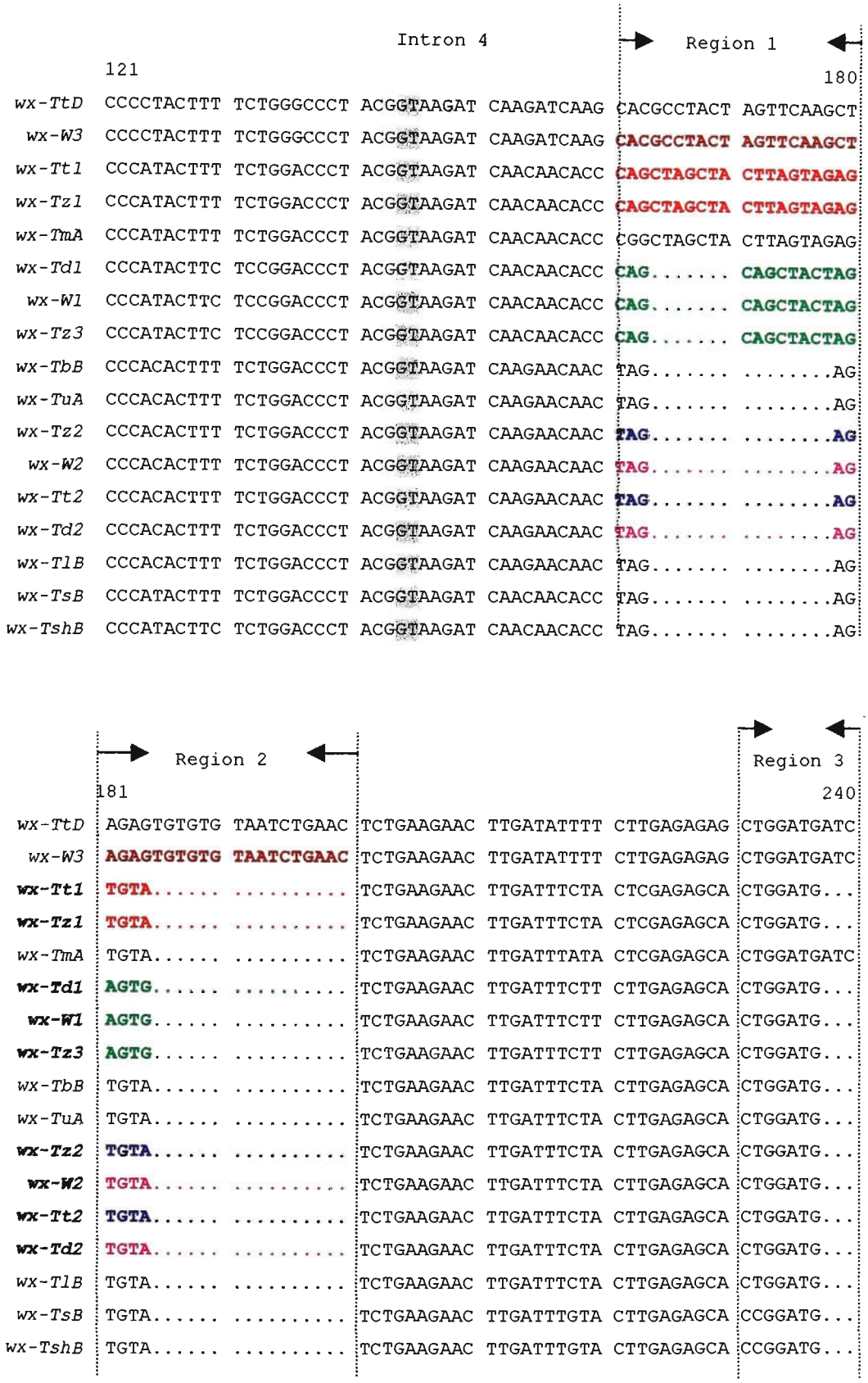


Fig. 6.1 *to be continued*

```
                                                                  Region 4
          241                                                                 300
wx-TtD    ACCAT..TTT TTTTTGTATC TGGGTGCCGT CGTCGTCCCT TGTTGCG... ..CGCCGCG.
 wx-W3    ACCAT..TTT TTTTTGTATC TGGGTGCCGT CGTCGTCCCT TGTTGCG... ..CGCCGCGC
wx-Tt1    ATCATCATTT TCCTTGTACC TGGGTGCTGC CATGCCGT.. .......... ...GCCGCGC
wx-Tz1    ATCATCATTT TCCTTGTACC TGGGTGCTGC CATGCCGT.. .......... ...GCCGCGC
wx-TmA    ATCATCATTT TCCTTGTACC TGGGTGCTGC CATGCCGTGC CGTGCCGCGT CGCGCCGCGC
wx-Td1    ATTATCATCT TCCTTGTGTC TTGGTGCTGC CACGCCATGC TATGCCGCGC CACGCCGCGC
 wx-W1    ATTATCATCT TCCTTGTGTC TTGGTGCTGC CACGCCATGC TATGCCGCGC CACGCCGCGC
wx-Tz3    ATTATCATCT TCCTTGTGTC TTGGTGCTGC CACGCCATGC TATGCCGCGC CACGCCGCGC
wx-TbB    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
wx-TuA    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
wx-Tz2    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
 wx-W2    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
wx-Tt2    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
wx-Td2    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
wx-TlB    ATTATCATCT TCCTTGTATC TTGGTGCTGC CATGCTATGC CGTGCCGTGC CGCGCCGCGC
wx-TsB    ATCATCATTC TCCTTGTATC TTGGTGCTGC CACGCCATGC TGTGCCGCGC CGCGCCGCGC
wx-TshB   ATCATCATCT TCCTTGTGTC TTGGTGCTGC CATGCCGTGC TGTGCCGCGC CGCGCCGCGC

          Exon 5                                                              360
wx-TtD    AGGGGAGGAC GTGGTGTTCG TGTGCAATGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
 wx-W3    AGGGGAGGAC GTGGTGTTCG TGTGCAATGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Tt1    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Tz1    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-TmA    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Td1    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
 wx-W1    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Tz3    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-TbB    AGGGGAAGAC GTGGTGTTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-TuA    AGGGGAAGAC GTGGTGTTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Tz2    AGGGGAAGAC GTGGTGTTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
 wx-W2    AGGGGAAGAC GTGGTGTTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Tt2    AGGGGAAGAC GTGGTGTTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-Td2    AGGGGAAGAC GTGGTGTTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-TlB    AGGGGAAGAC GTGGTGCTTG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
wx-TsB    AGGGGAAGAC GTGGTGTTCG TGTGCAACGA CTGGCACACA GGCCTTCTGG CCTGCTACCT
wx-TshB   AGGGGAAGAC GTGGTGCTCG TGTGCAACGA CTGGCACACG GGCCTTCTGG CCTGCTACCT
```

Fig. 6.1 *to be continued*

```
          361                                        399
wx-TtD    CAAGAGCAAC TACCAGTCCA ATGGCATCTA CAGGGCCGC
 wx-W3    CAAGAGCAAC TACCAGTCCA ATGGCATCTA CAGGGCCGC
wx-Tt1    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-Tz1    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-TmA    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-Td1    CAAGAGCAAC TACCAGTCCA GTGGCATCTA TAGGACGGC
 wx-W1    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-Tz3    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-TbB    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-TuA    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-Tz2    CAAGAGCAAC TACCAGCCCA ATGGCATCTA TAGGACGGC
 wx-W2    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-Tt2    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-Td2    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-TlB    CAAGAGCAAC TACCAGTCCA ATGGCATCTA TAGGACGGC
wx-TsB    CAAGAGCAAC TACCAGTCCA ATGGCATCTA CAGCACGGC
wx-TshB   CAAGAGCAAC TACCAGCCCA ATGGCATCTA TAGGACGGC
```

Fig. 6.1 Alignment of partial *waxy* genes from different types of wheat. The sequences of the partial *waxy* genes from different types of wheat were aligned using the Pileup program of ANGIS (2.6.3). These *waxy* gene fragments cover major sections of exons 4 and 5 and all of intron 4. The GT and AG splice junction of intron 4 is shaded. Short repeated DNA sequence motifs are in bold. Some sequences specific to individual *waxy* genes from different genomes of tetraploid and hexaploid wheat are shown in colour print. Such genome-specific sequences shared between different polyploid wheat are indicated the same colour. The four highly variable regions in intron 4 are indicated. Except for *wx-TtD* that was provided by R. Appels (CSIRO, Canberra), the other sixteen gene sequences were obtained in this project and their designations were shown in Table 6.2.

Table 6.2 The length of partial *waxy* genes from different genomes of wheat

waxy genes	Exon 4	Intron 4	Exon5	Total
T. monococcum (wx-TmA)	143	143	97	383
T. urartu (wx-TuA)	143	125	97	365
T. bicornis (wx-TbB)	143	125	97	365
T. longissima (wx-TlB)	143	125	97	365
T. sharonensis (wx-TshB)	143	125	97	365
T. speltoides (wx-TsB)	143	125	97	365
T. tauschii (wx-TtD)	143	151	97	391
T. turgidum-I (wx-Td1)	143	133	97	373
T. turgidum-II (wx-Td2)	143	125	97	365
T. aestivum-I (wx-W1)	143	133	97	373
T. aestivum-II (wx-W2)	143	125	97	365
T. aestivum-III (wx-W3)	143	152	97	392
T. timopheevii-I (wx-Tt1)	143	125	97	365
T. timopheevii-II (wx-Tt2)	143	125	97	365
T. zhukovskyi-I (wx-Tz1)	143	125	97	365
T. zhukovskyi-II (wx-Tz2)	143	125	97	365
T. zhukovskyi-III (wx-Tz3)	143	133	97	373

The length of exons 4 and 5 and intron 4 are deduced from the sequences of the respective *waxy* genes (Fig. 6.1). The AG/GT junction of intron of the *waxy* genes was based on that of the *waxy* gene of barley (Rohde *et al.*, 1988).

6.6 Sequences specific for the different *waxy* genes

6.6.1 Genome-specific sequences arising from insertions/deletions of intron 4

The variation in length of intron 4 among the different *waxy* genes appeared to have resulted mainly from insertions/deletions of 1-16 bp. The observed difference allowed identification of certain specific sequences representing the unique genomes of different wheat. As shown in Fig. 6.1, from positions 161 to 200, the three individual genes of *T. aestivum*, *wx-W1*, *wx-W2* and *wx-W3*, had insertions of 17 bp, 9 bp and 40 bp in the regions 1 and 2 of intron 4 together, respectively. The *wx-Tz1*, *wx-Tz2* and *wx-Tz3* of *T. zhukovskyi*, had 24 bp, 9 bp and 17 bp in the corresponding regions. In region 3, *wx-W3* of *T. aestivum* had three nucleotides (ATC) more than the other *waxy* genes in polyploid wheat. Specific sequences representing all three *waxy* genes, *wx-W3 of T. aestivum*, *wx-Tt1* of *T. timopheevi* and *wx-Tz1* of *T. zhukovskyi*, could be detected in region 4, resulting from minor deletions, relative to other *waxy* genes in these polyploid wheat.

6.6.2 Short repeated sequences mainly responding for variations in length of intron 4

The insertions/deletions of nucleotides in intron 4 were related to short repeated sequences in three main variable regions (Fig. 6.1). In region 1, the genes *wx-W1*, *wx-Td1* and *wx-Tz3* which had a 133 bp intron 4, had two AGC motifs in the 8 bp insertion that did not occur in *wx-TbB* and other eight *waxy* genes which had 125 bp intron 4. In region 2, *wx-TtD* and *wx-W3* had four copies of the nucleotide repeat GT in the 16 bp insertion, relative to other fifteen *waxy* genes. In region 4, *wx-TtD*, *wx-W3*, *wx-Tt1* and *wx-Tz1* appeared to have a deletion, as they had only one GCCGCG motif, while some others had two of these motifs (*wx-TmA*, *wx-Tz3*, *wx-W1* and *wx-Td1*), or three GCCGC motifs (*wx-TsB* and *wx-TshB*), or two GCCGC motifs in the other eight *waxy* genes.

6.7 Identities of the different partial *waxy* genes

6.7.1 Variation in sequences of the *waxy* genes from genomes of the *T. turgidum-T. aestivum* lineage

All seventeen *waxy* genes in Fig. 6.2 were analyzed in pairwise combinations for their sequence identity. In the *T. turgidum-T. aestivum* lineage of wheat, comparison of the two *waxy* genes in *T. turgidum* with the three *waxy* genes in *T. aestivum* (Table 6.3) showed that the *wx-Td1* had the best identity to the *wx-W1* and that the *wx-Td2* had the best identity to the *wx-W2*. Particularly, at positions 161 to 180 in region 1 of intron 4, *wx-Td1* and *wx-W1* contained the same specific sequence section, CAGCTACT, which could not be found in other *waxy* genes of *T. turgidum-T. aestivum* wheat. In this variable region, *wx-Td2* and *wx-W2* were identical in length. Therefore, *wx-Td1* and *wx-Td2* of *T. turgidum* most likely corresponded to the *wx-W1* and *wx-W2* of *T. aestivum*, respectively. Minor sequence variations were observed in the comparison of *wx-Td1* with *wx-W1* and *wx-Td2* with *wx-W2*. At position 381, there is a nucleotide difference between *wx-Td1* and *wx-W1*. There were three necleotide differences, at positions 22, 37 and 56, between *wx-Td2* and *wx-W2*. These differences accounted for the <100% identity between the pairs of *wx-Td1-wx-W1*, and *wx-Td2 -wx-W2* (Table 6.3).

Table 6.3 Identity (%) of the *waxy* genes between *T. turgidum* and *T. aestivum*

Genes	*wx-Td1*	*wx-Td2*	*wx-W1*	*wx-W2*	*wx-W3*
wx-Td1	100				
wx-Td2	92.60	100			
wx-W1	**99.73**	92.33	100		
wx-W2	92.88	**99.20**	93.15	100	
wx-W3	86.38	89.11	86.65	89.39	100

wx-Td1 and *wx-Td2*: the two *waxy* genes from *T. turgidum* ssp. *durum*; *wx-W1*, *wx-W2* and *wx-W3*: the three *waxy* genes from *T. aestivum*. The identities of sequences were determined by using the Bestfit program of ANGIS (section 2.6.2). The significantly high identities observed between certain pairs are shown in bold.

6.7.2 Variation in sequences of the *waxy* genes from genomes of the *T. timopheevi-T. zhukovskyi* lineage

The sequence identity of *wx-Tt1* of *T. timopheevi* to *wx-Tz1* of *T. zhukovskyi* was significantly higher than that to *wx-Tz2* and *wx-Tz3* of *T. zhukovskyi* (Table 6.4); there was only one nucleotide difference between *wx-Tt1* and *Tz1*, at position 37 in exon 4 (Fig. 6.1). The sequence identity of *wx-Tt2* to *wx-Tz2* was significantly higher than that to *wx-Tz1* and *wx-Tz3*; there was a difference of only two nucleotides between *wx-Tt2* and *wx-Tz2*, at position of 111 in exon 4 and position 377 in exon 5.

6.7.3 Relationship of the *waxy* genes between *T. turgidum-T.aestivum* and *T. timopheevi-T. zhukovskyi* lineages

To determine the possible relationship of genomes between the polyploid *T. turgidum-T. aestivum* and *T. timopheevi-T. zhukovskyi* lineages, the sequences of these *waxy* genes were compared (Table 6.5). The results indicated that both *wx-Tt2* of *T. timopheevi* and *wx-Tz2* of *T. zhukovskyi* were more related to *wx-Td2* of *T. turgidum* and *wx-W2* of *T. aestivum*, compared to the other *waxy* genes of these polyploid wheat. The *wx-Tz3* of *T. zhukovskyi* had the highest identity to *wx-Td1* of *T. turgidum* and *wx-W1* of *T. aestivum* among all the *waxy* genes of these polyploid wheat. A specific sequence in region 1 of intron 4, CAGCTACT, present in the *wx-Td1* and *wx-W1*, was also detected in the *wx-Tz3* but not in any of the other *waxy* genes (Fig. 6.1).

Table 6.4 Identity (%) of the *waxy* genes between *T. timopheevi* and *T. zhukovskyi*

Genes	*wx-Tt1*	*wx-Tt2*	*wx-Tz1*	*wx-Tz2*	*wx-Tz3*
wx-Tt1	100				
wx-Tt2	95.43	100			
wx-Tz1	**99.73**	95.71	100		
wx-Tz2	95.71	**99.18**	96.00	100	
wx-Tz3	94.69	92.06	94.41	92.33	100

wx-Tt1 and *wx-Tt2*: the two *waxy* genes from *T. timopheevi*; *wx-Tz1*, *wx-Tz2* and *wx-Tz3*: the three *waxy* genes from *T. zhukovskyi*. The identities of sequences were determined by using the Bestfit program of ANGIS (section 2.6.2). The significantly high identities observed between certain pairs are shown in bold.

Table 6.5 Identity (%) of the *waxy* genes between the *T. turgidum-T. aestivum* and *T. timopheevi-T. zhukovskyi* lineages

Genes	*wx-Td1*	*wx-Td2*	*wx-W1*	*wx-W2*	*wx-W3*
wx-Tt1	94.69	96.00	92.33	96.29	87.67
wx-Tt2	92.06	**98.63**	94.97	**99.18**	88.55
wx-Tz1	94.41	96.00	94.69	96.00	87.40
wx-Tz2	92.33	**98.90**	92.60	**99.45**	88.83
wx-Tz3	**99.46**	92.60	**99.73**	92.88	86.38

The identities of sequences between different *waxy* genes of *T. turgidum* (*wx-Td1* and *wx-Td2*)-*T. aestivum* (*wx-W1*, *wx-W2* and *wx-W3*) and *T. timopheevi* (*wx-Tt1* and *wx-Tt2*)-*T. zhukovskyi* (*wx-Tz1*, *wx-Tz2* and *wx-Tz3*) were determined by using the Bestfit program of ANGIS (section 2.6.2). The significantly high identities between certain pairs are shown in bold.

6.7.4 Relationship of the *waxy* genes between diploid and polyploid wheat

The relationships of genomes of polyploid and diploid wheat were analyzed utilizing the sequence identity of various partial *waxy* genes, given in Table 6.6. The following results were revealed:

1. The *wx-W3* fragment had 100% identity to the corresponding section of the *waxy* gene in *T. tauschii* and there was only a 1-base deletion in intron 4 of *T. tauschii*, relative to the *wx-W3*. However, the identity of *wx-W3* to all other diploid *waxy* genes was <89%, and the identity of *wx-W1* and *wx-W2* to the *waxy* gene of *T. tauschii* was 86.6% and 88.6%, respectively. Moreover, a total 26 bp and 34 bp deletion in three regions were detected in *wx-W1* and *wx-W2*, respectively, relative to the *waxy* gene in *T. tauschii* (Fig. 6.1).

2. The *wx-Td2* of *T. turgidum* and *wx-W2* of *T. aestivum* were more homologous to *wx-TuA* of *T. urartu* and *wx-TbB* of *T. bicornis* (>99%) than to the *waxy* genes of other diploid wheat species.

3. The sequence identity of *wx-Tt1* of *T. timopheevi* and *wx-Tz1* of *T. zhukovskyi* was much higher to *wx-TmA* of *T. monococcum* (>99%) than to the *waxy* genes of other diploid wheat species.

4. The *wx-Tt2* of *T. timopheevi* and *wx-Tz2* of *T. zhukovskyi* had much higher identity to *wx-TuA* of *T. urartu* and *wx-TbB* of *T. bicornis* (>99%) than to the *waxy* genes of other diploid wheat species.

5. The remaining *waxy* genes in the polyploid wheat, i.e., *wx-Td1* of *T. turgidum*, *wx-W1* of *T. aestivum* and *wx-Tz3* of *T. zhukovskyi*, had 92-98% identity to the *waxy* genes of all tested diploid wheat, except to the *wx-TtD* of *T. tauschii* (<89%).

Table 6.6 Identity (%) of the *waxy* genes between polyploid and diploid wheat

Genes	*wx-TmA*	*wx-TuA*	*wx-TbB*	*wx-TlB*	*wx-TsB*	*wx-TshB*	*wx-TtD*
wx-Td1	93.57	92.60	92.60	94.24	93.43	95.07	86.34
wx-Td2	95.34	**99.18**	**99.18**	97.81	92.88	93.70	88.27
wx-W1	93.83	92.88	92.88	94.52	93.70	95.34	86.61
wx-W2	95.62	**99.73**	**99.73**	98.08	92.60	93.43	88.55
wx-W3	88.53	89.11	89.11	88.55	88.06	86.38	**100.00**
wx-Tt1	**99.45**	96.00	96.00	96.00	94.57	95.43	87.64
wx-Tt2	94.80	**99.45**	**99.45**	97.26	92.06	92.88	87.71
wx-Tz1	**99.18**	96.29	96.29	95.71	94.57	95.43	87.36
wx-Tz2	95.07	**99.73**	**99.73**	97.53	92.33	93.70	87.99
wx-Tz3	93.57	92.60	92.60	94.25	93.43	95.07	86.34

The gene designations are as in Table 6.2: two genes (*wx-Td1* and *wx-Td2*) for *T. turgidum*; three genes (*wx-W1*, *wx-W2* and *wx-W3*) for *T. aestivum*; two genes (*wx-Tt1* and *wx-Tt2*) for *T. timopheevi*; three genes (*wx-Tz1*, *wx-Tz2* and *wx-Tz3*) for *T. zhukovskyi*; and *wx-TmA*, *wx-TuA*, *wx-TbB*, *wx-TlB*, *wx-TsB*, *wx-TshB* and *wx-TtD* for *T. monococcum*, *T. urartu*, *T. bicornis*, *T. longissima*, *T. speltoides*, *T. sharonensis* and *T. tauschii*, respectively. The identities of sequences between the ten *waxy* genes of polyploid and seven *waxy* genes of diploid wheat were obtained by using the Bestfit program of ANGIS (section 2.6.2). The significantly high identities shared between certain pairs are shown in bold.

6.7.5 Relationship of the *waxy* genes between different diploid wheat

The sequence identity of the *waxy* genes between different diploid wheat (Table 6.7) showed that *T. tauschii* had a unique sequence for the *waxy* gene, since its identity to all other *waxy* genes in diploid wheat was <89%. The two possible candidates for being the donor of A genome to polyploid wheat, *T. monococcum* and *T. urartu*, had 95% identity to each other (Table 6.7), and thus they were considered to be two different *waxy* genes (Fig. 6.2). This is consistent with our results showing that there were three gaps in the sequences of intron 4 between *T. monococcum* and *T. urartu* (Fig. 6.1). The four species of the *Sitopsis* section, *T. speltoides*, *T. sharonensis*, *T. bicornis* and *T. longissima*, had significant variations in the sequences of their *waxy* genes, with identities ranging from 92% to 98%. These genes had the same length of the sections of exons 4 and 5 covered in the clones, and also of intron 4; the sequence variation observed resulted from nucleotide substitutions, mainly in exon 4 (Fig. 6.1). Unexpectedly, *T. urartu* and *T. bicornis* had completely identical sequences.

6.8 Phylogenetic tree of different genomes of wheat

To determine the phylogenetic relationships of genomes in the different types of wheat, a phylogenetic tree was constructed using the method described in section 2. 6.8. The genome types of polyploid wheat were designated according to the *waxy* genes on different genomes (Fig. 6.2). Two clearly distinct clusters were identified in this phylogenetic tree. Cluster 1 included genomes II of *T. turgidum*, *T. aestivum*, *T. timopheevi* and *T. zhukovskyi* and the three diploid wheat, *T. urartu*, *T. bicornis* and *T. longissima*. Cluster 2 included genomes I of *T. turgidum* and *T. aestivum*, genome III of *T. zhukovskyi* and the diploids *T. speltoides* and *T. sharonensis*. The basal node in the tree separated *T. tauschii* and genome III of *T. aestivum* from the rest of the species. The remaining genomes, genomes I of *T. timopheevi* and *T. zhukovskyi* were more related to *T. monococcum*.

Table 6.7 Identity (%) of the *waxy* genes between different diploid wheat

Genes	*wx-TmA*	*wx-TuA*	*wx-TbB*	*wx-TlB*	*wx-TsB*	*wx-TshB*	*wx-TtD*
wx-TmA	100						
wx-TuA	95.34	100					
wx-TbB	95.34	100.00	100				
wx-TlB	95.34	97.81	97.81	100			
wx-TsB	94.25	92.60	92.60	93.15	100		
wx-TshB	95.34	93.43	93.43	94.80	95.62	100	
wx-TtD	88.50	88.27	88.27	87.71	86.91	85.24	100

The identities of sequences were determined by using the Bestfit program of ANGIS (section 2.6.2). The gene designations are as in Table 6.2.

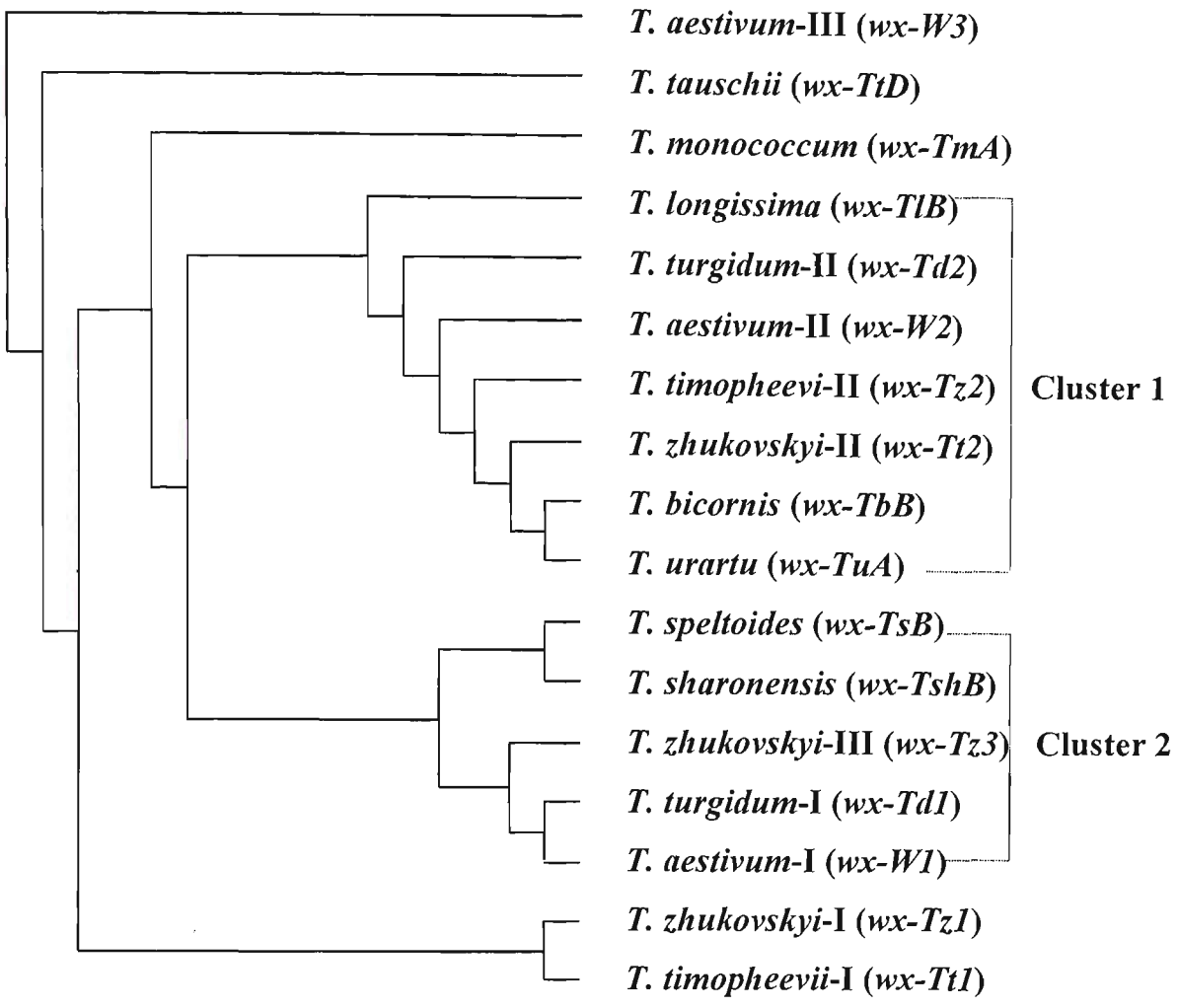


Fig. 6.2 The phylogenic tree of genomes of different types of wheat based on identities of the partial *waxy* genes. The phylogenetic tree was constructed using the Eneighbour program (section 2.6.8).

6.9 Cloning of the exon 1-exon 4 section of the *waxy* gene in *T. turgidum*

The *wx-Td1* of *T. turgidum* had <95% identity to the *waxy* genes in all the tested diploid wheat (Table 6.6). The donor of diploid progenitor to *T. turgidum*, could therefore, not be determined only on the basis of these sequence identity. To obtain a greater understanding of the relatedness between the two genes (*wx-Td1* and *wx-Td2*) in this tetraploid wheat and compare these with those of diploid *T. monococcum* and *T. speltoides* discussed in Chapter 4, the primers F1 and R1, which were then used to clone Fragment 1 of the *waxy* gene of *T. monococcum* and *T. speltoides* (section 2.3.3.1), were used to clone this fragment from the two *waxy* genes in *T. turgidum* ssp. *durum*. Three positive clones, du 6, du11 and du 33, were obtained and identified to show that all inserts had the same sequences. The insert of one of these (du11) was sequenced completely; it was found to be 938 bp in length. The sequence of du11 was identical to that of *T. speltoides* in both the exons and the introns included this region (Fig. 6.3), but it was different from *T. monococcum* (data not shown).

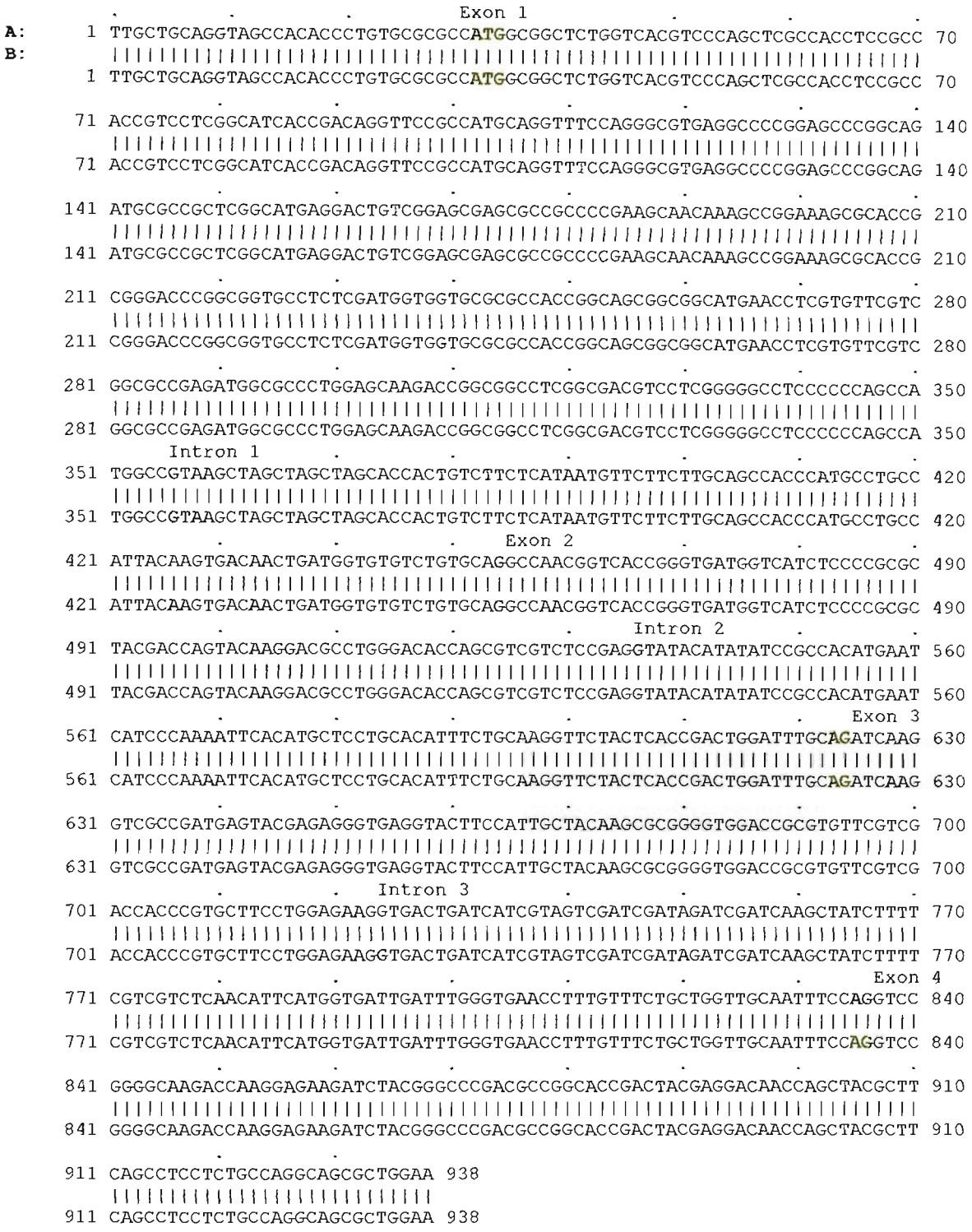


Fig. 6.3 Alignment of sequences of a partial *waxy* gene of *T. turgidum* and the corresponding region of the *waxy* gene of *T. speltoides*. Line A indicates the sequence of Fragment 1 of a *waxy* gene of *T. turgidum*; and line B indicates the sequence of the corresponding section of the *waxy* gene of *T. speltoides* (*wx-TsB*, Fig. 4.3). The GT/AG junction of introns and the first translation initiation codon in exon 1 are shaded. The two sequences were aligned using the Bestfit program of ANGIS (section 2.6.2) and indicate 100% identity.

6.10 Discussion

6.10.1 Sequences of the partial *waxy* genes representing different genomes of polyploid wheat

In this study, a series of the partial *waxy* genes from different diploid, tetraploid and hexaploid wheat have been cloned, sequenced and analyzed. The results indicated variations in lengths and sequences of the fragments cloned in this study, particularly in the intron region, allowing identification of the two individual genomes in tetraploid and the three individual genomes in hexaploid wheat. The genome-specific sequences were found to result from small insertions/deletions in intron 4 (Fig. 6.1). Moreover, these variations in intron 4 could be considered as phylogeneticallly informative, as these specific sections appeared to have been stable during the evolution of wheat from tetraploid *T. turgidum* to hexaploid *T. aestivum* and from tetraploid *T. timopheevi* to hexaploid *T. zhukovskyi*. These specific sequences representing different genomes of polyploid wheat could be utilized to develop molecular tools for identifying different mutants of the *waxy* genes in wheat breeding programs and distinguishing efficiently between different genomes for further understanding of the evolution of wheat.

The diploid species could be distinguished by insertions/deletions in intron 4 or minor replacement of nucleotides in the section cloned here, with the exception of *T. urartu* and *T. bicornis*, which had identical sequences of the section (Fig. 6.1). This exception might be due to the conservation of sequences in this section between these species. This needs to be confirmed by further investigating different accessions of these two species. The most significant observation was that *T. tauschii* was different from other diploids in length and sequence of the fragment cloned in this study. A difference in length of intron 4 between *T. monococcum* and *T. urartu*, the two proposed donors of A genome, has been also identified here. This result thus provides a tool for identification of A and A-like genomes superior to that described in previous reports (Jones *et al.*, 1982; Jones and Mak 1983; Kerby *et al.*, 1988). These investigations had indicated that *T. monococcum* and *T. urartu* contained the same sequences of the ß-purothionin form specified by *T. turgidum* and *T. aestivum*, and thus both *T. monococcum* and *T. urartu* might be possible donors of A genome to these polyploid wheat. The present work could thus offer a clearer understanding and

comparison of these two closely related species. The identical length of intron 4 was found in the four closely related diploid species in the *Sitopsis* section. The point mutations of nucleotides were detected in both exons and introns ; however, the sequence information obtained here is not sufficient to resolve readily the differences between these.

The extensive sequence data collected here on the *waxy* genes of the different types of wheat made it possible to construct a phylogenetic tree to analyse the evolutionary relationship of genomes between the various diploid species and the individual genomes of polyploid species. When the data regarding the genome-specific sequences and analysis of the phylogenetic tree were taken together, the following observations could be made and conclusions drawn.

6.10.2 Origin of the genomes of *T. aestivum*

6.10.2.1 The third set of genomes and the tetraploid progenitor of *T. aestivum*

The *wx-Td1* and *wx-Td2* of *T. turgidum* corresponded to *wx-W1* and *wx-W2* of *T. aestivum*, based on their sequence identities and specific sequences in intron 4 (Fig. 6.1, Table 6.3). This suggests that the genomes containing the *wx-W1* and *wx-W2* genes in *T. aestivum* would have originated from those containing *wx-Td1* and *wx-Td2* in *T. turgidum*, respectively, and that the *wx-W3* was on the third set of genomes in *T. aestivum*.

The *wx-W3* of *T. aestivum* had the highest identity to the *wx-TtD* of *T. tauschii*, compared to other diploid wheat (Table 6.6), indicating that the *wx-W3* might represent the *waxy* gene on D genome of *T. aestivum* and that a *T. tauschii* accession with this sequence is the most likely donor of the D genome of *T. aestivum*. This observation thus strongly supports previous investigations (Lagudah *et al.*, 1991; Hohmann *et al.*, 1993; Gill *et al.*, 1991; Gill *et al.*, 1993).

6.10.2.2 Origin of B genome

The *waxy* genes of all tested diploid species in the *Sitopsis* section tested here, which are proposed to be the donor of B genome of *T. turgidum* and *T. aestivum*, had

identical length of intron 4. It is thus difficult to determine which of these diploid species is the more likely donor of B genome. The *waxy* gene of another diploid species in the *Sitopsis* section, *T. searsii*, had only approximately 97% identity to the *wx-B* gene in *T. turgidum-T. aestivum* (Murai *et al.*, 1998), indicating that *T. searsii* might not be the donor of B genome.

The analysis of the phylogenetic tree (Fig. 6.2) showed that *wx-Td1* of *T. turgidum* and *wx-W1* of *T. aestivum* belong to the same cluster as *T. speltoides* and *T. sharonensis*, implying that *T. speltoides* or/and *T. sharonensis* might be the donor(s) of one genome containing *wx-Td1* in *T. turgidum* and *wx-W1* in *T. aestivum*. As *T. speltoides* and *T. sharonensis* have been proposed as the donors of B genome (Kerby and Kuspira, 1987), *wx-Td1* in *T. turgidum* and *wx-W1* in *T. aestivum* could be suggested to be from the B genome.

In order to determine whether *T. speltoides* or *T. sharonensis* was indeed the donor of B genome, we attempted to clone fragment 1 from these two species, so it could be compared to the corresponding section of to the *waxy* gene in *T. speltoides* (Chapter 4). Unfortunately, the cloning for *T. sharonensis* was unsuccessful. However, a 930 bp fragment from the exons 1-3, intron 1-3 and partial exon 4 of one *waxy* gene of *T. turgidum* was isolated successfully. This fragment had exactly the same sequence as the corresponding region of the *waxy* gene of *T. speltoides* (Fig. 6.3), and thus *T. speltoides* was more likely to be the candidate donor of the B genome of *T. turgidum* and *T. aestivum* than *T. sharonensis*, assuming that the sequence of the *waxy* gene of *T. sharonensis* had a minor variation in the exon 1-exon 4 region. Although some reports have suggested that more than one diploid species might be the likely donor of B genome (Kerby *et al.*, 1990; Takumi 1993) and that *T. speltoides* might not be the donor of B genome (Mori *et al.*, 1997), *T. speltoides* has been suggested as the donor of B genome by a majority of recent reports (Gill and Appels, 1988; Miyashita *et al.*, 1994; Talbert *et al.*, 1991 and 1995; Dau and Gustafson, 1996; Badaeva *et al.*, 1996; Ohsako *et al.*, 1996; Sasanuma *et al.* 1996; Blake *et al.*, 1998). The incomplete identity of the exon 4-intron 4-exon 5 region of the *waxy* gene between *T. speltoides* and *T. turgidum-T. aestivum* observed here indicates that this region has probably undergone significant variations compared to the D genome, during the evolution of B genome of wheat from diploid through tetraploid to hexaploid.

6.10.2.3 Origin of A genome

Based on the above suggestions that *wx-Td1* of *T. turgidum* and *wx-W1* of *T. aestivum* might be from B genomes and *wx-W3* might be from the D genome of *T. aestivum* (see section 6.10.2.2), *wx-Td2* of *T. turgidum* and *wx-W2* of *T. aestivum* should have derived from the A genome. In the phylogenetic tree, only *T. urartu*, a putative candidate of A genome, fell into the same cluster as *wx-Td2* of *T. turgidum* and *wx-W2* of *T. aestivum*. The *wx-TmA* of *T. monococcum*, the other proposed donor of A genome, was significantly different in sequences from *wx-Td2* of *T. turgidum* and *wx-W2* of *T. aestivum* and in a different cluster of the phylogenetic tree from *wx-Td2* and *wx-W2*. The identity of *wx-TmA* of *T. monococcum* to *wx-Td2* and *wx-W2* was approximately 95%; much less the 99% identity of *wx-TuA* of *T. urartu* to *wx-Td2* and *wx-TW2* (Table 6.6). These observations suggest that *T. urartu* and *T. monococcum* had different *waxy* genes and that only *T. urartu* might be the donor of A genome to *T. turgidum* and *T. aestivum*. This suggestion is consistent with the current theory of *T. urartu* being the donor of A genome to *T. turgidum* and *T. aestivum* (Nishikawa, 1983; Dvorak *et al.*, 1993; Takumi *et al.*, 1993; Jiang and Gill, 1994; Friebe and Gill, 1996; Randhawa *et al.*, 1997; Jaaska, 1997).

6.10.3 Origin of the genomes of *T. zhukovskyi*

6.10.3.1 The third set of genome and the tetraploid progenitor of *T. zhukovskyi*

The *wx-Tz1* and *wx-Tz2* of *T. zhukovskyi* showed the highest identity to *wx-Tt1* and *wx-Tt2* of *T. timopheevi*, respectively (Table 6.4). The *wx-Tt1–wx-Tz1* and *wx-Tt2–wx-Tz2* pairs were classified into the two separate clusters in the phylogenetic tree, respectively. These observations indicated that *wx-Tz1* and *wx-Tz2* of *T. zhukovskyi* were most likely derived from *wx-Tt1* and *wx-Tt2* of *T. timopheevi*, respectively. By deduction, the *wx-Tz3* should be on the third set of genomes in the hexaploid *T. zhukovskyi*. The phylogenetic tree indicated that the third genome containing *wx-Tz3* of *T. zhukovskyi* fell into the same cluster as *wx-Td1* of *T. turgidum* and *wx-W1* of *T. aestivum* (Fig. 6.2). Comparison of the sequences also indicated that *wx-Tz3* contained a specific sequences in intron 4, CAGCTACT, which was only detected in *wx-Td1* and *wx-W1*. *T. zhukovskyi* is a man-made hybrid of the tetraploid *T. timopheevi* and the diploid *T. monococcum*; i.e., the third set of genome of *T.*

zhukovskyi should be the A genome from *T. monococcum* (Jakubzizer, 1958, Upadhya and Swaminathan, 1963; Johnson, 1968 and 1972; Dvorak, 1991; Breiman and Graur, 1995; Friebe and Gill, 1996). However, *wx-Tz3* representing the third set of genome was significantly different from *wx-TmA of T. monococcum* in intron 4 region. Therefore, the present work suggest that intraspecies polymorphism existed in *T. monococcum* or that wild *T. monococcum* (*T. monococcum* ssp *boeoticum*) or might be the third set of genome *of T. zhukovskyi*. Any accession of *T. monococcum* that had the same sequences as *wx-Tz3* would be the best candidate of the donor of the third set of genome of *T. zhukovskyi*.

6.10.3.2 The donors of one genome to *T. timopheevi*

The *wx-Tt1* of *T. timopheevi* had greater identity to the *wx-TmA* of *T. monococcum* than to the *waxy* genes of other tested diploid species tested here (Table 6.6), although two short deletions/insertions in intron 4 were detected between them. This observation indicates that *T. monococcum* could be the donor of one genome to *T. timopheevi*, thus supporting other reports (Jakubzizer, 1958; Upadhya and Swaminathan, 1963; Johnson, 1968 and 1972; Appels *et al.*, 1988; Jaaska, 1997).

The second donor of the *T. timopheevi* could not be determined by the present study, because the single *waxy* genes from each of the three diploid wheat, *T. urartu*, *T. bicornis* and *T. longissima* were classified into the same cluster as *wx-Tt2* of *T. timopheevi*. It was not possible to determine which one of these was the more likely donor of *wx-Tt2*, although many reports have shown that *T. urartu* was one donor to *T. timopheevi* based on repeated sequence profiles (Dvorak, 1993; Friebe and Gill, 1996; Jiang and Gill, 1994; Breiman and Graur, 1995). Moreover, *wx-Tt2* should be from the G genome, because the *wx-Tt1* has been suggested to be from A genome of *T. monococcum*. *T. speltoides* (Johnson, 1975; Dvorak and Appels, 1982; Chen and Gill, 1983; Ogihara and Tsunewaki, 1988; Dvorak and Zhang, 1990; Jiang and Gill, 1994; Shands and Kimber, 1973; Jaaska, 1978; Tsunewaki and Ogihara, 1983; Talbert *et al.*, 1991) and *T. sharonensis* (Kushnir *et al.*, 1983) have been suggested to be the donor of G genome. The uncertain of G genome identified by intron 4 in this study is resulted from the fact that all the possible G genome donors had an identical length of

intron 4. Some variable regions need to be investigated to identify the donor of the G genome.

6.10.4 Intraspecies polymorphism of wheat

The DNA sequence information on the *waxy* genes obtained in this project provided some new and valuable information for phylogenetic and evolutionary relationships of different genomes of polyploid wheat. However, the origins of different genomes of wheat are complicated. The comparison of sequences of the *waxy* genes in *T. monococcum*, *T. speltoides* and *T. tauschii* constructed here and the recent reports by Mason-Gamer *et al.* (1998) indicated that significant variation in sequences of the *waxy* genes exist within *T. monococcum*, *T. speltoides* or *T. tauschii* species. These intraspecies variations were detected in both exons and introns. As discussed in section 5.3.4, the *T. speltoides* in this study was different from the *T. speltoides* studied by Fujita *et al.* (1996) based on the sequences of the partial region of the *waxy* proteins. Variations in RFLP patterns within diploid *T. monococcum* and *T. urartu* (Rodriguez-Quijiano *et al.*, 1997; Randhawa *et al.*, 1997) have been reported. Even though *T. tauschii* has been strongly suggested to contribute D genome to common wheat, some interspecies polymorphism has been found based on C-band (Gill *et al.*, 1994) and RFLP analysis (Talbert *et al.*, 1998).

Intraspecies polymorphism of common wheat has been reported by RFLP analysis (Chao *et al.*, 1989; Kam-Mogan and Gill, 1989; Liu *et al.*, 1990; Lubbers *et al.*, 1991; Frederiksen and Seberg, 1992; Monte *et al.*, 1993) and by sequence analysis of *Glu-A1* locus (Talbert *et al.*, 1998). Talbert *et al.* (1998) suggested that the D genomes in two types of common wheat were contributed by two different accessions of *T. tauschii*. These studies have thus suggested more than one progenitor of the single genome in polyploid wheat. In addition, translocation events between different genomes have been shown to exist in polyploid wheat (Chao *et al.*, 1989; Friebe and Gill, 1996). Polymorphism of genetics information in both diploid and common wheat makes it more complicated to determine the origins of the genomes of wheat. Therefore, more extensive studies on intraspecies and interspecies variation of diploid and polyploid wheat are needed to reveal the phylogenetic and evolutionary relationship of wheat.

CHAPTER 7

PRELIMINARY STUDIES TOWARDS DEVELOPING OF MOLECULAR MARKERS FOR THE *WAXY* A, B AND D LOCI AND POSSIBLE ALTERNATIVE TRANSCRIPTIONAL EVENTS OF *WAXY* GENES OF COMMON WHEAT

List of Figures in Chapter 7

Chapter 7 Development of Molecular Markers for the *waxy* A, B and D Loci and possible alternative transcriptional events of *waxy* genes of Common Wheat

ABSTRACT

Sequence analyses of the three partial genomic *waxy* genes (*wx-W1*, *wx-W2* and *wx-W3*) obtained in this project and a *waxy* cDNA (Clark *et al.*, 1991) of common wheat showed that the *wx-W2* gene had much higher identity than the *wx-W1* and *wx-W3* genes to this cDNA. However, in the reported cDNA a stretch of 33 bp encoding an additional 11 amino acids, appeared to be completely identical to a part of intron 4 in the *wx-W2*. 3 of 241 nucleotides in exons of this *wx-W2* were also different from the cDNA sequence. It is thus suggested that a possible alternative transcriptional event might be involved in expression of the *waxy* genomic gene in wheat. On the basis of the sequences of the three partial *waxy* genomic genes in common wheat, specific primers were designed and restriction enzyme sites were identified to represent the *wx-A*, *wx-B* and *wx-D* loci. A pair of primers was also applied successfully to amplify the fragment containing the variable length of intron 5 of the three *waxy* genes of common wheat in one PCR. Comparison of the amplified DNA bands between a line of normal wheat, a null-7A line and a null-4A line indicated that their patterns were identical, suggesting that the *waxy* gene had not been deleted completely in these null lines. The *waxy* locus-specific sequences, the deduced locus-specific restriction enzyme sites and the locus-specific PCR products could be developed further as molecular markers for identification of the *wx-A1*, *wx-B1* and *wx-D1* loci of hexaploid common wheat and their mutants.

7.1 Introduction

The three individual partial *waxy* genes, *wx-W1*, *wx-W2* and *wx-W3*, have been cloned from common wheat (Chapter 6). In view of the importance of *waxy* genes to starch quality (section 1.6.4) and for understanding the genetics and evolution of wheat, it is of great interest to identify any genome-specific properties of the three *waxy* genes and develop these as molecular markers for the respective *waxy* loci in common wheat. The homoeologous nature of the three genomes and the unique characteristics and contributions of the various waxy proteins to starch quality have been discussed at length in section 1.6; some of these features are revisited briefly here. Firstly, the different *waxy* genes have different extents of expression and make unequal contributions to the amylose content. The WX-B1 protein in common wheat is detected at a higher level in SDS-PAGE (Chapter 3), supporting reports that the WX-B1 protein is more abundant than WX-A1 and WX-D1 proteins (Nakamura *et al.*, 1993). Based on the analysis of the waxy proteins in monosomic lines of wheat, it has also been shown that the WX-B1 protein makes a larger contribution to the amylose content of starch than the other two proteins (WX-A1 and WX-D1) (Miura *et al.*, 1994). Miura and Sugawara (1996) reported that removal of chromosomes 4A containing the *wx-B* locus reduced amylose content twice as much as when the chromosomes carrying either *wx-A* or *wx-D* loci were removed. An understanding of the structures, functions and individual contributions of each these proteins and genes to amylose content will be obtained by the analysis of their individual biochemical and molecular properties. Identification of sequences specific to the three individual *waxy* loci in common wheat will provide the much-needed tools for cloning of the individual *waxy* genes for this purpose.

Secondly, since mutants of different low-amylose of waxy lines of common wheat have been revealed by SDS-PAGE (Nakamura *et al.*, 1993b), considerable attention has been paid to investigations and application of such lines in wheat breeding. Such mutants have been investigated from a number of countries, for example, Australia (Zhao and Sharp, 1996), US (Grayboasch, 1996; Zeng *et al.*, 1997; Grayboasch *et al.*, 1998), Canada (Demeke *et al.*, 1997) and Japan (Miura *et al.*, 1994; Yamamori *et al.*, 1994; Hoshino *et al.*, 1996; Yasui *et al.*, 1997). However, almost all mutants have been found to be partially waxy wheat, i.e., lack WX-A1 or WX-B1, or both

proteins, and only one line is reported to lack WX-D1 (Yamamori *et al.*, 1994). Numerous efforts have been made to develop completely waxy wheat, i.e., waxy protein-free wheat cultivars, by the crossing of different partial waxy lines in breeding programs in a number of countries including Japan (Nakamura *et al.*, 1995; Kiribuchi-Otobe *et al.*, 1997, 1998; Miura *et al.*, 1998), U.S.A. (Graybosch, 1998), Australia (Zhao and Sharp, 1996) and Canada (Demeke *et al.*, 1997). Molecular markers for different *waxy* loci are important for the rapid and clear identification of mutants at the early stage of wheat breeding programs. However, so far the only molecular marker for the null line of *wx-D* gene has been reported so far has used the repeated sequences at the 3' end of the *waxy* genes (Zhao *et al.*, 1998).

Finally, a number of theories have been proposed to explain the origin of different genomes of polyploid wheat (sections 1.8 and 1.9). As discussed in section 6.10.5, however, the polymorphism within and between the different diploid wheat species allows speculation for a more complicated relationship of different genomes between diploid, tetraploid and hexaploid wheat than the previous theories. The identification and analysis of sequences specific for individual *waxy* loci will provide essential tools for further understanding the intraspecies and interspecies variations of *waxy* loci and thus possible evolution of wheat.

This Chapter describes the preliminary efforts towards developing molecular markers for the *wx-A1*, *wx-B1* and *wx-D1* loci of common wheat.

7.2 Relationship of a published *waxy* cDNA and the three *waxy* genomic genes of common wheat

The three partial *waxy* genomic genes of common wheat, *wx-W1*, *wx-W2* and *wx-W3* (Chapter 6) were compared with the appropriate section of a published *waxy* cDNA of common wheat (Clark *et al.*, 1991). Sequence comparison showed that the identity of this cDNA to the *wx-W2* reached 98.9%, which was significantly higher than the 93.0% and 91.2% identity of *wx-W1* and *wx-W3* to this cDNA. The striking observation was that in this cDNA a 33 bp sequence, encoding an additional 11 amino acids, which did not exist in all other known waxy proteins in plants (Ainsworth *et al.*, 1993), was identical to sequences at the 3' end of intron 4 of *wx-W2* (Fig. 7.1). Six nucleotides of the *wx-W1* and 17 nucleotides of the *wx-W3* were different from the sequences of this cDNA in this 33 bp 'insert'. These comparisons indicated that *wx-W2* might be the genomic copy of the *waxy* gene corresponding to this cDNA and the 33 bp 'insert' is likely from a part of intron 4 of the *wx-W2* gene.

Further comparison of sequences of *wx-W2* and this cDNA showed that there were three different nucleotides, at positions 22, 47 and 56 of exon 4, between them (Fig. 7.1). It should be noted that the reported cDNA and the three partial genomic genes all were cloned from the same variety, Chinese Spring, in common wheat.

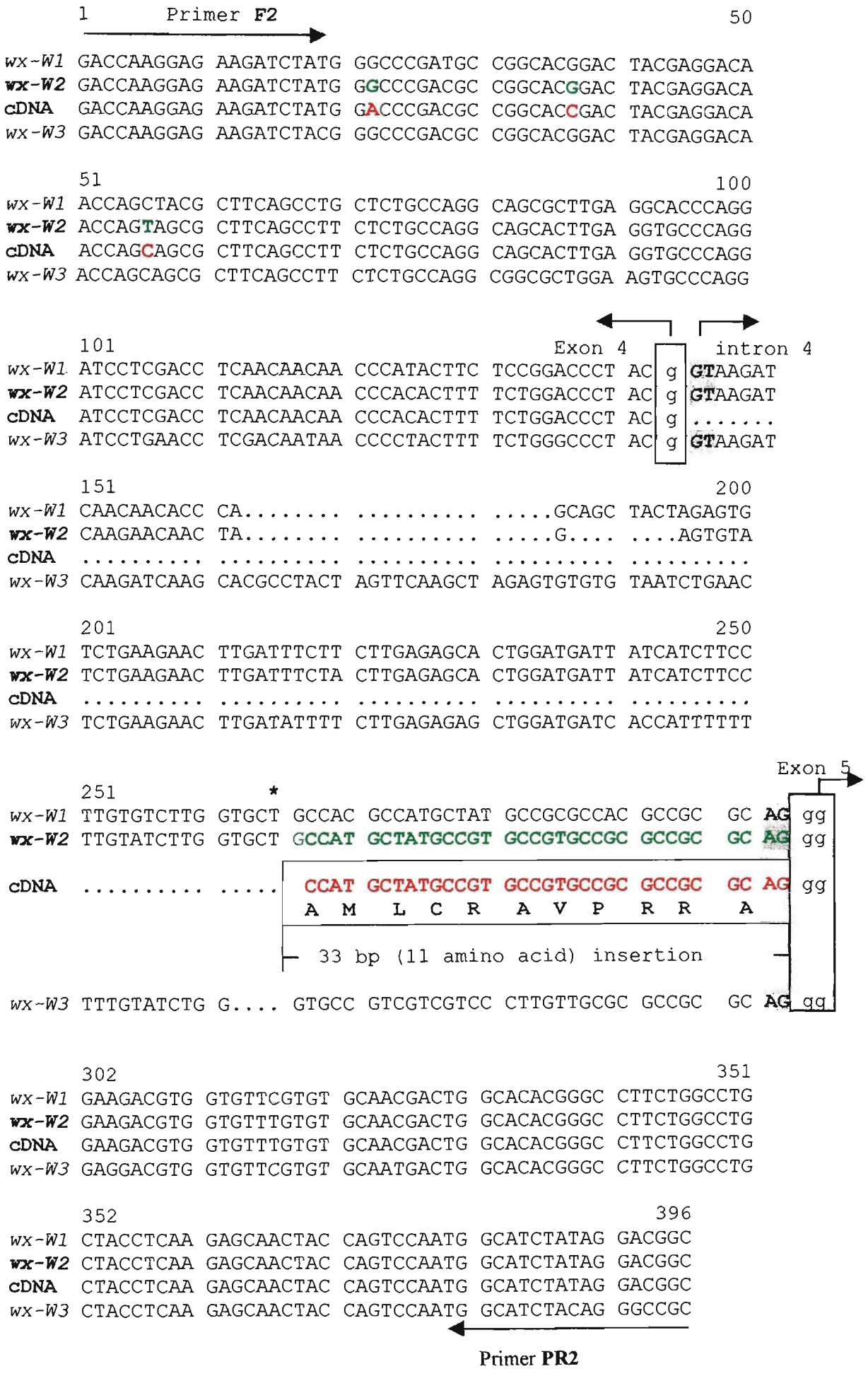

1 Primer F2 50
wx-W1 GACCAAGGAG AAGATCTATG GGCCCGATGC CGGCACGGAC TACGAGGACA
wx-W2 GACCAAGGAG AAGATCTATG GGCCCGACGC CGGCACGGAC TACGAGGACA
cDNA GACCAAGGAG AAGATCTATG GACCCGACGC CGGCACCGAC TACGAGGACA
wx-W3 GACCAAGGAG AAGATCTACG GGCCCGACGC CGGCACGGAC TACGAGGACA
51 100
wx-W1 ACCAGCTACG CTTCAGCCTG CTCTGCCAGG CAGCGCTTGA GGCACCCAGG
wx-W2 ACCAGTAGCG CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG
cDNA ACCAGCAGCG CTTCAGCCTT CTCTGCCAGG CAGCACTTGA GGTGCCCAGG
wx-W3 ACCAGCAGCG CTTCAGCCTT CTCTGCCAGG CGGCGCTGGA AGTGCCCAGG
Exon 4
intron 4
101
wx-W1 ATCCTCGACC TCAACAACAA CCCATACTTC TCCGGACCCT AC g GTAAGAT
wx-W2 ATCCTCGACC TCAACAACAA CCCACACTTT TCTGGACCCT AC g GTAAGAT
cDNA ATCCTCGACC TCAACAACAA CCCACACTTT TCTGGACCCT AC g ....
wx-W3 ATCCTGAACC TCGACAATAA CCCCTACTTT TCTGGGCCCT AC g GTAAGAT
151 200
wx-W1 CAACAACACC CA..... ....GCAGC TACTAGAGTG
wx-W2 CAAGAACAAC TA..... ....G.... ....AGTGTA
cDNA ....
wx-W3 CAAGATCAAG CACGCCTACT AGTTCAAGCT AGAGTGTGTG TAATCTGAAC
201 250
wx-W1 TCTGAAGAAC TTGATTTCTT CTTGAGAGCA CTGGATGATT ATCATCTTCC
wx-W2 TCTGAAGAAC TTGATTTCTA CTTGAGAGCA CTGGATGATT ATCATCTTCC
cDNA ....
wx-W3 TCTGAAGAAC TTGATATTTT CTTGAGAGAG CTGGATGATC ACCATTTTTT
Exon 5
251 *
wx-W1 TTGTGTCTTG GTGCT GCCAC GCCATGCTAT GCCGCGCCAC GCCGC GC AG gg
wx-W2 TTGTATCTTG GTGCT GCCAT GCTATGCCGT GCCGTGCCGC GCCGC GC AG gg
cDNA .... CCAT GCTATGCCGT GCCGTGCCGC GCCGC GC AG gg
A M L C R A V P R R A
33 bp (11 amino acid) insertion
wx-W3 TTTGTATCTG G.... GTGCC GTCGTCGTCC CTTGTTGCGC GCCGC GC AG gg
302 351
wx-W1 GAAGACGTG GTGTTCGTGT GCAACGACTG GCACACGGGC CTTCTGGCCTG
wx-W2 GAAGACGTG GTGTTTGTGT GCAACGACTG GCACACGGGC CTTCTGGCCTG
cDNA GAAGACGTG GTGTTTGTGT GCAACGACTG GCACACGGGC CTTCTGGCCTG
wx-W3 GAGGACGTG GTGTTCGTGT GCAATGACTG GCACACGGGC CTTCTGGCCTG
352 396
wx-W1 CTACCTCAA GAGCAACTAC CAGTCCAATG GCATCTATAG GACGGC
wx-W2 CTACCTCAA GAGCAACTAC CAGTCCAATG GCATCTATAG GACGGC
cDNA CTACCTCAA GAGCAACTAC CAGTCCAATG GCATCTATAG GACGGC
wx-W3 CTACCTCAA GAGCAACTAC CAGTCCAATG GCATCTACAG GGCCGC
Primer PR2

Fig. 7.1 Alignment of the three partial *waxy* genomic genes of common wheat and a reported *waxy* cDNA. The three partial *waxy* genomic genes of common wheat var. Chinese Spring, *wx-W1* , *wx-W2* and *wx-W3*, were obtained by PCR (Chapter 6). The sequence of the cDNA of common wheat var. Chinese Spring was reported by Clark *et al.* (1991) and analysed by Ainsworth *et al.* (1993). GT and AG junction of the intron 4 in the three *waxy* genes are shaded. Three nucleotides, the boxed "g" in exon 4 and the boxed "gg" in exon 5, denote the codon for glycine in all known waxy proteins: rice (Wang *et al.*, 1990), barley (Rohde *et al.*, 1988), maize (Shure *et al.*, 1983; Klösgen *et al.*, 1986) and potato (van der Leij *et al.*, 1991). The location of the extra 33 bp in the cDNA and 11 amino acids encoded thereby in this cDNA are shown in colour print and boxed. The corresponding sequence in intron 4 of *wx-W2* is shown in colour print. * denotes the nucleotide T at position -1 of the corresponding sequence hypothetically encoding 11 amino acids in *wx-W2*. Three mismatching nucleotides between *wx-W2* and this cDNA are shown in colour print.

7.3 Properties of intron 5 of the *wx-TdA* and *wx-TdB* genes in *T. turgidum*

The two partial *waxy* genes covering the exon 4-intron 4-exon 5-intron 5-exon 6 region, were amplified by using primers F2 and R2 (section 2.3.3.2) from *T. turgidum*. The sequence data revealed two classes of fragments, which were called *wx-TdA* and *wx-TdB*. Only a part, covering exon 4-intron 4-exon 5 from primer F2 to primer PR2 (Fig. 7.2), of the sequences of these were analysed in Chapter 6, for the alignment with other *waxy* genes amplified from different types of wheat. The *wx-TdB* containing the shorter intron 4 was designated as *wx-Td1*, and the *wx-TdA* containing the longer intron 4 was designated as *wx-Td2* in Chapter 6. DNA sequence of the 680 bp of the 5' end of the *wx-TdA* and *wx-TdB* of *T. turgidum* were also determined (Fig. 7.2). The sequence of intron 5 in *wx-TdA* and *wx-TdB* were analysed in this Chapter to develop the possible molecular markers for different *waxy* loci.

Based on the results in Chapter 6 and discussion in section 6.10.2, *T. aestivum* was suggested to be the hybrid of tetraploid *T. turgidum*, the donor of A and B genomes, and diploid *T. tauschii*, the donor of D genome. The sequences of the *wx-TdA* and *wx-TdB* of *T. turgidum* were, therefore, aligned with the sequences of the corresponding region of the *wx-TtD* of *T. tauschii* (R. Appels, pers. commun.). The results showed that there were significant differences in length of intron 5 between the three *waxy* genes, i.e., the *wx-TdA*, *wx-TdB* and *wx-TtD* had 99 bp, 69 bp and 141 bp of intron 5 in length (Fig. 7.2).

Fig. 7.2 *to be continued*

```
        631                                                                        700
wx-TdA  AGTCGTCCTT CGACTTCATC GACGGCTACG ACAAGCCGGT GGAGGGGCGC AAGATCAACT GGATGAAGGC
wx-TdB  AGTCGTCCTT CGACTTCATC GACGGCTACG ACAAGCCGGT GGAGGGGCGC AAGATCAACT GGATGAAGGC
wx-TtD  AGTCGTCCTT CGACTTCATC GACGGCTACG ACAAGCCGGT GGAGGGGCGC AAGATCAACT GGATGAAGGC
                             Primer GR

        701                                                             770
wx-TdA  CGGGATCCTG CAGGCCGACA AGGTGCTGAC TGTGAGCCCC TACTATGCTG AGGAGCTAAT...........
wx-TdB  CCGGATCCTG CAGGCCGACA AGGTGCTCAC GGTGAGCCCC TACTACGCGG AGGAGCTCAT...........
wx-TtD  CGGGATCCTG CAGGCCGACA AGGTGCTGAC GGTGAGCCCC TACTACGCGG AGGAGCTCAT...........
                                                                        Primer R2
```

Fig. 7.2 Alignment of sequences of the two partial *waxy* genes from *T. turgidum* and the corresponding section of the *waxy* gene of *T. tauschii.* The two *waxy* genes of *T. turgidum* (*wx-TdA* and *wx-TdB*) were amplified by primers F2 and R2 (section 2.3.3.2 and Table 2.5). The sequences of *wx-TdA* and *wx-TdB* were aligned with the sequence of the corresponding region of the *wx-TtD* gene of *T. tauschii* provided by Professor R. Appels (per. commun.). Primers GF and GR shown in this Figure were designed to amplify intron 5. The sequences in intron 4 that are suggested to design as primers for specific *waxy* loci are shown in colour print and boxed.

7.4 Preliminary development of the molecular markers for the *wx-A*, *wx-B* and *wx-D* loci of common wheat

7.4.1 Sequences specific for primers for the *wx-A*, *wx-B* and *wx-D* loci

Comparison of the *wx-W1*, *wx-W2* and *wx-W3* partial *waxy* genes of common wheat (Fig. 7.1) showed that the exons 4 and 5 were highly conserved in sequence and invariable in length. This would make it nearly impossible to distinguish the three *waxy* genes on the basis of their coding sequences. However, the existence of insertions/deletions in intron 4 of different *waxy* genes made it possible to design specific primers to represent the individual *waxy* loci which are located on different genomes in wheat. Three such sequences, which could be used for the purpose of distinguishing *wx-W2*, *wx-W1* and *wx-W3* genes in common wheat, are shown in Fig. 7.1. As discussed in section 6.10.2, *wx-W1*, *wx-W2* and *wx-W3* were suggested to be from genomes B, A and D of common wheat respectively. These sequences in intron 4, therefore, could be used for identification of *wx-A*, *wx-B* and *wx-D* loci of common wheat.

7.4.2 Restriction enzyme sites specific for the *wx-A*, *wx-B* and *wx-D* loci

The slight differences in DNA sequences of the *wx-W1*, *wx-W2* and *wx-W3* partial genes created different restrictive enzyme sites for 6 base cutters cutting only once in the three sections (Fig.7.3). Some of these restriction enzyme sites were common for two of the three partial genes, while some were unique to a single gene, i.e., five for *wx-A*, two for *wx-B* and 8 for *wx-D*. The restriction enzyme sites specific to one *waxy* locus would represent the particular chromosome type of common wheat.

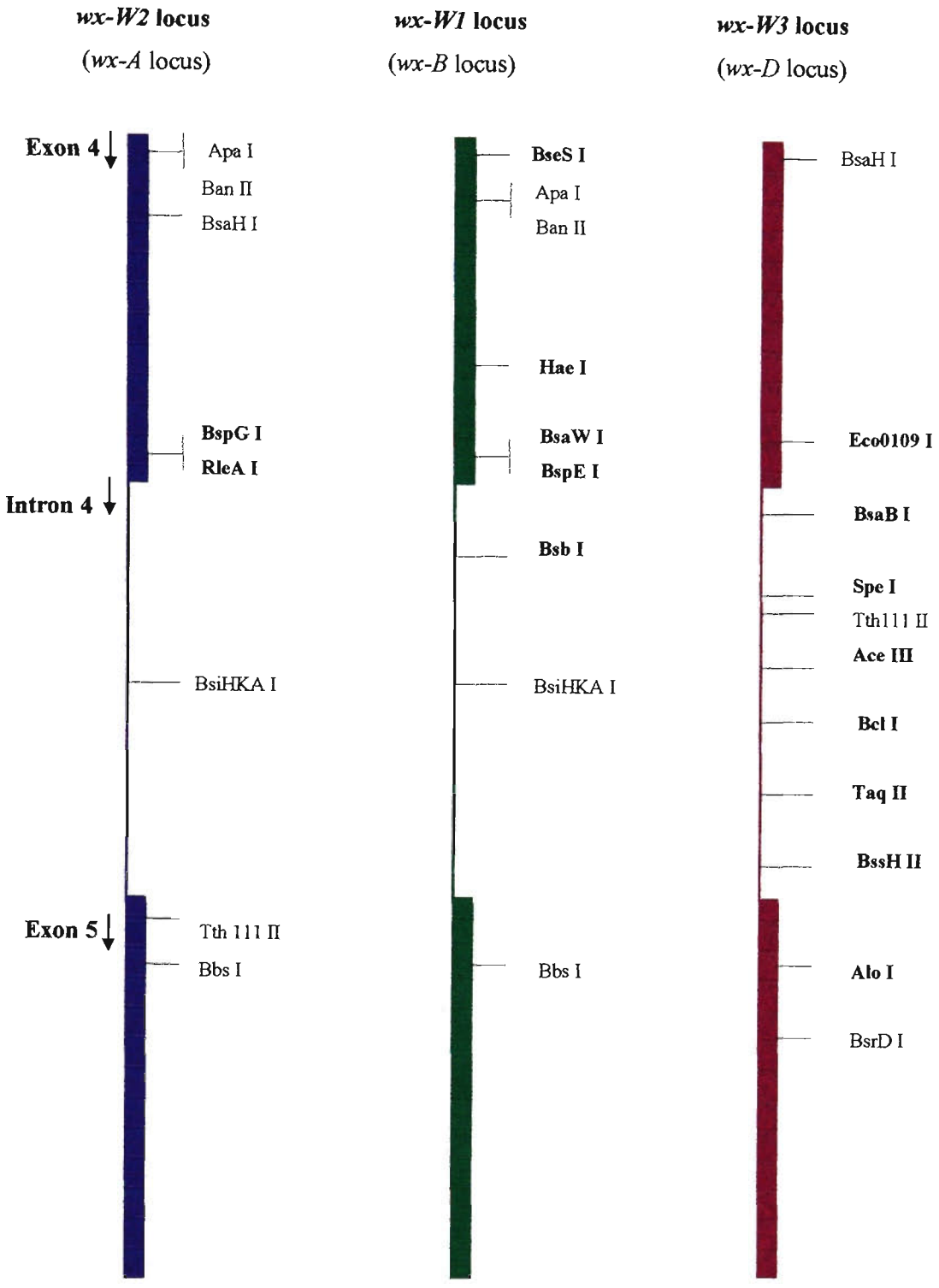


Fig. 7.3 The deduced restriction enzyme maps of the *wx-A*, *wx-B* and *wx-D* loci of common wheat. The partial *waxy* genes, *wx-W1*, *wx-W2* and *wx-W3*, are proposed to be from B, A and D genomes of common wheat as discussed in section 6.10.2. Restriction enzyme sites in *wx-W1*, *wx-W2* and *wx-W3* were predicted by using the Mapping program of ANGIS (section 2.6.4). The restriction enzyme sites were cut by 6 nucleotide enzymes and only once in these fragments. Only the sites differing between two or three loci are shown in this Figure, and the sites specific to a single locu are in bold. Eleven other putative restriction enzyme, which were the same on all the three loci, were not shown.

7.4.3 Specific PCR products for the *wx-A*, *wx-B* and *wx-D* loci

The significant difference in length of intron 5 between *wx-TdA*, *wx-TdB* and *wx-TtD* fragments was described in section 7.3. It has been suggested (section 6.10.2) that *T. turgidum* and *T. tauschii* might be derived from the A and B genomes, and D genome, respectively, to common wheat. There are 30 bp, 42 bp and 72 bp differences in length of intron 5 between *wx-TdA* and *wx-TdB*, *wx-TdA* and *wx-TtD*, and *wx-TdB* and *wx-TtD*. If the three *waxy* genes of common wheat corresponding to the *wx-TdA*, *wx-TdB* and *wx-TtD* were amplified in a single PCR, three PCR product bands would give the detectable size difference on high resolution 3% agarose gels. Based on these, a pair of primers was designed to amplify intron 5 of the *waxy* genes of common wheat to test whether the *wx-A*, *wx-B* and *wx-D* loci could be distinguished. The forward primer was GF (5'-CCTGCTACCTCAAGAGCAAC-3') located on exon 5 and the reverse primer was GR (5'-GGCTTGTCGTAGCCGTCGAT-3') located on exon 6 (Fig. 7.4). To obtain appropriate fragment sizes which could be separated clearly the primer smear in PCR products, it was necessary to include 53 bp of exon 5 and 122 bp of exon 6 in the amplification region (Fig. 7.4). This has been used to amplify the three specific PCR product bands representing *wx-A*, *wx-B* and *wx-D* loci in normal and null lines of common wheat, in order to identify the molecular basis of their *waxy* gene mutations (section 7.5).

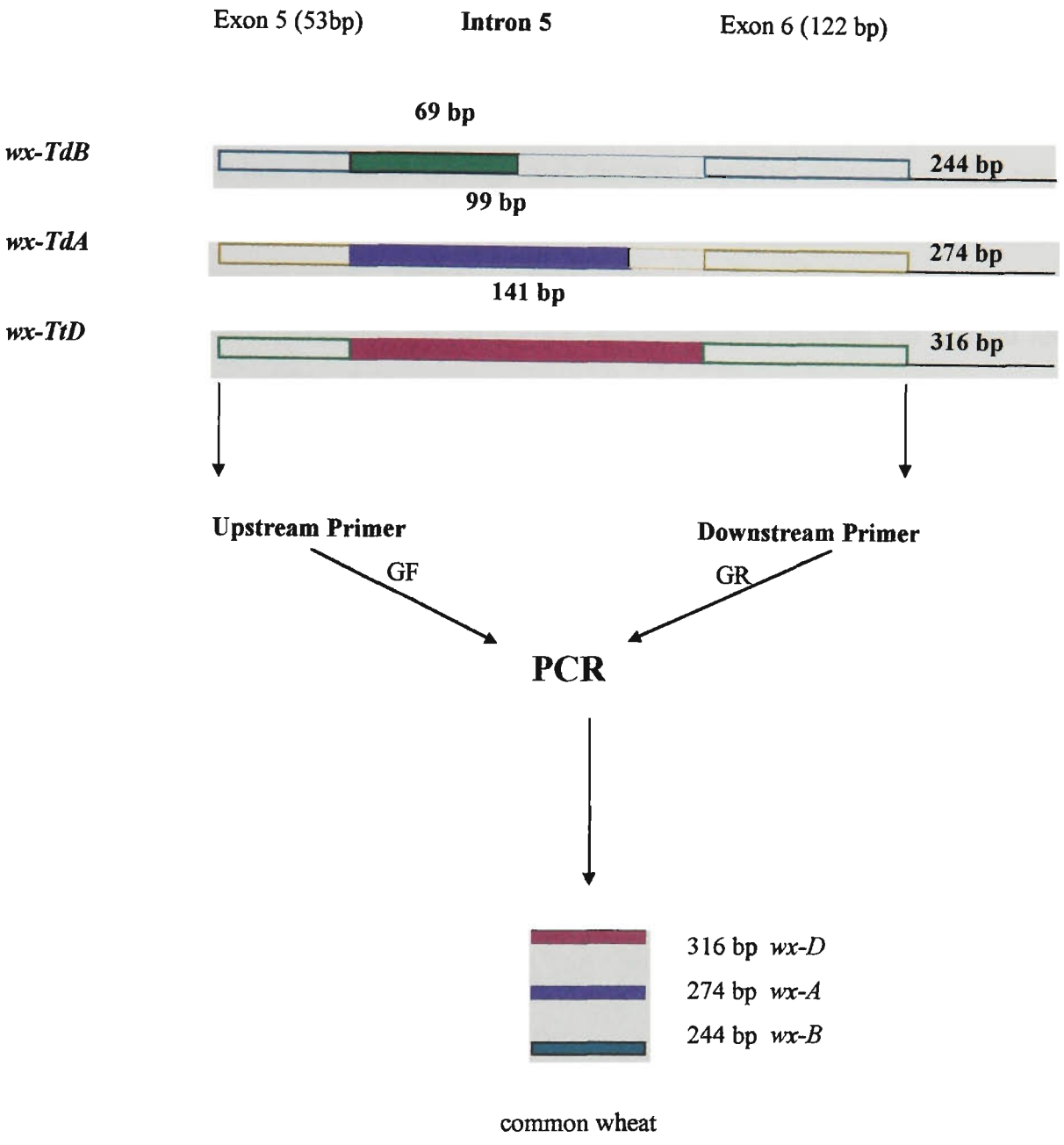


Fig 7.4. The predicted various lengths of PCR product bands specific for the *wx-A*, *wx-B* and *wx-D* loci in common wheat. The significant variations in size of fragments covering intron 5 of the three *waxy* genes in common wheat are predicted on the basis of the differences in sequences between *wx-TdA* and *wx-TdB* of *T. turgidum* and *wx-TtD* of *T. tauschii* (Fig. 7.2). The lengths of fragments of the *waxy* genes of common wheat were predicted using primers GF and GR to amplify the three partial *waxy* genes of common wheat.

7.5 Preliminary studies on *waxy* mutants

Primers GF and GR were used to amplify the *waxy* genes from common wheat, which included a normal line, Chinese Spring, and the two reported mutants of the *waxy* genes, variety Sturdy lacking the WX-7A protein and variety Rosella lacking the WX-4A protein (Yamamori *et al.*, 1995a). Three different sizes of PCR products were obtained in all tested lines; and no difference was observed between the normal and the null lines of wheat (Fig. 7.5).

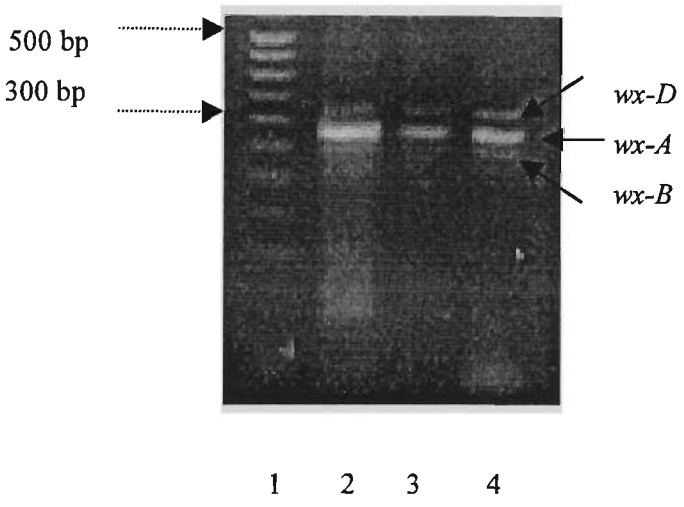


Fig. 7.5 The PCR products from the exon 5-intron 5-exon 6 region of the *wx-A*, *wx-B* and *wx-D* loci. The *waxy* genes were amplified from the genomic DNA using primers GF and GR (Fig. 6.2). The PCR conditions were as in section 2.3.8 and PCR products were electrophoresied in a 3% agarose gel. Expected sizes of fragments of the three *waxy* genes in common wheat, *wx-B*, *wx-A* and *wx-D*, were 244 bp, 274 bp and 316 bp, as described in Fig. 7.4. 1. DNA size marker (50 bp DNA ladder, Promega). 2. Normal wheat,, var. Chinese Spring; 3. Null-7A line, var. Sturdy; 4. Null-4A line, var. Rosella.

7.6 Discussion

7.6.1 Possible variation in mRNA processing events involved in the expression of the *waxy* gene in wheat

A reported waxy cDNA of common wheat contained an interesting 33 bp "insertion" encoding 11 additional amino acids relative to other plant waxy proteins (Ainsworth *et al.*, 1993). The 33 bp sequences in this cDNA could not be found in the exons of the *waxy* genes of *T. monococcum* and *T. speltoides*; however, it matched the partial intron 4 of the two diploid wheat (section 4.3.4). In order to explain the structure of the *waxy* genomic gene of common wheat corresponding to this cDNA and the additional "insertion" in this cDNA, the three partial genomic *waxy* genes that include the 33 bp region were compared further with this cDNA. The results indicated that the cDNA had a much higher homology to *wx-W2* than *wx-W1* or *wx-W3*. However, the 33 bp sequence encoding 11 extra amino acids in this cDNA was not a part of the coding region; it was present at the 3' end of intron 4 of *wx-W2*. Two possibilities can explain this unusual match.

1. One possibility is a T→A point mutation of the T nucleotide preceding this 33 bp sequence in the *wx-W2* genomic gene (Fig. 7.1), the transcript of this was used for this cDNA synthesis. This mutation would provide an alternative splice site and thus a longer coding sequence for exon 5. However, this mutation possibility could not be strong, because the genomic copy of the *waxy* genes cloned in this clone are from the same variety of common wheat, Chinese Spring, as the one from which this cDNA was obtained.

2. Alternatively, the extra coding sequence in this cDNA could have resulted from incorrect or incomplete splicing of the mRNA, or contamination of mRNA with nuclear material during library construction. The mismatch of three nucleotides in exon 4 of *wx-W2* and this cDNA could suggest mRNA editing events also.

Further work involving isolation and comparison of the *waxy* cDNAs transcribed from the individual *waxy* genes of common wheat is essential to test these possibilities.

7.6.2 Molecular markers for the *wx-A, wx-B* and *wx-D* loci

The *waxy* locus-specific primer, restriction enzyme sites and PCR products identified in this study provide the essential tools to identify the three homoeologous *waxy* genes in hexaploid common wheat. These molecular markers could be considered to be useful in analysing the genome-specific or chromosome-specific properties of the three *waxy* genes, *wx-A*, *wx-B* and *wx-D*, which have been located on chromosomes 7A, 4A and 7D of common wheat (Chao *et al.*, 1989).

7.6.2.1 Sequences for primers specific for the individual *waxy* loci

On the basis of analyses on the three *waxy* genomic genes of common wheat, specific sequences have been identified to represent the *wx*-A, *wx-B* and *wx-D* loci (Fig. 7.2). The striking advantage of these sequences in their utility is that it is now the feasible to synthesize primers for a PCR-based assay to distinguish between A, B and D genomes in common wheat or A and B genomes in tetraploid *T. turgidum*. Any one of these primers could be used as either forward or reverse primer, the other primer being from any conserved sequences in exons of the known *waxy* genes in wheat such as *T. monococcum* and *T. speltoides* (Chapter 4) and common wheat (Clark *et al.*, 1991). PCR products amplified by a locus-specific primer and a universal primer would be from the individual *waxy* gene of common wheat; thus the possible differences in the size of PCR products amplified by different primers could be used as molecular markers for the respective genomes of wheat. This approach could be also useful in identifying the null lines of different *waxy* genes in wheat.

7.6.2.2 Restriction enzyme sites specific for the individual *waxy* loci

Variations in sequences of introns and exons of the *wx-A, wx-B* and *wx-D* genes lead to different predicted restrictive enzyme sites (Fig. 7.3). These unique sites, particularly those in exons that are found to be less variable during the evolution of wheat, are useful for analyzing chromosome types of polyploid wheat and identifying

mutants of the *waxy* genes in common wheat. For example, DNA samples of normal lines of common wheat could be digested by the appropriate enzymes and run in agarose, followed by hybridization of Southern blotting using the waxy probe, different DNA bands could represent the *wx-A*, *wx-B* and *wx-D* genes on chromosomes 7A, 4A and 7D. DNA samples of waxy null lines of common wheat could be identified by such methods.

7.6.2.3 PCR products specific for the individual *waxy* loci

Based on the comparison of the *waxy* gene between *T. monococcum* and *T. speltoides* (Chapter 4) and *T. tauschii* (R. Appels, pers. commun.), intron 5 was found to be most variable in length. Because *T. turgidum* and *T. tauschii* have been suggested as the tetraploid and diploid progenitors, respectively, to common wheat (Chapter 6), we used the variation of the length of intron 5 between the two *waxy* genes of *T. turgidum* and the *waxy* gene of *T. tauschii* to develop molecular markers representing the *waxy* A, B and D loci in common wheat. Three distinct DNA bands from a single PCR were detected in the agarose gel, and this has been confirmed while identifying mutants of the different *waxy* genes in common wheat. The variation in the length of intron 5 would be thus an efficient molecular marker to distinguish between the three homoeologous *waxy* loci of common wheat.

7.6.3 Molecular basis of inactivation of the two *waxy* loci

The molecular marker based on the variation of the length of intron 5 indicated that the *wx*-A and *wx*-B null lines contained the same DNA bands as a normal line of common wheat. This result showed that at least this section of the three *waxy* genes existed in the tested null lines, suggesting that the *waxy* genes were not deleted completely in the null lines. The missing waxy protein in these null lines is thus possibly due to other types of mutations or deletions affecting other sections of the *waxy* gene. This result is consistent with a previous report that the structural genes are present in the three *waxy* loci of null 4A line (K107) and null 7D line (Bai Huo) as shown by Southern hybridization and Polymerase Chain Reaction (PCR) results (Nakamura *et al.*, 1995).

Nakamura *et al.* (1995) have shown that there is a small deletion (<30 bp) in the *wx-A* null line and both a small deletion (<50 bp) and insertion (<30 bp) in *wx-B* (*wx-4A*) null line of K107, but insertions or deletions were not detected in *wx-D* null line, when compared with the wild-type alleles of var. Chinese Spring. In maize, large insertions (150 bp-6.1 kbp) or deletions (>300 bp) were found to be associated with mutations among the *waxy* alleles, while single base changes or very small insertions or deletions were found to be responsible for waxy mutations in rice (Wessler and Varagona, 1985; Okagaki and Wessler, 1988). However, the mechanism of mutation of *waxy* alleles in common wheat could be more complicated than that in maize and rice, due to its genome constituents. The waxy mutations have not been reported to be associated with transposon activity in wheat. Except for the possible insertions or deletions which may have resulted in null alleles, polyploidization event of wheat may also cause silencing of functional genes so that polyploid species may not have significantly more expressed genes than diploid ones (Breiman and Graur, 1995). Different patterns of gene silencing have been suggested also as a mode of allopatric speciation at the polyploid level (Soltis and Soltis, 1993). Although examining the molecular basis of mutants was not a preliminary objective of this study, the sequence data made it possible to have carried out preliminary work. In addition, the locus-specific primers and the variable length of intron 5 together could be used to specifically clone the *waxy* genes at the null loci (if there are no deletions in these genes) and further work could be carried out to determine the molecular basis of gene inactivation using these lines.

CHAPTER 8

GENERAL CONCLUSIONS AND FURTHER DIRECTIONS

Chapter 8 General conclusions and further directions

The initial aim of this project was to clone and characterize the *waxy* gene in *T. monococcum*, a proposed donor of the A genome of wheat and the *waxy* gene from *T. speltoides* as a proposed donor of the B genome, to determine the functional, genetic and evolutionary relationship of different *waxy* genes and waxy proteins in wheat. After this aim had been achieved, we then took studies to clone a series of the partial *waxy* genes from the polyploid *T. turgidum-T. aestivum*, *T. timopheevi-T. zhukovskyi* and from other proposed diploid progenitors of wheat. Taken together, the large amount of sequence data obtained in this project has provided valuable molecular information which will increase our understanding the biochemical and physical properties of the different waxy proteins and the phylogenetic and evolutionary relationship of different genomes of wheat. The information would be useful for developing molecular markers for the individual *waxy* loci and analyses and identification of various mutants of the *waxy* genes of wheat.

8.1 General conclusions

8.1.1 Electrophoretic properties of the waxy proteins of different wheat

- The waxy protein of several accessions of the proposed diploid progenitors of A, B, D, G, A^t and A^z genomes of polyploid wheat indicated the presence of a 59 kDa protein band. However, there were slight differences in the molecular weight of the waxy proteins between some species of diploid wheat.

- The waxy proteins of the tetraploid *T. turgidum* (AABB) and hexaploid *T. aestivum* (AABBDD) were observed as a doublet of bands, the LMW and HMW proteins. However, the waxy proteins of tetraploid *T. timopheevi* (A^tA^tGG) and hexaploid *T. zhukovskyi* ($A^tA^tA^zA^zGG$) could not be separated into distinct bands.

An accession of *T. monococcum* that had a waxy protein corresponding to the HMW waxy protein of common wheat and *T. speltoides* that had a waxy protein corresponding to the LMW waxy proteins were chosen to clone their respective *waxy* genes. Each of the other diploid, tetraploid and hexaploid wheat was chosen with the view to cloning their respective partial *waxy* genes.

8.1.2 Starch granule proteins of different wheat

- The electrophoretic profile of 75 kDa soluble starch synthase (SSS) and 85 kDa starch branching enzyme (SBE) in different types of diploid, tetraploid and hexaploid wheat appeared to be indistinguishable in molecular weight on SDS-PAGE. However, variations in the >85 kDa starch granule proteins (SGP) were found in different genomes of wheat.

- Two SGP bands, 110 kDa and 100 kDa, were detected in the tetraploid *T. turgidum*, and three SGP bands, 110 kDa, 105 kDa and 100 kDa, were detected in the hexaploid *T. aestivum*. However, both the tetraploid *T. timopheevi* and hexaploid *T. zhukovskyi* were found to contain a 110 kDa SGP and a 95 kDa SGP.

- A single band, 110 kDa, 105 kDa, 100 kDa or 95 kDa, was detected in the proposed diploid progenitors. Intraspecies variation in molecular weight of a single SGP was found in diploid *T. monococcum*, *T. urartu* and *T. speltoides*.

This thesis is the first published document to investigate the SGP in endosperm of different genomes of wheat. The following conclusions were derived from the results obtained in this study.

1. This is the first report of a 95 kDa protein in starch granules in endosperm of wheat. The novel 95 kDa protein was observed in the tetraploid *T. timopheevi* and hexaploid *T. zhukovskyi* and some accessions of diploid *T. monococcum* and *T. speltoides*. It is possible that the 95 kDa protein belongs to the same

family as the 110 kDa, 105 kDa and 100 kDa proteins, which were reported to be starch synthases in common wheat (Rahman *et al.*, 1995; Denyer *et al.*, 1995; Yamamori *et al.*, 1998).

2. The genes encoding 110 kDa, 105 kDa and 100 kDa SGP have been located on chromosomes 7A, 7D and 7B, respectively, of common wheat (Denyer *et al.*, 1995). In this project, *T. tauschii* was observed to contain an unique 105 kDa SGP among all of the tested diploid wheat and is thus considered to be the donor of D genome to common wheat. However, the 110 kDa SGP appeared not only in the proposed donor of A genome, one accession of *T. monococcum* and *T. urartu*, but also in the proposed donor of B genome, *T. longissima*, *T. searsii*, *T. sharonensis* and one accession of *T. speltoides*. Similarly, the 100 kDa proteins appeared not only in the proposed donor of B genome, *T. bicornis* and one accession of *T. speltoides*, but also in *T. urartu*, which has been proposed recently to be the donor of the A genome (Breiman and Graur, 1995; Friebe and Gill, 1996; Dvorak, 1998). These observations suggest that the phylogenetic relationship of A and B genomes of polyploid wheat is much more complicated than expected.

3. This investigation has provided a basis to choose an appropriate diploid species to clone the individual genes encoding the 110 kDa, 105 kDa, 100 kDa or 95 kDa starch granule proteins from diploid species, in order to further analyse and compare the functions of these various starch granule proteins.

8.1.3 Molecular and structural properties of the *waxy* genes in wheat

The two *waxy* genomic genes covering the partial untranslated leader and the entire translated exons from the translation initiation codon to termination codon and their intervening introns were isolated from diploid *T. monococcum* and *T. speltoides* by the Polymerase Chain Reaction (PCR) approach. The molecular properties of the waxy genes could be summarised as follows:

- The *waxy* gene of both *T. monococcum* and *T. speltoides* consisted of 11 exons and 10 introns, from the translated initiation codon to the termination codon.

- The exon/intron structure of the *waxy* genes in these diploid wheat is the same as that of the barley *waxy* gene (Rohde *et al.*, 1988); however, it was different from that of the *waxy* gene in the rice, maize and potato *waxy* genes that consist of 13 exons and 12 introns. The fact that the *waxy* genes in wheat have lost two introns during plant evolution suggests a selective loss of introns of *waxy* genes after the divergence of wheat and rice.

- The lengths and sequences of exons of the *waxy* genes of *T. monococcum* and *T. speltoides* were highly conserved, but their introns appeared variable in length and sequences. Variation of introns mainly resulted from different lengths of insertion/deletion and from accumulation of some short repeated sequences.

This is the first report of the *waxy* genomic gene encoding the entire waxy proteins from *T. monococcum* and *T. speltoides* so far, although partial *waxy* genes have been cloned from different types of wheat (Mason-Gamer *et al.*, 1998; Murai *et al.*, 1998). The sequence data of the two *waxy* genes obtained in this project has been have been used for the following purposes:

1. The amino acid sequences of the waxy proteins of *T. monococcum and T. speltoides* were deduced from the DNA sequences of their *waxy* genes. The predicted biochemical and physical properties of the two deduced waxy proteins were analysed and compared.

2. The conserved exon sequences of the two *waxy* genes of diploid wheat have been used to design primers to clone the individual, partial *waxy* genes in other diploid and polyploid species of wheat.

3. On the basis of the sequence comparison of the *waxy* genes of diploid wheat, introns 4 and 5 of the *waxy* gene were identified as being a particularly important region for sequence variations, to distinguish between different genomes of wheat or to develop molecular markers for identification of waxy mutants.

8.1.4 Variation in amino acid sequences of the waxy proteins in wheat

- The two mature proteins of *T. monococcum* and *T. speltoides* deduced from their respective *waxy* genes were found to have the same number of amino acids but slightly different calculated molecular weights and isoelectric points. These differences in the predicted physical properties most probably accounted for electrophoretic differences of the two proteins revealed in SDS-PAGE.

- Slight differences in the predicted physical properties of waxy proteins of *T. monococcum* and *T. speltoides* were found to result from only 3% amino acid substitutions between them; particularly, a key amino acid substitution could resulted in an major change of molecular weight and isoelectric points of the waxy proteins.

- The amino acid sequences conserved at the N-terminus of plant waxy protein, putative ADP-glucose binding site and the conserved domains for starch synthases were all found in the mature waxy proteins of *T. monococcum* and *T. speltoides*. However, it was found that the predicted secondary structures and the number of glycosylation sites appeared variable in different waxy proteins of *Triticum*. The altered secondary structures of the proteins could be suggested to relate to differences in the expressions of the proteins and their different contributions to amylose synthesis in wheat.

- The transit peptides of the waxy protein were observed to be more variable, relative to the mature waxy proteins. A high content of small amino acids

arginine, or threonine and serine, and a net positive charge were found to exist in the transit peptides of the waxy proteins, although homology of amino acid sequences between some of the transit peptides of plant waxy proteins was only around 20%. It could thus be suggested that the transporting property of the transit peptides would rely on their amino acid composition but not on the amino acid sequence.

8.1.5 Sequences specific for the *waxy* loci of different genomes in polyploid wheat

In this project, partial *waxy* genes have been cloned from the two *waxy* loci of tetraploid *T. turgidum* (AABB) and *T. timopheevi* (A^tA^tGG), the three *waxy* loci of hexaploid *T. aestivum* (AABBDD) and *T. zhukovskyi* ($A^tA^tA^zA^zGG$), and the single *waxy* locus each of the proposed diploid progenitors of different genomes. This study is the first to report that the partial *waxy* gene in the same region of a structural gene from these genomes have been determined simultaneously. The sequence data have revealed the following:

- The length and sequence of intron 4 were variable in the two *waxy* genes of tetraploid wheat and the three *waxy* loci of hexaploid wheat, while lengths and sequences of exons 4 and 5 fragments cloned in this study were highly conserved. The variable region of intron 4 enabled us to identify certain locus-specific sequences to the A, B, D, G, A^t or A^z genomes of polyploid wheat. The specific sequences were found to be almost invariable in a particular genome during the evolution of wheat from tetraploid to hexaploid.

- Diploid *T. tauschii* and *T. monococcum* were found to be distinguishable in length and sequence of intron 4. Other diploid wheat species could be distinguished only by the point mutations of nucleotides in their partial *waxy* genes, with the exception of *T. urartu* and *T. bicornis*, where the sequences were identical.

8.1.6 Phylogenetic and evolutionary relationship between different genomes of diploid and polyploid wheat

A phylogenetic tree was constructed to determine the evolutionary relationship of different genomes between the various diploid and polyploid species of wheat. Following the analysis of the genome-specific sequences and the phylogenetic tree, the following suggestions were made:

1. *T. tauschii* (DD) most likely contributed D genomes to *T. aestivum*, and *T. urartu* (AA) and *T. speltoides* (BB) were the possible donors of A and B genomes, respectively, to *T. turgidum* and *T. aestivum*.

2. *T. monococcum* would be A^t genome donor of *T. timopheevi* and thus *T. zhukovskyi*.

3. Based on the above two suggestions, *T. turgidum* and *T. timopheevi* would have originated from two independent lineages of diploid progenitors. This is consistent with our previous results that *T. turgidum* and *T. timopheevi* had two significantly different electrophoretic profiles of waxy proteins and starch granules proteins.

This thesis is the first published document to report the DNA sequences of the same region of a structural gene from the individual A, B, D, G, A^t or A^z genomes of polyploid wheat and their different diploid progenitors. The striking differences between these results and the conventional theory (Kerby and Kuspira, 1987) and current theory (Friebe and Gill, 1996) on the phylogenetic relationships of different genomes are:

1. *T. timopheevi* might have originated from *T. monococcum* rather than the conventional diploid species containing A genome, *T. urartu*.

2. The third set of genomes (A^z) of *T. zhukovskyi* was different from diploid *T. monococcum*. On the available data this study is the first to provide evidence of a linkage of the third set of genome in *T. zhukovskyi* with other accessions of *T. monoccoccum* or other diploid wheat.

The difference between the partial *waxy* genes of G genome of *T. timopheevi* and A^z genome of *T. zhukovskyi* and those of the tested diploid progenitors were minor in the cloned fragments because they had same lengths of intron 4. Single nucleotides were found variable, but it is not reliable to make any suggestion for the progenitors of G and A^z genomes. However, the separation of intron 4 of *waxy* genes of G and A^z from other *waxy* genes in these polyploid wheat has provided the possibility for identifying further the donors of G and A^z genomes.

8.1.7 Expression of the waxy proteins

- The unequal expressions of different waxy proteins in *T. turgidum* and *T. aestivum* were detected in SDS-PAGE. Generally, the WX-B protein was expressed much more than the WX-A protein in the *T. turgidum-T. aestivum* lineage of wheat, while the WX-D protein was expressed slightly more than the WX-A protein in *T. aestivum*.

- Our results also raised the possibility that an alternative transcriptional event might be involved in the expression of the *waxy* gene in wheat. This proposal was based on the sequence comparison of the three *waxy* genomic fragments from common wheat and a reported *waxy* cDNA of common wheat (Clark *et al.*, 1991). A stretch of 33 bp nucleotides encoding an additional 11 amino acids in this cDNA of wheat could not be observed in any exon of the three *waxy* genomic genes obtained in this project but appeared to be a part of intron 4 of one *waxy* genomic copy. There was also a small degree of variation in the sequence of this cDNA and the corresponding genomic fragment. These observations imply that some

variations might have arisen during the expression of the *waxy* gene from the genomic DNA sequence to mRNA.

8.1.8 Molecular markers for identification of the different *waxy* loci and mutants of common wheat

Preliminary work was carried out to develop molecular markers for the three homoeologous *waxy* genes, *wx-A*, *wx-B* and *wx-D* in common wheat, and the following observations were obtained and suggestions were made.

- Specific sequences in intron 4 were observed for the *wx*-A, *wx-B* and *wx-D* loci; these sequences were proposed to design primers for isolation of individual complete *waxy* genes and identification of mutants of common wheat.

- The special restriction enzyme sites for the *wx*-A, *wx-B* or *wx-D* loci were predicted on the basis of variation in sequences of three *waxy* genes and could be used for identification chromosomes or genomes of wheat.

- The three partial *waxy* genes of common wheat were obtained in a single amplification reaction; the three separated PCR product bands were identified to be from different *waxy* loci and thus used to identify the *waxy* gene of null lines.

8.1.9 Molecular basis of inactivation of the *waxy* gene

The variable length of intron 5 of the *wx*-A, *wx-B* and *wx-D* genes in normal and null-4A and null-7A lines of common wheat was amplified by a pair of primers based on the putative sequences of the *waxy* genes in wheat. The three separated DNA bands did show variability but no difference was detected between the null lines and normal wheat, suggesting that all three structural genes existed in null lines and that the *waxy* gene in the null line did not involve a whole deletion.

8.2 Further research directions

In order to reveal more molecular properties of the *waxy* genes for regulation of starch composition and for understanding wheat evolution using the sequence information obtained in this study, the following directions were suggested for further research.

8.2.1 Cloning of the complete individual *waxy* genes of common wheat

The partial *waxy* genes have been cloned from common wheat, but the functional relationship between the three waxy proteins in common wheat cannot be revealed by their amino acid sequences alone. Although the *waxy* genes have been cloned from two possible diploid progenitors (the third one has been cloned by a collaborating laboratory and used here for comparative purposes), amino acid data could not be considered to be the ideal for understanding the properties of three *waxy* genes in common wheat. Therefore, the cloning of three *waxy* genes in common wheat would be necessary in order to provide more accurate information on the proteins encoded by them, and thus information on the unequal expression of three waxy proteins.

The cloning work could be carried out in two ways. Firstly, designed primers F1 and R4 in this project (section 2.3.3.1) would be appropriate for cloning the complete section of the *waxy* gene from the translation initiation codon to the termination codon to clone the cDNA using RT-PCR. The Long polymerase (Promega) has provided the possibility to amplify the approximately 2.8-3.0 kb of the *waxy* genes or cDNA by one PCR reaction. Secondly, primers specific to the three *waxy* loci can be used together with conserved sequences as upstream or downstream primers to amplify the individual genes by appropriate amplification reactions, and the sequences of the individual genes could be obtained by directly sequencing PCR products.

8.2.2 Cloning and expression of the individual *waxy* cDNAs

One of the main objectives for cloning genes in plants is to improve the agricultural traits for specific crops including wheat. Following the reported technique breakthrough in transformation of transgenic wheat, it is anticipated that the antisense RNA technology would allow changes to be made for the pathway of starch synthesis and the production of waxy wheat. However, before this could be taken it would be neccessary to clone the individual *waxy* cDNAs. It also would be important to clone the respective *waxy* cDNAs in common wheat in order to understand the interactions of the three waxy proteins, as our results suggest that possible mRNA processing events might be involved in the expression of the waxy proteins in common wheat.

8.2.3 Further understanding of evolution of wheat

The origins of different genomes of wheat have been described in this project, based on electrophoretic properties of the waxy proteins and SGP on SDS-PAGE and on the sequences of the partial *waxy* genes in different types of wheat. However, polymorphism of the SGP within species of diploid wheat, *T. monococcum*, *T. urartu* and *T. speltoides*, which are considered to be the three important diploid species to the genomes of polyploid wheat (section 8.1.2), suggests that identification of diploid progenitors could be much more complicated than suggested so far. Analysis of our sequence data of the *waxy* genes of diploid *T. monococcum* and *T. speltoides* showed significant differences compared to those obtained recently by Mason-Gamer *et al.* (1998) from the same species, implying intraspecies polymorphism of *T. monococcum* and *T. speltoides* may exist at the DNA level. These observations make it very difficult to determine the origin of the genomes of wheat unequivocally. Since most of the previous researches has been focussed on interspecific variability for studies on phylogenetic relationships, and the origins of different genomes of wheat are still in debate. In future, more attention should be paid to genetic variability within species. However, the sequences, PCR products and restrictive enzyme sites specific for the *waxy* loci on different genomes of polyploid wheat have provided the potential to develop molecular markers to determine the origin of the different genomes of wheat.

REFERENCES

Ainsworth C. C., Gale M. D. and Baird S. (1984) The genetic control of grain esterases in hexaploid wheat. Theor. Appl. Genet. **68:** 219-226.

Ainsworth C. C., Clark J. and Balsdon J. (1993) Expression, organisation and structure of the genes encoding the waxy protein (granule-bound starch synthase) in wheat. Plant Mol. Biol. **22:** 67-82.

Amano E. (1981) Genetic and biochemical characterisation of waxy mutants in cereals. Environ. Health Perspect. **37:** 35-41.

Appels R., Reddy P., McIntyre C. L., Moran L. B., Frankel O. H. and Clarke B. C. (1989) The molecular-cytogenetic analysis of grasses and its application to studying relationships among species of the *Triticeae*. Genome **31:** 123-133.

Asins M. J. (1988) On the phylogeny of cultivated wheats. Evol. Trends in Plants **2:** 81-83.

Ayres N. M., McClung A. M., Larkin P. D., Bligh H. F. J. and Park W. D. (1997) Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor. Appl. Genet. **94:** 773-781.

Baba T., Yoshii M. and Kainuma K. (1987) Acceptor molecule of granule-bound starch synthase from sweet-potato roots. Starch **2:** 52-56.

Baba T., Kimura K., Mizuno K., Etoh H., Ishida Y., Shida O. and Arai Y. (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme I. Biochem. Biophys. Res. Commun. 181: 87-94.

Baba T., Nishihara M., Mizuno K., Kawasaki T., Shimada H., Kobayashi E., Ohnishi S., Tanaka K. and Arai Y. (1993) Identification, cDNA cloning, and gene expression of solule starch synthase in rice (*Oryza sativa* L) immature seeds. Plant Physiol. **103:** 565-573.

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.

Baga M., Mallard C. and Chibbar R. N. (1998) Alternative promoters control expression of a starch branching enzyme gene in wheat. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 85-88.

Ball S., Guan H-P, James M., Myers A., Keeling P., Mouille G., Buleon A., Colonna P. and Preiss J. (1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell **86:** 349-352.

Bernatzky R. B. and Tanksley S. D. (1986) Toward a satured linkage map in tomato based on isozymes and random cDNA sequence. Genetics. **112:** 887-898.

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. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **2:** 14-16.

Bowden W. M. (1959) The taxonomy and nomenclature of the wheats, barleys, and ryes and their wild relatives. Can. J. Botany **37:** 657-684.

Braun H. J., Payne T. S., Morgounov A. I., van Ginbel M., Rajaram S. (1998) The Challenge: one billion tons of wheat by 2020. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 33-40.

Breathnach R. and Chambon P. (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:** 349-83.

Breiman A. (1987) Mitochondrial DNA diversity in the genera of *Triticum* and *Aegilops* revealed by Southern blot hybridisation. Theor. Appl. Genet. **73:** 563-570.

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

Briggs R. W. Amano E. and Smith H. H. (1965) Genetic recombination with ethy-methanesulphonate-induced waxy mutants in maize. Nature **207:** 891-892.

Bremer K. (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution **42:** 795-803.

Buleon A., Gallant D. J., Bouchet S., Mouille G. D., Hulst C., Kossmann J. and Ball S. (1997) Starches from A to C. Plant Physiol. **115:** 949-957.

Burton R. A., Bewley J. D., Smith A. M., Bhattacharya M. K., Tatge H., Ring S., Bull V., Hamilton W. D. O. and Martin C. (1995) Starch branching enzymes belonging to distinct enzyme families are differentially expressed during pea embryo development. The Plant J. **7:** 3-15.

Chao S., Sharp P. J., Worland A. J., Warham E. J., Koebner R. M. D. and Gale M. D. (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor. Appl. Genet. **78:** 494-504.

Chen K. Gray J. C. and Wildman S. G. (1975) Fraction 1 protein and the origin of polyploid wheat. Science **198:** 1304-1306.

Chen P. D. and Gill B. S. (1983) The origin of chromosome 4A and genomes B and G of tetraploid wheats. Proc. of the 6th Int. Wheat Genet. Symp., Kyoto, Japan pp 39-48.

Chen P. D., Liu D. J., Pei G. Z., Qi. I. L. and Huang L. (1988) The chromosome constitution of three endemic hexaploid wheats in Western China. Proc. 0f the 7th Int. Wheat Genet. Symp, IPSR, Cambridge, England pp 75-80.

Cheng M. Fry J. E., Pang S., Zhou H., Hironaka C. M., Duncan D., Conner T. W. and Wan Y. (1997) Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. Plant Physiol. **115:** 971-980.

Chibbar R. N., Baga M., Caswell K., Repellin A., Leung N., Abdel Aal E. and Huel P. (1998) Genetic transformation strategies to alter starch structure in wheat. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 167-170.

Chou P. Y. and Fasmas G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequences. Adv. Enzymol **47:** 45-47.

Clark J. R., Robertson M. and Ainsworth C. C. (1991) Nucleotide sequence of a wheat (*Triticum aestivum* L.) cDNA clone encoding the waxy protein. Plant Mol. Biol. **16:** 1099-1101.

Clarke B. C., Taylor W., Morell M., Li Z., Rahman S., Ali S. and Appels R. (1998) Gene expression in wheat: The quality genes in grain endosperm. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 65-80.

Cornell H. J. and Hoveling A. W. (1998) Wheat chemistry and utilisation. Technomic Publishing AG, Pennsylvania, USA pp. 1-50.

Crosbie G. B. (1991) The relationship between starch swelling properties, paste viscosity and boiled noodle quality in wheat flours. J. Cereal Sci. **13:** 145-150.

Crosbie G. B., Lambe W. J., Tsutsui H. and Gilmour R. F. (1992) Further evaluation of the flour swelling volume test for identifying wheat potentially suitable for Japanese noodles. J. Cereal Sci. **15:** 271-280.

Crow J. F. and Dove W. F. (1987) Anecdotal, historical and critical commentaries on genetics. Genetics **116:** 339-342.

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.

Delrue B., Fontaine T., Routier F., Decq A., Weiruszeski J-M., Koornhuyse N. V. D., Maddelein M-L., Fournet B. and Ball S. (1992) Waxy *Chlamydomonas reinhardtii*: monocellular algal mutants defective in amylose biosynthesis and granule-bound starch synthase activity accumulate a structurally modified amylopectin. J. Bacteriol. **174:** 3612-20.

Demeke T., Hucl P., Nair R. B., Nakamura T. and Chibbar R. N. (1997) Evaluation of Canadian and other wheats for waxy proteins. J. Cereal Chem. **74:** 442-444.

Denyer K., Hylton C. M., Jenner C. F. and Smith A. M. (1995) Identification of multiple isoforms of soluble and granule-bound starch synthase in developing wheat endosperm. Planta **196:** 256-265.

Denyer K., Clarke B., Hylton C., Tatge H. and Smith A. M. (1996) The elongation of amylose and amylopectin chains in isolated starch granules. The Plant J. **10:** 1135-1143.

Denyer K. Smith A. M. (1992) The purification and characterisation of the two forms of soluble starch synthase from developing pea embryos. Planta **186:** 659-667.

Devos K. M. and Gale M. D. (1992) The use of randomly amplified DNA markers in wheat. Theor. Appl. Genet. **84:** 567-572.

Devos K. M., Dubcovskyi J., Dvorak J., Chinoy C. N. and Gale M. D. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor. Appl. Genet. **91:** 282-288.

Devos K. M. and Gale M. D. (1997) Comparative genetics in the grasses. Plant Mol. Biol. **35:** 3-15.

Doane W. M. (1994) Opportunities and challenges for new industrial uses of starch. Cereal Foods World **39:** 556-557.

Doehlert D. C. and Knutson C. A. (1991) Two classes of starch debranching enzymes from developing maize kernels. J. Plant Physiol. **138:** 566-572.

Doehlert D. C., Kuo T., Tuvik J. A., Beers E. P. and Duke S. H. (1993) Characteristics of carbohydrate metabolism in sweet corn (*sugaryl*) endosperms. J. Amer. Soc. Hortic. Sci. **188:** 661-66.

Dry I., Smith A., Edwards A., Bhattacharyya M., Dunn P. and Martin C. (1992) Characterisation of cDNAs encoding two isoforms of granule-bound starch synthase which differential expression in developing storage organs of pea and potato. The Plant J. **2:** 193-202.

Dvorak J. and Appels R. (1982) Chromosome and nucleotides sequence differentiation in genomes of polyploid *Triticum* species. Theor. Appl. Genet. **63:** 349-362.

Dvorak J. (1983) The origin of wheat chromosomes 4A and 4B and their genome reallocation. Can. J. Genet. Cytol. **25:** 210-214.

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

Dvorak J. and Zhang H. B. (1990) Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc. Natl. Acad. Sci. U.S.A. **87:** 9640-9644.

Dvorak J. Resta P. and Kota R. S. (1990) Molecular evidence on the origin of wheat chromosomes 4A and 4B. Genome **33:** 30-39.

Dvorak J. and Zhang H. B. (1992) Application of molecular tools for study of the phylogeny of diploid and polyploid taxa in *Triticeae*. Heredias **116:** 37-42.

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. (1998) Genome analysis in the *Triticum-Aegilops* alliance. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 8-11.

Echt C. S. and Schwartz D. (1981) Evidence for the inclusion of controlling elements within the structural gene at the *waxy* locus in maize. Genetics **99:** 275-284.

Edwards A., Marshall J., Denter K., Sidebottom C., Visser R. G. F., Martin C. and Smith A. M. (1996) Evidence that a 77 kDa protein from the starch of pea embryos is a isoform of starch synthase that is both soluble and granule bound. Plant Physiol. **112:** 89-97.

Ekaterina D., Friebe B. 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.

Feldman M. (1978) New evidence on the origin of the B genome of wheat. Proc. of the 5th Int. Wheat Genet. Symp. New Delhi, Kapoor Art Press, New delhi, India **1:** 120-132.

Ferguson V. (1994) High amylose and waxy corns. In: A.R. Hallauer (Ed.), Specialty corns, CRC Press, Boca Raton, USA pp. 55-57.

Fisher D. K., Boyer C. D. and Hannah L. C. (1993) Starch branching enzyme II from maize endosperm. Plant Physiol. **102:** 1045-1046.

Fisher D. K., Gao M., Kim K. N., Boyer C. D. and Guiltinan M. J. (1996) Allelic analysis of the maize amylose-extender locus suggests that independent genes encode starch branching enzyme Iia and Iib. Plant Physiol. **110:** 611-619.

Flispe E., Keetels M., Jacobsen E. and Visser P. G. F (1996) The dosage effect of the wild type GBSS allele is linear for GBSS activity but not for amylose content: absence of amylose has a distinct influence on the physio-chemical properties of starch. Theor. Appl. Genet. **92:** 121-127.

Frederiksen S. and Seberg O. (1992) Phylogenetic analysis of the *Triticeae* (*Poaceae*) Hereditas **116:** 15-19.

Friebe B., Mukai Y. and Gill B. S. (1992) C-banding polymorphisms in several accessions of *Triticum tauschii* (*Aegilops squarrosa*). Genome **35:** 192-199.

Friebe B. and Gill B. S. (1996) Chromosome banding and genome analysis in diploid and polyploid cultivated wheat. In: Methods of Genome Analysis in Plants. (ed. Jauhevr P. P.), CRC Press, Boca Raton, Fla. pp: 39-60.

Fujita N., Wadano A., Kozaki S., Takaoka K., Okabe S. and Taira T. (1996) Comparison of the primary structure of proteins (granule-bound starch synthase) between polyploid wheats and related diploid species. Biochem. Genet. **34:** 403-413.

Furukawa K., Tagaya M., Inouye M., Preiss J. and Fukui T. (1990) Identification of lysine 15 at the active site in *Escherichia coli* glycogen synthase. J. Biol. Chem. **265:** 2086-2090.

Gao M. and Chibbar R. N. (1998) Molecular and biochemical characterisation of soluble starch synthases in wheat kernels. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 151-153.

Gautier M. F., Aleman M. E., Guirao A., Marion D. and Jourdier P. (1994) *Triticum aestivum* puroindolines, two basic cystine rich seed proteins: cDNA sequence analysis and developmental gene expression. Plant Mol. Biol. **25:** 43-58.

Galdwell K. A. and Kasarda D. D. (1978) Assessment of genomic and species relationships in *Triticum* and *Aegilops* by PAGE and by differential staining of seed albumins and globulins. Theor. Appl. Genet. **52:** 273-280.

Gerlach W. L., Appels R., Dennis E. S. and Peacock W. J. (1978) Evolution and analysis of wheat genomes using highly repeated DNA sequences. Proc. of the 5th Int. Wheat Genet. Symp. New Delhi, Kapoor Art Press, New delhi, 81-89.

Gielly L. and Taberlet P. (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. Mol. Biol. Evol. **11:** 769-777.

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

Gill B. S. and Chen P. D. (1987) Role of cytoplasm-specific introgression in the evolution of the polyploid wheats. Proc. Natl. Acad. Sci., U.S.A. **84:** 6800-6804.

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

Gill B. S., Browder L. E., Hatchett J. H., Harvey T. L., Martin T. J., Raupp W. J., Sharma H. C. and Waines J. G. (1993) Disease and insect resistance in wild wheat. Proc. of the 6th Int. Wheat Genet. Symp. Kyoto, pp. 785-792.

Gill K. S., Lubbers E. L., Gill B. S., Raupp W. J. and Cox T. S. (1991) A genetic linkage map of *Triticum tauschii* (DD) and its relationship of the D genome of bread wheat (AABBDD). Genome **34:** 362-376.

Gill K. S., Gill B. S. and Endo T. R. (1993). A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma **102:** 374-381.

Graybosch R. A. Peterson C. J., Hansen L. E., Rahman S., Hill A. and Skerritt J. H. (1996) Identification and Characterisation of US wheats carrying null alleles at the *wx* loci. Cereal Chem. **75:** 162-165.

Graybosch R. A. (1998) Waxy wheats: origin, properties, and prospects. Trends Food Sci. Technol. **9:** 135-142.

Greenwell P. and Schofield J. D. (1986) A starch granule protein associated endosperm softness in wheat. Cereal Chem. **63:** 379-380.

Greta B. A. (1997) ANGIS Bioinformatics handbook - University of Sydney, Australia.

Hanley B. A. and Schuler M. A. (1988) Plant intron sequences: evidence for distinct groups of introns. Nucl. Acids Res. **16:** 23-33.

Hegstad J. B., Kianian S. F., McMullen M. S. and Doehlert D. C. (1998) Development of waxy (low amylose) durum cultivars. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 172-174.

Heun M., Schafer-Pregl R., Klawan D., Castagna R., Accerbi M., Borghi B. and Salamini F. (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science **278:** 1312-1314.

Hirano H. and Sano Y. (1991) Molecular characterisation of the *waxy* locus of rice (*Oryza sativa*). Plant Cell Physiol. **32:** 989-997.

Hohmann U. and Lagudah E. S. (1993) C-banding polymorphism and linkage of nonhomoeologous RFLP loci in the D genome progenitor of wheat. Genome **36:** 235-243.

Hovenkamp-Hermelink J. H. M., Bijmolt E. W., de Vries J. N., Witholt B. and Feenstra W. J. (1987) Isolation of an amylose-free starch mutant of the potato (*Solanum tuberosum L.*). Theor. Appl. Genet. **75:** 217-221.

Jaaska V. (1978) NADP-dependent aromatic alcohol dehydrogenase in polyploid wheats and their diploid relatives. On the origin and phylogeny of polyploid wheats. Theor. Appl. Genet. **53:** 209-217.

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.

Jakubziner M. M. (1958) New wheat species. Proc. of the 1st Int. Wheat Symp. Winnipeg, Canada pp: 207-217

Jenner C. F., Deyer K. and Guerin J. (1995) Thermal characteristics of soluble starch synthase from wheat endosperm. Aust. J. Plant Physiol. **22:** 703-709.

Jiang J. and Gill B. S. (1994) New 18S-26S ribosomal RNA gene loci: chromosomal landmarks for the evolution of polyploid wheats. Chromosoma **100:** 71-78.

Johnson B. L. (1968) Electrophoretic evidence on the origin of *Triticum zhukovskyi*. Proc. of the 3rd Int. Wheat Symp. Canberra, Australia. 105-110.

Johnson B. L. (1975) Identification of the apparent B-genome donor of wheat. Can. J. Genet. Cytol. **17:** 21-39.

Johnson B. L. and Dhaliwal H. S. (1976) Reproductive isolation of *Triticum boeoticum* and *Triticum urartu* and the origin of the hexaploid wheats. Amer. J. Bot. **63:** 1088-1094.

Johnson B. L. and Dhaliwal H. S. (1978) *Triticum urartu* and genome evolution in the tetraploid wheats. Amer. J. Bot. **65:** 907-918.

Jones B. L., Mak L A. and Cooper D. B. (1982) Sequences of purothionins and their inheritance in diploid, tetraploid, and hexaploid wheats. J. Heredity **73:** 143-144.

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.

Kam-Morgan M. and Gill B. S. (1989) DNA restriction fragment length polymorphisms: a strategy for genetic mapping of the D genome of wheat. Genome **32:** 724-732.

Kanzaki K. and Noda K. (1988) Glutinous (waxy) endosperm starch mutant of *Triticum monococcum* L. Japan. J. Breed. **38:** 423-427.

Kawasaki T., Mizuno K., Baba T. and Shimada H. (1993) Molecular analysis of the gene encoding a rice starch branching enzyme. Mol. Gen. Genet. **237:** 10-16.

Keeling P. L., Wood J. R., Tyson R. H. and Bridges I. G. (1988) Starch synthesis in developing wheat grain. Evidence against the direct involvement of triose phosphates in the metabolic pathway. Plant Physiol. **87:** 311-311.

Kellogg E. A., Appels R. and Mason-Gamer R. J. (1996) When genes tell different stories: the diploid genera of *Triticeae* (*Gramineae*). System. Bot. **21:** 321-347.

Kerber E. R. (1964) Wheat: reconstitution of tetraploid component (AABB) of hexaploid wheat. Science **143:** 253-255.

Kerber E. R. and Green G. J. (1980) Suppression of stem rust resistance in the hexaploid wheat cv. Canthach by chromosome 7DL. Can. J. Bot. **58:** 1347-1350.

Kerby K. and Kuspira J. (1987) The phylogeny of the polyploid wheats *Triticum aestivum* (bread wheat) and *Triticum turgidum* (macaroni wheat). Genome **29:** 722-737.

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.

Keskin S., Eser V., Gocmen B. and Erisen S. (1998) Biochemical variation in *Ae. speltoides spel.* and *Ae. speltoides lig.* Populations. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **2:** 53-55.

Kimber G. and Athwal R. S. (1972) A reassessment of the course of evolution of wheat. Proc. Natl. Acad. Sci. USA **69:** 912-915.

Kimber G. (1984) Evolutionary relationships and their influence on plant breeding. In: Gene manipulation in plant improvement (ed: J. P. Gustafson). Plenum Press, New York, pp 281-293.

Kimber G. and Sears E. R. (1987) Evolution in the genus *Triticum* and the origin of cultivated wheat. In: Wheat and wheat improvement. (2nd ed. E. G. Heyne). Amer., Society of Agron, Crop Science Society of Amer. and Soil Sci. Society of Amer. Publishers Madison, Wis., USA pp. 154-164.

Kiribuchi-Otobe C., Nagamine T., Yanagisawa T., Ohnishi M. and Yamaguchi I. (1997) Production of hexaploid wheats with waxy endosperm character. Cereal Chem. **74:** 72-74.

Kiribuchi-Otobe C., Yasui T., Yanagisawa T. and Yoshida H. (1998) Allelism test of waxy hexaploid wheats from different sources. Breed. Sci. **48:** 93-94.

Klösgen R. B., Gierl A., Schwarz-Sommer Z. S. and Saedler H. (1986) Molecular analysis of the *waxy* locus of *Zea mays.* Mol. Gen. Genet. **203:** 237-244.

Klösgen R. B., Saedler H. and Weil J. H. (1989) The amyloplast-targeting peptide of the waxy protein of maize also mediates protein transport *in vitro* into chloroplasts. Mol. Gen. Genet. **217:** 155-161.

Klösgen R. B. and Weil J. H. (1991) Subcellular location and expression level of a chimeric protein consisting of the maize waxy transit peptide and the ß-glucuronidase of *Escherichia coli* in transgenic potato plants. Mol. Gen. Genet. **225:** 297-304.

Koßmann J., Buttcher V., Abel G.J.W., Duweig E., Emmermann M, Frohberg C., Lod J. R., Springer F., Welsh T. and Willmitzer L. (1997) Starch Biochemistry and modification of starch structure in transgenic plants. Macromol Symp. **120:** 29-38.

Kobrehel K. and Feillet P. (1975) Identification of genomes and chromosomes involved in peroxidase synthesis of wheat seeds. Can. J. Bot. **53:** 2336-2344.

Konarev V. G. (1983) The nature and origin of wheat genomes on the data of grain protein immunochemistry and electrophoresis. Proc. of the 6th Int. Wheat Genet. Symp. Kyoto, Japan. 65-75.

Konik C. M., Miskelly D. M. and Gras P. W. (1992) Contribution of starch and non-starch parameters to the eating quality of Japanese white salted noodle. J. Sci. Food Agric. **58:** 403-406.

Konik C. M. Miskelly D. M. and Gras P. W. (1993) Starch swelling power, grain hardness and protein: Relationship of sensory properties of Japanese noodles. Starch/Staerke **46:** 292-299.

Konishi Y., Nojima Y., Okuno K., Asaoka M. and Fuwa H. (1985) Characterisation of starch granules from waxy, nonwaxy, and hybrid seeds of *Amaranthus hypochondriacus* L. Agric. Biol. Chem. **49:** 1965-1971.

Kronstad W. E. (1998) Agricultural development and wheat breeding in 20th century. In: H.-J. Braun, F. Altay, W. E. Kronstad, S. P. S. Beniwal and A. McNab (eds) Wheat: prospects for global improvement. Proc. of the 5th Int. Wheat. Conf. Ankara, Turkey. pp: 1-10.

Kumar I. and Khush G. S. (1988) Inheritance of amylose content in rice (*Oryza sativa* L.). Euphytica **38:** 261-269.

Kumar A., Larsen C. E. and Preiss J. (1986) Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADP-glucose: a 1,4-glucan, 4-glycosyl transferase as deduced from the nucleotide sequence of the *glg*A gene. J. Biol. Chem. **261:** 16256-16259.

Kushnir U. and Halloran G. M. (1981). Evidence for *Aegilops sharonensis* as the donor of the B genome of wheat. Genetics **99:** 495-512.

Lagudah E. S., Clarke B. C. and Appels R. (1989) Phygenetic relationships of *Triticum tauschii*, the D-genome donor to hexaploid wheat. 4. Variation and chromosomeal location of 5S DNA. Genome **32:** 1017-1025.

Lagudah E. S., Appels R. and Brown A. H. D. (1991). The molecular genetic analysis of *Triticum tauschii*, the D genome donor to hexaploid wheat. Genome **34:** 375-386.

Law C. N. (1966) The location of genetic factors affecting a quantitative character in wheat. Genetics **53:** 487-498.

Law C. N. (1967) The location of genetic factors controlling a number of quantitative characters in wheat. Genetics **56:** 445-461.

Li Z., Rahman S., Kosar-Hashemi B., Mouille G., Appels R. and Morell M. K. (1999) Cloning and characterisation of a gene encoding wheat starch synthase I. Theor. Appl. Genet. *In press*.

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

Liu C. J., Atkinson C. N., Chinoy C. N., Devos K. M. and Gale M. D. (1990) Nonhomoeologous translocations between group 4, 5 and 7 chromosomes within wheat and rye. Theor. Appl. Genet. **83:** 305-312.

Liu X. Wang R. R.-C., Jia J. and Dong Y. (1998) Genome relationships among *Sitopsis* species of *Aegilops* and the B/G genome of *Triticum* assessed by RAPD markers. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **2:** 79-81.

Liu Y. G. and Tsunewaki K. (1991) Restriction fragment length polymorphism (RFLP) analysis in wheat. II Linkage maps of the RFLP sites in common wheat. Jpn. J. Genet. **66:** 617-633.

Lomedico P., Rosenthal N., Efstratiadis A., Gilbert W., Kolodner R., Tizard R., 1979. Cell **18:** 545-58.

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.

MacDonald F. D. and Preiss J. (1983) Solubilization of the starch-granule-bound starch synthase of normal maize kernels. Plant Physiol. **73:** 175-178.

MacDonald F. D. and Preiss J. (1985) Partial purification and characterisation of granule-bound starch synthases from normal and waxy maize. Plant Physiol. **78:** 849-852.

Mackey J. (1954) Neutron and X-ray experiments in wheat and a revision of the *speltoides* problem. Hereditas **40:** 65-180.

McCormick K. M., Panozzo J. F., Hong S. H. (1991) A swelling power test for selecting potential noodle quality wheat. Aus. J. Agric. Research **42:** 317-323.

Maddelein M-L., Libessart N., Bellanger F., Belrue B., D'Hulst C. van den Koornhuyse N., Fontaine T., Wieruszeski J-M., Decq A. and Ball S. (1994). Towards an understanding of the biogenesis of the starch granule: determination of granule-bound and soluble starch synthase functions in amylopectin synthesis. J. Biol. Chem. **268:** 25150-57.

Martin C. and Smith A. M. (1995) Starch synthesis. The Plant Cell. **7:** 971-985.

Mason-Gamer R. J., Weil C. F. and Kellogg E. A. (1998) Granule-bound starch synthase: structure, evolution and phylogenetic utility. Mol. Biol. Evol. **15:** 1658-1673.

Matus M. A., Hucl P. J. and Chibbar R. N. (1998) Isolation and Characterisation of a granule-bound starch synthase enzyme I cDNA from wheat. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 201-202.

May C. E. and Appels R. (1987) The molecular genetics of wheat: toward an understanding of 16 billion base pairs of DNA. Wheat and Wheat Improvement (2nd ed. E. G. Heyne), pp 165-198.

McIntosh R. A., Hart G. E., Devos K. M., Gale M. D. and Rogers W. J. (1998) Catalogue of gene symbols for wheat. In: Proceedings of the 9th International Wheat Genetics Symposium, Saskatoon, Canada, Ed., A. E. Slinkard. pp 115-116.

McKenzie K. S. and Rutger J. N. (1983) Genetic analysis of amylose content, alkali spreading score and grain dimension in rice. Crop Sci. **23:** 306-313.

Merrit N. R. (1967) A new strain of barley with starch of high amylose content. J. Inst. Brew. **73:** 583-585.

Miura H. and Tanii S. (1994) Endosperm starch properties in several wheat cultivars preferred for Japanese noodles. Euphytica **72:** 171-175.

Miura H., Tanii S., Nakamura T. and Watanabe N. (1994) Genetic control of amylose content in wheat endosperm starch and differential effects of three *wx* genes. Theor. Appl. Genet. **89:** 276-280.

Miura H. and Sugawara A. (1996) Dosage effects of the three *wx* genes on amylose synthesis in wheat endosperm. Theor. Appl. Genet. **93:** 1066-1070.

Miura H., Tarui S., Araki E. and Nakagawa Y. (1998) Production of Wx-protein deficient lines in wheat cv. Chinese Spring. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 208-210.

Miyashita N., Mori N. and Tsunewaki K. (1994) Molecular variation in chloroplast DNA regions in ancestral species of wheat. Genetics **137:** 883-889.

Mizuno K., Kawasaki T., Shimada H., Satoh H., Koyabashi E. M., Okumura S., Arai Y. and Baba T. (1993) Alteration of the structural properties of starch components by the lack of an isoform of starch branching enzyme in rice seeds. J. Biol. Chem. **268:** 19084-19091.

Monte J. V., McIntyre C. L. and Gustafson I. P. (1993) Analysis of phylogenic relationship relationships in the *Triticeae* tribe using RFLPs. Theor. Appl. Genet. **86:** 845-855.

Mori N., Liu Y.-G., Tsunewaki K. (1995) Wheat phylogeny determined by RFLP analysis of nuclear DNA. 2. Wild tetraploid wheat. Theor. Appl. Genet. **90:** 129-134.

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.

Morell M. K., Rahman S., Alrahams S. L. and Appels R. (1995) The biochemistry and molecular biology of starch synthesis in cereals. Aust. J. Plant Physiol. **22:** 647-60.

Morell M., Blennow A., Kosar-Hashemi B., Samuel M. (1997) Differential expression and properties of starch branching isoforms in developing wheat endosperm. Plant Physiol. **113:** 201-208.

Morris C. F. (1998) Genetic determinant of wheat grain quality. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 245-253.

Morris R. and Sears E. R. (1967) The cytigenetics of wheat and its relatives. In: K. S. Quisenberry and L. P. Reitz (eds), Wheat and wheat improvement. Amer. Society of Agron. Madison, WI. USA pp. 19-87.

Mu C., Harn C., Ko Y. T., Singletary G. W., Keeling P. L. and Wasserman B. P. (1994) Association of a 76 kDa polypeptide with soluble starch synthase I activity in maize endosperm. Plant J. **6:** 151-159.

Mu-Froster C. and Wasserman B. P. (1998) Surface location of zein storage proteins in starch granules from maize endosperm. Plant Physiol. **116:** 1563-1571.

Muller-Rober B.and Koβmann J. (1994) Approaches to influence starch quantity and starch quality in transgenic plants. Plant Cell Environ. **17:** 601-613.

Murai J., Moriyama Y. and Taira T. (1998) Phylogenetic analysis based on the nucleotide sequences of the waxy genes in polyploid wheats and related diploid species. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **2:** 95-97.

Nair R., Baga M., Schles G., Kartha K. and Chibbar R. (1997) Isolation, characterisation and expression analysis of a starch branching enzyme II cDNA from wheat. Plant Sci. **122:** 153-163.

Nagamine T., Kato T., Yoshida H. and Komae K. (1998) Genetic variation and developmental changes of the multiple isoforms of starch branching enzyme in wheat endosperm and their influence on the amylopectin structure. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 225-228.

Nakamura T., Yamamori M., Hidaka S and Hoshino T. (1992) Expression of HMW Wx protein in Japanese common wheat (*Triticum aestivum L.*) cultivars. Jpn. J. Breed. **42:** 681-685.

Nakamura T., Yamamori M., Hirano H. and Hidaka S. (1993a) Identification of three Wx proteins in wheat (*Triticum aestivum L.*). Biochem. Genet. **31:** 75-86.

Nakamura T., Yamamori M., Hirano H. and Hidaka S. (1993b) Decrease of waxy (Wx) protein in two common wheat cultivars with low amylose content. Plant Breed. **111:** 99-105.

Nakamura T., Yamamori M., Hirano H., Hidaka S. and Nagamine T. (1995) Production of waxy (amylose-free) wheats. Mol. Gen. Genet. **248:** 253-259.

Nakamura T., Yamamori M., Hirano H. and Hidaka S. (1995) In: Z. S. Li and Z. Y. Xin (eds), Proc. of the 8th Int. Wheat Genet. Symp. Beijing, China.

Nakamura Y., Umemoto T., Takahata Y., Komae K., Amano E. and Satoh H. (1996a) Changes in the structure of starch and enzyme activities affected by sugary mutations in developing rice endosperm. Possible role of starch debranching enzyme (R-enzyme) in amylopectin biosynthesis. Physiol. Plant **97:** 491-498.

Nakamura Y. Umemoto T., Ogata N., Kuboki Y., Yano M and Sasaki T. (1996b) Starch debranching enzyme (R-enzyme or pullulanse) from developing rice endosperm: purication, cDNA and chromosomal localization of the gene. Planta **199:** 209-218.

Naranjo T., Roca A., Goicoecha P.G. and Giraldz R. (1987) Arm homoeology of wheat and rye chromosomes. Genome **29:** 873-882.

Nelson O. E. and Rhines H. W. (1962) The enzymatic deficiency in the waxy mutant of maize. Biochem. Biophys. Res. Commun. **9:** 297-300.

Nelson O. E., Chourey P. S. and Chang M. T. (1978) Nucleoside diphosphate sugar-starch glucosyl transferase activity of wx starch granules. Plant Physiol. **62:** 383-386.

Nelson O. E. (1987) The *waxy* locus in maize twenty-five years later. Genetics **116:** 339-342.

Netrphan S., Khandelwal R. L. and Chibbar R. N. (1998) Starch debranching enzyme activities in developing wheat (*Triticum aestivum L.*) kernels. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 232-233.

Nishikawa K. (1983) Species relationship of wheat and its putative ancestors as viewed from isozyme variation. Proc. of the 6th Int. Wheat Genet. Symp. Kyoto, Japan, 59-63.

Nishikawa K., Furuta Y., Yamada T. and Kudo S. (1992) Genetic studies of α-amylase isozymes in wheat. VII. Variation in diploid ancestral species and phylogeny of tetraploid. Jpn. J. Genet. **67:** 1-15.

Oda M., Yasuda Y., Okazaki S., Yamauchi Y. and Yokoyama Y. (1980) A method of flour quality assessment for Japanese noodles. Cereal Chem. **57:** 253-254.

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.

Okagaki R. J. and Wessler S. R. (1988) Comparison of non-mutant and mutant waxy genes in rice and maize. Genetics **120:** 1137-1143.

Okagaki R.J. (1992) Nucleotide sequence of a long cDNA from the *waxy* gene. Plant Mol. Biol. **19:** 513-516.

Okita T.W. (1992) Is there an alternative pathway for starch synthesis? Plant Physiol. **100:** 560-564.

Ogihara Y. and Tsunewaki K. (1988) Diversity and evolution of chloroplast DNA in *Triticum* and *Aegilops* as revealed by restriction fragment analysis. Theor. Appl. Genet. **76:** 321-332.

Pan D. and Nelson O. E. (1984) A debranching enzyme deficiency in endosperms of the sugary-1 mutants of maize. Plant Physiol. **74:** 324-328.

Payne P. I. and Lawrence T. (1983) Catalogue of alleles for the complex gene loci, *Glu-A1*, *Glu-B1* and *Glu-D1* which code for high molecular weight subunits of glutenin in hexaploid wheat. Cereal Res. Commun. **11:** 29-35.

Preiss J. (1991) Biology and molecular biology of starch synthesis and its regulation. (ed. Miflin B. J.). Oxford Surveys of Plant Molecular and Cell Biology, Vol. 7 Oxford University Press, Oxford, UK pp 59-114.

Rabbi M. F. and Wilson K. G. (1993) The mitochondrial *coxII* intron has been lost in two different lineages of dicots and altered in others. Amer. J. Bot. **80:** 1216-1223.

Poulsen P. and Kreiberg J. D. (1993) Starch branching enzyme cDNA from *Solanum tuberosum*. Plant Physiol. **102:** 1053-1054.

Hoshino T., Ito S., Hatta K., Nakamura T. and Yamamori M. (1996) Development of waxy common wheat by haploid breeding. Breed. Sci. **46:** 185-188.

Rahman S., Behjat K. H., Samuel M. S., Hill A., Abbott D. C., Skerritt J. H., Preiss J., Appels R. and Morell M. K. (1995) The major proteins of wheat endosperm starch granules. Aust. J. Plant Physiol. **22:** 793-803.

Rahman S., Abrahams S., Abbott D., Mukai Y., Samuel M., Appels R. (1997) A complex arrangement of genes at a starch branching enzyme I locus in the D-genome donor of wheat. Genome **40:** 465-474.

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.

Rayburn A. L. and Gill B. S. (1985) Molecular evidence for the origin and evolution of chromosome 4A in polyploid wheats. Can. J. Genet. Cytol. **27:** 246-250.

Repellin A., Nair R., Baga M. and Chibbar R. (1997) Isolation of a starch-branching enzyme I cDNA from a wheat endosperm library (accession No. Y12320). Plant Physiol. **114:** 1135-1136.

Repellin A., Weir B. J. and Chibbar R. N. (1998) Isolation and characterisation of partial cDNAs encoding starch debranching enzyme (limit dextrinase or pullulanase) from developing wheat kernels. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 265-267.

Rodriguez-Quijano M., Neito-Taladriz M. T. and Carrello J. M. (1997) Variation in B-LMW glutenin subunits in Einkorn wheats. Genet. Resource and Crop Evol. **44:** 539-543.

Rohde W., Becker D. and Salamini F. (1988) Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucl. Acids Res. **16:** 7185.

Rohde W., Becker D., Kull B. and Salamini F. (1990) Structural and functional analysis of two *waxy* gene promoters from potato. J. Genet. and Breed. **44:** 311-315.

Ross A.S., Quail K. J. and Crosbie G. B. (1997) Physicochemical properties of Australian flours influencing the texture of yellow alkaline noodles. Cereal Chem. **74:** 824-820.

Saitou N. and Nei M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406-425.

Sambrook J., Fritsch E. F. and Maniatis (1989) Molecular Cloning: A laboratory Manual. Cold Spring Harbor, New York, USA.

Sano Y. (1984) Differential regulation of *waxy* gene expression in rice endosperm. Theor. Appl. Genet. **68:** 467-473.

Sano Y., Maekawa M. and Kikuchi H. (1985) Temperature effects on the Wx protein level and amylose content in the endosperm of rice. J. Heredity **76:** 221-222.

Sano Y., Katsumata M. and Okuno K. (1986) Genetic studies of speciation in cultivated rice. 5. Inter-and intraspecific differentiation in the *waxy* gene expression of rice. Euphytica **35:** 1-9.

Sasamuma T., Miyashita N. T., Tsunewaki K. (1996) Wheat phylogeny determined by RFLP analysis of nuclear DNA. 3. Intra- and interspecific variations of five *Aegilops Sitopsis* species. Theor. Appl. Genet. **92:** 928-934.

Sax, K. (1922), Sterility in wheat hybrids. II. Chromosome behaviour in partially sterile hybrids. Genetics **7:** 513-552

Sears E. R. (1972) The nature of mutation in hexaploid wheat. Symp. Biol. Hung **12:** 73-82.

Schwartz D. and Echt C.S. (1982) The effect of *Ac* dosage on the production of multiple forms of Wx protein by the *wxm-9* controlling element mutation in maize. Mol. Gen. Genet. **187:** 410-413.

Schofield J. D. and Greenwell P. (1988) Wheat starch granule proteins and their technological significance Cereals in a European Context (I. D. Morton, ed.) pp.407-420.

Schondelmaier J., Jacobi A., Fischbeck G and Jahoor A. (1992) Genetic studies on the ode of inheritance and localisation of the *amol* (high amylose) gene in barley. Plant Breed. **109:** 274-290.

Sears E. R. (1972) The nature of mutation in hexaploid wheat. Symp Biol. Hung **12:** 73-82.

Shands H. and Kimber G. (1973) Reallocation of the genomes of *Triticum timopheevii* Zhuk. In: Proceedings of the 4th Int. Wheat Genet. Symp. Columbia, Missouri, Columbia, USA pp. 101-108.

Shure M., Wessler S.R. and Federoff N. (1983) Molecular identification and isolation of the *waxy* locus in maize. Cell **35:** 225-233.

Shimada H., Tada Y., Kawasaki T. and Fujimura T. (1993) Antisense regulation of the rice *waxy* gene expression using a PCR-amplified fragment of the rice genome reduces the amylose content in grain starch. Theor. Appl. Genet. **86:** 665-672.

Siedler H., Messmer M. M., Schachermayr G. M., Winzeler H., Winzeler M. and Keller B. (1994) Genetic diversity in European wheat and spelt breeding material based on RFLP data. Theor. Appl. Genet. **88:** 994-1003.

Siedler H., Messmer M. M., Schachermayr G. M., Winzeler M., Keller B. (1994) Genetic diversity in European wheat and spelt breeding material based on RFLP data. Theor. Appl. Genet. **88:** 994-1003.

Simpson G. G. and Filipowicz (1996) Slicing of precursors to mRNA in higher plants: mechanism, regulations and sub-nuclear organisation of the spliceosomal machinery Plant Mol. Biol. **32:** 1-41.

Sivak M.N., Wagner M. and Preiss J. (1993) Biochemical evidence for the role of the waxy protein for pea (*Pisum sativum* L.) as a granule-bound starch synthesis. Plant Physiol. **103:**1355-1359.

Smith A. M. (1988) Major differences in isoforms of starch branching enzyme between developing embryos of round-and winkled-seeded peas (*Pisum Sativum* L.). The Plant J. **175:** 270-279.

Smith A. M., Bettey M. and Bedford I. D. (1989) Evidence that the *rb* locus alters the starch content of developing pea embryos through an effect on ADP glucose pyrophosphorylase. Plant Physiol. **89:** 1279-1284.

Smith A.M. (1990) Evidence that the waxy protein of pea (*Pisum sativum* L.) is not the major starch-granule-bound starch synthase. *Planta* **182:** 599-604.

Smith A. M. and Martin C. (1993) Starch biosynthesis and the potential for its manipulation. In: Biosynthesis and manipulation of plant products, Plant biotechnology (ed. Grierson D). Blackie Academic and Professional Press, Glasgow. Vol. 3, pp 1-54.

Smith A. M., Denyer K. and Martin C. R. (1995) What control the amount and structure of starch in storage organs? Plant Physiol. **107:** 673-677.

Smith A. M. Denyer K. and Martin C. (1997) The synthesis of the starch granule. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48:** 67-87.

Soltis, D. E and Soltis P. S. (1993) Molecular data and the dynamic nature of polyploidy. Crit. Rev. Plant Sci. **12:** 243-273.

Soulaka A. B. and Morrison W. R. (1985) The bread baking quality of six wheat starches differing in composition and physical properties. J. Sci. Food Agric. **36:** 719-727.

Taira T., Fujita N., Takaoka K., Uematsu K., Wadano A., Kozaki S. and Okabe S. (1995) Variation in the primary structure of waxy proteins (granule-bound starch synthase) in diploid cereals. Biochem. Genet. **33:** 269-280.

Takaoka M., Watanabe S., Sassa H., Yamamori M., Nakamura T., Sasakuma T. andd Hirano H. (1997) Structural characterisation of high molecular weight starch granule-bound proteins in wheat (*Triticum aestivum* L.). J. Agric. Food Chem. **45:** 2929-2934.

Takeda Y. and Hizukuri S. (1987) Structures of rice amylopectin with low and high affinities iodine. Carbohydrate Res. **168:** 79-88.

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. Amer. J. Bot. **78:** 340-349.

Talbert L. E., Kimber G., Magyar G. M. and Buchanan C. B. (1993) Repetitive DNA variation and pivotal-differential evolution of wild wheats. Genome **36:** 14-20.

Talbert L. E., Blake N. K., Chee P. W. and Magyer G. M. (1994) Evaluation of "sequence-targged-site" PCR products as molecular markers in wheat. Theor. Appl. Genet. **87:** 789-794.

Talbert L. E., Blake N. K., Storlie E. W. and M. Lavin (1995) Variation in wheat based on low-copy DNA sequence comparisons. Genome **38:** 951-957.

Talbert L. E., Smith L. Y. and Blake N. K. (1998) More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome **41:** 402-407.

Tanaka M., Kawahara T. and Sano J. (1978) The evolution of wild tetraploid wheats. Proc. of the 5th Int. Wheat Genet. Symp. New Delhi, India. **1:** 73-80.

Terachi T., Ogihara Y., and Tsunewaki K. (1993) The *rbcL* gene in wheat and several *Aegilops* species with divergent chloroplast genomes Proceedings of the 6th Int. Wheat Genet. Symp. Kyoto pp. 789-795.

Terachi T. and Tsunewaki K. (1992) The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. VII. Mitochondrial RFLP analysis using cloned genes as probes. Mol. Biol. Evol. **9:** 917-931.

Tetlow I. J., Blissett K. J. and Emes M. J (1994) Starch synthesis and carbohydrate oxidation in amyloplasts from developing wheat endosperm. Planta **194:** 454-460.

Thompson P. J. and Nath J. (1986) Elucidation of the B-genome donor to *Triticum turgidum* by unique-and repeated-sequence DNA hybridisations. Biochem. Genet. **24:** 39-50.

Toyokawa H., Rubenthaler G. L., Powers J. R. and Schanus E. G. (1989) Japanese noodle qualities. II. Starch components. Cereal Chem. **66:** 387-391.

Tsunewaki K. and Ogihara Y. (1983) The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops* species. II. On the origin of polyploid wheat cytoplasms as suggested by chloroplast DNA restriction fragment payyerns. Genetics **104:** 155-171.

Tsunewaki K. (1993) Genome-plasmon interactions in wheat. Jpn. J. Genet. **68:** 1-34.

Upadhya M. D. and Swaminathan M. S. (1963) Genome analysis in *Triticum zhukovskyi*, a new hexapoid wheat. Chromosoma **14:** 589-600.

Van Campenhout S., Sagi L., Vander Stappen J. and Volckaert G. (1998) Characterisation of type-I thionin loci from the A, B, D and R genomes of wheat and rye. Theor. Appl. Genet. **96:** 80-86.

Vasil V., Castillo A. M., Fromm M. E. and Vasil I. K. (1992) Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenetic callus. Bio/Technology **10:** 667-675.

van der Leij F. R., Visser R. G. F., Ponstein A. S., Jacobsen E. and Feenstra W. J. (1991) Sequence of structural gene for granule-bound starch synthase of potato (*Solanum tuberosum* L.). Mol. Gen. Genet. **228:** 240-248.

Vershinin A., Svitashev S., Gummesson P.-O., Salomon B., Bothmer R. von and Bryngelsson T. (1994) Characterisation of a family of tandemly repeated DNA sequences in *Triticeae.* Theor. Appl. Genet. **89:** 217-225.

Visser R. G. F., Hergersberg M., van der Leij F. R., Jacobsen E., Witholt B. and Feenstra W. J. (1989) Molecular cloning and partial characterisation of the gene for granule-bound starch synthase from a wildtype and an amylose-free potato (*Solanum tuberosum* L.). Plant Sci. **64:** 185-192.

Visser R.G.F. and Jacobsen E. (1993) Toward modifying plants for altered starch content and composition. Trends Biotech. **11:** 63-68.

Visser R. G. F., Somhort I., Kuipers G. J., Ruys N. J., Feenstra W. J. and Jacobsen E. (1991) Inhibition of the expression of the gene for granule-bound starch synthase in potato by antisense constructs. Mol. Gen. Genet. **225:** 296-298.

Vos-Scheperkeuter G. H., Boer W., Visser R. G. F., Feenstra W. J. and Witholt B. (1986) Identification of granule-bound starch synthase in potato tubers. Plant Physiol. **82:** 411-416.

Waines J. G. and Barnhart D (1992) Biochemistic research in *Aegilops* and *Triticum*. Hereditas **116:** 207-212.

Wang T. L., Bogracheva T. Y and Hedley C. L. (1998) Starch: as simple as A, B, C? J. Experi. Bot. **49:** 481-502.

Wang Z., Wu Z., Xing Y., Zheng F., Guo X., Zhang W. and Hong M. (1990) Nucleotide sequence of rice *waxy* gene. Nucl. Acids Res. **18:** 5898.

Wang Z., Zheng F., Shen G., Gao J., Snustad P., Li M., Zhang J. and Hong M. (1995) The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. The Plant J. **7:** 613-622.

Wasserman B. P. (1995) Progress towards genetically modified starches. Cereal Foods World **40:** 810-17.

Watanabe N., Noda Y. and Miura H. (1998) Variation for apparent amylose content of endosperm starch in *Triticum durum* and *Aegilops squarrosa*. Euphytica **101:** 283-286.

Weeks J. T., Anderson O. D. and Blechl A. E. (1993) Rapid production of multiple independent lines of fertile transgenic wheat. Plant Physiol. **102:** 1077-1084.

Wessler S. R., Baran G., Varagona M. and Dellaporta S. L. (1986) Excision of *Ds* produces waxy proteins with a range of enzymatic activities. The EMBO J. **5:** 2427-2432.

Wolfe K. H., Gouy M., Yang Y. W., Sharp P. M. and Li W. H. (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86:** 6201-6205.

Wright C. S. and Raikhel N. (1989) Sequence variability in three wheat germ agglutinin isolectins: products of multiple genes in polyploid wheat. J. Mol. Evol. **28:** 327-336.

Yamamori M., Nakamura T. and Kuroda A. (1992) Variations in the content of starch-granule bound proteins among several Japanese cultivars of common wheat (*Triticum aestivum* L.). Euphytica **64:** 215-219.

Yamamori M., Nakamura T., Endo T. R. and Nagamine T. (1994) Waxy protein deficiency and chromosomal location of coding genes in common wheat. Theor. Appl. Genet. **89:** 179-184.

Yamamori M., Nakamura T. and Nagamine T. (1995) Polymorphism of two waxy proteins in the emmer group of tetraploid wheat, *Triticum dicocoides, T. dicoccum* and *T. durum.* Plant Breed. **114:** 215-218.

Yamamori M. (1998) Selection of a wheat lacking a putative enzyme for starch synthesis, SGP-1. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 300-302.

Yanagisawa T., Kiribuchi-Otobe C. and Yoshida H. (1998) Wx protein and high-molecular-weight glutenin identification in waxy wheat lines. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **4:** 303-305.

Yasui T., Sasaki T., Matsuki J. and Yamamori M (1997) Waxy endosperm mutants of bread wheat (*Triticum aestivum* L.) and their starch properties. Breed. Sci. **47:** 161-163.

Yu Y., Mu H. H., Mu-Forster C. and Wasserman B. P. (1998) Polypeptides of the maize amyloplast stroma. Plant Physiol. **116:** 1451-1460.

Yun S. H., Quail K. J. and Moss R. (1996) Physiocochemical properties of Australian wheat flours for white salted noodles. J. Cereal Sci. **23:** 181-189.

Zeng M., Morris C. F., Batey I. L. and Wrigley C. W. (1997) Sources of variation for starch gelatinisation, pasting and gelation properties in wheat. Cereal Chem. **74:** 63-71.

Zhao X. C. and Sharp P. J. (1996) An improved 1-D-PAGE method for the identification of three bread wheat 'waxy' proteins. J. Cereal Sci. **23:** 191-193.

Zhao X. C., Batey I. L., Sharp P. J., Crosbie G., Barelay I., Wilson R., Morell M. K. and Appels R. (1998) A single genetic locus associated with starch granule properties and noodle quality in wheat. J. Cereal Sci. **27:** 1-13.

Zhao X. C., Shariflou M. R., Good G. and Sharp P. J. (1998) Developing waxy wheat cultivars: Wx null alleles and molecular markers. Proc. of the 9th Int. Wheat Genet. Symp., Saskachewan, Canada **1:** 254-256.

APPENDICES

APPENDIX I

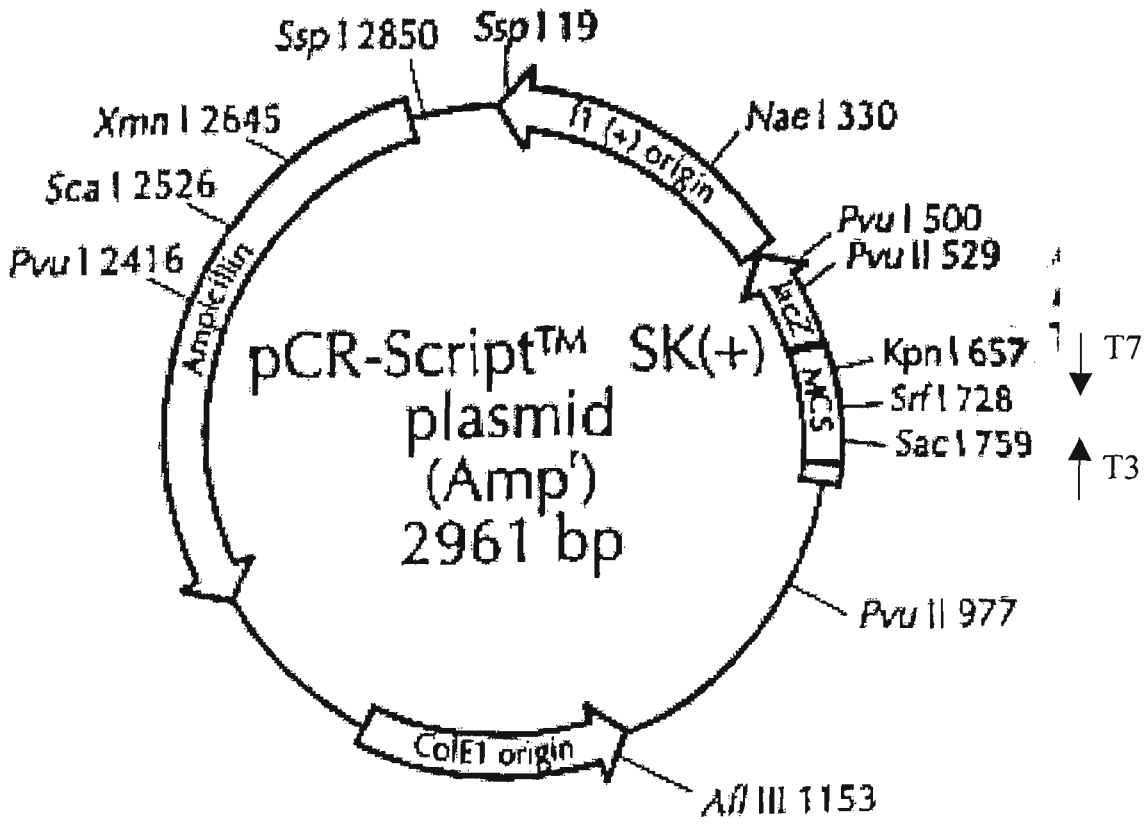


Figure I. Map of the pCR-Script Amp SK(+) cloning vector. The pCR-Script Amp SK(+) cloning vector (Stratagene) was designed by incorporating a *Srf* I site into the polylinker region (MCS, multiple cloning sites) of the pBluescript II SK(+) phagemid. The fragments of the *waxy* genes were cloned into a *Srf* I site in this project.

APPENDIX II

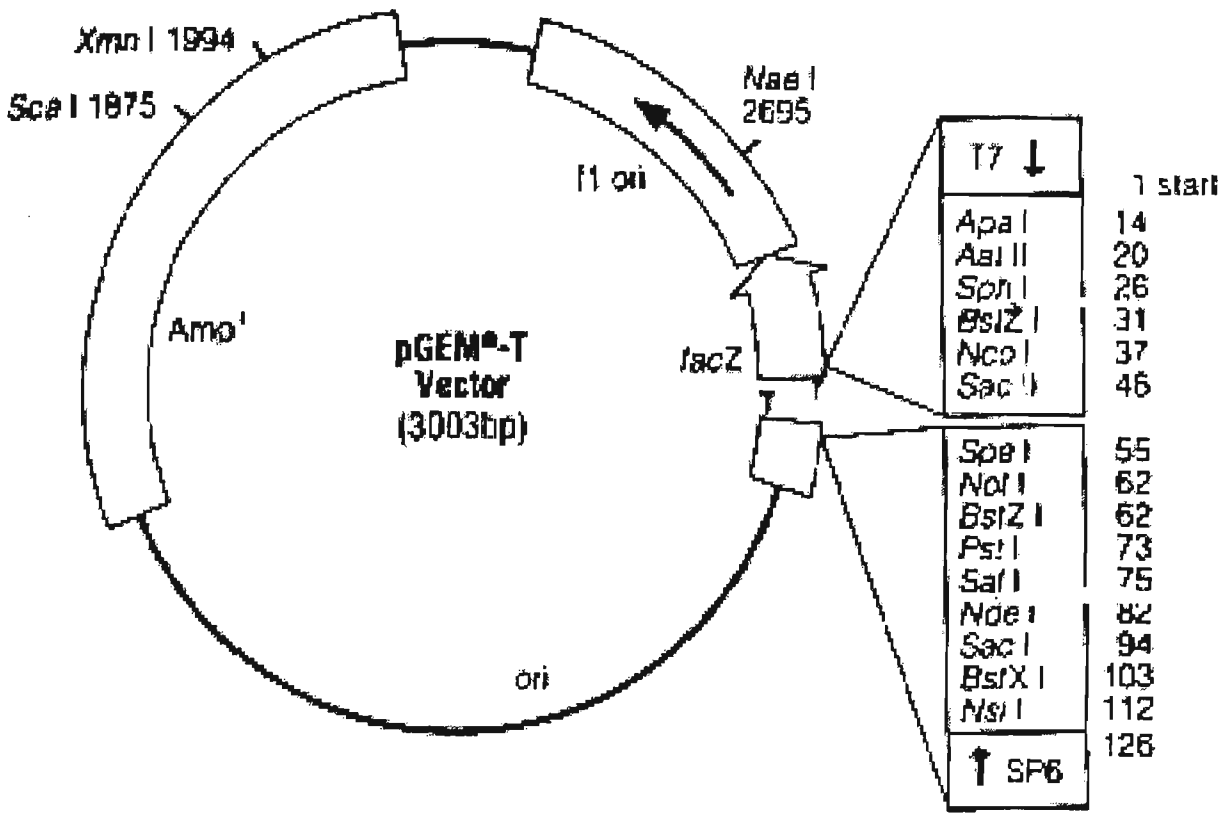


Figure II. Map of the pGEM-T cloning vector. The pGEM-T cloning vector (Promega) was designed by cutting the polylinker region (MCS, multiple cloning sites) in the vector by *Eco*R V between *Sac* II and *Spe*I as shown in the Figure, and then adding a 3' terminal thymidine to both ends. The fragments of the *waxy* genes were cloned by ligation to the extended 3' T termini.

APPENDIX III

CHEMICALS AND REAGENTS

A. Enzymes

A large number of enzymes used in this project were provided with the Kits for PCR amplifications, cloning and sequencing reactions. These enzymes included:

Srf I for digesting the pCR-Script Amp SK(+) cloning vector which was provided by Stratagene;

Polymerases for PCR amplification which was provided by Perkin Elmer, Qiagen and Promega;

T4 DNA Ligase for ligation reaction of fragments of the waxy genes obtained in this project and the pCR-Script Amp SK(+) cloning vector and the pGEM-T cloning vector which were provided by provided by Stratagene and Promega respectively;

Polymerases for sequencing reactions which were provided by Perkin Elmer.

The majority of other restriction enzymes in this project was supplied by Progen and MBI and used according to the instruction of the supplier. Some enzymes were purified and prepared:

- Purification and Preparation of RNase A (10μg/μl)
  100 mg of pancreatic RNase A raw extracts (Progen) was dissolved in 10 mL of sterile 10 mM Tris·Cl buffer (pH 7.5) containing 15 mM NaCl and heated in boiling water bath for 5 minutes to inactivate DNase in the RNase A. This DNaes-free RNase A was clod slowly to room temperature and dispensed into aliquots (20 μl/Eppendorf tube) and stored at -20^0C

- Purification and Preparation of Proteinase K (20μg/μl)

  100 mg of Proteinase K (Sigma) was dissolved in 50 mL of 10 mM Tris·Cl and 10 mM NaCl and incubated at 37^0C for 1 hour. This Proteinase K was dispended into aliquoted (20μl/Eppendorf) and stored at -20^0C.

- Preparation of lysozyme (50 mg/mL)

  100 mg lysozyme (Boehrin) was dissolved in 2 mL of solution (50 mM Tris·Cl, pH 8.0, 10 mM EDTA and 50 glucose) was prepared fresh.

C. Buffers

A great number of buffers for PCR amplifications, ligation reactions, and digestion reactions were supplied with Kits or enzymes used in this project. The following buffers were prepared:

1. TE, 10 mM Tris·Cl (pH 8.0), 1 mM EDTA (pH 8.0).
   High-salt TE buffer 1 M NaCl was prepared by adding 1 M NaCl. This high-salt was used to dissolve the raw genomic DNA extracted by CTAB.

2. STE: 0.1 M NaCl, 10 mM Tris·Cl, 1 mM EDTA.

3. STET: 0.1 M NaCl, 10 mM Tris·Cl, 1 mM EDTA (pH 8.0), 5% Triton X-100.

4. TAE 50X: Tris base, 242 g Glacial acetate acid 57.1 mLEDTA (0.5 M, pH 8.0) 100 mL, These components were dissolved and made up to 1000 mL TAE buffer sterile H_2O.

5. Extraction buffer for preparation of starch granules and proteins

0.55 M Tris-Cl pH 6.8	10 mL
10% SDS	26 mL
Glycerol	10 mL

These three components were mixed and made to 90 mL of extraction buffer and stored at room temperature. 5% β-mercaptoethanol and 5% 1 M DTT stock were added before use.

6. Electrode buffer (pH 8.3)

0.025 M Tris

0.192 M glycine

0.1% SDS.

7. Protein Sample Buffer (loading buffer).

0.125 M Tris·Cl (pH 6.8)

10% glycerol

5% β–mercaptoethanol

2% SDS

0.013% bromophenol blue

8. Genomic DNA extraction buffer from plant leaves

100 mM Tris·Cl pH 8.0,

20 mM EDTA, pH 8.0

1.4 M NaCl,

There three components were made to 96% of the final volume of the buffer and stored at room temperature. 2% β-mercaptoethanol and 2% CTAB (hexadecyltrimethylammonium bromide) was added before use.

C. General solutions and stocks

1. Acrylamide:biscrylamide stock

30 g acrylamide (Electran grade, BioRad) and 0.135 g bisacrylamide (Electran grade, BioRad) were added to 60 mL dH_20 and incubated at 37^0C until dissolved. The solution was made up to 100 mL with d H_20 and sterilised by 0.45 μm filter. This 30:0.135 of acrylamide:bisacrylamide stock was stored in dark bottles at room temperature. The 29 g acrylamide and 1 g bisacrylamide

were dissolved in 100 mL of d H_20 and 29:1 of acrylamide:bisacrylamide stock.

2. 0.5 M EDTA

186.1 g of disodium ethylenediaminetetra-acetate·$2H_2O$ was dissolved to 800 mL of dH_2O by stirring on a magnetic stirrer. pH was adjusted to 8.0 by NaOH pellets (approximately 20 g), autoclaved for at 121^0C for 20 minutes and stored at room temperature.

3. Ethium Bromide (10 mg/mL)

0.5 g ethidium bromide (Sigma) was dissolved into 50 mL of dH_2O by stirring with a magnetic stirrer for 2 hours to ensure that the solid dye has been dissolved. This stock was transferred into a 50 mL tube, which was wrapped in aluminum foil and stored at room temperature.

4. 10% CTAB

10 g CTAB (hexadecyltrimethylammonium bromide, Sigma) was dissolved in 100 mL of 0.7 N NaCl, and stored at room temperature.

5. 10% SDS

10 g sodium dodecyl sulfate (SDS) (electrophoretic grade, BioRad) was dissolved in 90 mL dH_2O and heated at 68^0C until SDS was completely dissolved. This solution was adjusted pH to 7.2 using dilute acetic acid and made up to 100 mL with dH_2O, followed by filtration with 0.22 μm filter. This 10% SDS solution was stored at room temperature.

5. 3 M Sodium acetate

40.8 g of sodium acetate·$3H_2O$ was dissolved in dH_2O to make 100 solution, followed by the adjusting the pH to 5.2 with glacial acetic acid. This solution was autoclaved and stored at room temperature.

7. 5 N NaCl

29.2 g of NaCl was dissolved in dH_2O to make 100 mL solution and autoclaved and stored at room temperature.

8. 1M DDT

3.09 g of dithiotheitol (DTT) (Sigma) was dissolved in 20 mL of 0.1 M sodium acetate (pH 5.2). The solution was sterilised by 0.45 μm filter, dispended into aliquots and stored at -20^0C.

9. 20% IPTG

2 g of isopropylthio-β-D-galactoside (IPTG) (Progen) was dissolved in dH_2O to make a 10 mL solution. This solution was sterilised by 0.22 μm filter, then aliquoted and stored at -20^0C.

10. X-gal (20 mg/mL) .

100 mg of 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (Progen) was dissolved in dimethyformamide to make a 5 mL of solution. This solution was wrapped with aluminum foil and stored at -20^0C.

11. 1 M $MgCl_2$

20.33 g of $MgCl_2.H_2O$ (AR grade, BDH) was dissolved in 80 mL of dH_2O and made up to 100 mL with dH_2O. This solution was sterilised by autoclaving and stored at room temperature.

12. 1 M $CaCl_2$

27 g of $CaCl_2.6H_2O$ was dissolved in 60 mL and made up to 100 mL. This solution was sterilised by 0.22 μm filter and stored at room temperature.

13. Ampicillin (50mg/mL)

200 mg of Amipicillin (Sigma) was dissolved in dH_2O to make 4 mL of solution. This solution was sterilised by 0.22 μm filter and aliquoted into light-tight containers.

14. Phenol

Phenol (Novachem) was equilibrated with Tris·Cl before use. Phenol was melted at 68^{0}C and then an equal volume 0.5 M Tris·Cl (pH 8.0) was added. The mixture was stirred for 15 minutes and then the upper phase was discarded. An equal volume of 0.1 M Tris·Cl (pH 8.0) was added to phenol (lower phase) and mixed up by stirring, followed by the upper phase was discarded. This process was repeated several times until the pH of the aqueous layer was >8.0. This phenol was added 0.1 volume of 0.1 M Tris·Cl (pH 8.0) and a little hydroxyquinoline and stored at -4^{0}C.

C. Bacterial medium

1. LB (Luria Broth)

Trypton	10 g
Yeast extract	50 g
Sodium chloride	10 g

This medium was made up to 1000 mL, followed by autoclaving at 121^{0}C for 20 minutes. 20 g agar was added such LB medium to make LB solid medium and autoclaved..

2. SOB

20.0 g Trypton
5.0 g yeast extract
0.5 g of NaCl

This medium was made up to 1000 mL and autoclaved at 121^{0}C for 20 minutes. 10 mL of 1 M $MgCl_2$ and 10 mL of 1 M $MgSO_4$ were added prior to use.

3. SOC

2 mL of 20% sterilised glucose was added to the SOB medium before use.