6. SUMMARY

The present study in wheat was undertaken with the following three major components: (i) Association mapping (AM) for some important agronomic traits with major emphasis on pre-harvest sprouting tolerance (PHST) using simple sequences repeat (SSR) markers, (ii) Association analysis of SNPs within the gene TaGW2 for grain weight in Indian wheat genotypes, and (iii) Evaluation and characterization of improved bread wheat lines that were derived following marker-assisted selection (MAS) and carried the high grain protein content (GPC) gene Gpc-B1 and two leaf rust resistance genes (Lr24 and Lr28) derived following MAS.

6.1. Association mapping of some important traits including PHST

(a) A set 230 Indian wheat genotypes were used for AM of 15 traits including PHST, morphological [plant height (PH), peduncle length (PL), flag leaf length (FLL), awn length (AL), days to heading (DTH) and days to maturity (DTM)], agronomic [spike length (SL), number of spikelet/spike (SKS), number of seeds/spike (SS) and 1000-grain weight (TGW)] and quality traits [Grain protein content (GPC), hardness index (HI), hectoliter weight (HW) and sedimentation volume (SV)]. These 230 Indian genotypes were mostly susceptible to PHS and therefore, 12 exotic wheat genotypes, which were tolerant to PHS were included in the AM panel along with 230 Indian wheat genotypes to increase the diversity for PHS trait.

(b) For AM, a set of 250 SSR markers spread over all the 21 wheat chromosomes was selected from the consensus genetic map of wheat (Somers et al. 2004). For the study of population structure a set of 42 SSRs, one SSR from each arm of the 21 individual chromosomes, was selected from previous published wheat genetic maps.

(c) Data on PHST was recorded for 242 wheat genotypes for 2-4 years during 2006-07 to 2009-10. Data on morphological, agronomic and quality traits of the 230 Indian wheat genotypes were available from DWR, Karnal (Kundu et al. 2006) and were utilized for AM during the present study.

(d) Model-based cluster analysis was performed to infer population structure in the dataset using the software STRUCTURE version 2.2 (Pritchard et al. 2000a). The number of sub-populations was determined following ΔK method (Evanno et al. 2005). The relationship between population structure and phenotypic traits (if any) was examined by means of multiple regression analysis.
(e) The best fit model, for marker trait association analysis for different traits, was selected using the lowest value of mean square difference (MSD) among the four available models viz. Naive, Q, K and Q+K. The expected P values used for MSD calculation were obtained by dividing the rank of an observed P value with the total number of markers (Stich et al. 2008).

(f) The software program TASSEL version 3.0 was used to identify associations between individual markers and traits, employing the following models: (i) naive model, (ii) Q-model, (iii) K-model, and (iv) Q+K-model. Significance of marker-trait associations was determined at p \( \leq 0.05 \). False discovery rate (FDR) criterion (Benjamini and Hochberg 1995) was used to reduce the number of false positive associations.

(g) Taken together, the extent of variability and normal distribution available for the different traits suggested suitability of Indian wheat genotypes included in the present study for the study of market-trait associations.

(h) Model-based cluster analysis identified 15 distinct sub-populations among 242 genotypes and 13 distinct sub-populations among 230 genotypes.

(i) For PHST, a total of 26 MTAs following GLM (The \( \text{R}^2 \) ranged from 0.56 to 4.48%), and a total of 17 MTAs following MLM were identified. As many as 30 MTAs were identified with 14 common MTAs from both the models. All these marker loci (including those detected by GLM and MLM) were located on 17 different chromosomes (1A, 1B, 2A, 2B, 2D, 3B, 3D, 4A, 4B, 4D, 5B, 5D, 6B, 6D, 7A, 7B and 7D). However, only three of these 30 QTL were found significant, when FDR criteria were applied.

(j) For PHST, association analysis was also conducted for two of the four individual environments, for which data on all the 242 genotypes were available. Seven MTAs were consistent over the environments (four were identified using MLM and three were identified using GLM). Out of the seven MTAs, 2 (Xgwm16 and Xcfd152) also fulfilled the FDR criteria.

(k) For morphological, agronomic and quality traits, altogether 213 MTAs were identified involving 129 SSRs on 21 wheat chromosomes. Maximum (24) SSRs were associated with AL and minimum 9 SSRs were associated with DTM (Table 3.12). Out of 129 SSRs, 72 SSRs were involved in trait-specific MTAs and 57 SSRs were involved in multi-trait MTAs. \( \text{R}^2 \) value of these MTAs ranged from 1.75 to 16.44%. Out of 213 MTAs, only 10 MTAs (involving 9 SSR markers; one SSR marker was associated with two traits) could qualify the FDR criterion.
(I) Some of the MTAs confirmed the QTL that were reported earlier, while others were novel for each of the 15 different traits. This confirms that AM is a powerful approach for identification of MTAs. Novel MTAs that were identified during the present study may prove useful for MAS after revalidation, perhaps using joint linkage and AM analysis, which could be the best approach for detecting reliable MTAs to be used for molecular breeding.

6.2. Association mapping for grain weight involving TaGW2-6A

(a) The plant material for this study comprised 207 Indian wheat genotypes (Appendix I). A pair of A-genome specific primer (Hap-6A-P1) was used to amplify the promoter region of gene TaGW2-6A including forward primer: 5' CGTTACCTCTGGTTGGGTGTCGTG 3' and reverse primer: 5'CACCTCTCGAAAATCTT CCCAA TTA3'.

(b) The amplified products were sequenced using ABI 3730xl sequencers at University of Delhi South Campus (UDSC), New Delhi. Sequence alignment and SNP detection were performed using software CLC genomics/DNA workbench.

(c) Association analysis for grain weight was conducted using General linear model (GLM) with 1000 permutation with the help of TASSEL (http://www.maizegenetics.net). Significance of the association test was determined by p-value (≤ 0.05). Analysis of variance was conducted by PROC MIXED in the Statistical Analysis System (SAS Institute, 1997) to test significance levels of grain weight of different haplotypes. Analysis of promoter region (regulatory element) involving sequences that were ~1kb upstream of the TaGW2-6A gene was conducted by PLACE database.

(d) Sequencing of 207 Indian wheat genotypes allowed identification of four SNPs in the promoter region of TaGW2-6A, at the positions -988bp, -739bp, -593bp and -494bp. All the four SNPs were biallelic and all were G-A substitutions. Minor allele frequencies of these SNPs ranged from 6.3 to 15.0%. Out of four, two SNPs (at -988bp and at -494bp) were novel.

(e) Out of four SNPs, only one SNP (at -494bp), that was associated with TGW with a p-value of 0.006 was found to be the causal SNP. Mean TGW (41.1 g) of 13 genotypes having allele A was significantly higher than the mean TGW (38.6 g) of the remaining 194 genotypes containing G allele.

(f) Using four SNPs, we could identify five haplotypes. Out of 207 genotypes, Hap1, Hap2, Hap3, Hap4 and Hap5 were present in 2 (0.97%), 15 (7.25%), 14 (6.76%), 163 (78.74%) and 13
wheat genotypes, respectively. Hap 5 had significantly higher TGW than other haplotypes.

(g) Out of the four SNPs, the SNP (SNP -494) showing significant association with higher TGW was located in a motif ‘CGCG’. The causal SNP (SNP-494) was used for developing a cleaved amplified polymorphism sequence (CAPS) marker to distinguish the TaGW2-6A alleles. After digestion with FauI, length polymorphism (363- vs 418-bp) was observed which could be easily resolved on agarose gels.

6.3. Field evaluation of MAS-derived lines with high grain protein content

(a) A set of 124 BC$_3$F$_5$/BC$_3$F$_6$ MAS-derived lines with high grain protein content gene Gpc-B1 in the background of 10 different wheat genotypes was evaluated for GPC and other agronomic traits in multi-location replicated trials for three consecutive years (2009-10 to 2011-12).

(b) In Crop season 2009-10, all 124 lines were evaluated along with their respective parental genotypes at three different locations i.e. Meerut, Pantnagar and Ludhiana. However, in crop seasons 2010-11 and 2011-12, A set of 75 MAS-derived lines (selected on the basis high GPC in 2009-10) were evaluated only in the zones for which the corresponding recipient parents were released.

(c) The analysis of variance (ANOVA) was conducted for different yield traits including GPC using data for all locations. Significance of differences between means was tested using Tukey’s test.

(d) Mean squares due to locations, blocks, genotypes and parents/derived lines were significant for all traits including GPC (%) with some exceptions. For the contrast, with and without Gpc-B1, the mean squares for GPC (%) were significant, but those for grain yield were not significant, suggesting that protein content in derived lines with Gpc-B1 was higher than in parents lacking Gpc-B1, and further suggesting that there is no yield penalty due to increase in protein content. Background effects due to recipient genotypes were also significant suggesting that the genetic background had some effect on the expression of Gpc-B1.

(e) In crop-season 2009-2010, only three lines one at each location showed significantly higher GPC (%) without any yield penalty relative to their respective recipient parental genotypes. A perusal of pooled data from three locations, however, showed that five lines involving three of the 10 recipient parents [two each belonging to genotypes HD2329 (Lr24 + Lr28) and Raj3765
and one belonging to HI977] had significantly higher GPC (%) with no significant reduction in yield (Table 5.10). One of the above five MAS derived lines, also had significantly higher GPC at one of the three locations, so that altogether there were seven lines, which either had higher GPC at one of the three locations or exhibited higher GPC in pooled data.

(f) In 2010/11, we identified three lines with high GPC and no loss in yield. Out of these three lines, two lines were identified at Ludhiana location in the background of HD2329-\(Lr24+Lr28\) and one line was identified at Pantnagar location in the background of Raj3765. These three lines were different from the above mentioned seven lines identified in 2009/10. In these three lines, GPC ranged from 14.67 to 15.99%.

(g) In 2011/12, we identified twelve lines with high GPC and no loss in yield. These lines were present in the background of HD2329 (\(Lr24+Lr28\)), PBW343 (\(Lr24\)), PBW343 and HI977. Although these lines were different from the above mentioned lines identified in 2009/10 and 2010/11 except for one (HD2329(\(Lr24+Lr28\))-Line No. 392) which was identified in 2010/11 as well as 2011/12. GPC ranged from 13.28 to 15.79% in the above mentioned 12 lines.