

S. Olmos · A. Distelfeld · O. Chicaiza · A. R. Schlatter ·
T. Fahima · V. Echenique · J. Dubcovsky

Precise mapping of a locus affecting grain protein content in durum wheat

Received: 19 February 2003 / Accepted: 13 June 2003 / Published online: 16 August 2003
© Springer-Verlag 2003

Abstract Grain protein content (GPC) is an important factor in pasta and breadmaking quality, and in human nutrition. It is also an important trait for wheat growers because premium prices are frequently paid for wheat with high GPC. A promising source for alleles to increase GPC was detected on chromosome 6B of *Triticum turgidum* var. *dicoccoides* accession FA-15-3 (DIC). Two previous quantitative trait locus (QTL) studies found that the positive effect of DIC-6B was associated to a single locus located between the centromere and the *Nor-B2* locus on the short arm of chromosome 6B. Microsatellite markers *Xgwm508* and *Xgwm193* flanking the QTL region were used in this study to develop 20 new homozygous recombinant substitution lines (RSLs) with crossovers between these markers. These 20 RSLs, plus nine RSLs developed in previous studies were characterized with four new RFLP markers located within this chromosome segment. Grain protein content was determined in three field experiments organized as randomized complete block designs with ten replications each. The

QTL peaks for protein content were located in the central region of a 2.7-cM interval between RFLP markers *Xcdo365* and *Xucw67* in the three experiments. Statistical analyses showed that almost all lines could be classified unequivocally within low- and high- protein groups, facilitating the mapping of this trait as a single Mendelian locus designated *Gpc-6B1*. The *Gpc-6B1* locus was mapped 1.5-cM proximal to *Xcdo365* and 1.2-cM distal to *Xucw67*. These new markers can be used to reduce the size of the DIC chromosome segment selected in marker-assisted selection programs. Markers *Nor-B2* and *Xucw66* flanking the previous two markers can be used to select against the DIC segment and reduce the linkage drag during the transfer of *Gpc-6B1* into commercial bread and pasta wheat varieties. The precise mapping of the high GPC gene, the high frequency of recombinants recovered in the targeted region, and the recent development of a tetraploid BAC library including the *Gpc-6B1* DIC allele are the first steps towards the map-based cloning of this gene.

Communicated by J. Dvorak

S. Olmos · A. Distelfeld · O. Chicaiza · V. Echenique ·
J. Dubcovsky (✉)
Department of Agronomy and Range Science,
University of California,
One Shields Avenue, Davis, CA 95616-8515, USA
e-mail: jdubcovsky@ucdavis.edu
Tel.: +1-530-7525159
Fax: +1-530-7524361

A. Distelfeld · T. Fahima
Institute of Evolution,
University of Haifa,
Mount Carmel, Haifa, 31905, Israel

A. R. Schlatter
INTA Pregamino,
Argentina

Present address:

S. Olmos · V. Echenique, CONICET
and Departamento de Agronomía,
Universidad Nacional del Sur,
San Andrés 800, 8000 Bahía Blanca, Argentina

Keywords *Triticum turgidum* var. *dicoccoides* ·
Grain protein content · QTL · Genetic map · Wheat

Introduction

High grain protein content (GPC) is one of the most important factors determining pasta and breadmaking quality, and is also important to human nutrition. In addition, high protein determines premium prices for wheat in many regions of the world, making high GPC a primary target for hard common wheats and durum wheat breeding-programs. The recent increase in the emphasis on end-use quality in the current export markets has increased the value of this trait (Dohlman and Hoffman 2000).

In spite of its economic importance, genetic improvement of GPC by conventional breeding has been slow because of the complex genetic system governing this trait, the high influence of the environment, and the

existence of a negative correlation between GPC and yield (Simmonds 1995). However, there are exceptional varieties that combine excellent yields with high levels of GPC, and simultaneous increases in yield and GPC have been reported in target populations (Cox et al. 1985). Efforts to improve GPC without selecting for low yield can be accelerated by the identification of the genes that affect GPC and the direct selection of the alleles with positive effects.

A promising source of high GPC content was detected in a survey of wild populations of tetraploid *Triticum turgidum* var. *dicoccoides* (accession FA-15-3) referred as *dicoccoides* or DIC hereafter (Avivi 1978). Cantrell and Joppa (1991) developed complete sets of disomic substitution lines from each of the chromosomes of this *dicoccoides* accession into the tetraploid durum cultivar Langdon (LDN). The substitution of the complete chromosome 6B of DIC into LDN [LDN(DIC-6B)] showed the highest protein yield. This significant increase in GPC was not associated with a significant decrease in grain yield or kernel weight (Cantrell and Joppa 1991). Moreover, the increase in protein content in LDN(DIC-6B) was associated with an increase in the quality of the pasta products (Joppa et al. 1991). When introduced into adapted durum germplasm, the DIC allele was also associated with an increase in protein content (1.5%) with non-significant effects on protein quality, plant height, heading date or yield (Chee et al. 2001). The introgression of the DIC allele into common wheat varieties also resulted in increases of GPC (Mesfin et al. 1999).

The evaluation of a mapping population of recombinant substitution lines (RSLs) from the cross between LDN(DIC-6B) × LDN, showed that the increase in GPC in LDN(DIC-6B) was associated with a quantitative trait locus (QTL) located on the short arm of chromosome 6B within the *Xmwg79–Xabg387* interval (Joppa et al. 1997). This QTL, designated *QGpc.ndsu.6Bb*, accounted for approximately 66% of variation in GPC in this cross, suggesting that a single gene or a closely linked group of genes was segregating in this population. The previous study was able to simplify the complex genetic regulation of GPC by using RSLs segregating for a single chromosome in the isogenic genetic background of LDN.

Other QTL studies have shown significant effects on GPC associated with markers located in similar regions of the short arms of homoeologous group 6. Blanco et al. (2002) found that the *Xpsr167* marker on chromosome arm 6AS ($R^2=18.4\%$) and the *Nor-B2* marker on chromosome arm 6BS ($R^2=6.5\%$) were significantly associated with GPC in a cross between the durum wheat cultivar 'Messapia' and a wild *dicoccoides* accession MG4343. Sourdille et al. (2003) reported a significant QTL for GPC on chromosome arm 6AS in hexaploid wheat associated with AFLP marker *XE38M60*₂₀₀ ($R^2=17.1\%$). Finally, See et al. (2002) and Mickelson et al. (2003) found QTLs on chromosome arm 6HS in barley associated with GPC. These studies suggest that a major gene affecting GPC in the *Triticeae* is probably located in

the short arms of the chromosomes from homoeologous group six.

PCR-based markers, including AFLP, microsatellite and PCR-specific markers were developed for the *QGpc.ndsu.6Bb* region to facilitate the transfer of the DIC FA-15-3 allele into commercial tetraploid and hexaploid varieties (Khan et al. 2000). The microsatellite markers *Xgwm508* and *Xgwm193* identified in that study, and which encompass the peak of the *QGpc.ndsu.6Bb* QTL, are being used to introgress the DIC allele in the US National Program for Wheat Marker Assisted Selection (<http://maswheat.ucdavis.edu/>). However, these two microsatellite markers are approximately 12-cM apart, and the simultaneous selection for these two markers can result in the transfer of relatively large segments of the *dicoccoides* 6B chromosome.

The aim of this work was to provide a more precise location of the high GPC gene within the *Xgwm508* and *Xgwm193* interval by: (1) mapping new markers in the *QGpc.ndsu.6Bb* region, and (2) developing new RSLs with additional crossover events within this region. The long-term objective of this project is to clone this gene to provide a better understanding of the physiological processes involved in the differences in GPC associated with this locus.

Materials and Methods

Plant materials

LDN(DIC-6B) RSLs developed from the cross LDN(DIC-6B) × LDN were provided by L.R. Joppa. In this study we included nine RSLs with at least one recombination event between microsatellite markers *Xgwm508* and *Xgwm193* (RSLs: 8, 11, 19, 28, 50, 54, 58, 59 and 77). In addition, we included five lines with complete LDN (RSLs: 5, 14, 48 and 78, LDN) or DIC (RSLs: 47, 63, 65, 67 and 68) chromosome segment in the same chromosome interval as controls (Table 1).

Additional RSLs were generated from crosses between RSL65 × LDN. The RSL65 carries a complete DIC segment in the *Xgwm508* and *Xgwm193* interval. One hundred and thirty four F₂ plants were screened with microsatellite markers *Xgwm508* and *Xgwm193*, and 20 plants showing recombination between these markers were self-pollinated. Ten to sixteen F₃ plants from each of the recombinant F₂ plants were analyzed and homozygous recombinant lines were selected for each of the 20 F_{2,3} families. Seeds from the new selected RSLs were increased for the replicated field experiments.

Probes and RFLP procedures

DNA extraction and procedures for Southern blots and hybridization were as described before (Dvorak et al. 1988; Dubcovsky et al. 1994). Clones used in this study were selected from wheat ESTs from contigs mapped on deletion BIN 6BS5 located proximal to the *Nor* secondary constriction (<http://wheat.pw.usda.gov>), from RFLP probes previously mapped in this region (Marino et al. 1996; Mingeot and Jacquemin 1999), and from wheat ESTs similar to the colinear region in rice (A. Distelfeld et al. in preparation). EST clones WHE1104_F01_L02 (GenBank# BE443401), WHE0929_D01_G01 (GenBank# BF473522), WHE2958_B02_D04 (GenBank# BG606570), WHE0031.D07 (GenBank# BE399163) and WHE0613_H01_O01

Table 1 Genotypes and GPC content of the RSLs included in the three experiments

Line	<i>Xgwm508</i>	<i>Nor-B2</i>	<i>Xcdo365</i>	<i>XGpc</i>	<i>Xucw67</i>	<i>Xucw66</i>	<i>Xgwm193</i>	Experiment I	Experiment II	Experiment III
LDN	L	L	L	L	L	L	L	13.57 ^L	14.14 ^L	15.69 ^L
5	L	L	L	L	L	L	L	14.34 ^L	-	-
14	L	L	L	L	L	L	L	13.88 ^L	-	-
48	L	L	L	L	L	L	L	13.55 ^L	-	-
78	L	L	L	L	L	L	L	13.40 ^L	-	-
47	D	D	D	D	D	D	D	15.35 ^H	-	-
63	D	D	D	D	D	D	D	15.11 ^H	-	-
65	D	D	D	D	D	D	D	15.57 ^H	15.02 ^H	-
67	D	D	D	D	D	D	D	14.80 ^H	-	-
68	D	D	D	D	D	D	D	15.32 ^H	-	-
11	D	L	L	L	L	L	L	13.90 ^L	-	-
19	D	L	L	L	L	L	L	13.82 ^L	-	-
58	L	D	D	D	D	D	D	15.23 ^H	14.37 ^{NS}	-
101	D	L	L	L	L	L	L	-	13.96 ^L	16.76 ^{NS}
102	D	D	L	L	L	L	L	-	13.64 ^L	-
59	D	D	L	L	L	L	L	13.57 ^L	-	-
122	D	D	L	L	L	L	L	-	13.59 ^L	15.80 ^L
106	D	D	D	L	L	L	L	-	14.41 ^{NS}	16.53 ^L
113	L	L	L	D	D	D	D	-	15.27 ^H	17.49 ^H
116	L	L	L	D	D	D	D	-	14.69 ^H	16.98 ^H
8	L	L	L	D	D	D	D	14.72 ^H	-	-
121	L	L	L	D	D	D	D	-	14.94 ^H	-
110	D	D	D	D	L	L	L	-	14.53 ^H	17.45 ^H
117	D	D	D	D	L	L	L	-	15.06 ^H	18.01 ^H
119	D	D	D	D	L	L	L	-	15.32 ^H	17.68 ^H
28	D	D	D	D	L	L	L	14.92 ^H	-	-
50	D	D	D	D	D	L	L	14.91 ^H	-	-
54	D	D	D	D	D	L	L	15.01 ^H	-	-
104	D	D	D	D	D	L	L	-	14.81 ^H	-
111	D	D	D	D	D	L	L	-	14.72 ^H	17.58 ^H
114	D	D	D	D	D	L	L	-	15.10 ^H	16.91 ^{NS}
115	D	D	D	D	D	L	L	-	15.35 ^H	18.51 ^H
118	D	D	D	D	D	L	L	-	15.34 ^H	17.44 ^H
107	L	L	L	L	L	D	D	-	13.55 ^L	15.72 ^L
108	L	L	L	L	L	D	D	-	13.62 ^L	-
112	L	L	L	L	L	D	D	-	13.63 ^L	16.09 ^L
109	D	D	D	D	D	D	L	-	14.96 ^H	17.22 ^H
77	L	L	L	L	L	L	D	13.61 ^L	-	-
120	L	L	L	L	L	L	D	-	13.54 ^L	15.85 ^L
						Average L group		13.74	13.71	15.99
						Average D group		15.16	14.96	17.53

D= DIC alleles and L= LDN alleles. A change between gray and white cells indicates a recombination event. The last three columns indicate the average protein content in each experiment. Superscript "L:" indicates lines classified as low-GPC and superscript "H." a line classified as high-GPC by the statistical analysis

(GenBank# BE515435), were provided by Olin Anderson (Grain-Genes Probe Repository, USDA ARS, Albany, Calif., USA); FBA clones (FBA67, FBA152, FBA328, FBA344, FBA357 and FBA399) by P. Leroy (Station d'Amélioration des Plantes de Clermont-Ferrand, INRA, France); and GBX clones (GBXR004, GBXG138, GBX3321 and GBX3327) by J.M. Jacquemin (Centre de Recherches Agronomiques, Belgium). These clones were used as probes to screen for polymorphism between DNAs from parental genotypes LDN(DIC-6B) and LDN digested with 24 different restriction enzymes (*AscI*, *ApaI*, *AvaII*, *BamHI*, *BfaI*, *BglII*, *BstEII*, *BstNI*, *DdeI*, *DraI*, *EcoRI*, *EcoRV*, *HaeIII*, *HhaI*, *HindIII*, *MspI*, *NcoI*, *NdeI*, *SacI*, *Sau3AI*, *SspI*, *StuI*, *StyI* and *XbaI*). New loci mapped with polymorphic EST clones were assigned UCW numbers (*Xucw64*: BE443401, *Xucw65*: BF473522, *Xucw66*: BG606570, *Xucw67*: BE515435, Table 1). The *Nor-B2* locus was mapped using probe pTA250.15 that includes the 750-bp *HhaI* fragment from the spacer of the ribosomal gene (Appels and Dvorak 1982). Marker *Xcdo365* (Mesfin et al. 1999) previously mapped in the *QGpc.ndsu.6Bb* region was also used to characterize the RSLs in this study.

PCR procedures

Primer sequence and PCR conditions for the *Xgwm508* and *Xgwm193* microsatellites were as described by Röder et al. (1998). PCR reactions were performed in a 20- μ l vol. Eight microliters of formamide dye were added to the 20- μ l PCR reaction and amplification products were separated on denaturing 6% (w/v) polyacrylamide gels at 40 W for 3 h and detected by silver staining (Silver Sequence Staining Reagents kit, Promega, Madison, Wis., USA).

Field experiments

Field experiments were conducted in the experimental field of the University of California, Davis, USA (Sacramento Valley, 38°32' N, 121°46' W). Planting dates for the three experiments were November 20, 2000 (Exp. I); January 15, 2002 (Exp. II); and February 20, 2002 (Exp. III, late planting).

Soils at Davis (Yolo Loam soils) were fertilized with 88 kg ha⁻¹ of N prior to planting for Exp. I, and with 111 kg ha⁻¹ of N for Exp. II and Exp. III. Experiments received enough rainfall during the winter, maintaining adequate soil moisture without irrigation. Two flood irrigations (120 mm) were applied to all plots after anthesis. Two additional irrigations were required for Exp. III (late planting).

The three experiments were organized in a Randomized Complete Block Design (RCBD) with ten blocks. Experimental units consisted of 1-m row plots, spaced 0.3 m apart, and with adjacent RSL rows separated by a row of dwarf common wheat to avoid mixing.

Eighteen RSLs developed by L. Joppa (Joppa et al. 1997) were evaluated in the first experiment. Twenty two and 16 RSLs were evaluated in experiments II and III respectively (Table 1). In addition, all three experiments included the parental line LDN as a control. Sixteen of the 20 RSLs developed in this study were tested in Exp. II and Exp. III, but four RSLs were included only in Exp. II due to a limited seed supply.

Nitrogen determinations

Individual rows were hand-harvested at physiological maturity (approximately 6 months after planting), and threshed using a Vogel plot thresher. Grains were dried to constant weight at 60°C before grinding. GPC determinations were performed on whole-meal flour (Udy Cyclodone Mill 1-mm mesh sieve) using a near-infrared reflectance equipment (Technicon InfraAlyzer model 400). Six standard samples with protein contents ranging from 9% to 15.5% protein were used to calibrate the near-infrared reflectance equipment. Nitrogen content in the standard samples was deter-

Table 2 Contrasts between the lines with critical cross-overs in the *Xcdo365*–*Xucw67* interval

RSL	Exp.	<i>P</i> vs DIC group	<i>P</i> vs LDN group	Classification
8	I	0.3003	0.0017	High
28	I	0.3172	0.0015	High
106	II	0.0465	0.0112	NS
106	III	0.0002	0.0450	Low
110	II	0.1380	0.0047	High
110	III	0.7626	<0.0001	High
113	II	0.2566	<0.0001	High
113	III	0.8747	<0.0001	High
116	II	0.3280	0.0004	High
116	III	0.0453	0.0003	High
117	II	0.7112	<0.0001	High
117	III	0.0674	<0.0001	High
119	II	0.1882	<0.0001	High
119	III	0.6024	<0.0001	High
121	II	0.9463	<0.0001	High

mined using a Combustion Nitrogen Analyzer. Grain nitrogen content was expressed as grams of N per grams of dry grains and then converted to the percent of protein by multiplying the percent of nitrogen content \times 5.7. These values are referred as GPC throughout the text.

Statistical analysis

Average GPC from the ten replications were used in separate QTL analyses for each of the three experiments. Data from Joppa's experiments at Fargo, ND (two replications), and Prosper, ND (four replications), were included in the QTL analysis for comparison (Joppa et al. 1997). Linkage maps were constructed with the aid of the program Mapmaker (Lander et al. 1987) using the Kosambi function. Mapmaker QTL (Lander et al. 1987) was used for the QTL analyses to identify the most-likely interval including the locus responsible for the observed differences in GPC.

Additional analyses of variance were performed using the SAS Version 8.0 program (SAS Institute 2003). Lines with recombination events outside the critical interval for the GPC gene identified in the QTL analysis were classified in two groups according to the alleles present at both flanking markers. Lines with LDN alleles at both loci were included in the low-protein control group and those with DIC alleles at both loci in the high-protein control group. Each of the lines with crossovers in the critical interval for the GPC gene (Table 2) were compared against these two groups using contrasts. The lines included in each of the control groups were also tested against the two control groups (excluding the tested sample) to validate the consistency of the classification.

A line was classified as a low-GPC if it was simultaneously significantly lower ($P<0.01$) than the high-GPC group, and non-significantly (NS) different from the low-GPC group. A line was classified as high-GPC if it was simultaneously significantly higher ($P<0.01$) than the low-GPC control group and NS different from the high-GPC control group. Lines that were NS different from both groups were not classified. Data from each experiment were analyzed separately (Table 1). A low Type-I error ($\alpha=0.01$) was used to compensate for the possibility of a higher experiment-wise error rate generated by the use of non-orthogonal contrasts.

Results

Production and characterization of recombinant lines

Twenty new recombination events between microsatellite markers *Xgwm508* and *Xgwm193* were identified among 134 F₂ plants from the cross between LDN and RSL65. The selected plants were self-pollinated and 20 homozygous RSLs carrying the recombinant chromosomes were recovered using the microsatellite markers.

These 20 new RSLs, plus nine available RSLs with crossovers between the same flanking microsatellites (Joppa et al. 1997; Khan et al. 2000), were characterized with the probe for the *Nor-B2* locus and the new polymorphic RFLP markers identified in this study. RSLs with numbers lower than 85 correspond to the lines produced by L. Joppa and those with higher numbers were produced in this study (Table 1). The two polymorphic GBX markers and the three polymorphic FBA markers were mapped outside the critical region between *Nor-B2* and microsatellite marker *Xgwm193* (Fig. 1). Four polymorphic EST markers were mapped within the critical region and were used to characterize all the recombinant lines (Fig. 1, Table 1).

Graphical genotypes (Young and Tanksley 1989) are presented in Table 1. White cells with an "L" indicate Langdon alleles, whereas gray cells with a "D" indicate DIC alleles. Recombinants were recovered among most markers in the *Nor-B2-Xgwm193* region with the exception of the pairs *Xucw64-Xcdo365* and *Xucw65-Xucw67* that were completely linked (Fig. 1). Three chromosomes showed recombination between the *Nor-B2* locus and *Xcdo365*, nine between *Xcdo365* and *Xucw67*, ten between *Xucw67* and *Xucw66*, and three between the last RFLP marker and microsatellite marker *Xgwm193*. The resulting map (Fig. 1, without the GPC gene) was used as a frame for the QTL analysis.

Fig. 2 QTL analysis for Grain Protein Content. Dotted lines correspond to previous experiments performed at Fargo and Prosper, North Dakota (Joppa et al. 1997). Distances between markers are proportional to the genetic distances in Fig. 1

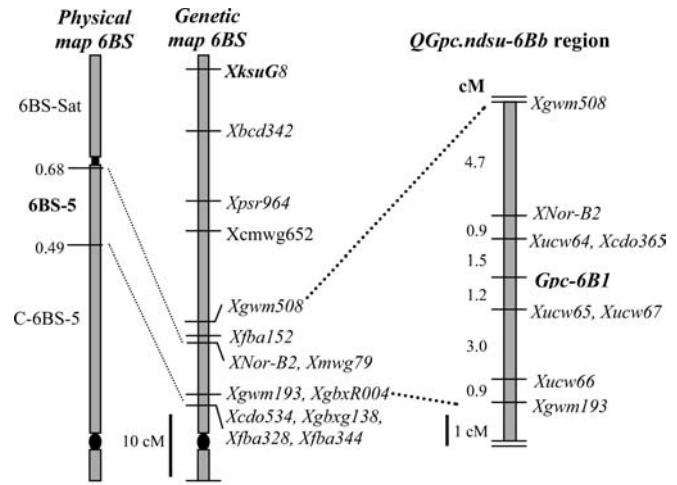
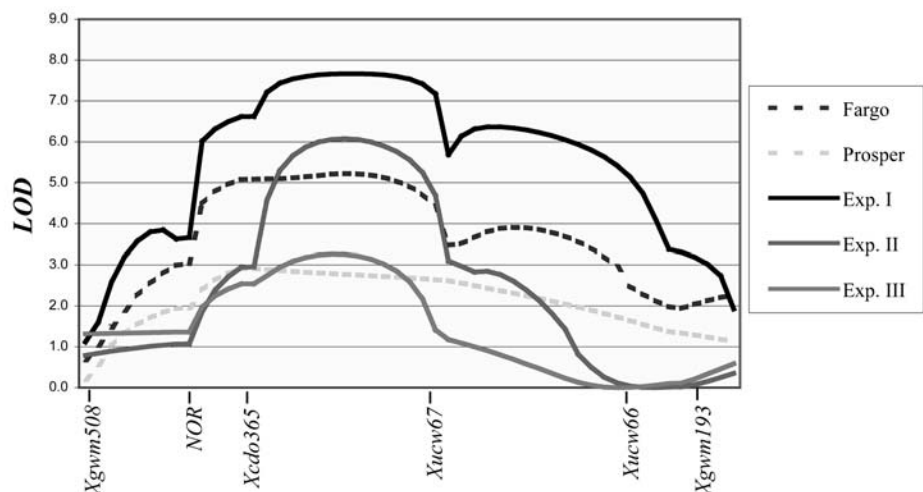


Fig. 1 Location of the *QGpc.ndsu.6Bb* locus based on QTL analysis of the *Xgwm508-Xgwm193* interval. Each recombination event represents approximately 0.3 cM. The position of the centromere is indicated with a black oval

QTL analysis

The GPC from each RSL, determined as the least square mean of all the available replicates in each experiment, is indicated in Table 1. Least square means were used instead of regular means because a few lines were lost or did not produce enough grain for the protein determination (11 missing rows in Exp. I, 4 in Exp. II, and 34 in Exp. III). The average grain protein content ranged from 13.4% to 15.6% in Exp. I and II, and from 15.7% to 18.5% in Exp. III.

QTL analysis of the GPC-means from each of the three experiments performed in California, and the re-analysis of the data from the two locations from North Dakota (Joppa et al. 1997), are presented in Fig. 2.

Results from the re-analysis of Joppa's experiments at Fargo (two replications) and Prosper (four replications) including only the lines used in our Exp. I, are indicated

with dotted lines in Fig. 2. These experiments showed peaks within the *Xcdo365*–*Xucw67* interval, but the Prosper peak was close to the *Xcdo365* border of the interval. The QTL peaks were relatively flat across the *Nor-B2*–*Xucw67* interval.

Experiment II and III, including the new lines produced in this study, showed better-defined peaks within the *Xcdo365* and *Xucw67* interval. All three experiments suggest a relatively central location of the QTL peak within this interval. Peaks for the QTLs detected in the first two experiments showed higher LOD scores (Exp. I, LOD=7.7 and Exp. II, LOD=6.1) than Exp. III (LOD=3.3) from the late planting.

This QTL explained a large proportion of the variation in GPC in all three experiments. The percent of variation explained by this QTL (R^2) was 87.0% for Exp. I, 79.7.2% for Exp. II and 71.2% for Exp. III. The large proportion of variation explained by this QTL and the perfect overlap of the QTL peaks from the three experiments are indicative of the strong effect of this locus. This overlap is even more valuable considering that Exp. II and III were performed in different years and with a different set of RSLs than Exp. I. The similar map locations obtained in the three experiments supported the location of the gene responsible for the increase in GPC between RFLP markers *Xcdo365* and *Xucw67*.

Statistical analysis and mapping of the high protein-content gene

The previous QTL analysis was useful to determine the approximate location of the high GPC gene between *Xcdo365* and *Xucw67*, but did not use all the power of the experimental design. In order to determine more precisely the location of the high GPC gene we analyzed each experiment separately using ANOVA.

Differences among RSLs in GPC were highly significant in all three experiments ($P < 0.0001$) with R^2 values of 0.52, 0.46 and 0.64 for Exp. I, II and III, respectively. These R^2 values are smaller than those obtained in the QTL analysis because the ANOVA includes the variation among individual rows, whereas the QTL analysis includes only the variation among the averages. The Coefficients of Variation of the three experiments were low, ranging from 4% in experiments III to 6% in experiments I and II.

To determine the precise location of the gene affecting GPC within the *Xcdo365* and *Xucw67* interval it was particularly important to determine the classification of the nine lines with recombination events within these markers into the high- and low-GPC groups (Table 1, lines 8, 28, 106, 110, 113, 116, 117, 119 and 121). These critical lines were compared with the lines included in the two control groups without recombination within the *Xcdo365*–*Xucw67* interval. The average GPC content for each of the control groups (LDN = low-GPC and DIC = high-GPC) is indicated at the bottom of Table 1.

The difference between these two averages provided an estimate of the effect of the DIC allele on GPC. In all three experiments the lines with the DIC alleles at the *Xcdo365*–*Xucw67* interval showed significantly higher ($P < 0.0001$) levels of GPC than the lines with LDN alleles. The increases in GPC due to the DIC allele were 1.25% in experiment II, 1.42% in experiment I and 1.55% in experiment III, resulting in an average increase of 1.4% in GPC across experiments.

The average between the low- and high-GPC control-group averages was considered the midpoint between low and high protein-values for a particular experiment. All the GPC values higher than the midpoint of the respective experiment are indicated in bold letters in Table 1. This calculation provided a preliminary classification of all the RSLs but did not indicate the significance of that classification.

To test this classification statistically, each line was compared to the lines with LDN alleles in the *Xcdo365*–*Xucw67* interval, and then to the lines with DIC alleles in the same interval using contrasts. Twenty three GPC values were significantly lower than the DIC group ($P < 0.01$) and not significantly higher than the LDN group from their respective experiment. Therefore, these lines were classified in the low-protein group and assigned an “L” superscript to the protein value (Table 1). Thirty-two GPC values were significantly higher than the LDN group ($P < 0.01$) and not significantly lower than the DIC group. These lines were classified in the high-protein group and assigned an “H” superscript to the GPC value (Table 1). Only four GPC averages showed intermediate values between the two groups that were not significantly different from the LDN or DIC groups. These lines were indicated with an “NS” superscript to the GPC value (Table 1).

Two of the “NS” lines were from Exp. II and two from Exp. III. In Experiment II, RSLs 58 and 106 did not significantly differ from both the LDN and DIC group. Fortunately, RSL 58 was classified as a high-GPC in Exp. I (P vs LDN group = < 0.001 ; P vs DIC group = 0.792), and RSL 106 was classified as a low-GPC in Exp. III (Table 2). RSL 106 was the only line without clear-cut results. The average GPC for this line was above the average between low- and high- groups in Exp. II but below that number in Exp. III. However, the first result was not significant, suggesting that there was no real contradiction between the two data. Two RSLs from Exp. III, 101 and 114, were also not significantly different from the LDN and DIC group in Exp. III but were classified in Exp. II as a low-GPC (RSL 101, P vs LDN group = 0.32; P vs DIC group = 0.0006), and a high-GPC (RSL 114 P vs LDN group < 0.0001 ; P vs. DIC group = 0.56).

Significance values for the contrasts for the nine RSLs with crossover events within the *Xcdo365*–*Xucw67* interval are provided in Table 2. In all cases, except the previously mentioned RSL 106, these lines were classified unequivocally into one of the phenotypic groups. RSL 106 was classified in the low-GPC group in Experiment III (Table 2), but because of its intermediate

values and borderline statistical tests, additional field tests are being performed to confirm its classification. Five of the other critical lines were tested in two experiments with identical results. Among the three lines tested only in one experiment in this study, lines 8 and 28 were also tested in Joppa's experiments. The GPC values of these two lines were above the average between the low- and high-GPC groups in both the Fargo and Prosper experiments confirming the classification of these lines in the high-GPC group (data not shown).

Based on the statistical analysis of the critical lines, the high-GPC gene was mapped 1.5 cM proximal to *Xcdo365* and 1.2 cM distal to *Xucw67* (Table 1, Fig. 1) as a single Mendelian locus designated *Gpc-6B1*. All the other lines with recombination events outside the *Xcdo365*–*Xucw67* interval showed consistent results with the previous map location. All lines carrying the LDN allele in this interval showed protein values below the mean of their respective experiment, and were significantly included in the low-protein group in at least one experiment. On the contrary, all the lines carrying the DIC allele in this interval showed protein values above the mean of their respective experiment (bold letters), and were significantly included in the high-protein group in at least one experiment.

Discussion

Effect of the *XGpc-B1* locus

The GPC values observed in Exp. III, planted beyond the date for normal plantings in the Sacramento Valley (15.7% to 18.5%), were higher than those obtained in the first two experiments (13.4% to 15.6%). The increase in GPC in Exp. III may be a consequence of the non-optimal environmental conditions for wheat growing during the summer season in the Sacramento Valley. Elevated evapotranspiration rates during grain filling might have increased grain-protein concentration by preventing dilution of protein by starch (Stoddart and Marshall 1990). In spite of the different planting dates and the different total GPC, all three experiments showed a consistent increase of 1.3 to 1.6% GPC in the lines carrying the DIC allele relative to the lines carrying the LDN allele. The similar effect of the DIC allele independently of the differences in planting date and in absolute GPC suggests that this gene will be effective across different environments.

This hypothesis is also supported by the similar increases in GPC observed in our experiments in California and in those performed by Joppa et al. (1997) in two locations in North Dakota. Considering only the lines included in Exp. I (Table 1), the increase in GPC in the lines with the DIC allele was 1.2% in Prosper, and 1.6% in Fargo. Similar increases (1.4% to 1.6%) were reported in RILs from the cross LDN(DIC-6B)/Vic/Renville in experiments performed at Langdon and Prosper (ND) (Chee et al. 1998). Relative smaller increases of 0.6% GPC were associated with the DIC alleles in a comparison between DT367 and LDN(DIC-

6B)/3*DT367 in experiments performed at three locations in Canada (Kovacs et al. 1998). All these results indicate a consistent effect of the DIC allele across environments.

Precise mapping of the *Gpc-6B1* locus

The map location of the *Gpc-B1* locus (Fig. 1) reported in this study agrees with the previous position of the peak of the QTL identified by Joppa et al. (1997) between loci *Xmwig79* (linked to *Nor-B2*, Khan et al. 2000) and *Xabg387* (linked to *Xgwm193*, Khan et al. 2000). There is a small difference between our mapping results and those reported by Chee et al. (2001). In our study, the *Gpc-B1* locus was mapped 1.5 cM proximal to the *Xcdo365* locus, whereas in their study the most-likely location of the QTL peak was indicated 0.7 cM distal to *Xcdo365*. However, inspection of the graphical genotypes of the lines with crossover events close to the GPC gene (Chee et al. 2001) shows that the GPC gene is completely linked to *Xcdo365*. The complete linkage between the GPC gene and *Xcdo365* detected in Chee et al. (2001) does not differ significantly from the 1.5-cM genetic distance found in our study. The additional recombinants found in our study were expected based on the larger mapping population, and provided a better resolution to map the *Gpc-B1* locus.

The conversion of the initial QTL mapping (*QGpc.ndsu.6Bb*) into a precise mapping of *Gpc-6B1* as a single Mendelian locus was possible by combining different strategies. First, genetic variation was reduced to a minimum by using homozygous isogenic recombinant lines that differed only in a small chromosome region of chromosome arm 6BS, whereas the rest of chromosome 6B and all other 13 chromosome pairs were identical to LDN. Second, the use of complete rows as experimental units was necessary to eliminate the large variation observed among individual plants in the field. Preliminary experiments using individual plants instead of complete rows as experimental units failed to differentiate statistically most of the lines (data not shown). Finally, a large level of replication (ten blocks per experiment) was necessary to differentiate statistically most of the lines.

Marker-assisted selection (MAS)

Many laboratories are using the *Nor-B2*-derived PCR markers (Khan et al. 2000) to select for the high-GPC allele from DIC in MAS programs (<http://maswheat.ucdavis.edu/>). However, the genetic distance estimated from this study suggests that selection using only the *Nor-B2* marker would result in the loss of the high-GPC allele in 2.4% of the progeny in each generation. In a complete backcrossing program of six generations followed by selection of homozygous BC₆F₂ lines, the probability of losing the high-GPC allele by recombination with the targeted markers increases to approximately 14%.

To eliminate this risk, other breeding programs are using microsatellite markers *Xgwm508* and *Xgwm193*

(Khan et al. 2000). In this case the probability of losing the high-GPC allele is small because double-crossovers are unlikely in the 12-cM interval between the two microsatellites. However, this strategy would result in the introgression of a relatively large segment of DIC in the commercial varieties.

The new markers identified in this study can be used to minimize the amount of DIC chromatin transferred during the backcrossing program. Flanking loci *Xucw67* and *Xcdo365* (or the linked *Xucw65* and *Xucw64* markers) can be used to select for the transfer of the 2.7-cM DIC segment, whereas the more external PCR markers *Nor-B2* and *Xucw66* or *Xgwm193* can be used to select against the DIC alleles. Selection of recombination events between the flanking markers *Xucw67* and *Xcdo365*, and the more external markers, can be used to minimize the linkage drag during the MAS program. An alternative simpler strategy would be the use of the PCR specific marker for the DIC allele at the *Nor-B2* locus followed by selection every other generation by *Xucw67* (or *Xucw65*).

The DIC allele for high-GPC was also transferred into hexaploid bread wheat variety 'Glupro' by Dr. R. Froberg. This variety carries a large segment of DIC-6B including RFLP markers *Xcdo534* and *Xcmwg652* (Mesfin et al. 1999; Khan et al. 2000). A smaller DIC-6B segment was transferred to ND683 ('Stoa sib'/ND645) (Mesfin et al. 1999). ND683 has a high-GPC and RFLP bands with similar mobility to the mapped DIC allele at the *Xcdo365* (Mesfin et al. 1999) and *Xucw65* loci (data not shown). However, flanking loci *Nor-B2* and *Xgwm193* did not show the DIC alleles (Mesfin et al. 1999; Khan et al. 2000). These results suggest that the DIC segment incorporated into ND683 is between 2.7 and 7.5 cM, making it an attractive starting material to transfer the high-GPC allele to bread wheat. The only disadvantage of this approach is that there are currently no PCR markers for this smaller segment. The RFLP markers identified in this study will be valuable to develop new PCR based markers for this smaller DIC chromosome segment.

Progress towards the positional cloning of the *Gpc-6B1* gene

For the map-based cloning of the *Gpc-6B1* gene it is important to establish the gene density and the ratio of physical to genetic distances in the targeted region. Average estimates for the target regions can be obtained from the results of the consortium of 13 USA-laboratories funded by US-NSF that has mapped more than 15,000 loci in a subset of 101 deletion stocks (<http://wheat.pw.usda.gov/cgi-bin/westsq/locus.cgi>). Two hundred and seventy five ESTs have been mapped to the short arm of chromosome 6B (February 2003), which has been divided into three bins: 6BS-SAT, distal to 30% of the arm including 88 ESTs; 6BS-5, 20% proximal to the *Nor-B2*, including 69 ESTs; and C-6BS-5, the 50% centromeric half of the arm including 118 ESTs (Fig. 1). Based on the number of ESTs and the length of each bin it can

be concluded that the 6BS5 bin has the highest gene density and that it can be considered a relatively gene-rich region, as pointed out previously for the distal half of the 6BS chromosome arm (Weng et al. 2000).

This bin includes loci *Xucw64* and *Xucw65* that flank the *Gpc-6B1* locus (Fig. 1). Therefore, the high-GPC gene can be assigned to the 6BS-5 bin. Assuming that the total DNA content of a haploid genome of hexaploid wheat is 16,000 Mb (Arumuganathan and Earle 1991), and that the length of the 6BS arm represents 2.4% of the length of the chromosome complement (Endo and Gill 1996), the DNA content of the 6BS arm and the 6BS-5 bin can be estimated to be 380 Mb and 76 Mb respectively. Markers included in the 6BS-5 bin cover a region of approximately 8 to 9 cM of the genetic map of 6BS (Fig. 1). A similar genetic distance can be estimated from the 6BS map of Weng et al. (2000), in which the *Xcmwg79/Nor-B2* loci were 8.3 cM distal from the *Xpsr141* locus assigned to the centromeric C-6BS-5 bin (Weng et al. 2000). Based on these two values, the average physical to genetic distance ratio in this bin can be estimated as approximately 8.4–9.5 Mb cM⁻¹. This result suggests that a positional cloning effort within the 6BS-5 bin would be very difficult. However, using very large mapping populations (>6,000 gametes) we have been successful in the positional cloning of the wheat vernalization gene *Vrn1* in a region of chromosome 5A (fraction length 0.68–0.78) with a similar ratio of physical to genetic distances (Yan et al. 2003), suggesting that the positional cloning of *Gpc-6B1* will be difficult but not impossible.

One positive result from this study was the demonstration that the *Gpc-6B1* gene can be clearly separated by recombination from the *Nor-B2*. This is important because it has been shown that recombination is reduced in the vicinity of the *Nor* (Luo et al. 1998). The five recombinants detected in this study showed that the *Gpc-6B1* gene is far enough from the region where recombination is suppressed by the *Nor-B2*, and will not preclude a positional cloning effort in this region.

The precise mapping of the *Gpc-6B1* gene reported in this study, and the recent completion of a 5 × BAC library of RSL65 (Cenci et al. 2003) represent two fundamental steps in the efforts of our laboratory towards the positional cloning of this important agronomic gene.

Acknowledgements This research was supported by Research Grant US-3224-0IR from BARD (The United States -Israel Binational Agricultural Research and Development Fund) and USDA-IFAFS competitive grant 2001-04462. We thank Dr. O. Anderson, P. Leroy and J.M. Jacquemin for providing the probes used in this study, and to Dr. S. Huang and the California Wheat Commission for their help with the protein-content determinations.

References

- Appels R, Dvorak J (1982) The wheat ribosomal DNA spacer: its structure and variation in populations and among species. *Theor Appl Genet* 63:337–348

- Blanco A, Pasqualone A, Troccoli A, Di Fonzo N, Simeone R (2002) Detection of grain protein content QTLs across environments in tetraploid wheats. *Plant Mol Biol* 48:615–623
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Avivi L (1978) High protein content in wild tetraploid *Triticum dicoccoides* Korn. In: Ramanujam S (ed) 5th Int Wheat Genet Symp, Indian Soc Genet, Plant Breed, New Delhi, pp 372–380
- Cantrell RG, Joppa LR (1991) Genetic analysis of quantitative traits in wild emmer (*Triticum turgidum* L. var *dicoccoides*). *Crop Sci* 31:645–649
- Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half-million clones BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). *Theor Appl Genet* (in press)
- Chee PW, Elias SF, Kianinan SF, Anderson JA (1998) Introgression of a high protein gene from LDN(DIC-6B) substitution line. In: Slinkard A (ed) 9th Int Wheat Genetics Symposium, Vol. 2. University Extension Press, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, pp 179–181
- Chee PW, Elias EM, Anderson JA, Kianian SF (2001) Evaluation of a high grain protein QTL from *Triticum turgidum* L. var. *dicoccoides* in an adapted durum wheat background. *Crop Sci* 41:295–301
- Cox MC, Qualset CO, Rains DW (1985) Genetic variation for nitrogen assimilation and translocation in wheat III. Nitrogen translocation in relation to grain yield and protein. *Crop Sci* 26:737–740
- Dohlgan E, Hoffman L (2000) The new agricultural trade negotiations: background and issues for the U.S. wheat sector. Wheat yearbook. Economic research service, USDA
- Dubcovsky J, Galvez AF, Dvorak J (1994) Comparison of the genetic organization of the early salt-stress response gene system in salt-tolerant *Lophopyrum elongatum* and salt-sensitive wheat. *Theor Appl Genet* 87:957–964
- Dvorak J, McGuire PE, Cassidy B (1988) Apparent sources of the A genomes of wheats inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* 30:680–689
- Endo TR, Gill BS (1996) The deletion stocks of common wheat. *J Hered* 87:295–307
- Joppa LR, Hareland GA, Cantrell RG (1991) Quality characteristics of the Langdon Durum-Dicoccoides chromosome substitution lines. *Crop Sci* 31:1513–1517
- Joppa LR, Du C, Hart GE, Hareland GA (1997) Mapping a QTL for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. *Crop Sci* 37:1586–1589
- Khan IA, Procinier JD, Humphreys DG, Tranquilli G, Schlatter AR, Marcucci-Poltri S, Frohberg R, Dubcovsky J (2000) Development of PCR based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat. *Crop Sci* 40:518–524
- Kovacs MIP, Howes NK, Clarke JM, Leisle D (1998) Quality characteristics of durum wheat lines deriving high protein from *Triticum dicoccoides* (6b) substitution. *J Cereal Sci* 27:47–51
- Lander ES, Green P, Abrahamson J, Barlow A, Daly M, Lincoln SE, Newburg L (1987) MAPMAKER: an integrated computer package for construction of primary linkage maps of experimental and natural populations. *Genomics* 1:174–181
- Luo MC, Yang ZL, Dvorak J (1998) Position effects of ribosomal RNA multigene loci on meiotic recombination in wheat. *Genetics* 149:1105–1113
- Marino CL, Nelson JC, Lu YH, Sorrells ME, Leroy P, Lopes CR, Hart GE (1996) RFLP-based linkage maps of the homoeologous group-6 chromosomes of hexaploid wheat (*Triticum aestivum* L. em. Thell.). *Genome* 39:359–366
- Mesfin A, Frohberg RC, Anderson JA (1999) RFLP markers associated with high grain protein from *Triticum turgidum* L. var. *dicoccoides* introgressed into hard red Spring Wheat. *Crop Sci* 39:508–513
- Mickelson S, See D, Meyer FD, Garner JP, Foster CR, Blake TK, Fischer AM (2003) Mapping of QTL associated with nitrogen storage and re-mobilization in barley (*Hordeum vulgare* L.) leaves. *J Exp Bot* 54:801–812
- Mingeot D, Jacquemin JM (1999) Mapping of RFLP probes characterized for their polymorphism on wheat. *Theor Appl Genet* 98:1132–1137
- Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M, Leroy P, Ganal MW (1998) A microsatellites map of wheat. *Genetics* 149:2007–2023
- SAS Institute Inc. (2003) SAS user's guide, version 8. SAS Institute, Inc., Cary, New York
- See D, Kanazin V, Kephart K, Blake T (2002) Mapping genes controlling variation in barley grain protein concentration. *Crop Sci* 42:680–685
- Simmonds NW (1995) The relation between yield and protein in cereal grain. *J Sci Food Agric* 67:309–315
- Sourdille P, Cadalen T, Guyomarc'h H, Snape JW, Perretant MR, Charmet G, Boeuf C, Bernard S, Bernard M (2003) An update of the Courtot × Chinese Spring intervarietal molecular marker linkage map for QTL detection of agronomic traits in wheat. *Theor Appl Genet* 106:530–538
- Stoddart FL, Marshall DR (1990) Variability in grain protein in Australian hexaploid wheats. *Aust J Agric Res* 41:277:88
- Weng Y, Tuleen NA, Hart GE (2000) Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L. em. Thell.). *Theor Appl Genet* 100:519–527
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of wheat vernalization gene *Vrn1*. *Proc Natl Acad Sci USA* 100:6263–6268
- Young ND, Tanksley SD (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor Appl Genet* 77:95–101

Note added in proof Recent re-testing of DNA from RSL#106 showed that this line was heterozygous for the *Gpc-B1* locus, explaining the difficulty to classify it within the low or high-protein groups in the field experiments.