EST-SNPs in Bread Wheat: Discovery, Validation, Genotyping and Haplotype Structure

Sachin RUSTGI[1], Rajib BANDOPADHYAY[1,2], Harindra S. BALYAN[1] and Pushpendra K. GUPTA[1]

[1]Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut, India; [2]Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi, India

Abstract: The present study involves discovery, validation and use of single-nucleotide polymorphisms (SNPs) in bread wheat utilizing 48 EST-contigs (individual contigs having 20-89 ESTs, derived from 2 to 11 different genotypes). In order to avoid a problem due to homoeologous relationships, the ESTs in each contig were classified into 175 sub-contigs (3.7 sub-contigs/EST-contig) using characteristic homoeologue sequence variants (HSVs), which had a density of 1 HSV every 136.7 bp. *In silico* analysis of sub-contigs led to the discovery of 230 candidate EST-SNPs with a density of 1SNP/273.9 bp. Locus specific primers (each primer pair flanking 1–18 SNPs) were designed utilizing one sub-contig each from 42 EST-contigs that contained SNPs, the remaining 6 contigs having no SNPs. To provide locus specificity to the PCR products, each primer was tagged with an HSV at its 3' end. Only 10 primer pairs, which gave each a characteristic solitary band, were utilized to validate EST-SNPs over 30 diverse bread wheat genotypes; 7 SNPs were validated through resequencing the PCR products. Allele specific primers were designed and utilized for genotyping of 50 diverse bread wheat accessions (including 30 bread wheat genotypes previously used for validation of SNPs), with an aim to test their utility in genotyping and map construction. The allele specific primers allowed the classification of 50 genotypes in two alternative allele groups for each SNP as expected, thus suggesting their utility for genotyping. Of the above 7 validated SNPs, 4 belonged to a solitary locus (PKS37); 7 haplotypes were available at this locus. Altogether, the results suggested that EST-SNPs constitute an important source of molecular markers for studies on wheat genomics.

Keywords: contigs; DNA polymorphism; DNA sequencing; homoeologue sequence variants (HSVs); *Triticum aestivum* L.

Bread wheat (*Triticum aestivum* L.) is a hexaploid crop species having three genetically related sub-genomes ($2n = 6x = 42$; AABBDD) and a large genome size (17 Gb). A variety of molecular markers including RFLPs, SSRs and AFLPs have been developed and used in this crop for different purposes including the construction of molecular maps (Somers *et al*. 2004; Goyal *et al*. 2005; Liu *et al*. 2005; Song *et al*. 2005). More recently, EST database has also been utilized in this crop for development of a variety of functional markers, particularly the EST-SSRs (Gupta & Rustgi 2004; Gupta *et al*. 2005). An extensive study for assigning > 16 000 EST loci to chromosome bins has also been completed recently (Qi *et al*. 2004; Peng & Lapitan 2005), although these mapped ESTs cannot be directly used as markers. However, taken together, only ~7000 molecular markers have so far been placed on different maps (Somers *et al*. 2004; Gupta *et al*. 2005; Liu *et al*. 2005; Song *et al*. 2005; Sourdille *et al*. 2005) although according to some estimates, as many as 10 000 to 20 000 markers are needed on one map to have a reasonably high-density map in bread wheat (Appels 2003). Therefore, keeping in view the need for developing and mapping additional markers, emphasis is now shifting to a new class of abundant markers, the single nucleotide polymorphisms (SNPs). In 2002,

an International Wheat SNP Consortium (IWSC) was also constituted to accelerate the pace of work on SNPs in bread wheat (http://wheat.pw.usda.gov/ITMI/WheatSNP/). A number of studies for developing SNP markers in bread wheat have been conducted (Mochida *et al.* 2003; Schwarz *et al.* 2003; Somers *et al.* 2003; Yanagisawa *et al.* 2003; Zhang *et al.* 2003; Blake *et al.* 2004; Caldwell *et al.* 2004; Massa *et al.* 2004; Beales *et al.* 2005; Huang & Röder 2005; Ravel *et al.* 2006; Chao *et al.* 2009) but in most of these studies, SNPs have been utilized for characterizing polymorphism within specific genes (Schwarz *et al.* 2003; Yanagisawa *et al.* 2003; Zhang *et al.* 2003; Blake *et al.* 2004; Caldwell *et al.* 2004; Massa *et al.* 2004; Beales *et al.* 2005; Huang & Röder 2005). Only in a few of these studies, SNPs have been developed from ESTs (Mochida *et al.* 2003; Somers *et al.* 2003; Ravel *et al.* 2006; Chao *et al.* 2009), and no study on random genomic SNPs is available. Thus, there is a need for further work on development and use of SNPs in bread wheat, where triplicate loci due to three closely related sub-genomes and paralogues due to further duplication of loci make the development of SNPs rather difficult. The present study is another effort to develop SNP markers in this crop, and the uniqueness of this work lies in the genotyping platform used during the study, which is extremely user-friendly, cost efficient and can be implemented in any marker-assisted plant breeding program.

MATERIALS AND METHODS

Seed material

A set of 50 diverse bread wheat genotypes originating from 25 countries was used for SNP validation and genotyping (Table 1). None of these genotypes was the source of ESTs available in the database that were used for SNP discovery, so that a subset of these genotypes proved useful for validation of SNPs.

EST-contigs and detection of candidate SNPs

Under the aegis of IWSC, ESTs were assembled into contigs and 48 (belonging to 19 genotypes) contigs, each having 20–89 ESTs, were used for the detection of candidate SNPs (http://wheat.pw.usda.gov/ITMI/WheatSNP). The EST-contigs were visualized using JalView (M. Clamp; http://www.ebi.ac.uk/~michele/jalview), following Somers *et al.* (2003), to identify homoeologue specific EST clusters (hereafter called EST-subcontigs), which were defined by the presence of homoeologue sequence variants (HSVs) at specific nucleotides. Subsequently in each homoeologue specific EST-subcontig, SNPs were detected manually.

Designing and synthesis of STS primers

Individual pairs of STS (sequence tagged site) primers were designed for each of the 42 EST-subcontigs in such a manner that each selected subcontig had the maximum number of ESTs within the corresponding contig. The two primers of an individual pair flanked a part of the sequence of EST-subcontig containing SNPs, and were tagged at their 3' ends on HSVs that were used to define the homoeologue specific EST-subcontig. The length of STS primers ranged from 18 to 32 bases and annealing temperatures ranged from 55°C to 69°C (Table 2). The primers were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and synthesized on contract from Illumina Inc., USA (http://www.illumina.com).

PCR amplification, purification and sequencing of PCR products

DNA amplification was carried out in a total of 1260 PCR reactions (42 STS primers × template DNA of 30 genotypes). The PCRs were carried out in 25 µl reaction mixtures, each containing 50 ng template DNA, 2.5µM of each of the two STS primers, 200µM dNTPs, 2.5mM $MgCl_2$, 1× PCR buffer and 0.5 U Taq DNA polymerase (Life Technologies, USA) in a DNA Mastercycler, Eppendorf, Germany using the following PCR profile: initial denaturation at 95°C for 5 min followed by 40 cycles at 95°C for 1 min, 55°C to 69°C for 1 min (according to primer's annealing temperature), 72°C for 1 min, with a ramp at the rate of 0.5°C per second and final extension at 72°C for 10 min. Small aliquots of PCR products were resolved on 10% polyacrylamide denaturing gels (PAGE), following silver staining. The remaining aliquots of PCR products that gave a solitary band each were purified using AuPreP™ PCR purification kit (Life Technologies

Table 1. Fifty bread wheat genotypes and their countries of origin

S. no. & genotype	Country of origin	S. no. & genotype	Country of origin
1. E3249*	Sweden	26. E3389*	Lebanon
2. E3111*	Portugal	27. E3840*	Columbia
3. E3877*	Brazil	28. E3274*	The Netherlands
4. E3901*	Ecuador	29. E3839*	Columbia
5. E4205*	Africa	30. E3273*	The Netherlands
6. E3275*	The Netherlands	31. E3086	USA
7. E549*	Canada	32. E3860	Guatemala
8. E3083*	S. Africa	33. E3896	Ecuador
9. E585*	Kenya	34. E4813	Kenya
10. E661*	Argentina	35. E4229	Spain
11. E3876*	Brazil	36. E4328	Czechoslovakia
12. E146*	Kenya	37. E581	Kenya
13. E149*	Kenya	38. E336	Canada
14. E680*	Argentina	39. E288	Poland
15. E780*	Brazil	40. E319	Australia
16. E1003*	Cyprus	41. E271	Poland
17. E965*	Yugoslavia	42. E677	Argentina
18. E2336*	USA	43. E784	Brazil
19. E1000*	Cyprus	44. E2055	Poland
20. E2724*	Israel	45. E2161	USA
21. E2990*	Turkey	46. E3066	Mexico
22. E2401*	USA	47. E3070	Mexico
23. E3068*	Mexico	48. E3859	Guatemala
24. E2602*	Italy	49. E3492	USA
25. E3414*	Ethiopia	50. E3387	Lebanon

* Genotypes used for validation of SNPs; – Pedigrees of the genotypes are described elsewhere (Gupta *et al*. 2003)

India Pvt. Ltd.), and were sequenced from both the directions on contract from MACROGEN, Seoul, South Korea (http://www.macrogen.com). The alignment of sequences and detection of SNPs was carried out using Sequencher (http://www.genecodes.com).

Validation of SNPs and development of PCR based assay

Seven (7) putative SNPs validated through direct sequencing of PCR products were used to develop allele specific primers using Primer3 (for primer details, see Table 3) following the strategy described elsewhere (Jeong *et al*. 2004; Chiapparino *et al*. 2004). The allele specific primers varied in length from 18 to 30 bases and the annealing temperatures ranged from 58.3°C to 64.2°C (Table 3). The primers were synthesized on contract from Illumina Inc., USA. The two allele specific primers were used along with two corresponding STS primers (each tagged with an HSV on either side of the SNP(s)), previously used for validation. The size of PCR products in such a reaction would depend on the specific SNP allele (see Figure 1a, b).

PCR reactions using allele specific primers were set up as described earlier for homoeologue specific STS primers. The PCR products obtained using allele specific primers were subjected to agarose gel (2.0%) electrophoresis followed by ethidium bromide staining.

Table 2. Details of 10 homoeologue specific STS primers each amplifying a solitary band in bread wheat genotypes

Primer designation	Forward Primer (5'→3')	Reverse Primer (5'→3')	Am (°C)	Product size (bp)
PKS2	TTTCTTCCGCATCAAGAGATCC	CCTCAGGCTATGGCACAGAAT	57.20	325
PKS4	CCGCTATGCTCCTCGTAGCCTG	CAACGTACTCAGCGTAGATCGCTG	61.40	366
PKS6	CACGAAGAGATATCACCCCGAG	GGATGTCTGCGAGCCTTTCATAT	58.70	242
PKS8	TTCGCACATGCGCATAGATACA	ACGCGGTCAGACAAACATGC	59.61	390
PKS10	CAACACGCCACCAACAATACTTCTG	TGCTTCTCCTGCGGCGAGTC	62.61	249
PKS16	GCTTGTTGAAGGTTTGAAGCGT	CCCAGCCAAACTCCTCAGAC	57.50	372
PKS30	GCTGCCTACAACAATGGTTCTGC	CGTACCTCGTCGGGCTGTTC	60.21	360
PKS35	CAACCCGATTCACCGCCTAC	CTAACGGCCTCCGTTTGCTG	60.00	480
PKS36	TCCTCTCCATGGCCAACGCC	AACCAAGGCCGCCGGACGC	66.88	325
PKS37	TTGAGCGGCACAGCTCATCG	TGCAAGTGATCCTCCCGTGCTA	62.75	530

Am – annealing temperature

Haplotype construction

In one case, four SNPs were available at the same locus. Using SNP alleles at each of these SNP loci, haplotypes were assembled.

RESULTS

In silico mining of SNPs

When 48 EST-contigs were subjected to multiple alignments, as many as 462 homoeologue sequence variants (HSVs) were identified in 47 contigs. The number of HSVs per contig ranged from 2 to 26, giving an average density of 1 HSV every 136.7 bp (for HSV, Somers *et al*. 2003). Utilizing information on HSVs, ESTs in 48 EST-contigs were grouped into 175 subcontigs, with an average of 3.70 subcontigs per contig; the number of ESTs in individual subcontigs ranged from 2 to 30. Subcontigs containing 6 or more ESTs were considered suitable for the visual detection of SNPs. As many as 230 putative SNPs (giving an average density of 1 SNP every 273.9 bp) were detected in 155 such subcontigs belonging to 42

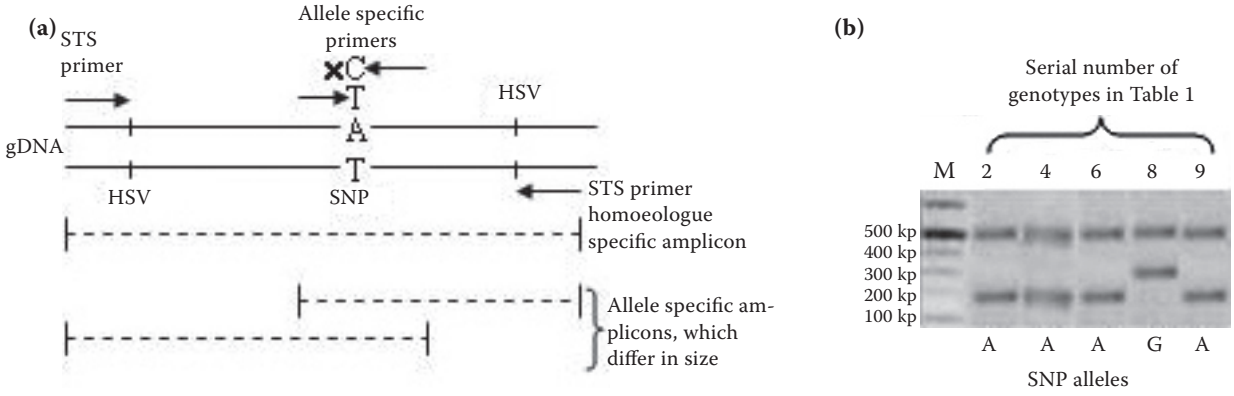

Figure 1. PCR-based assay for genotyping of SNPs; (a) diagrammatic representation of the strategy used for genotyping of SNPs; the two allele specific primers were designed in such a way that their 3' ends carried different SNP allele; STS primers were designed in such a way that their 3' ends carried each an HSV whereas the distance of two HSVs on either side of SNP differed; consequently two PCR products (one invariant band and the other variable band) will be available, and the variable band will help genotyping; (b) PCR amplification profiles obtained using allele specific primers at locus PKS37b in five representative bread wheat genotypes, showing an alternative allele in genotype 8 (the other four genotypes have the same allele); M = 100 bp marker

Table 3. Details of 7 allele specific primers, their product sizes (when used in combination with respective homoeologue specific STS primers, see Table 2) and annealing temperatures

Contig ID	Primer name	SNP position (bp)	Allele specific SNP primer	Product size (bp)	Am (°C)
1636-2458	PKS8b	C/T (396)	GATGTGCAGCACCTCGCA**A**/**C**	113	60.33
			CCAAGGTCCCGAAGTTCTGC**T**/**A***	313	59.82
	PKS16	C/T (1220)	GGGGACTTGCTCATAACTG**C**/**G***	234	59.14
			TGTTCTGGAGAAGTCCTGCCA**T**	179	58.35
1636-3749	PKS30	A/G (266)	TACTCCAACAACCATATTTCTTCTACAC**C**/**G**	242	60.97
			CTAATTGCCGTGCTAGGTCTATCTTCT**C**/**T***	148	60.14
1636-935	PKS37b	G/A (601)	AAGGGCCACGACCCTTTATG/**G**	390	61.34
			TTGGTCTCGGATGCTGCTG**G**/**T***	180	62.53
	PKS37c	T/G (806)	CGCAAGCACTCTTAATCCAGCG**T**/**T**	188	62.63
			CCCTCGTGTGGCTTCG**C**/**C***	383	62.59
	PKS37e	G/A (809)	AAGCACTCTTAATCCAGCGAT**G**/**C**G	185	62.66
			CCGTCCCTCGTGTGGCT**G**/**T***	386	62.01
	PKS37f	C/A (838)	TTGTACTGCGGCCGACTA**G**/**C**G*	417	62.73
			CGTTGTACTGCGGCCGACTA**T**/**C**T*	419	61.64

*Reverse primer; underlined base represents non-template specific mismatch; Am – annealing temperature

EST-contigs. In the remaining 20 subcontigs belonging to six contigs, no SNPs could be detected. Out of 230 SNPs, 123 (53.5%) represented transitions and the remaining 107 (46.5%) represented transversions. The frequencies of different types of substitutions ranged from 0.65% for A/C or C/A to 25.3% for G/A.

Validation of SNPs

Only 39 (92.8%) of the 42 homoeologue specific STS primers each amplified products in 15 to 30 genotypes. Ten (10) of these 39 STS primers gave each a solitary band (representing a single homoeolocus) in each of 237 (79%) of the 300 possible (10 primers × 30 genotypes) primer-genotype combinations (Table 2); the remaining 29 primer pairs amplified 2 to 8 fragments each in 22 to 30 genotypes. Only the above 10 primers amplifying the solitary fragment were considered suitable for validation of 30 putative SNPs that were detected *in silico* in the amplifiable regions (amplicon) of the corresponding EST-contigs (Table 4).

The solitary PCR fragments amplified in each of the above 237 primer-genotype combinations were sequenced from both the directions. The sequence analysis validated only 7 (23%) of the 30 SNPs available in three contigs (for details, see Table 4). Five of these 7 SNPs belonged to ORFs (no annotations for these ORFs were available in the database), 3 of which represented substitutions resulting in nonsynonymous codons and the remaining two represented substitutions that resulted in synonymous codons. Also, 5 (71.43%) of the 7 SNPs represented transitions and the remaining 2 (28.57%) represented transversions.

Development of PCR based assay

Allele-specific primers developed for the seven validated SNPs were tried on a set of 50 diverse bread wheat genotypes to test their utility as SNP markers. The test allowed not only the confirmation of already validated SNP alleles in 20 to 25 of the 30 bread wheat genotypes (previously used for validation of SNPs) but also the characterization of alleles in the remaining 20 previously untested genotypes (Figure 1a, b). Thus, suggesting that these allele specific primers can be used as markers to faithfully genotype individuals at these loci. However,

Table 4. Summary of the results of putative and validated EST-SNPs, and haplotypes in 10 EST-contigs

Contig ID	No. of subcontigs	Primer designations	No. of putative SNPs in contig	No. of putative SNPs in amplicon	Validated SNPs
1636-2208	4	PKS4	5	5	0
1636-2377	3	PKS6	3	1	0
1636-2458	3	PKS8	3	3	2
1636-2602	3	PKS10	2	1	0
1636-3749	4	PKS16	17	9	1
1636-709	4	PKS30	2	1	0
1636-923	3	PKS35	1	1	0
1636-935	4	PKS37	4	4	4
1636-203	3	PKS2	3	2	0
1636-924	3	PKS36	3	3	0
Total			43	30	7

a homoeologue (represented by an EST-subcontig) specific band (outer amplicon) was observed only with 37 to 42 genotypes out of 41 to 45 genotypes amplified using 6 of the above 7 allele specific primers when used in combination with corresponding homoeologue specific STS primers.

Haplotype structure

The occurrence of 4 out of the 7 validated SNPs at a single locus amplified by primer pair PKS37 allowed haplotype analysis at this locus. The analysis of genotyping data belonging to these 4 SNPs over 42 (out of 50) bread wheat genotypes allowed the detection of seven haplotypes. Haplotype 1 was detected in 29 (69.05%) genotypes followed by haplotypes 3 and 4 in 4 (9.52%) genotypes each, haplotype 2 in two (4.76%) genotypes, and haplotypes 5, 6 and 7 each in a single (2.38%) genotype (Figure 2).

Annotation of EST-contigs

In view of assigning putative functions to the three EST-contigs harbouring seven validated SNPs, their sequences were compared with the nonredundant (nr)-protein sequence database maintained at NCBI using blastx. The analysis allowed the assignment of putative functions to two of the three EST-contigs, numbered 1636-2458 and 1636-3749, which showed similarity with trypsin/

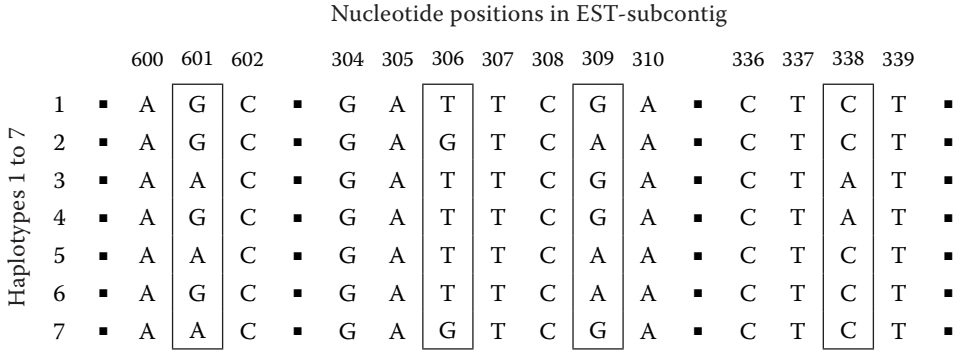

Figure 2. Seven haplotypes detected at the locus characterized by STS primer pair PKS37 in 42 bread wheat genotypes; haplotypes 1 was detected in 29 genotypes (2, 6, 9, 11, 13, 15, 17, 18, 20, 22, 23, 24, 26, 27, 29, 30, 32, 33, 36, 37, 38, 39, 40, 41, 44, 45, 47, 49, 50), haplotype 2 in 2 genotypes (8, 25), haplotype 3 in 4 genotypes (1, 14, 19, 21), haplotype 4 in 4 genotypes (31, 35, 42, 43) and haplotypes 5 (34), 6 (46) and 7 (48) each in a solitary genotype (genotype numbers in parenthesis correspond to the serial numbers in Table 1); the SNPs are indicated by hollow rectangular boxes

alpha-amylase inhibitor CMX1/CMX3 (89%; e-value 8e-56; score 220) and elongation factor 2 (95%; e-value 0.0; score 1223), respectively.

DISCUSSION

In bread wheat, the development of SNPs is difficult due to a high level of similarity among EST sequences representing homoeologous triplicate loci and due to the occurrence of multigene families. Consequently, an individual EST-contig contains ESTs belonging to three or more homoeologues/paralogues. The differences among sequences representing homoeoloci are minor, so that despite stringent assembly conditions (95% similarity over a 40-bp window) homoeologous ESTs cannot be assembled into independent contigs. In view of this, if the EST contigs are directly used for discovery of SNPs, single nucleotide variations due to homoeoloci will confound the results. Therefore, for SNP discovery in bread wheat, EST contigs are first separated into subcontigs, each representing a locus on an individual chromosome. This is made possible by identifying polymorphisms, described as homoeologue sequence variants (HSVs), which characterize each of the triplicate homoeoloci within an EST-contig (Somers *et al.* 2003). The HSVs could also be utilized for designing primers in such a manner that each primer pair allows the amplification of only one of the three homoeologous loci. This strategy was successfully utilized by Somers *et al.* (2003) and the present study confirms the utility of this approach for SNP discovery in bread wheat. However, surprisingly, in the present study, 1 HSV occurred every 136.7 bp as against 1 HSV every 24 bp reported by Somers *et al.* (2003). This is intriguing, although it may partly be attributed to a small sample used in the present study and to different sets of contigs used in the two studies.

SNP discovery and validation

During the present study, only 10 out of the 42 primers designed were found suitable for discovery of locus specific SNPs. Complete failure of amplification by three primers and amplification of more than one products by each of the remaining 29 primers may be attributed to several reasons: first, there may be possible errors in single pass EST sequences organized into contigs; second, a single nucleotide corresponding to an HSV at the 3'-end of the primers may provide insufficient specificity, and the third, there are also paralogues within the genome due to duplications [one-quarter of genes in the bread wheat genome are believed to have two or more duplicate loci (paralogues), so that some genes can have as many as ~6 loci (Akhunov *et al.* 2003a, b)]. An important example of the occurrence of homoeologues and paralogues is provided by the *Glu-1* loci encoding HMW-glutenin, which are located on the long arms of each of the homoeologous group 1 chromosomes of bread wheat (Shewry *et al.* 1992). In view of the above, the amplification of up to 8 fragments by an individual STS primer is not surprising. Nevertheless, this causes a major problem in the study of SNPs in individual sequences representing a particular gene in bread wheat. In a recent study, only 12% of the 5 762 informative EST probes detected 1–2 loci on the wheat genome (Qi *et al.* 2004). Considering the magnitude of the problem, an NSF funded wheat SNP project was initiated recently. The above project will undertake validation and subsequent genotyping of 1 800 SNPs derived from ESTs representing unique loci on the wheat physical map (Powell & Langridge 2004; http://wheat.pw.usda.gov/SNP/). More recently, the successful completion of a pilot project (Caldwell *et al.* 2004) where genome specific primers were designed from the sequences of two genes involved in starch biosynthesis (*GSS* and *Xwye838*) gave further support to the desirability/feasibility of the above NSF funded project.

Density of SNPs

In the present study, estimates of the density of SNPs were done through both the *in silico* analysis of the ESTs available as contigs and the analysis of sequences of PCR products that were obtained by the use of 10 selected STS primer pairs with 30 exotic wheat genotypes. The *in silico* analysis of 48 EST-contigs gave a density of 1 SNP/273 bp, but when this *in silico* analysis was restricted only to the amplicons flanked by the 10 selected STS primers, this density was much higher (1 SNP every 121.3 bp). This was not surprising, because while designing the primers, segments having a higher density of SNPs were really selected. However, the sequences of PCR products obtained using 10 primers (see above) gave a lower density (1 SNP/520 bp), which is also understandable in view of the fact

that only 7 out of 30 SNPs could be validated. The density of SNPs reported during the present study, however, falls within the range (1SNP/21bp to 1SNP/8500bp) of densities of SNPs reported in earlier studies on different tree and plant species (Bryan *et al*. 1999; Schmid *et al*. 2003; Blake *et al*. 2004; Bundock & Henry 2004; Feltus *et al.* 2004; Gupta & Rustgi 2004; Le Dantec *et al*. 2004; Morales *et al*. 2004; Russell *et al*. 2004; Shen *et al.* 2004; Yang *et al*. 2004; Rostoks *et al*. 2005), but is little above the density reported for bread wheat in an earlier study (1 SNP/540 bp; Somers *et al.* 2003). We also noticed that in bread wheat the density of SNPs scored in EST databases (1SNP/144.9 bp; 1 SNP/540 bp) was higher than that reported in sequences of genes of economic importance (1SNP/1000bp to 1SNP/1700bp; Bryan *et al*. 1999; Mochida *et al.* 2003; Somers *et al.* 2003; Zhang *et al*. 2003; Blake *et al*. 2004). This variation in the density of SNPs may be attributed to (i) the number and type of genotypes sampled and (ii) the variable selection pressure on the genomic regions under study. Also, in the present study, transitions were more frequent than transversions, which is in conformity with the results of earlier studies in bread wheat (Caldwell *et al.* 2004) and sugar beet (Schneider *et al*. 2001).

Development of PCR based assay for SNP genotyping

PCR based assays developed for 7 validated SNPs allowed faithful genotyping of 40–45 bread wheat genotypes, suggesting the probable use of SNPs in constructing genetic maps and in studying genetic diversity available in the wheat germplasm (Gupta *et al*. 2003; Somers *et al*. 2003). However, with six out of the above seven allele specific primers, a homoeologue specific band (large fragment) was observed in 37 to 42 genotypes. The above observation is in conformity with earlier studies conducted in barley (Chiapparino *et al*. 2004; Bundock *et al.* 2005), where the failure of genome specific amplification was attributed to the competition among one of the locus specific primers and allele specific primers used during the PCR reaction.

Haplotype structure

During the present study, the availability of 4 SNPs at a solitary locus allowed studying the haplotype structure at this locus. It was shown that allelic variants at this locus consisted of haplotypes rather than single base replacements. Similar studies on haplotyping were conducted earlier in bread wheat (Caldwell *et al*. 2004; Beales *et al*. 2005), barley (Bundock & Henry 2004; Russell *et al*. 2004) and maize (Ching *et al.* 2002; Palaisa *et al.* 2003). Taken together, these studies suggest that on average there are fewer SNPs/haplotype in inbreeders like wheat and barley than in outbreeders like maize (Clark *et al*. 2004; Jung *et al*. 2004; Palaisa *et al.* 2004; Buntjer *et al*. 2005). This has been attributed to selective constraints, absence of recombination and narrow genetic base of the inbreeders (Kanazin *et al*. 2002; Zhu *et al*. 2003; Bundock & Henry 2004; Caldwell *et al*. 2004; Russell *et al*. 2004). During the present study, out of the seven haplotypes, haplotype 1 (GTGC) was the most abundant. However, once a function is assigned to the concerned locus, it would be interesting to find the association of haplotypes and phenotypes.

Annotation of EST-contigs

By extensive homology searches, putative functions could be assigned to two of the three EST-contigs containing validated SNPs. One of these EST-contigs showed similarity with the trypsin/alpha-amylase inhibitor CMx1/CMx3, which is known to inhibit the production of germination specific *alpha*-amylase, thus regulating the amount of sugars released from starchy endosperm during germination, which not only helps in preventing precocious germination (i.e. pre-harvest sprouting) but also protects seeds from predators (Taufel *et al.* 1997). Interestingly, the map locations of *CMx* (*CMx1*, *CMx2* and *CMx3*) genes also coincide well with the known location of three major dormancy QTLs detected on group 4 chromosomes of bread wheat (Sanchez De La Hoz *et al.* 1994; Mori *et al*. 2005), thus suggesting the usefulness of these SNP markers in gene-assisted selection (GAS) for pre-harvest sprouting tolerance (PHST).

CONCLUSIONS

Single-nucleotide polymorphisms (SNPs) represent an important source of polymorphisms in bread wheat like in other plant systems, although most of the allelic variants for different genes of economic importance in wheat (like in other crops) may consist of haplotypes rather than single base replacements. Furthermore, only a few SNPs can potentially define a haplotype and such SNPs can be used for the study of linkage disequilibrium (LD) and association analysis. The minimal number of SNPs defining an allele may also prove useful for gene-assisted selection (GAS) programmes for the crop of interest. Extensive studies on the discovery of SNPs followed by genotyping of the germplasm with these SNPs are needed for this purpose.

Acknowledgements. The financial support received from the Department of Biotechnology, Government of India for carrying out this study, from the Council of Scientific and Industrial Research, New Delhi in the form of a Senior Research Fellowship to SR and from the Indian National Science Academy, New Delhi in the form of a Senior Scientist Position to PKG is gratefully acknowledged.

References

Akhunov E.D., Akhunova A.R., Linkiewicz A.M., Dubcovsky J., Hummel D., Lazo G.R., Chao S., Anderson O.D., David J., Qi L., Echalier B., Gill B.S., Miftahudin, Gustafson J.P., La Rota M., Sorrells M.E., Zhang D., Nguyen H.T., Kalavacharla V., Hossain K., Kianian S.F., Peng J., Lapitan N.L.V., Wennerlind E.J., Nduati V., Anderson J.A., Sidhu D., Gill K.S., McGuire P.E., Qualset C.O., Dvorak J. (2003a): Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. Proceedings of National Academy of Sciences, USA, **100**: 10836–10841.

Akhunov E.D., Goodyear A.W., Geng S., Qi L.-L., Echalier B., Gill B.S., Miftahudin, Gustafson J.P., Lazo G., Chao S., Anderson O.D., Linkiewicz A.M., Dubcovsky J., La Rota M., Sorrells M.E., Zhang D., Nguyen H.T., Kalavacharla V., Hossain K., Kianian S.F., Peng J., Lapitan N.L.V., Gonzalez-Hernandez J.L., Anderson J.A., Choi D.-W., Close T.J., Dilbirligi M., Gill K.S., Walker-Simmons M.K., Steber C., McGuire P.E., Qualset C.O., Dvorak J. (2003b): The organization and the rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Research, **5**: 753–763.

Appels R. (2003): A consensus molecular genetic map of wheat-a cooperative international effort. In: Pogna N.E. (ed.): Proc. 10th Int. Wheat Genetics Symposium. Paestum, 211–214.

Beales J., Laurie D.A., Devos K.M. (2005): Allelic variation at the linked *AP1* and *PhyC* loci in hexaploid wheat is associated but not perfectly correlated with vernalization response. Theoretical and Applied Genetics, **110**: 1099–1107.

Blake N.K., Sherman J.D., Dvořák J., Talbert L.E. (2004): Genome-specific primer sets for starch biosynthesis genes in wheat. Theoretical and Applied Genetics, **109**: 1295–1302.

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

Bundock P.C., Henry R.J. (2004): Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. Theoretical and Applied Genetics, **109**: 543–551.

Bundock P.C., Cross M.J., Shapter F.M., Henry R.J. (2005): Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. Theoretical and Applied Genetics, **112**: 358–365.

Buntjer J.B., Sørensen A.P., Peleman J.D. (2005): Haplotype diversity: the link between statistical and biological association. Trends in Plant Science, **10**: 466–471.

Caldwell K.S., Dvořák J., Lagudah E.S., Akhunov E., Luo M.-C., Wolters P., Powell W. (2004): Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. Genetics, **167**: 941–947.

Chao S., Zhang W., Akhunov E., Sherman J., Ma Y., Luo M.-C., Dubcovsky J. (2009): Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Molecular Breeding, **23**: 23–33.

Chiapparino E., Lee D., Donini P. (2004): Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS-PCR. Genome, **47**: 414–420.

Ching A., Caldwell K.S., Jung M., Dolan M., Smith O.S., Tingey S., Morgante M., Rafalski A.J. (2002): SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genetics, **3**: 19.

Clark R.M., Linton E., Messing J., Doebley J.F. (2004): Patterns of diversity in the genomic region near the maize domestication gene *tb1*. Proceedings of National Academy of Sciences, USA, **101**: 700–707.

Feltus F.A., Wan J., Schulze S.R., Estill J.C., Jiang N., Paterson A.H. (2004): An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. Genome Research, **14**: 1812–1819.

Goyal A., Bandopadhyay R., Sourdille P., Endo T.R., Balyan H.S., Gupta P.K. (2005): Physical molecular maps of wheat chromosomes. Functional and Integrative Genomics, **5**: 260–263.

Gupta P.K., Rustgi S. (2004): Molecular markers from the transcribed/expressed region of the genome in higher plants. Functional and Integrative Genomics, **4**: 139–162.

Gupta P.K., Rustgi S., Sharma S., Singh R., Kumar N., Balyan H.S. (2003): Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Molecular Genetics and Genomics, **270**: 315–323.

Gupta P.K., Kulwal P.L., Rustgi S. (2005): Wheat cytogenetics in the genomics era and its relevance to breeding. Cytogenetics and Genome Research, **109**: 315–327.

Huang X.-Q., Röder M.S. (2005): Development of SNP assays for genotyping the Puroindoline b gene for grain hardness in wheat using pyrosequencing. Journal of Agriculture and Food Chemistry, **53**: 2070–2075.

Jeong S.C., Saghai Maroof M.A. (2004): Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv1* and *Rsv3*, of soybean. Plant Breeding, **123**: 305–310.

Jung M., Ching A., Bhattramakki D., Dolan M., Tingey S., Morgante M., Rafalski A. (2004): Linkage disequilibrium and sequence diversity in a 500-kbp region around the *adh1* locus in elite maize germplasm. Theoretical and Applied Genetics, **109**: 681–689.

Kanazin V., Talbert H., See D., Decamp P., Nevo E., Blake T. (2002): Discovery and assay of single-nucleotide polymorphisms in barley. Plant Molecular Biology, **48**: 529–537.

Le DAntec L., Chagné D., Pot D., Cantin O., Garnier-Géré P., Bedon F., Frigerio J.M., Chaumeil P., Léger P., Garcia V., Laigret F., De Daruvar A., Plomion C. (2004): Automated SNP detection in expressed sequence tags: statistical considerations and application to maritime pine sequences. Plant Molecular Biology, **54**: 461–470.

Liu Z.H., Anderson J.A., Hu J., Friesen T.L., Rasmussen J.B., Faris J.D. (2005): A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. Theoretical and Applied Genetics, **111**: 782–794.

Massa A.N., Morris C.F., Gill B.S. (2004): Sequence diversity of puroindoline-a, puroindoline-b, and the grain softness protein gene in *Aegilops tauschii* Coss. Crop Science, **44**: 1808–1816.

Mochida K., Yamazaki Y., Ogihara Y. (2003): Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Molecular Genetics and Genomics, **270**: 371–377.

Morales M., Roig E., Monforte A.J., Arús P., Garcia-Mas J. (2004): Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). Genome, **47**: 352–360.

Mori M., Uchino N., Chono M., Kato K., Miura H. (2005): Mapping QTLs for grain dormancy on wheat chromosome 3A and the group 4 chromosomes, and their combined effect. Theoretical and Applied Genetics, **110**: 1315–1323.

Palaisa K.A., Morgantem M., Williams M., Rafalski A. (2003): Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. The Plant Cell, **15**: 1795–1806.

Palaisa K., Morgante M., Tingey S., Rafalski A. (2004): Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proceedings of National Academy of Sciences, USA, **101**: 9885–9890.

Peng J.H., Lapitan N.L.V. (2005): Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Functional and Integrative Genomics, **5**: 80–96.

Powell W., Langridge P. (2004): Unfashionable crop species flourish in the 21st century. Genome Biology, **5**: 233.

Qi L.L., Echalier B., Chao S., Lazo G.R., Butler G.E., Anderson O.D., Akhunov E.D., Dvorak J., Linkiewicz A.M., Ratnasiri A., Dubcovsky J., Bermudez-Kandianis C.E., Greene R.A., Kantety R., La Rota C.M., Munkvold J.D., Sorrells S.F., Sorrells M.E., Dilbirligi M., Sidhu D., Erayman M., Randhawa H.S., Sandhu D., Bondareva S.N., Gill K.S., Mahmoud A.A., Ma X.-F., Miftahudin, Gustafson J.P., Conley E.J., Nduati V., Gonzalez-Hernandez J.L., Anderson J.A., Peng J.H., Lapitan N.L.V., Hossain K.G., Kalavacharla V., Kianian S.F., Pathan M.S., Zhang D.S., Nguyen H.T., Choi D.-W., Fenton R.D., Close T.J., McGuire P.E., Qualset C.O., Gill B.S. (2004): A chromosome bin map of 16,000 EST loci and distribution of genes among the three genomes of polyploid wheat. Genetics, **168**: 701–712.

Ravel C., Praud S., Murigneux A., Canaguier A., Sapet F., Samson D., Balfourier F., Dufour P.,

Chalhoub B., Brunel D., Beckert M., Charmet G. (2006): Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome, **49**: 1131–1139.

Rostoks N., Mudie S., Cardle L., Russell J., Ramsay L., Booth A., Svensson J.T., Wanamaker S.I., Walia H., Rodriguez E.M., Hedley P.E., Liu H., Morris J., Close T.J., Marshall D.F., Waugh R. (2005): Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Molecular Genetics and Genomics, **274**: 515–527.

Russell J., Booth A., Fuller J., Harrower B., Hedley P., Machray G., Powell W. (2004): A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome, **47**: 389–398.

Sanchez De La Hoz P., Castagnaro A., Carbonero P. (1994): Sharp divergence between wheat and barley at loci encoding novel members of the trypsin/α-amylase inhibitors family. Plant Molecular Biology, **26**: 1231–1236.

Schmid K.J., Sörensen T.R., Stracke R., Törejék O., Altmann T., Mitchell-Olds T., Weisshaar B. (2003): Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. Genome Research, **13**: 1250–1257.

Schneider K., Weisshaar B., Borchardt D.C., Salamini F. (2001): SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. Molecular Breeding, **8**: 63–74.

Schwarz G., Sift A., Wenzel G., Mohler V. (2003): DHPLC Scoring of a SNP between promoter sequences of HMW glutenin x-type alleles at the *Glu-D1* locus in wheat. Journal of Agriculture and Food Chemistry, **51**: 4263–4267.

Shen Y.-J., Jiang H., Jin J.-P., Zhang Z.-B., Xi B., He Y.-Y., Wang G., Wang C., Qian L., Li X., Yu Q.-B., Liu H.-J., Chen D.-H., Gao J.-H., Huang H., Shi T.-L., Yang Z.-N. (2004): Development of genome-wide DNA polymorphism database for map-based cloning of rice genes[1w]. Plant Physiology, **135**: 1198–1205.

Shewry P.R., Halford N.G., Tatham A.S. (1992): High molecular weight subunits of wheat glutenin. Journal of Cereal Science, **15**: 105-120.

Somers D.J., Kirkpatrick R., Moniwa M., Walsh A. (2003): Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome, **49**: 431–437.

Somers D.J., Isaac P., Edwards K. (2004): A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theoretical and Applied Genetics, **109**: 1105–1114.

Song Q.J., Shi J.R., Singh S., Fickus E.W., Costa J.M., Lewis J., Gill B.S., Ward R., Cregan P.B. (2005): Development and mapping of microsatellite (SSR) markers in wheat. Theoretical and Applied Genetics, **110**: 550–560.

Sourdille P., Gandon B., Nicot N., Chiquet V., Zhang L., Specel S., Foisset N., Dufour P., Murigneux A., Bernard M. (2005): Development of microsatellite markers for wheat genetic mapping improvement: outcomes from the wheat genoplante program. In: ITMI Summer Workshop. May 29–31, 2005, Bozeman, Montana.

Taufel A., Bohm H., Flamme W. (1997): Protein inhibitors of alpha-amylase in mature and germinating grain of rye (*Secale cereale*). Journal of Cereal Science, **25**: 267–273.

Yanagisawa T., Kiribuchi-Otobe C., Hirano H., Suzuki Y., Fujita M. (2003): Detection of single nucleotide polymorphism (SNP) controlling the waxy character in wheat by using a derived cleaved amplified polymorphism sequence (dCAPS) markers. Theoretical and Applied Genetics, **107**: 84–88.

Yang W., Bai X., Kabelka E., Eaton C., Kamoun S., Van Der Knaap E., Francis D. (2004): Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. Molecular Breeding, **14**: 21–34.

Zhang W., Gianibelli M.C., Ma W., Rampling L., Gale K.R. (2003): Identification of SNPs and development of allele-specific PCR markers from *γ*-gliadin alleles in *Triticum aestivum.* Theoretical and Applied Genetics, **107**: 130–138.

Zhu Y.L., Song Q.J., Hyten D.L., Van Tassell C.P., Matukumalli L.K., Grimm D.R., Hyatt S.M., Fickus E.W., Young N.D., Cregan P.B. (2003): Single nucleotide polymorphisms in soybean. Genetics, **163**: 1123–1134.

Received for publication February 23, 2009
Accepted after corrections August 5, 2009

Corresponding author:

Dr. Pushpendra K. Gupta, Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, 250 004 Meerut, India
tel.: + 91 121 276 81 95, e-mail: pkgupta36@gmail.com