Theor Appl Genet (1997) 95: 822–827

D. Mingeot · J. M. Jacquemin

A wheat cDNA coding for a thaumatin-like protein reveals a high level of RFLP in wheat

Received: 27 January 1997 / Accepted: 18 April 1997

Abstract A cDNA clone that reveals a high level of polymorphism between wheat varieties was isolated from a wheat cDNA library. When hybridized to *Dra*I-, *Eco*RV- and *Hin*dIII digested DNA this clone, gbx3832, enables us to distinguish 42 different patterns among 48 varieties: 37 varieties are clearly identified, the remaining 11 are divided into five groups. Base-sequence analysis of the clone reveals 72–74% sequence identity to mRNAs encoding thaumatin-like proteins from different cereals.

Key words RFLP · Thaumatin-like protein · Wheat

Introduction

Molecular probes revealing DNA restriction fragment length polymorphism (RFLP) are being used as tools to improve practical plant breeding. The main applications of molecular markers are the characterization of breeding lines and varieties, as well as marker-assisted selection. However, these applications are hampered in wheat due to the low degree of polymorphism displayed by this crop (Chao et al. 1989; Kam-morgan and Gill 1989; Liu et al. 1990). Moreover, much of the current breeding in wheat involves crosses among lines that share common ancestors.

RFLP markers useful in selection were selected by testing cDNAs and genomic DNAs for their polymorphism on 13 commercial wheat varieties that constitute potential parents in breeding programs. A molecular map of hexaploid wheat was developed with the markers selected. Among these, gbx3832 revealed a very high level of polymorphism. The characterization of this clone is the subject of this report.

Communicated by J. W. Snape

D. Mingeot (✉) · J. M. Jacquemin
Centre de Recherches Agronomiques, Station d'Amélioration des Plantes, 4 rue du Bordia, 5030 Gembloux, Belgium

Materials and methods

Origin and preparation of gbx3832

gbx3832 is a 810-pb *Eco*RI fragment cloned from a cDNA library made from wheat leaf poly(A)$^+$ RNA. The insert was removed from the plasmid pUC18 by *Eco*RI-digestion and was isolated in low-melting-point agarose, following electrophoresis, for use as a probe.

Plant material

The Chinese spring nullisomic-tetrasomic (NT) stocks produced by Prof. E. R. Sears (Sears 1954, 1966) were provided by T. E. Miller, John Innes Center, Norwich, UK. The 14 wheat varieties used to select probes suitable for application to wheat breeding are: Appolo, Camp Remy, Sperber, Token, Odeon, Capitaine, Castell, Soisson, Courtot, Moulin, Genial, Pernel, Franco and Chinese Spring. In the case of probe gbx3832, the study was extended to the 48 varieties listed in Table 1. Mapping was carried out on a population derived from the cross W-7984 (a synthetic wheat) with Opata85 (Nelson et al. 1995 a). The plants were grown in a culture room. About 2 g of leaf material was frozen in liquid N_2, ground to a fine powder and stored at $-70°C$.

RFLP procedures

Genomic DNA was extracted as described by Sharp et al. (1988). Restriction-enzyme digestion and 0.8% agarose-gel electrophoresis followed standard procedures. Southern blotting on Nylon membranes (Hybond N+, Amersham) was as described by the manufacturer. Probe was ^{32}P labelled using the random primed method (Feinberg and Vogelstein 1983). Pre-hybridization and hybridization were performed as described by Sharp et al. (1988). The membranes were washed at 65°C, twice in 2 × SSC/0.1% SDS, twice in 1 × SSC/0.1% SDS and once in 0.5 × SSC/0.1% SDS for 15 min each. Membranes were exposed to X-ray films with one intensifier screen at $-70°C$ for 10–15 days.

Table 1 Pedigree information of the 48 wheat varieties used in the RFLP analysis

Line	Origin	Pedigree	Lane in Fig. 1
Ami	F	Gala/54//Fleuron	
Apollo	D	M.Beacon/Clement//Kronjuwel	18
Arminda	NL	Carstens 854/Ibis	
Bul 94 5	Bulgaria		24
Cadenza	GB	Axona/Tonic	
Camp Remy	F	362/Atou//Hardi	2
Cappelle-Desprez	F	Vilmorin 27/Hybride du joncquois	1
Capitaine	B	B5780//Norda/Cama	20
Castell	D	Caribo/Diplomat	19
Chinese Spring	PRC	Chinese land variety	29
Compair	GB	Chinese spring/Aeco (p)	17
Courtot	F	Mexique 50/B21	8
Eiffel	NL	Arminda/Cebeco 323	
Escorial	B	composite cross	14
Estica	NL	Arminda/Virtue	12
Eureka	F	F1(Mironovskaya 808/M. Huntsmann)/ F1 ((VPM/Moisson)1.5/Courtot)	9
Fidel	F	W33/Capitol//Major	
Franco	D	Carimulti/Monopol//Monopol	
Galahad	GB	Durin/Joss Cambier//Hobbit	
Galibier	F	42.2.3/5684//Ducat	
Genesis	GB	Arminda/Moulin	
Genial	F	Arminda/Abo	
Hardi	F	Capelle/Tatcher	
Highbury	GB	Jufy I/Svenno//Sona 227	16
Lovrin 14	Roumania		22
Maris Huntsman	GB	Capelle/Hybride 46//Prof Marshall	15
Michigan amber	USA		25
Minaret	NL	Bastion/Mironovskaya 808	13
Norin 27	JP	Iga chikugo/Oregon//Jarl	27
Oasis	USA	Atr/5/Atrx3/3/Ribox//RLYx2/RLY67	26
Odeon	B	Zemon/Gemma	
Opata	Mex	BJY/JUP	21
Osu 94 90	USA	650B077/5/ND/P101//BB/GLL	
Paradis	F	7305113/Bizel//743211	
Ramses	GB	Alcedo/Avalon//Moulin	
Renan	F	F1(Mironovskaya 808/M. Huntsmann)/ F1 ((VPM/Moisson)1.5/Courtot)	7
Ritmo	NL	Hobbit/1320//Wizzard/Marksman	11
Roazon	F	Ae. ventricosa/T. persicum//Marne*3 /3/Moisson	5
Rusalka	Bulgaria	S13/Ban 54	23
Shan 482	PRC		28
Sicco	NL	Ring//Opal/Selkirk	10
Soissons	F	Iena/HN35	3
Sperber	D	Robert/Merlin//Kormoran	
Supral	F	Festival/Avalon//Moulin	6
Token	GB	Bounty/Armada//Flanders	
Tremie	F	SE32/Moulin	
Vivant	F	Boxer/Gawain	4
88 15	B	Rendez vous/Moulin	

Informativeness of clones

The polymorphism information content (PIC) described by Botstein et al. (1980) was calculated using the formula $PIC_i = 1 - \Sigma p_{ij}^2$ where p_{ij} is the frequency of the jth RFLP pattern revealed by the probe i (combined with one given enzyme).

Mapping

A subset of 71 F_8 lines derived by single-seed descent from the cross W-7984 with Opata85 (Nelson et al. 1995 a) was used for segregation analysis. Markers were mapped using the computer program Mapmaker v2.0 (Lander et al. 1987). The Kosambi mapping function was used to convert recombination frequency to centimorgans.

DNA sequence analysis

The base sequence of gbx3832 was determined using the dideoxynucleotide chain-termination method (Sanger et al. 1977) as recommended for Sequenase version 2.0. The sequences of both DNA strands were completed by primer walking. DNA sequence data

were analyzed with the GCG Package (Genetics Computer Group, Inc. Wisconsin).

Results

RFLP analysis

In order to select probes suitable for application to wheat breeding, 499 random wheat cDNA and genomic DNA clones were hybridized to the genomic DNA of 14 wheat varieties (13 commercial varieties that constitute potential parents in breeding programs and the reference variety Chinese Spring). DNA was digested by four different restriction enzymes (*Eco*RI, *Hin*dIII, *Eco*RV and *Dra*I). Among the 499 clones, 275 provide a low-copy pattern with strong and easy to read signals. Among these, 131 probes detected polymorphism between at least two varieties for one given enzyme. The polymorphism information content (PIC) has been calculated for each polymorphic probe. The average probe PIC was 0.47 with values ranging from 0.14 to 0.88. The PIC observed for probe gbx3832 (0.88) was the highest recorded, with ten different patterns observed with *Dra*I. To further explore the potential use of gbx3832, the study was extended to the 48 varieties listed in Table 1. When hybridized to *Dra*I-, *Eco*RV- and *Hin*dIII-digested DNA, clone gbx3832 generates 29, 25 and 21 different patterns out of 45, 45 and 46 varieties respectively. The 29 different hybridization patterns for *Dra*I digestions are shown in Fig. 1. From two to seven hybridizing fragments were detected among the varieties; some varieties had the same number of fragments but the molecular weights were different.

If we group information obtained with the three enzymes, 42 different patterns were distinguished among the 48 varieties: 37 varieties are clearly identified, the remaining 11 are divided into five groups (Ramses and 88–15; Hardi and Camp Remy; Eiffel and Estica; Genesis and Paradis; Sperber, Odeon and Arminda). We note that varieties from the first three groups are very closely related.

gbx3832 was hybridized to genomic DNA of one variety each of Barley (Truitel), rye (Malo) and oat (Margot) (Fig. 1). Two clear bands are observed on rye and two and six weak signals are respectively observed for barley and oat, suggesting that gbx3832 is at least partially homologous to a sequence present in the genome of those cereals.

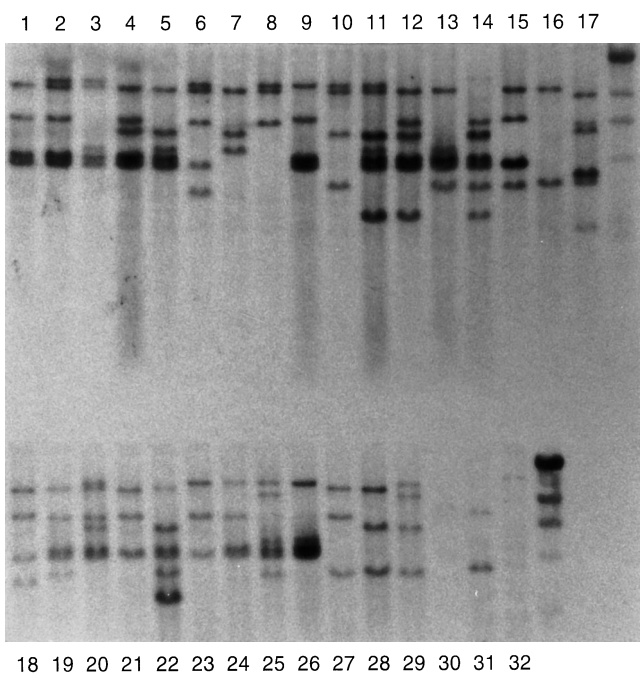

Fig. 1 Southern blot obtained by hybridization of gbx3832 to *Dra*I-digested DNAs from 29 wheat varieties. The identity of the variety in each numbered lane from *1 to 29* is given in Table 1. *Lane 30* Truitel (barley), *31* Malo (rye) and *32* Margot (oat)

Chromosomal location

The carrier chromosomes of the sequences homologous to gbx3832 were determined by nulli-tetrasomic analysis: gbx3832 when hybridized to *Eco*RI-digested DNA from 21 nulli-tetrasomic lines of Chinese Spring displays a pattern of four bands corresponding to related sequences on chromosomes 2A, 2D, 4A and 6B (Fig. 2). The localization of gbx3832 on the genetic map of wheat was determined for sequences on chromosomes 2A, 2D and 4A. Segregation analysis was carried out with F_8 lines from the synthetic × bread wheat cross. Linkages with other loci (Nelson et al. 1995 a, b) are shown in Fig. 3. For the sequence on chromosome 6B, no polymorphism was found between the parents of the population with the 15 enzymes tested.

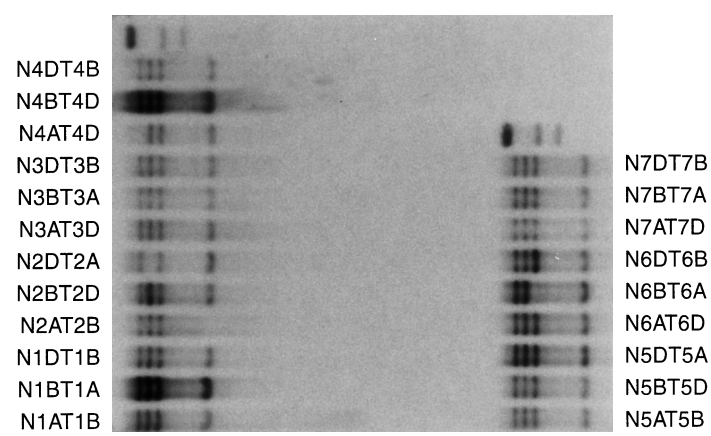

Fig. 2 Autoradiograph after hybridization of *Eco*RI-digested genomic DNA of 21 Chinese spring nullisomic-tetrasomic lines with gbx3832

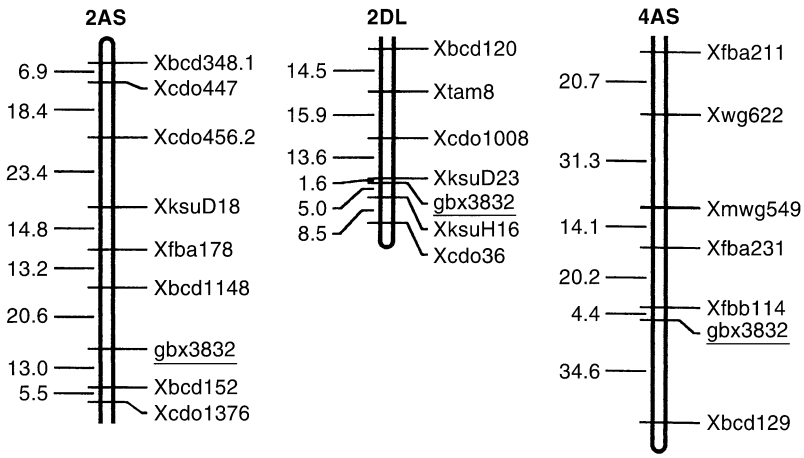

Fig. 3 Location of the fragments detected by probe gbx3832 on the RFLP linkage maps (Nelson et al. 1995 a, b) of chromosomes 2A, 2D and 4A. *Symbols* to left of chromosome maps denote distances in cM

Fig. 4 Nucleotide sequence of the gbx3832 cDNA insert. The amino-acid sequence translated from the open reading frame is given in the single-letter code below the nucleotide sequence. The *arrowhead* marks the putative cleavage site of the signal peptide

```
  1   CG TTG CTG TCG GAC ACA GAA GCA AAG CAA CTG GAA ATG GCT TCG TCC
  1                                               M   A   S   S

 48  CAC CTG GCA GCA GCT GCC TCC ATG GTC CTC TTC CTT GCC GTG TTC GCT
  5   H   L   A   A   A   A   S   M   V   L   F   L   A   V   F   A

 96  GCC AGC ACG AAC GCG GCG ACG TTC AAC ATC AAG AAC AAC TGC CCC TAC
 21   A   S   T   N   A   A   T   F   N   I   K   N   N   C   P

144  ACG GTG TGG CCG GCG GCC ACC CCG ATC GGC GGC GGT CGG CAG CTC AAC
 37   T   V   W   P   A   A   T   P   I   G   G   G   R   Q   L   N

192  ACC GGC GAG ACG TGG ACC CTC GAC GTC CCC GCG AAC ACG CCC TCC GGC
 53   T   G   E   T   W   T   L   D   V   P   A   N   T   P   S   G

240  AGG GTG TGG GGC CGC ACG GGC TGC AAC TTC AAT GGC AAC TCC GGG AGC
 69   R   V   W   G   R   T   G   C   N   F   N   G   N   S   G   S

288  TGC CAG ACT GCC GAC TGC GGC GGC GCG CTG TCG TGC ACG CTG TCC GGG
 85   C   Q   T   A   D   C   G   G   A   L   S   C   T   L   S   G

336  CAG CCG CCG CTG ACC CTG GCC GAG TTC ACC ATC GGC AAC GGC CAG GAC
101   Q   P   P   L   T   L   A   E   F   T   I   G   N   G   Q   D

384  TTT TAC GAC ATC TCT GTC ATC GAC GGC TTC AAC GTG CCG TTG TCA TTC
117   F   Y   D   I   S   V   I   D   G   F   N   V   P   L   S   F

432  TCC TGC AGC AAC GGG CCC AAC CTG GTG TGC CAG GCC GAC AAG TGC CCC
133   S   C   S   N   G   P   N   L   V   C   Q   A   D   K   C   P

480  GAC GCC TAC CTC TTC CCG ACC GAT GAC ACC AAG AAC CAC GCC TGT AAC
149   D   A   Y   L   F   P   T   D   D   T   K   N   H   A   C   N

528  GGC AAC AAC AAC ACC TAC CAG GTT ACC TTC TGC CCA TGA GGA AGA AGG
165   G   N   N   N   T   Y   Q   V   T   F   C   P   *

576  TAT CAT CGT AGC TAG TAG CGG ACG ATA CCA CCA CCA GCA TAA TAC GCG

624  TAC ATA CAA TGA GTG TGG AGA ATA AAG TTT GTT ACA TAA GAT GAA TAA

672  GAG CGA TGT CGT GAA AAT AAG GCT GTC CGT GCG TGC AGT GTA CGC ATT

720  TAT ACT ATA TGT ATC TTA TAA TTT AAT TTA TGT ACC CCT CAA ATT AAA

768  TAC TCC CTC CGT TTC AAA AAA AAA AAA AAA AAC GAC AGC AAC G
```

Nucleotide sequence of gbx3832

The complete nucleotide sequence of gbx3832 has been determined and deposited in GeneBank (accession no. x97687). The clone is 810 bp in length and contains an open reading frame for a protein of 176 amino acids flanked by 5′- and 3′-noncoding regions of 35 and 244 nucleotides respectively (Fig. 4). The first 25 amino acids have the typical features of a signal peptide indicating that the protein enters the secretory pathway. The sequence of gbx3832 was compared, by computer analysis, with plant sequences stored in GeneBank. Extensive homology was found with the sequences coding for the thaumatin-like protein (TLP) family. This family includes the sweet-tasting thaumatin from *Thaumatococcus daniellii* (Edens et al. 1982), antifungal permatins (Vigers et al. 1991), osmotins (Singh et al. 1987) and some pathogenesis-related proteins. Pathogenesis-related TLP genes of certain cereals are clustered into a small group that is phylogenetically separated from the major group of TLP genes of several plant species (Lin et al. 1996). The gbx3832 nucleotide sequence displays 72–76% identity with the sequences of this small group. The protein sequence encoded by gbx3832 shares 61% identity and 75% similarity with the previously characterized wheat TLP PWIR2 (Rebmann et al. 1991), 63 to 67% identity and 75–78% similarity with four TLPs encoded by a oat gene family (Lin et al. 1996), 66% identity and 80% similarity with the PIR2 TLP of rice (Reimmann and Dudler 1993), and 60–62% identity and 73–75% similarity with three PRHv TLPs of barley (Hahn et al. 1993) (Fig. 5).

Discussion

DNA fingerprinting of wheat has several practical applications such as the registration of new cultivars, the evaluation of wheat germ plasm and the estimation of relationships among cultivars. Using RFLP markers, it is theoretically expected that even closely related cultivars can be distinguished if a large number of probes and restriction enzymes are employed. The availability of hypervariable probes should enable the use of only a few of them. As non-coding regions may be less conserved, it is expected that genomic libraries demonstrate more RFLPs than cDNA libraries (Helentjaris and Gesteland 1983). Thus, hypervariable wheat genomic DNA probes containing moderately repeated elements have been characterized by Hartcourt and Gale (1991) and by Liu et al. (1992).

We describe here the characterization of a coding wheat cDNA detecting a high level of RFLP between wheat cultivars. gbx3832 displays the highest PIC among 221 cDNA and 278 genomic DNA wheat clones randomly tested. When used with three restriction enzymes (*Dra*I, *Hin*dIII and *Eco*RV), it enables us to distinguish 42 different patterns among 48 varieties: 37 varieties are clearly identified, the 11 remaining are divided into five groups. The nucleotide sequence of

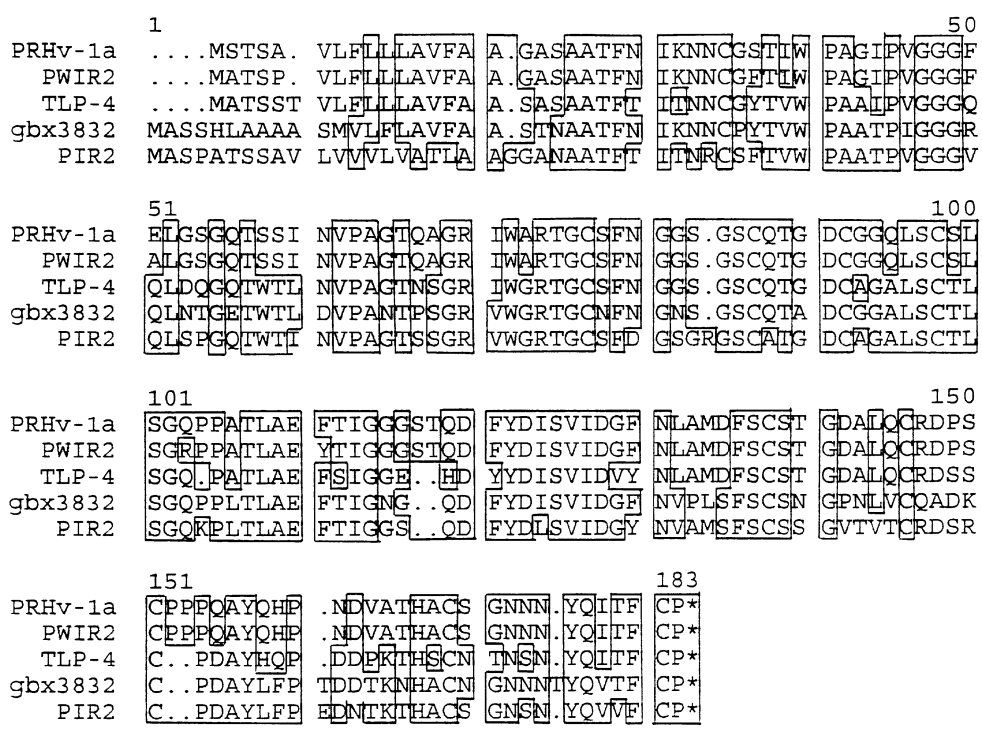

Fig. 5 Alignment of the deduced gbx3832 amino-acid sequence with other cereal thaumatin-like proteins. The sequences were aligned by using the 'pileup' program from the GCG Package (Genetics Computer Group, Inc. Wisconsin). *Dots* indicate gaps introduced for optimal alignment. *PWIR2* is a thaumatin-like pathogenesis protein from wheat infected by barley powdery mildew fungus (Rebmann et al. 1991); *PIR2* is a pathogen induced thaumatin-like protein of rice (Reimmann et al. 1993); *PRHv-1a* is an elicitor-induced thaumatin-like pathogenesis-related protein from barley (Hahn et al. 1993); *TLP-4* is a thaumatin-like pathogenesis-related protein from oat infected by an incompatible oat stem rust fungus (Lin et al. 1996).

gbx3832 does not contain the repeated sequences often associated with hypervariable loci. The high degree of polymorphism revealed by gbx3832 is due to variation in the molecular weights of the fragments as well as in the number of loci detected: hybridization of gbx3832 to nulli-tetrasomic lines demonstrates, for Chinese Spring, the existence of four different loci while only two and three hybridizing fragments are detected for Highbury and Courtot respectively.

The sequence of gbx3832 shares extensive homology to the thaumatin-like protein family and especially to pathogenesis-related thaumatin-like proteins of barley, rice and oat. The sequence encodes a protein distinct from the previously characterized wheat TLP PWIR2. On the other hand, unlike most wheat low-copy clones that detect sequences in the three homeologuous genomes, gbx3832 hybridizes to sequences carried by chromosomes 2A, 2D, 4A and 6B. This suggest the existence of a TLP gene family in wheat and supports previous reports that TLPs are encoded by a gene family in barley (Hahn et al. 1993) and oat (Lin et al. 1996).

Acknowledgments We wish to thank Roberte Baleux, Laurence Kutten, Maryse Larouillère and Agnès Mélard for technical assistance. We also thank Fabienne Delporte for critical reading of the manuscript. This work was supported by IRSIA under project no. D1/4-5125/5496A and by the Jorion et fils, Clovis Matton and Lochow Petkus Belgium companies.

References

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction length polymorphisms. Am J Hum Genet 32 : 314–331

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic map of wheat homeologous group-7 chromosomes. Theor Appl Genet 78 : 495–504

Edens L, Heslinga L, Klok R, Ledeboer AM, Maat J, Toonen MY, Visser C, Verrips CT (1982) Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in *Escherichia coli*. Gene 18 : 1–12

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132 : 6–13

Hahn M, Jüngling S, Knogge W (1993) Cultivar-specific elicitation of barley defense reactions by the phytotoxic peptide NIP1 from *Rhynchosporium secalis*. Mol Plant-Microbe Interact 6 : 745–754

Hartcourt RL, Gale MD (1991) A chromosome-specific DNA sequence which reveals a high level of RFLP in wheat. Theor Appl Genet 81 : 397–400

Helentjaris T, Gesteland R (1983) Evaluation of random cDNA clones as probes for human restriction fragment polymorphisms. J Mol Appl Genet 2 : 237–247

Kam-Morgan LNW, Gill BS (1989) DNA restriction fragment length polymorphisms: a strategy for genetic mapping of the D genome of wheat. Genome 32 : 724–732

Lander ES, Green P, Abrahamson J (1987) An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1 : 174–181

Lin KC, Bushnell WR, Szabo LJ, Smith AG (1996) Isolation and expression of a host-response gene family encoding thaumatin-like proteins in incompatible oat-stem rust fungus interactions. Mol Plant-Microbe Interact 9 : 511–522

Liu YG, Mori N, Tsunewaki K (1990) Restriction fragment length polymorphism (RFLP) analysis in wheat. I. Genomic DNA library construction and RFLP analysis in common wheat. Jpn J Genet 65 : 367–380

Liu YG, Ikeda TM, Tsunewaki K (1992) Moderately repeated, dispersed and highly variable (MRDHV) genomic sequences of common wheat usable for cultivar identification. Theor Appl Genet 84 : 535–543

Nelson JC, Sorrels ME, Van Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995 a) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5 and 7. Genetics 141 : 721–731

Nelson JC, Van Deynze AE, Autrique E, Sorrels ME, Lu YH, Merlino M, Atkinson M, Leroy P (1995 b) Molecular mapping of wheat. Homoeologous group 2. Genome 38 : 516–524

Rebmann G, Mauch F, Dudler R (1991) Sequence of a wheat cDNA encoding a pathogen-induced thaumatin-like protein. Plant Mol Biol 17 : 283–285

Reimmann C, Dudler R (1993) cDNA cloning and sequence analysis of a pathogen-induced thaumatin-like protein from rice. Plant Physiol 101 : 1113–1114

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74 : 5463–5467

Sears ER (1954) The aneuploids of common wheat. Mo Agric Exp Stn Res Bull 572 : 1–58

Sears ER (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis KR (eds) Chromosome manipulation and plant genetics. Oliver and Boyd, London, pp 29–45

Sharp PJ, Kreis M, Shewry PR, Gale MD (1988) Location of B-amylase sequences in wheat and its relatives. Theor Appl Genet 75 : 286–290

Singh NK, Bracker CA, Hasegawa PM, Handa AK, Buckel S, Hermodson MA, Pfankoch E, Regnier FE, Bressan RA (1987) Characterization of osmotin. A thaumatin-like protein associated with osmotic adaptation in plant cells. Plant Physiol 85 : 529–536

Vigers AJ, Roberts WK, Selitrennikoff CP (1991) A new family of plant antifungal proteins. Mol Plant-Microbe Interact 4 : 315–323