2. REVIEW OF LITERATURE

The grass family Poaceae (formerly known as Gramineae) represents the world's single most important source of food that includes major crop plants such as wheat, barley, rice, oat, rye, sorghum, maize, etc. One of its tribes, Triticeae is a fairly big group of grasses with 330 to 360 species (Kawahara 2009), which comprises several cereal crops including wheat, barley and rye. After the discovery of a polyploid series in wheat by Sakamura (1918), the genome analysis in wheat and its wild relative was undertaken (Kihara 1954). Through advances in cereal genomics, 5× coverage of wheat genome has been sequenced and its characterization is underway. In the present study, we reviewed wheat genomics concerning the accessible genomic resources and their impact in various molecular studies.

2.1. EST and Genomic Sequences Databases

Expressed sequence tags (ESTs) are short cDNA sequences that serve to tag genes from which messenger RNA (mRNA) originated. Typically, anonymous ESTs are each a single-pass (either from 5' or 3') sequence 200-700 bp long (Adams et al. 1991). EST databases represent a potentially valuable resource for development of molecular markers belonging to the transcribed region of a genome, so that they are likely to be conserved across a broader taxonomic range (Gupta and Rustgi 2004). Also, EST analysis has opened exciting prospects for gene discovery, genome annotation and comparative genomics in all organisms, irrespective of their genome size (Ewing et al. 1999; Fernandes et al. 2002). In plants including wheat, ~28 million ESTs have been generated and deposited in the National Center for Biotechnology Information (NCBI) databases (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). In bread wheat alone, the number of ESTs has swelled from merely 8 in May, 2000 to more than one million as on October 1, 2011. Wheat ESTs available in databases constitute an essential resource for development of molecular marker (Gupta and Rustgi 2004). Using *in silico* approaches, SSRs and SNPs have been identified in the EST sequences leading to the development of EST-SSR and EST-SNP markers (Rafalski 2002a, b; Gupta and Rustgi 2004; Rustgi et al. 2009). The available numbers of ESTs for some important crop plants is listed in Table 1.

Similarly, random genomic sequences available in the genome survey sequence database (dbGSS) of GenBank (http://www.ncbi.nlm.nih.gov) are similar to ESTs in the EST databases

and represent a potential source for marker development (Table 1). At present more than 50,000 GSSs of wheat are available in NCBI database (http://www.ncbi.nlm.nih.gov/nucgss/?term=wheat), in which ~13000 genomic sequences (representing ~14Mb) belong to bread wheat and its wild relatives. Using these sequences, a total of 286 SSRs were recently searched in our own laboratory for development of SSR markers for use in physical mapping studies (unpublished results). In future, sequences in dbGSS will increase in number and will be increasingly used for marker development. A summary of the number of genomic sequences available in the dbGSS database (http://www.ncbi.nlm.nih.gov/dbGSS/dbGSS_summary.html) belonging to genomes of some important plant species is presented in Table 1.

2.2. DNA Based Molecular Markers

2.2.1. Sources and properties of DNA-based molecular markers

DNA-based molecular markers are the most powerful diagnostic tools to detect DNA polymorphism at individual loci and whole genomes. In the initial phase, these DNA-based molecular markers were developed from random genomic sequences using either the polymerase chain reaction (PCR) amplification or genomic DNA libraries. Later, the availability of genomic DNA (gDNA) and cDNA sequences (or ESTs) in the public domain emphasis has shifted towards the development of sequence-based molecular markers.

DNA-based molecular markers have a number of desirable characteristics, which include the following: (i) unlimited in number; (ii) highly polymorphic in nature; (iii) dominant and co-dominant inheritance; (iv) randomly or frequently distributed throughout the genome; (v) simple, quick and inexpensive, (vi) need small amount of DNA; (vii) provide adequate resolution of genetic differences; (viii) not influenced by the environment; (ix) detectable at all stages of plant growth; (x) high reproducibility.

However, except in some cases, none of the molecular marker system would have all the desirable attributes, so that depending on the type of study, a marker system can be selected that would carry the required attributes.

2.2.2. Different classes of molecular markers

DNA-based molecular markers may be classified into hybridization- and PCR-based molecular markers.

Table 1. A summary of the number of ESTs and GSSs belonging to genomes of some important plant species available in NCBI database (updated on October 1, 2011)

Common name (Species name)	Family	ESTs	GSSs
Einkorn wheat (*Triticum monococcum*)	Poaceae	11,190	37
Durum wheat (*T. turgidum*)	Poaceae	19,723	4
Bread wheat (*T. aestivum*)	Poaceae	1,073,845	51,846
Barely (*Hordeum vulgare*)	Poaceae	501,620	2,160
Purple False Brome (*B. distachyon*)	Poaceae	128,092	266,612
Rye (*Secale cereale*)	Poaceae	9,298	2,947
Rice (*Oryza sativa*)	Poaceae	1,252,989	408,438
Maize (*Zea mays*)	Poaceae	2,019,114	2,092,533
Sorghum (*Sorghum bicolor*)	Poaceae	209,828	796,840
Pearl millet (*Pennisetum glaucum*)	Poaceae	2,915	377
Foxtail millet (*Setaria italica*)	Poaceae	2,741	96,975
Ryegrass (*Lolium perenne*)	Poaceae	17,822	303
Oats (*Avena sativa*)	Poaceae	25,344	2,671
Thale cress (*Arabidopsis thaliana*)	Brassicaceae	1,529,700	531,757
Mustard (*Brassica rapa*)	Brassicaceae	44,575	18,833
Potato (*Solanum tuberosum*)	Solanaceae	249,761	141,471
Tomato (*Solanum lycopersicon*)	Solanaceae	297,115	413,489
Onion (*Allium cepa*)	Liliaceae	20,159	10,725
Cotton (*Gossypium hirsutum*)	Malvaceae	274,247	53,582
Soybean (*Glycine max*)	Fabaceae	1,461,624	368,581
Chickpea (*Cicer arietinum*)	Fabaceae	41,984	50,853
Groundnut (*Arachis hypogaea*)	Fabaceae	150,177	3,677
Pea (*Pisum sativum*)	Fabaceae	18,552	204

2.2.2.1. Hybridization-based markers: Hybridization-based markers include the following: (i) Restriction fragment length polymorphisms (RFLPs) are the first generation molecular markers based on the differential hybridization of a labeled cloned DNA (used as a probe) to DNA fragments in a sample of restriction enzyme digested DNAs; each RFLP marker is specific to a single probe/restriction enzyme combination. RFLPs were initially developed and used for human genome mapping and later used in various crop species including wheat for the

preparation of genetic maps, gene tagging and genetic diversity analysis (Chao et al. 1989; Williams et al. 1994; Prasad et al. 2000). RFLP markers have also proved useful in comparative mapping studies because the DNA probes belonging to one species could be readily hybridized to related species (Devos et al. 1993). Low-throughput nature of hybridization, low level of polymorphism, high cost of genotyping, etc. discouraged the frequent use of RFLP markers as a tool for molecular breeding. (ii) Array-based molecular marker systems permit the simultaneous assessment of numerous molecular markers, by which 'target' sequence in solution is interrogated by the arrayed 'probe' in an analogous manner. This system is well suited to the analysis of thousands of markers in hundreds of individuals. Although, Flavell et al. (2003) described a reverse arrangement of 'target' and 'probe', by which amplified PCR products from genomic DNA arrayed on a chip and then interrogated by hybridization with a labeled probe to determine the allelic state of locus; array-based approach is particularly well suited for the analysis of thousands of markers in hundreds of individuals. In this way tens of thousands of genomic samples can be scored for a few markers. This system is particularly useful for screening large populations for important markers (e.g. large plant populations for resistance genes). Array-based marker systems include single feature polymorphism (SFPs), diversity array technology (DArT) and restriction site-associated DNA (RAD) markers which are now available in various crops and facilitated medium to ultra high-throughput genotyping at low cost (Gupta et al. 2008a).

2.2.2.2. PCR-based markers: PCR-based markers include the following: (i) Random amplified polymorphic DNA (RAPD) markers based on the differential PCR amplification of a sample of DNA from short oligonucleotide sequences. In the early years of marker applications, RAPDs were the preferred types of markers for mapping and tagging of genes in crops (Williams et al. 1990). Due to their dominant nature and poor reproducibility and locus specificity, these markers could not become popular in molecular marker technologies. (ii) Amplified fragment length polymorphism (AFLP) markers are generated by selective PCR amplification of the restriction fragments from a total digest of genomic DNA (Vos et al. 1995). In this marker system, specific double stranded DNA adapters are ligated to DNA restriction fragments so that the sequences of adapters and the adjacent restriction sites serve as primer binding sites. The primers are designed to contain the sequences that are complementary to those of adapters and the restriction sites, along with one to three selective bases added at their 3' ends. The use of

selective bases allows amplification of only a subset of the restriction fragments, which still generate a large number of bands facilitating the detection of polymorphisms (Gupta et al. 1999). AFLP marker analysis has been used in several crops including wheat to construct genetic maps to study phylogenetic relationships and for gene tagging (Parker et al. 1999; Bohn et al. 1999). (iii) Microsatellite or simple sequence repeat (SSR) markers are tandemly repetitive short (1-6 bases) sequences that are ubiquitous and widely distributed in all eukaryotic genomes. SSR markers offer several advantages over the other markers because they are co-dominant, reproducible and highly polymorphic (multiallelic). Detailed description of SSR markers are given in the next section. (iv) Single nucleotide polymorphism (SNP) markers are the smallest unit of genetic variation and based on the variation of a specific nucleotide at a given sequence position between individuals that represent the most common type of sequence polymorphism in plant as well as animal genomes. SNPs are usually discovered *in silico* from preexisting datasets of genomic sequences or ESTs (Kota et al. 2001, 2003; Rostoks et al. 2005). SNP markers are biallelic, but most frequent of all markers. In fact, all marker types are based on SNPs (sequences polymorphism) except SSR markers, which are based on length polymorphism (Hayden et al. 2009).

The general characteristics of important molecular marker systems and their applications have been extensively reviewed (Gupta et al. 1999; Prasad et al. 2000; Langridge et al. 2001; Korzun and Ebmeyer 2003; Mohler and Schwarz 2004; Rakoczy-Trojanowska and Bolibok 2004; Roder et al. 2004; Varshney et al. 2005; Khlestkina and Salina 2006b; Rustgi et al. 2009). A comparison of DNA-based molecular markers involving RFLP, RAPD, AFLP, SSR and SNP is presented in Table 2. However, among these molecular markers, only SSR markers (both EST-SSR and gSSR) were used for physical mapping during the present study. Therefore, literature on EST and gSSR markers will be reviewed in some detail.

Table 2. Comparison of commonly used DNA-based molecular markers in crops

Feature	RFLPs	RAPDs	AFLPs	SSRs	SNPs
DNA required (µg)	10	0.02	0.5-1.0	0.05	0.05
DNA quality required	High	High	Moderate	Moderate	High
PCR-based or not	No	Yes	Yes	Yes	Yes
Inheritance	Co-dominant	Dominant	Dominant	Co-dominant	Dominant
Level of polymorphism	High	Medium	High	Very high	High

No. of polymorphic loci analyzed/assay	1-5	1-10	20-100	1-3	1.0
Ease to use	Not easy	Easy	Easy	Easy	Easy
Amenability to automation	Low	Moderate	Moderate	High	High
Reproducibility	High	Unreliable	High	High	High
Development cost	Low	Low	Moderate	High	High
Cost per analysis	High	Low	Moderate	Low	Low

2.2.3. Simple sequence repeats (SSRs): The markers of choice

The first report of SSRs in plants was made by Condit and Hubbel (1991), suggesting their abundance in plant systems. Although, the term microsatellite (simple sequence repeat) was coined by Litt and Lutty (1989) and refers to tandemly repeated DNA sequences of short motifs (e.g. poly CA, poly CT, poly AT and other repeated sequences of 1-6 bases). Only with the advent of microsatellite or simple sequence repeat (SSR) markers, the use of molecular markers gained momentum during the last 15 years and thus SSRs were recognized as highly informative and locus-specific markers. Initially, Akkaya et al. (1992) reported length polymorphisms of SSRs in soybean, which opened a new source of PCR-based molecular markers for other plant genomes also. Later, SSR markers were described in hexaploid wheat in 1995/1997 (Devos et al. 1995; Roder et al. 1995; Bryan et al. 1997). They are present in both coding and non-coding regions and are usually characterized by a high degree of length variation (Gupta and Varshney 2000; Zane et al. 2002). The variability of microsatellite sequences in genome is not based on point mutations but on the variation in the number of the simple sequence repeat motifs. Such variation occurs approximately 10 times more frequently, and originates through processes such as slippage during replication or unequal crossing over (Hancock 1999).

The flanking sequences of SSRs are often unique, allowing primers to be designed, so that each SSR marker represents a single locus. Moreover, high information content (estimated as either the polymorphic information content or as genetic diversity index or as expected heterozygosity), locus specificity and co-dominant nature make SSRs the most suitable polymorphic marker system in plant species. In wheat, it has been shown that SSRs show a much higher level of polymorphism and informativeness than any other marker system (Plaschke et al. 1995; Roder et al. 1995; Ma et al. 1996; Bryan et al. 1997; Korzun et al. 1997a,

b; Gupta et al. 2002a). For instance, in wheat ~3000 SSR markers have been derived from genomic clones generated through a variety of enrichment procedures (Roder et al. 1998a, b; Pestsova et al. 2000; Song et al. 2002; Somers et al. 2004; Nitta and Nasuda 2008). Later, a large number of wheat SSR markers were also identified within EST sequences (mRNA transcripts contain repeated motifs) through bioinformatics data mining (Gupta et al. 2003; Gao et al. 2004; Peng and Lapitan 2005; Goyal et al. 2005; Mohan et al. 2007).

Based on their origin, SSRs can be classified into two classes, (i) EST-SSRs, which are derived from EST sequences representing the expressed portion of the genome, and (ii) genomic SSRs (gSSRs), which are developed from genomic DNA sequences,

2.2.3.1. Expressed sequence tag-simple sequence repeats (EST-SSRs)

Development and use of expressed sequence tags (ESTs) provided new opportunities for gene discovery, genome annotation and comparative genomics in all organisms, irrespective of their genome size (Adams et al. 1991; Ewing et al. 1999; Fernandes et al. 2002). ESTs are obtained by partial sequencing of random cDNA clones. Once generated, they are useful in cloning specific genes of interest and synteny mapping of functional genes in various related organisms. With the availability of programs like *MISA* (MIcroSAtellites), written in Perl 5 script and simple sequence repeat information tool (SSRIT) available at Gramene website, it is possible to search for SSRs in EST sequences. SSRs derived from ESTs (EST-SSRs) can be used for designing locus specific primers to amplify SSR loci located within the genes. They are sorted *in silico* with ease, unbiased in repeat type, present in genic regions of the genome, and are more likely to produce amplicon from the coding region of the genome than those designed from non-coding sequences (Yu et al. 2004; Zhang et al. 2005; Parida et al. 2006). However, the generation of EST-SSRs is obviously limited to those species for which sufficient numbers of ESTs are available in the database.

EST-SSR markers have been reported in several plant species including grape, wheat, *Arabidopsis*, soybean, rice, maize, barley, cotton, sorghum (Scott et al. 2000; Scotti et al. 2000; Pillen et al. 2000; Eujayl et al. 2002; Holton et al. 2002; Kantety et al. 2002; Morgante et al. 2002; Saha et al. 2004; Han et al. 2006), etc. The EST-SSRs are particularly useful due to their physical association with coding regions of the genome, which can enhance the role of molecular markers in mapping agronomically important loci. The frequencies of EST-SSRs vary from species to species. Cardle et al. (2000) reported frequencies of EST-SSRs in rice (1 in

3.4 kb), maize (1 in 8.1 kb), soybean (1 in 7.4 kb), tomato (1 in 11.1 kb), poplar (1 in 14.0 kb), *Arabidopsis* (1 in 13.8 kb) and cotton (1 in 20.0 kb). In bread wheat, the frequency of EST-SSRs ranged from 1 in 9.20 kb to 1 in 17.42 kb (Gupta et al. 2003; Gao et al. 2003; Parida et al. 2006). The difference in frequency of EST-SSRs in a species may also depend on the criteria used to identify SSRs in the sequence database. Among different repeat types, trinucleotides have been most abundant in most of the crop species including bread wheat (Morgante et al. 2002; Holton et al. 2002; Gao et al. 2003) with GC rich trinucleotide repeat motifs as the most abundant (Gao et al. 2003; Gupta et al. 2003; Gupta and Rustgi 2004).

During the last decade, EST-SSRs have been used for the study of genetic diversity, transferability and mapping in a number of plant species, which include wheat, rice, maize, oats, barley, sorghum, rye and *Arabidopsis* (Cho et al. 2000; Cardle et al. 2000; Hackauf and Wehling 2002; Holton et al. 2002; Theil et al. 2003; Gao et al. 2004; Nicot et al. 2004; Bandopadhyay et al. 2004; Yu et al. 2004; Zhang et al. 2005; Peng and Lapitan, 2005; Torada et al. 2006; Mohan et al. 2007). Particularly in wheat, EST-SSR based genetic maps (Gao et al. 2004; Nicot et al. 2004; Yu et al. 2004; Xue et al. 2008) and physical maps (Yu et al. 2004; Qi et al. 2004; Goyal et al. 2005; Peng and Lapitan 2005; Mohan et al. 2007) have been developed recently. EST-SSRs were also used to study the impact of modern plant breeding on the allelic diversity in wheat suggesting that allelic diversity has reduced in the transcribed portion of wheat (Fu et al. 2005). Due to their origin from coding regions, EST-SSRs exhibit higher level of transferability (relative to gSSRs) across closely related genera (Holton et al. 2002). This high transferability of EST-SSRs across species implies their considerable potential for comparative mapping (Scott et al. 2000; Eujayl et al. 2002; Bandopadhyay et al. 2004). In view of this, EST-SSRs were used for comparative mapping in rice and wheat (Yu et al. 2005). Mapping of EST-SSRs also provide map locations for genes of known functions and important agronomic traits (Holton et al. 2002). These EST-SSRs served as a valuable source for a variety of studies including comparative mapping, evolutionary studies, genetic diversity estimates, gene mapping and tagging, marker-assisted selection (MAS), positional cloning of genes, etc. However, EST-SSR markers have been shown to be less polymorphic than the SSRs derived from non-coding part of genome e.g. genomic SSRs (Eujayl et al. 2001).

2.2.3.2. Genomic SSRs (gSSRs)

The genome survey sequence (GSS) division of GenBank contains nucleotide sequences that are genomic in origin. These genomic sequences may be: (i) random "single pass read" genome survey sequences; (ii) cosmid/BAC/YAC end sequences; (iii) exon trapped genomic sequences; (iv) Alu PCR sequences and (v) transposon-tagged sequences. The methodology/tools for the identification of SSRs in above genomic sequence are same as in EST sequences. Edwards et al. (1996) successfully developed SSRs-enriched genomic libraries by hybridizing genomic DNA fragments with membrane bound SSRs in many plant species including wheat. During SSRs development from the genomic libraries, clone sequences often contain SSRs close to one of the two ends, such that normal flanking primers could not be designed for the SSR. For utilizing such SSR-containing sequences, anchored primers have been designed in the past (Varghese et al. 2000; Singh et al. 2006).

The gSSRs have been developed in all important cereal species including wheat (Roder et al. 1995; Edwards et al. 1996; Bryan et al. 1997; Song et al. 2005; Torada et al. 2006, Goyal 2007); rice (Wu and Tanksley 1993; McCouch et al. 1997; 2002); barley (Saghai-Maroof 1994; Liu et al. 1996; Ramsay et al. 2000) and maize (Senior et al. 1996; Sharapova et al. 2002). Like most other cereal species, in bread wheat also, gSSRs were exploited for three main investigations: (i) gene/QTL mapping, (ii) genetic/physical mapping, and (iii) diagnostic and analysis of genetic diversity. Due to their locus-specificity, gSSR markers have been used for QTL/gene tagging and mapping for important agronomic traits like pre-harvest sprouting tolerance (Roy et al. 1999; Kulwal et al. 2004, Kulwal et al. 2005a, b; Kumar et al. 2009), grain protein content (Prasad et al. 1999, 2003), grain weight (Varshney et al. 2000; Eloufi and Nachit 2004; Kumar et al. 2006), plant height (Huang et al. 2003), vernalization sensitivity (Kato et al. 2003), fungal resistance genes/QTLs (Ericksen et al. 2003; Suenaga et al. 2003; Singh et al. 2004; Vikal et al. 2004), and insect resistance genes/QTLs (Arzani et al. 2003; Malik et al. 2003). In addition to the above, the locus specific gSSRs have been used for the preparation of genetic maps (Roder et al. 1998; Pestosova et al. 2000; Gupta et al. 2002; Somers et al. 2004; Song et al. 2005; Torada et al. 2006) and physical maps (Sourdille et al. 2004; Goyal et al. 2005).

The gSSRs have also been used for the evaluation of germplasm collections to identify potential duplicates and genetic relationship among the accessions (Huang et al. 2002). Several studies also reported the use of gSSRs to identify the effects of modern plant breeding on the genetic diversity by comparing old and new varieties (Manifesto et al. 2001; Christiansen et al.

2002; Roussel et al. 2004; Reif et al. 2005). The numerous advantages of the gSSRs, including their abundance and distribution throughout the genome, polyallelic nature, co-dominant inheritance, high reproducibility and locus specificity, are all well documented (Morgante and Olivieri 1993; Powell et al. 1996). However, most of the gSSRs neither have a genic function nor have close linkage to coding regions (Metzgar et al. 2000); they are very time- and cost-expensive to develop and show only a limited transferability to related species (Sourdille et al. 2001; Guyomarch et al. 2002).

2.2.4. Uses of molecular markers

Molecular markers have been extensively used for germplasm characterization, genetic/physical mapping, diversity analysis, gene tagging, genome-wide QTL mapping, association analysis, marker-assisted selection and map based cloning.

2.2.4.1. Genetic mapping

A major use of molecular markers is the construction of genetic maps and their use in studying marker-trait associations by analyzing co-segregation of markers and phenotypes or traits in defined populations. These populations come from a variety of sources - F_2 populations resulting from narrow or wide crosses; single seed descent derived recombinant inbred (RI) populations, or doubled haploid (DH) populations. The last two types of populations are immortal and have the advantage that they can be permanently maintained.

During the last two decades, using a variety of molecular markers, high-density molecular linkage maps have been constructed for all major crops. In wheat also, molecular markers mainly including RFLP and SSR were extensively used for genetic mapping. A detailed account on mapping of chromosomes of individual homoeologous group and that of whole-wheat genome is available elsewhere (Gupta et al. 1999; Gupta et al. 2002a; Torada et al. 2006; Peleg et al. 2008; Xue et al. 2008); an updated version is also available at GrainGenes (http://wheat.pw.usda.gov/gpages/maps.shtml). Details of the types and numbers of markers mapped on individual chromosomes or whole genome in wheat are summarized in Table 3.

2.2.4.2. Physical mapping

Physical maps of genomes are based on physical distances between genes or molecular markers or between a gene and a marker either in terms of base pairs (i. e. kb) or in terms of physical length (proportion of an arm or whole chromosome) of chromosome segment. Physical mapping

is important for localization and isolation of genes (e.g., positional cloning), study of genome organization and for clone-to-clone hierarchical genome sequencing.

Methods for constructing physical maps can be classified into two basic categories. (i) Cytogenetically based, and (ii) molecular based. Cytogenetically based methods include use of chromosome aberrations for physical mapping (for example pseudodominance and aneuploids), *in situ* hybridization (Jiang and Gill 1994) and fluorescence *in situ* hybridization (Jiang and Gill 2006). Molecular methods include the construction of contigs using libraries of genomic fragments, and long-range restriction mapping using rare cutting enzymes (Cheung et al. 1991). These molecular methods are useful for fine mapping of small areas of the genome. Currently strategies are being developed for the preparation of physical maps involving BAC contigs for sequencing of

Table 3. Summary of the types and number of marker loci genetically mapped in bread wheat

Homoeologous group /whole genome	Marker type	Number of loci mapped	Reference
Group 1	RFLP	98	Van Deynze et al. (1995)
Group 2	RFLP	287	Nelson et al. (1995b)
Group 3	RFLP	220	Nelson et al. (1995c)
Group 4	RFLP	98	Nelson et al. (1995a)
Group 5	RFLP	168	Xie et al. (1993); Nelson et al. (1995a)
Group 6	RFLP	216	Marino et al (1996)
Group 7	RFLP	109	Nelson et al. (1995a)
Whole genome	RFLP	197	Liu and Tsunewaki 1991
Whole genome	RFLP	264	Cadalen et al. 1997
Whole genome	SSR	279	Roder et al. (1998a)
Whole genome	SSR	53	Stephenson et al. (1998)
Whole genome	SSR	79	Korzun et al. (1999)
Whole genome	SSR	55	Pestsova et al. (2000)
Whole genome	SSR	172	Harker et al. (2001)
Whole genome	SSR	66	Gupta et al. (2002a)
Whole genome	SSR	1,235	Somers et al. (2004)
Whole genome	SSR	222	Song et al (2005)
Whole genome	SSR	464	Taroda et al. (2006)

Whole genome	AFLP	426	Penner et al. (1998)
Whole genome	AFLP	140	Hazen et al. (2002)
Whole genome	RFLP, SSR	230	Messmer et al. (1999)
Whole genome	RFLP, SSR, AFLP	355	Chalmers et al. (2001)
Whole genome	RFLP, SSR, AFLP	659	Sourdille et al. (2003)
Whole genome	RFLP, SSR	188, 208	Paillard et al. (2003)
Whole genome	RFLP, AFLP, SSR	123, 194, 242	Quarrie et al. (2005)
Whole genome	SSR, TRAP, STS	646	Liu et al. (2005)
Whole genome	DArT, AFLP, SSR	189, 165, 270	Semagn et al. (2006)
Whole genome	ISSR, SRAP, TRAP	381	Li et al. (2007)
Whole genome	SSR, DArT	690	Peleg et al. (2008)
Whole genome	EST-SSR	887	Xue et al. (2008)

whole genome of bread wheat [Devos et al. 2005; (http://www.international.inra.fr)]. In plants physical mapping have been used for location of genes or markers with the help of chromosomes deficiencies (deletions) since 1931. Different workers used different techniques for physical localization of important genes or markers (see next section).

(A) Marker-based physical maps

(a) Physical mapping using pseudodominance: Recessive marker stocks may be crossed as female with irradiated pollen from a male carrying the dominant allele or alternatively, the F_1 heterozygous seeds may be irradiated. If the male gamete carries a deficiency or a deficiency is induced in seed, the recessive allele of the female will express itself. The position of deficiencies can be identified through the study of pachytene chromosomes in the heterozygotes. This phenomenon of pseudodominance followed by pachytene analysis was used as a criterion for physical mapping in corn (McClintok 1931, 1933, 1934; Stadler 1933, 1935; Singleton 1939) and in tomato (Rick and Khush 1966; Khush and Rick 1967, 1968).

(b) Physical mapping using acrotrisomics and metatrisomics: Acrotrisomics, a cytgentical variant with an acrocentric chromosome in addition to normal chromosome complement is one of the efficient tools for chromosome mapping. Similarly in metatrisomics, meta 3A, meta 3B and meta 6A meant trisomics for these chromosomes with terminal deficiencies on both the ends. Acrotrisomics and metatrisomics have been utilized in association with Giemsa banding technique, for physical mapping of genes in barley. T. Tsuchiya and his

coworkers obtained six such acrotrisomics in barley. These acrotrisomics had terminal deficiencies that can be identified by Giemsa banding. Tsuchiya (1991) described results of genetic analysis of 28 genes located on six chromosomes using acrotrisomics and metatrisomics.

(c) Physical mapping using translocation break-points (TBs): Break point analysis has been used in maize (Harper and Cande 2000), wheat (Gill et al 1996a), rye (Alonso-Blanco et al. 1994) and rice (Reyes et al. 1998; Singh et al. 1996) to generate cytogenetic maps. TBs have also been used for mapping of more than 300 genetically mapped RFLPs in barley (Kunzel et al. 2000). As many as 240 TBs were integrated as physical landmarks into linkage maps of the seven barley chromosomes. A total of 120 reciprocal translocations were used in this study. PCR analyses were performed until the two markers most closely flanking the TBs were identified. Then, two to three markers on either side of TBs were checked by PCR in addition to the TB-flanking markers to ensure that the TBs were correctly localized onto the genetic maps.

(d) Physical mapping using *in-situ* hybridization (ISH): Physical maps can also be generated through *in-situ* hybridization (ISH), where chromosome sites that are homologous to a known radioactively labeled DNA probes can be directly visualized under the microscope (Pardue and Gall 1969). The technique initially proved useful for DNA probes that were at least a few kilobases in length (Maluszynska 2002). Although the original technique was highly sensitive, it was time consuming and cumbersome. These limitations were overcome by the development of non-radioactive ISH technique (Langer-Safer et al. 1982). Rayburn and Gill (1985) were the first to use non-radioactive ISH in plants.

In recent years, use of fluorochromes for signal detection in ISH became common and technique was described as FISH. Using this approach one could label more than one DNA probes with different haptens for simultaneous detection of several DNA sequences leading to their physical mapping (Lawrence et al. 1990; Lichter et al. 1990). Highly repeated DNA sequences (pSc119.2 and pAs1) and a low-copy 3BS-specific RFLP sequence (PSR907) enabled determination of the physical positions of the breakpoints (BPs) along 3BS and 3DS arms of bread wheat using FISH (Biagetti et al. 1999). FISH was also used for physical mapping of large DNA segments that were available as BACs. This approach, called as BAC-FISH, has been an effective approach for physical mapping of specific DNA sequences and identifying individual chromosomes in humans and in plants with small genomes such as rice (Jiang et al.

1995; Cheng et al. 2002), cotton (Hanson et al. 1995), sorghum (Gomez et al. 1997; Kim et al. 2002), *Arabidopsis thaliana* (Fransz et al. 2000), and potato (Dong et al. 2000). In wheat, 56 RFLP specific BAC clones from libraries of *Aegilops tauschii* and *Triticum monococcum* were localized to centromeric, telomeric, or several interstitial regions of wheat chromosomes (Zhang et al. 2004). Eleven (11) BAC clones were mapped physically using FISH in a study in tomato (Wang et al. 2006). An advanced technique, called fiber-FISH, has also been used for physical mapping. In this technique fluorescent probes are hybridized onto extended chromatin DNA fibers. The fiber technique proved highly versatile in permitting fine physical mapping of segments carried on YACs, cosmids, lambda and plasmid vectors, thus permitting direct mapping of DNA fragments ranging from a few kilobases (Florijin et al. 1995) to several hundred kilobases (Haaf and Ward 1994). DNA fiber fish was successfully utilized for physical mapping in rice (Dong et al. 1998; Cheng et al. 2002), *Arabidopsis* (Jackson et al. 1998) and potato (Tak et al. 2005). In wheat 13 transgenic wheat plants were analyzed using fiber-FISH to determine more accurate physical structure of the transgene loci (Jackson et al. 2001; Svitashev and Somers 2002). High resolution physical mapping by extended DNA fiber-FISH has also been successfully achieved in *Arabidopsis* and tomato (Fransz et al. 1996).

(e) Physical maps based on deletion stocks: (I) *Wet-lab approach*: Endo (1988) introduced a method for systematic production of terminal chromosomal deletions in wheat chromosomes. The method involved the use of gametocidal system; which had the following features: (i) Deletions occur in gametes lacking the alien chromosome and are thus recovered in pure wheat background. (ii) Most deletions arise from single breakages with concomitant loss of distal segments as indicated from C-banding analysis. (iii) Deletions are stable from generation to generation and broken chromosomes are healed by the addition of telomeric sequences at the physical ends (Werner et al. 1992). (iv) Breakages along a chromosome occur regardless of a euchromatin or heterochromatin region. Using this approach 430 deletion lines involving all 21 wheat chromosomes with well ordered break points (BPs) were produced (Endo and Gill 1996). These lines facilitated preparation of physical maps for all the 21 chromosomes in wheat (Werner et al. 1992; Gill et al. 1993a, b, 1996a, b; Hohmann et al. 1994; Delaney et al. 1995a, b; Mickelson-Young et al. 1995; Qi et al. 2003, 2004; Sourdille et al. 2004; Song et al. 2005, Goyal et al. 2005; Mohan et al. 2007).

A summary of wheat physical maps prepared using wheat deletion lines involving different types of molecular markers such as RFLPs, EST-SSRs/gSSRs, RAPDs, ESTs, etc. is given in Table 4. Besides the above physical maps involving a variety of molecular markers, wheat deletion stocks have also been used for construction of physical maps involving ESTs. In a major exercise, a consortium of 13 laboratories in USA funded by National Science Foundation (NSF), USA, assigned 16,000 unique ESTs to physical locations on chromosomes through Southern hybridization by using these deletion lines (http://wheat.pw.usda.gov/NSF/progressmapping.html; Qi et al. 2004). Another major project on physical mapping using deletion stocks was undertaken in France (GENOPLANTE) to establish the relationship between genetic and physical distances in wheat, where small genetic distances over large physical regions around the centromeres and large genetic to physical map ratios close to the telomeres were described (Sourdille et al. 2004). Later, similar results were obtained in an integrated physical map of SSRs being constructed in our own laboratory (unpublished data).

(II) *In silico approach*: *In silico* physical mapping of SSR-containing wheat sequences to individual bins was also carried out recently. For this purpose, Parida et al. (2006) identified 1,671 unigene-derived microsatellites (UGMS) in 24,854 wheat unigene sequences that were retrieved from GenBank using FASTA sequence retrieval system. *In silico* physical mapping of SSR containing sequences were then carried out by BLASTN (Basic Local Alignment Search Tool) against GrainGenes databases (http://www.wheat.pw.usda.gov/GG2/blast.html) of wheat EST sequences that were bin mapped earlier using deletion stocks. Out of a total of 1,671

in wheat. These UGMS have each a unique identity and position in wheat genome.

(f) Radiation hybrid mapping (RHM): Radiation hybrid mapping technique was developed by Goss and Harris (1975), who used X-rays-induced chromosome rearrangements to map genes on human chromosome X. In this approach, tissue culture cells from a donor species are treated with a lethal dose of irradiation to induce random chromosomal breaks. These broken chromosomal fragments are then rescued by fusion with a suitable recipient cell line from a different species. Derived individual cell lines which contains, each a fraction of the donor genome retained as a collection of large genomic fragments integrated into the recipient genome, are termed radiation hybrid lines. A set of approximately 100 radiation hybrid lines

representing the entire donor genome is produced to make a mapping panel or radiation hybrid (RH) panel. In plant, RH panels were constructed first in maize by irradiating maize chromosome 9 in an oat background (using oat–maize addition lines) with 30, 40, and 50-krad gamma-rays and subsequent characterization of these lines with maize-specific molecular markers (Riera-

Table 4. A summary of the available deletion based physical maps of bread wheat

Homoeologous group/ chromosome/ arm/ whole genome	Marker loci mapped	No. of deletion lines used	Reference
Group 1chromosomes	19 RFLP	18	Kota et al. (1993)
Group 1chromosomes	50 RFLPs	56	Gill et al. (1996a)
Group 2 chromosomes	30 RFLPs	21	Delaney et al. (1995a)
Group 2 chromosomes	43 SSRs	25	Roder et al. (1998b)
Group 3 chromosomes	29 RFLPs	25	Delaney et al. (1995b)
Group 4 chromosomes	40 RFLPs	39	Mickelson-Young et al. (1995)
Group 5 chromosomes	155 RFLPs	65	Gill et al. (1996b)
Group 5 chromosomes	245 RFLPs, 3 SSRs	36	Faris et al. (2000)
Group 5 short arm	100 RFLPs	17	Qi and Gill (2001)
Chromosome 5A	22 RFLPs	19	Ogihara et al. (1994)
Group 6 chromosomes	24 RFLPs	26	Gill et al. (1993a)
Group 6 chromosomes	210 RFLPs	45	Weng et al. (2000)
Group 6 short arm	82 RFLPs	14	Weng and Lazar (2002)
Group 7 chromosomes	16 RFLPs	41	Werner et al. (1992)
Group 7 chromosomes	91 RFLPs, 6 RAPDs	54	Hohmann et al. (1995b)
Chromosomes 6B, 2D and 7D	16 SSRs	13	Varshney et al. (2001)
Chromosome arm 1BS	24 AFLPs	8	Zhang et al. (2000)
Chromosome arm 4DL	61 AFLPs, 2 SSRs, 2 RFLPs	8	Milla and Gustafson (2001)
Chromosome arm 1BS	22 ESTs	2	Sandhu and Gill (2002a)
Whole genome	305 SSRs	106	Cadalen et al. (2002)
Whole genome	16,000 ESTs	101	Qi et al. (2004)
Whole genome	725 SSRs	118	Sourdille et al. (2004)

Whole genome	266 eSSR	105	Peng and Lapitan et al. (2005)
Whole genome	260 BARC loci	117	Song et al. (2005)
Whole genome	313 SSRs	164	Goyal et al. (2005)
Whole genome	210 EST-SSRs	101	Mohan et al. (2007)
Whole genome	439 SSRs	192	(Unpublished data)

Lizarazu et al. 1996; 2000). Subsequently, the maize chromosome 1 RH panel was developed and used to map 45 simple sequence repeat (SSR) markers (Kynast et al. 2004). Similar efforts have also been initiated in some cereals. For instance, transgenic barley protoplasts harboring the *bar* transgene as a selectable marker was fused with tobacco protoplasts to produce radiation hybrid panels (Wardrop et al. 2002, 2004).

Radiation hybrids have also been used for physical mapping in wheat. As many as 100 radiation hybrid lines were characterized and 57 radiation-induced chromosome breaks were confirmed through 23 markers in order to facilitate fine mapping of markers to chromosome 1D. The RH population and molecular markers for 1D were used to localize the gene scs^{ae} (species cytoplasm specific), which confers fertility restoration capability between the cytoplasm of *Ae. longissimum* and the nucleus of durum wheat (Hossain et al. 2004a, b). Later, a high-resolution radiation hybrid map of 378 molecular markers (including RFLPs, SSRs, AFLPs, and ESTs) provided the resolution of ~199 kb, in wheat chromosome 1D (Kalavacharla et al. 2006). This work was extended to chromosome 3B in a collaborative effort to develop a complete physical map of this chromosome. Three RH panels of varying resolution (15, 25 and 35 Krad) were generated and then used for mapping of ISBP (Insertion Site-Based Polymorphism) markers to determine physical resolution and utility of this approach in BAC contig alignment. Results to date indicate marker retention frequencies of about 74% involving many critical lines showing numerous breaks in chromosome 3B (Michalak et al. 2008, Paux et al. 2008). A summary of radiation hybrid mapping reports in crop plants is given in Table 5.

(B) BAC-based physical maps for whole genome sequencing

The availability of genome wide DNA-contigs has been a prerequisite for high quality sequencing of the model genomes of *Arabidopsis* and rice (The *Arabidopsis* Genome Initiative 2000; Sasaki and Burr 2000). DNA contigs can be assembled using large insert DNA clones such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs).

These contigs are then assigned to chromosomes by using major landmarks from existing maps. The term 'physical map' is often also used for such contig maps. Physical maps generated through contig assembly are then used to find the minimum tiling path for sequencing. However, contig assembly is dependent on a high density genetic map and/or high-quality large insert DNA libraries. The BAC based whole-genome physical maps have been constructed for *Caenorhabditis elegans* (Hodgkin et al. 1995), *A. thaliana* (Mozo et al. 1999; Chang et al. 2001), and soybean (Wu et al. 2004). In cereals crops like rice (Tao et al. 2001; Chen et al. 2002), maize (Zhou et al. 2009) and grass like brachypodium (Gu et al. 2009). BAC-based physical maps have been constructed to facilitate analysis of genome structure, comparative genomics, and assembly of the entire genome sequence. To facilitate complete genome sequencing in bread wheat an International Wheat Genome Sequencing Consortium (IWGSC) was established in 2005, which now involves more than 15 countries.

Table 5. A summary of radiation hybrid mapping studies in plants

Crop	Lines used for mapping	Chromosome used for mapping	Source of radiation	Dose of radiation	Reference
Wheat	Durum wheat carrying chromosome 1D of hexaploid wheat	1D	rays	15, 25 and 35 Krad	Kalavacharla et al. 2006; Hossain et al. 2004
	LDN – 3D (3B); Langdon D-genome substitution line	3B	□ rays	35 Krad	Michalak et al. 2008; Paux et al. 2008
	Bupleurum scorzonerifolium/ Triticum aestivum asymmetric somatic hybrids	WG-RH	UV rays	380 $\mu W/cm^2$	Zhou et al. 2006
Maize	Oat-maize addition line	9	□ rays	30, 40 and 50 Krad	Lizarazu et al. 2000
Barley	Barley genome fragments into tobacco protoplast	Whole Genome	□ rays	5 Krad	Wardrop et al. 2002
Cotton	Interspecific hybrids of tetraploid cotton (*Gossypium barbadense* × *Gossypium hirsutum*)	WG-RH	□ rays	1.5 and 5 Krad	Gao et al. 2004

WG-RH: Whole genome radiation hybrid, Krad: Kilorad

2.2.5. Use of genetic/physical maps

The molecular genetic and physical maps are used for the study of colinearity of genomic DNA sequences among related species, for localization and isolation of genes (e.g. positional cloning), for the study of genome organization and constructing a framework for systematic whole genome sequencing, as detailed later in this chapter.

2.2.5.1. Analysis of synteny and colinearity

Relationship of gene arrangements among taxa along with chromosomal allocation, also known as synteny and colinearity, has become valuable frameworks for inference of shared ancestry of genes and for transfer of knowledge from a species to another related species (Tang et al. 2008). Particularly, many comparative studies have been carried out in order to characterize the genes that determine morphological and functional similarities of the grasses. It can also improve our understanding of the evolutionary mechanisms that have led to the current structure of grass genomes. The idea behind these studies is that the knowledge gained in one species can be easily transferred to other grass species (Paterson et al. 1995; Bennetzen 2000), which will be of great interest for genes controlling key agronomic traits.

Among cereals using molecular maps, colinearity was first reported among A, B and D sub-genomes of wheat (Chao et al. 1989; Devos et al. 1992), and later between the high-density genic regions of wheat and barley. Comparison of rice, wheat and barley have led to the identification of colinear regions and established the genetic correspondence of the seven homoeologous groups of the Triticeae genomes with the 12 rice chromosomes (Ahn et al. 1993; Kurata et al. 1994; Vandeynze et al. 1995; Sherman et al. 1995; Devos and Gale 1997; Sorrels et al. 2003). A consensus map of 12 grass genomes including wheat represents chromosome segments of each genome relative to those in rice on the basis of mapping of anchor DNA markers (Devos and Gale 2000). Some of the immediate applications of comparative genomics in wheat include a study of evolution (Wicker et al. 2007) and that involving isolation of genes using a small model genome such as rice genome in order to predict and isolate genes in a larger and more complex genomes like that of wheat. For instance, detailed characterization of wheat chromosome pairing gene, *Ph1,* was possible through an analysis of syntenous region in rice (Foote et al. 1997; Roberts et al. 1999). Synteny and colinearity based comparative mapping among rice, wheat and maize indicated that genome colinearity is observed even in species

belonging to different subfamilies of Poaceae (Hulbert et al. 1990; Moore et al. 1995b; Devos and Gale 1997).

Low molecular weight (LMW) and high molecular weight (HMW) glutenin loci as well as the *Lr10* locus of diploid wheat, when compared with their orthologs from tetraploid and hexaploid wheat were found to be largely conserved except some changes that took place in intergenic regions (Wicker et al. 2003; Isidore et al. 2005). Synteny analysis of genes for glume coloration and pubescence in hexaploid wheat revealed that the major gene *Bg* (for black glume color), *Rg1* and *Rg3* (for red glume), and a locus for smokey-grey colored glume represent a set of homoeoloci, which are designated as *Rg-A1*, *Rg-B1*, and *Rg-D1* (Khlestkina et al. 2006a). Microcolinearity has also been examined using small genomic regions associated with specific genes that have been fully sequenced. For instance, Chantret et al. (2004) sequenced a 101 kb BAC clone encompassing *Hardness* locus from *T. monococcum* to study microcolinearity between diploid wheat and rice.

More recently, the grass genus brachypodium (*Brachypodium distachyon*) is emerging as a better model system for wheat because of a more recent divergence of these two genera (35-40 million years) relative to wheat-rice divergence (Draper et al. 2001; Hasterok et al. 2006; Vogel et al. 2006a). Also, the available sequence of brachypodium may help further detailed analyses of colinearity and synteny among grass genomes. This has already been demonstrated through a comparison of 371 kb sequence of *Brachypoium sylvaticum* with orthologous regions from rice and wheat (Bossolini et al. 2007). In this region, brachypodium and wheat showed perfect macro-colinearity, but rice was shown to contain a ~220 kb inversion relative to brachypodium sequence. Reports based on internal transcribed spacer (ITS) and 5.8 s rDNA sequence (Hsaio et al. 1994), genomic RFLP and RAPD markers (Catalan et al. 1995), and ITS sequence plus the chloroplast *ndhF* gene (Catalan and Olmstead 2000), all placed brachypodium between rice and a clade containing temperate grains like wheat, barley and rye. Usefulness of brachypodium for comparative genomics and map-based cloning of genes in wheat has also been shown through isolation and characterization of *Ph1* locus for chromosome pairing in polyploid wheat (Griffiths et al. 2006). In the *Ph1* region, more orthologous genes were identified between the related species *B. sylvaticum* and wheat than between wheat and rice, thus once again demonstrating relative utility of brachypodium genome as a better model than rice genome for wheat comparative genomics (Huo et al 2006; Kumar S et al. 2009).

2.2.5.2. Identification of gene-rich and gene-poor regions

Genetic and physical maps of the wheat genome described above have been utilized for a study of gene distribution in this genome (Gill et al. 1996a, b; Sandhu et al. 2001; Qi et al. 2003). In wheat, gene distribution was resolved by physical mapping of gene markers with the help of single-break deletion lines (Endo and Gill, 1996; Gill et al. 1996a, b; Sandhu et al. 2001). High-density consensus physical maps containing ~3000 gene markers mapped on 352 deletion lines revealed that about 30% of the wheat genome contains >90% of the genes (Erayman et al. 2004) that are present in clusters spanning small regions. These are described as gene-rich regions (GRRs). A total of 48 GRRs were identified with an average of about seven GRRs per homoeologous group of chromosomes. Of these, 20 were major GRRs spanning 13% of the genome but accounting for ~66% of the genes. The GRRs are interspersed by gene-poor regions (GPRs) predominantly consisting of retrotransposon-like repetitive DNA sequences (SanMigul et al. 1996; Barakat et al. 1997, 1998; Feuillt and Keller 1999) and pseudogenes. The GRRs vary in size and have a gene-density with a general trend of increased gene-density towards the distal parts of the chromosome arms. About 60% of the genes are present in the distal one-third of the chromosomes (Gill et al. 1993a; Sandhu and Gill 2002a, b; Akhunov et al. 2003; Erayman et al. 2004).

2.2.5.3. Gene tagging

A direct application of molecular maps has been in tagging genes of economic importance with molecular markers. This is achieved by a variety of approaches including bulked segregant analysis (BSA) followed by genotyping and regression analysis (Michelmore et al. 1991) or by QTL analysis. In both cases, robust, reliable phenotypic data has a strong bearing on the success of the efforts to tag a gene controlling a component of a trait. BSA is very effective for tagging a qualitative trait, which has Mendelian inheritance, and where a PCR-based marker system such as RAPD, SSR and AFLP is used. QTL analysis is, however, more appropriate, although not always necessary, where the traits are quantitative in nature and apparently do not exhibit Mendelian inheritance. In this case, full genetic maps are needed to scan the genome for associations of specific bins with the trait. A number of gene tagging studies have been conducted in bread wheat for more than a dozen traits, including grain protein content (Singh et al. 2001; Prasad et al. 1999, 2003), pre-harvest sprouting tolerance (Roy et al. 1999), grain weight (Varshney et al. 2000; Kumar et al. 2006), resistance against various diseases, including

powdery mildew (Hartl et al. 1999; Huang et al. 2000), and stripe rust (Bariana et al. 2002) and resistance against various insects/pests including hessian fly (Ma et al. 1994; Willimas et al. 1994; Feuillet et al. 1995; Dweikat et al. 1997; Nelson et al. 1997; Gupta and Varshney 2004; Varshney et al. 2006; for reviews, see Gupta et al. 1999, 2008b).

2.2.5.4 Map-based cloning

In the last few years, map-based cloning was facilitated by the extensive use of markers exhibiting synteny and colinearity among the grass genomes (for reviews, see Peters et al. 2003; Paran and Zamir 2003; Sasaki and Antonio 2004; Stein and Graner 2004; Keller et al. 2005; Salvi and Tuberosa 2005). More recently, a physical map of chromosome 3B has been developed and it is currently used for map-based cloning, recombination and linkage disequilibrium (LD) studies as well as to analyze the wheat genome composition, organization, function and evolution (Paux et al. 2008). Genes/QTLs, which have already been isolated or are likely to be isolated in the near future through map-based cloning in wheat, are listed in Table 6.

2.3. Wheat Genome Sequencing

The International Wheat Genome Sequencing Consortium (IWGSC) was launched in 2005 with the aim of *"advancing agricultural research for wheat production and utilization by developing DNA-based tools and resources that result from the complete sequence of the common (hexaploid) wheat genome while ensuring that these tools and the sequence are available for all to use without restriction and without cost".* The mid-term goals of IWGSC were to develop a physical map of the 21 chromosomes of bread wheat cv. Chinese Spring to accelerate map-based cloning and to develop new markers, to increase our knowledge about wheat genome organization, and to assess sequencing technologies.

Table 6. A list of genes already cloned through map-based cloning in wheat

Gene/QTL	Trait	Reference
Cre3	Cereal cyst nematode resistance	Lagudah et al. (1997)
Lr1	Leaf rust resistance	Ling et al. (2003)
Lr10	Leaf rust resistance	Feuillet et al. (2003)
Lr21	Leaf rust resistance	Huang L et al.(2003)
VRN1	Vernalization response	Yan et al. (2003)
VRN2	Vernalization response	Yan et al. (2004)

VRN3	Vernalization response	Yan et al. (2006)
Q	Free threshing character	Faris et al. (2003); Simons et al. (2005)
Pm3b	Powdery mildew resistance	Yahiaoui et al. (2004); Brunner et al. (2005)
GPC-B1	High grain protein content	Uauy et al. (2006)
Qfhs.Ndsu-3bs	Fusarium head blight resistance	Liu et al. (2005)
Yr5	Resistance to stripe rust	Ling et al. (2005)
B	Boron tolerance	Schnurbusch et al. (2005)
Fr2	Frost resistance	Galiba et al. (2009)
EPS-1	Flowering time	Lewis et al. (2008)
Tsn1	Host-selective toxin *Ptr ToxA*	Lu et al. (2006)

On January 12, 2010, IWGSC organized a workshop to develop and discuss protocols and standards for the physical mapping of the hexaploid wheat genome. In addition, the workshop surveyed sequencing efforts undertaken within the consortium to coordinate the studies carried out in member laboratories. The goal was to ensure homogeneity in the procedures used for constructing the wheat physical maps by providing guidelines developed in expert laboratories and distributing these to the groups participating in the physical mapping and sequencing of bread wheat chromosomes under the auspices of the IWGSC.

Moreover, the road map for achieving a high-quality reference sequence of the bread wheat genome established by the IWGSC includes, the construction of physical maps in hexaploid wheat using a chromosome specific strategy (http://www.intl-pag.org/18/18-iwgsc.html). This approach relies on recent improvements in chromosome sorting and BAC library construction technologies that have allowed the construction of chromosome specific BAC libraries (Dolezel et al. 2007). Construction of the first sequenced-based physical map has been constructed for the largest wheat chromosome, 3B (~1 Gb in size) (Paux et al. 2008; http://urgi.versailles.inra.fr/projects/Triticum/index.php). In addition to the above, physical mapping and sequencing project leaders have been secured for all the bread wheat chromosomes by IWGSC; the current status is presented in Fig. 2. Updated information related to wheat genome sequencing is available on IWGSC website (www.wheatgenome.org).

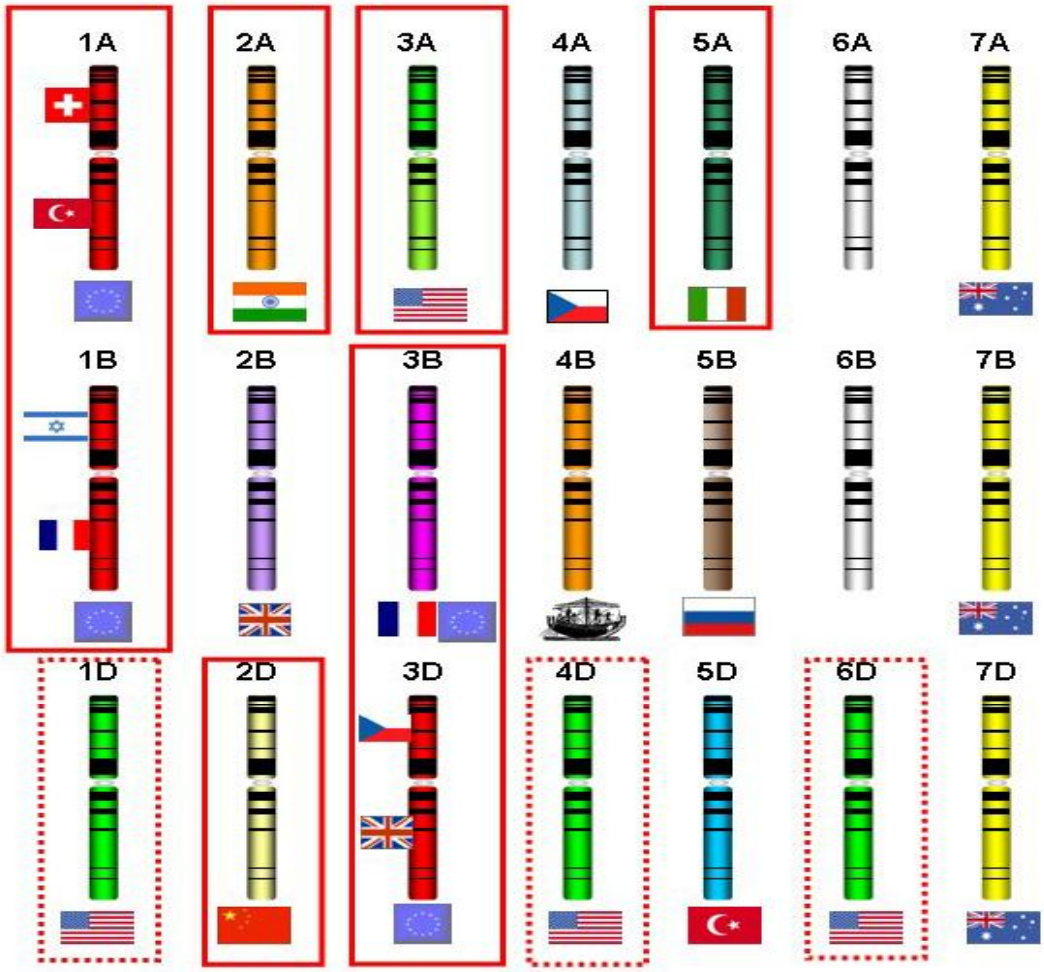

Figure 2. Status of the IWGSC physical mapping projects as of January 2011. Chromosomes boxed in red continuous line represent funded and chromosomes boxed in red interrupted line represent pending projects as on August 2008.