

# Advances in cereal genomics and applications in crop breeding

Rajeev K. Varshney<sup>1</sup>, David A. Hoisington<sup>1</sup> and Akhilesh K. Tyagi<sup>2</sup>

<sup>1</sup> Applied Genomics Laboratory, Global Theme on Biotechnology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, A.P., India

<sup>2</sup> Interdisciplinary Centre for Plant Genomics (ICPG), Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110021, India

**Recent advances in cereal genomics have made it possible to analyse the architecture of cereal genomes and their expressed components, leading to an increase in our knowledge of the genes that are linked to key agronomically important traits. These studies have used molecular genetic mapping of quantitative trait loci (QTL) of several complex traits that are important in breeding. The identification and molecular cloning of genes underlying QTLs offers the possibility to examine the naturally occurring allelic variation for respective complex traits. Novel alleles, identified by functional genomics or haplotype analysis, can enrich the genetic basis of cultivated crops to improve productivity. Advances made in cereal genomics research in recent years thus offer the opportunities to enhance the prediction of phenotypes from genotypes for cereal breeding.**

## Potential of cereal genomics

Cereals, including rice, maize, wheat, barley, rye, sorghum, oats and millets, have constituted the staple food of the world since their domestication ~10 000 years ago. Cereals are also the most important group of cultivated plants for food production and acreage covered, providing >60% of the calories and proteins in our daily diet. In the past, cereals have been the subject of intensive cytogenetic investigations, and these are now further extended using the powerful tools of molecular biology in the genomics era. The structural and functional genomics research on cereal genomes, which during the past two decades has covered both basic and applied aspects, deepens our understanding about gene networks for cereal development and agronomy through the available molecular maps, genomic and expressed sequence tag (EST) sequences and the interaction of gene products, and information about QTLs (quantitative trait loci or genomic regions that are associated with a phenotypic trait exhibiting continuous variation) (Table 1). Furthermore, comparative genomics studies have transformed grasses (cereals) into a single genetic system; therefore, information gained from one cereal crop, such as colinearity (see Glossary) and gene function, might also benefit the improvement of other cereals. More interestingly,

genomics is revolutionizing breeding methodology through marker-assisted selection (MAS) (see Glossary) and directed mutagenesis, which are significantly enhancing the efficiency of breeding for the improvement of agronomic traits. Genomics can also enhance cereal genetic engineering by identifying the functions of native genes. Many genes controlling important agronomic traits can be mapped and cloned based on their position on genetic maps (map-based or positional cloning; see Glossary). The cloned genes, containing their own exons, introns and regulatory elements, are good candidates for transformation into other varieties of the same crop, or into other cereals, without additional modification. In addition, the cloned and characterized genes for the trait of interest can be used for mining potentially favourable alleles and haplotypes (see Glossary), which can be transferred into the high-yielding varieties preferred by farmers. In this article, we discuss recent advances in cereal genomics and outline the utility of the generated information for crop improvement programs.

## Molecular markers and applications

Owing to advances in the area of molecular genetics and automation, dense molecular genetic maps are now available for the major cereal species [1]. However, there is still a need to integrate more markers in the genetic maps of rye, oats and millet species. Furthermore, among the different classes of molecular markers (see Glossary), simple sequence repeat (SSR or microsatellite) markers (see Glossary) have proven to be the marker of choice for a variety of applications, particularly in breeding [2]. Single nucleotide polymorphisms (SNPs) are another important class of molecular markers that are more abundant in the genome and amenable to automation for high-throughput genotyping. In addition, diversity array technology (DART) markers (see Glossary) represent another high-throughput marker system, which can be used to prepare the whole genome map even without the availability of sequence data (as it is the case for SNP markers) for the crop [3]. Such genome-wide maps will be useful for establishing marker-trait associations, not only through linkage analysis but also through association mapping (see Glossary).

The availability of sequence data for genes through genome- and/or EST-sequencing projects has enabled the development of molecular markers from the transcribed region of the genome that are commonly referred to as

Corresponding authors: Varshney, R.K. (r.k.varshney@cgiar.org), (varshney.raj@gmail.com).

Available online 7 September 2006.

## Glossary

**Association mapping:** also known as LD (linkage-disequilibrium) mapping or association analysis is a population-based survey used to identify trait-marker relationships based on linkage disequilibrium.

**Binning:** sorting and/or compiling of haplotypes, based on data obtained by sequencing (which is expensive) or EcoTILLING (which is inexpensive).

**Bi-parental mapping populations:** are the progenies derived after crossing two genotypes used as male and female parents, for example, F<sub>2</sub> lines generated from selfing F<sub>1</sub> progenies; DH (doubled haploid) lines generated after doubling haploids obtained from F<sub>1</sub> plants through another egg cell or ovule culture or distant hybridization; and RILs (recombinant inbred lines) that are derived by single seed descent for at least five or more generations by repeated selfing or sibling mating.

**Colinearity:** can be defined as the conservation of the linear order of marker or gene loci along the ortho- or homologous chromosomes in related plant species. Early comparative genetic mapping studies in cereals have indicated a high colinearity among different grass genomes, despite the large differences in ploidy level, chromosome number and haploid DNA content in these species.

**COS:** conserved orthologous set of markers that are used for comparative mapping between closely related species – for a given group of species, COS markers are developed by identifying genes from each species that are orthologous to genes of other species in the set.

**DART:** marker is a segment of genomic DNA, the presence of which is polymorphic in a defined genomic representation. DART markers are biallelic and behave in a dominant (present versus absent) or co-dominant (2 doses versus 1 dose versus absent) manner.

**Epistasis:** is a form of gene interaction, whereby one gene interferes with the phenotypic expression of another non-allelic gene or genes. Gene X is said to be epistatic to gene Y if an allele of gene X alters the encoded effects of gene Y. In the case of epistasis, the combined phenotypic effect of two or more genes is either less than (negative epistasis) or greater than (positive epistasis) the sum of effects of individual genes.

**Gene space:** refers to long gene-rich regions that contain the vast majority of genes separated by long gene-poor regions in a genome of given species. Occurrence of 'gene space' is a common feature of cereal species (e.g. wheat, barley), having a large genome size due to the abundance of repetitive DNA (retro and/or transposons) in their genome.

**Haplotype:** SNPs are the most abundant form of DNA variation at a given gene locus. Combination of two or more than two SNPs at a locus is called a haplotype. Haplotypes are selected on the basis of a subset of common SNPs that is maximally informative. Such haplotypes or the haplotype maps are useful resources to identify regions of the genome associated with traits of interest in populations with high LD as well as candidate genes in populations with low LD.

**Linkage disequilibrium (LD):** refers to non-random association between markers, genes or QTLs in a population. When variants of two genetic loci are in strong LD, the variant seen at one locus is predictive of the variant found at the other.

**Marker-assisted selection (MAS):** is a method that uses molecular markers for indirect selection of difficult traits at the seedling stage, thus speeding up the process of conventional plant breeding and facilitating the improvement of traits that can not be easily selected using conventional methods.

**Map-based cloning (MBC):** involves identification of a mutant phenotype for the trait of interest (obtained by mutagenesis or from natural variation) and genetic fine mapping using a large number of progeny plants. This map is then used for chromosome walking or landing, with the help of large-insert DNA libraries or physical maps to isolate the gene [23].

**Massively parallel signature sequencing (MPSS):** is an approach to gene profiling that uses a novel sequencing method to identify gene signatures in a cDNA population. The method generates several million signatures of 16 to 20 bases that can be used to identify the corresponding coding sequence and

determine transcript abundance. It offers a greater depth of analysis than many other methods and provides information on absolute amounts of particular transcripts.

**Molecular markers:** are the set of DNA-based genetic markers that can detect DNA polymorphism both at the level of specific loci and at the whole genome level. There are many types of molecular markers: the earliest are RFLPs (restriction fragment-length polymorphisms) and others include RAPDs (random amplification of polymorphic DNAs), CAPS (cleaved amplified polymorphic sites), SSRs (simple sequence repeats) and AFLPs (amplified fragment length polymorphisms). The latest includes SNPs (single nucleotide polymorphisms) and SFPs (single feature polymorphisms).

**Pseudomolecule:** is a complete sequence or a virtual contig for a given chromosome, assembled by using sequence data of overlapping genomic DNA clones (e.g. BAC, PAC) spanning the entire chromosome. For example, rice has 12 chromosomes and therefore 12 pseudomolecules are expected for rice. A list of the ordered BAC and/or PAC clones for each of the 12 chromosomes can be obtained from the website of the International Rice Genome Sequencing Project (IRGSP, (<http://rgp.dna.affrc.go.jp/E/IRGSP>) or The Institute of Genome Research (TIGR, <http://www.tigr.org/>).

**Serial analysis of gene expression (SAGE):** is a technique designed to take advantage of high-throughput sequencing technology to obtain a quantitative profile of gene expression that measures not the expression level of a gene, but quantifies a 'tag' that represents the transcriptome product of a gene. A tag, for the purpose of SAGE, is a nucleotide sequence of a defined length directly adjacent to the restriction site for a particular restriction enzyme closest to the 3' end. The data product of the SAGE technique is a list of tags with their corresponding count values – a digital representation of cellular gene expression. Based on the length of tags, several modified forms of SAGE, such as, MicroSAGE, MiniSAGE, LongSAGE and SuperSAGE, have been developed.

**SSRs:** simple sequence repeats (commonly called microsatellites) consist of simple tandem repeated di- to penta-nucleotide sequence motifs. Because they are abundant, hypervariable, multi-allelic and evenly distributed throughout the nuclear genomes of a majority of organisms they provide a valuable source of polymorphism, making them an important class of genetic markers. The exceptionally high levels of polymorphism detected by SSRs are due to the variability in the number of tandem repeats at a particular locus.

**Tiling microarray:** involves the generation and immobilization on a glass slide of nucleic acid probes that represent a target genomic region. These probes can either overlap, lay end-to-end or be spaced at a predefined average distance in genomic space. A sequence of probes spanning a genomic region is called a 'tile path', or a 'tiling', and the average distance, in nucleotides, between the centers of neighboring probes is termed the 'step' or 'resolution' of the tiling. Each probe on a tiling array interrogates the presence of a sequence in a nucleic acid population by hybridization. Tiling arrays can be developed for sequenced organisms only, and among plant species, *Arabidopsis* and rice enjoy the development and application of such arrays.

**TILLING:** Targeting-induced local lesions in genomes is a reverse genetic method that combines random chemical mutagenesis with PCR-based screening of genes of interest. This provides a range of allele types, including missense and knock-out mutations. By comparing the phenotypes of isogenic genotypes differing in single sequence motifs, TILLING provides direct proof-of-function of both induced and natural polymorphisms without the use of transgenic modifications.

**Unigenes:** are the collections of non-redundant sets of genes for a given tissue or an organism. Unigenes or a unigene set are defined based on cluster analysis of sequence data generated through EST or shotgun sequencing approach. When preparing the cDNA arrays, generally, cDNA clones corresponding to the unigenes are spotted on a glass slide or nylon membrane for gene expression profiling.

'genic' or 'functional' markers (FMs) because a putative function can generally be deduced for such markers [4]. The FMs are generally superior to random DNA markers (derived from genomic DNA) because of complete linkage with potential trait-locus alleles. Although functional markers were developed early, in the form of cDNA-RFLP [5], their functions could not be predicted at that time. Nevertheless, some efforts were made to sequence these early cDNA clones to determine the genes and their functions [6]. In recent years, ESTs or gene sequences have been used to identify SSRs [7] and SNPs [8], and genic molecular-markers have been developed in several cereal species. In addition to being useful for identifying the 'perfect' or 'ideal' markers in marker-assisted selection, FMs are an

important resource for estimating functional variation in natural or breeding populations and for studying genome evolution, through comparative mapping.

### Marker-assisted selection

Marker-assisted selection (MAS) is a powerful tool for the indirect selection of difficult traits at an early stage before production of the next generation, thus speeding up the process of conventional plant breeding and facilitating the improvement of traits that cannot be improved easily by conventional methods [9]. Using MAS, a large number of genes and QTLs controlling agronomic traits and conferring tolerance to both abiotic and biotic stresses in cereals have been identified and tagged using molecular

**Table 1. Genome size, structure and genomic resources of major cereal species**

Crop (Botanical name)	Ploidy level and chromosome number	Genome size (bp) [50]	Repetitive DNA [51]	Genetic map	Large insert library	Physical map (contig based)	Genomic sequence	EST sequences in public domain <sup>a</sup>
Rice ( <i>Oryza sativa</i> )	2n = 2x = 24	0.4 × 10 <sup>9</sup>	35%	+++	+++	Map available for whole genome <sup>b</sup>	Complete sequence available <sup>b</sup>	1 183 584
Maize ( <i>Zea mays</i> )	2n = 2x = 20	3.0 × 10 <sup>9</sup>	78%	+++	++	(++) <sup>c</sup>	(++) <sup>d</sup>	734 267
Wheat ( <i>Triticum aestivum</i> )	2n = 6x = 42	17.9 × 10 <sup>9</sup>	83%	++	++	(+) <sup>e</sup>	(+) <sup>e</sup>	853 316
Barley ( <i>Hordeum vulgare</i> )	2n = 2x = 14	5.5 × 10 <sup>9</sup>	76%	++	++	(+) <sup>f</sup>		432 321
Sorghum ( <i>Sorghum bicolor</i> )	2n = 2x = 20	0.8 × 10 <sup>9</sup>		++	+	(++) <sup>g</sup>	(++) <sup>h</sup>	208 466
Pearl millet ( <i>Pennisetum glaucum</i> )	2n = 2x = 14	2.5 × 10 <sup>9</sup>		+	+			2848
Rye ( <i>Secale cereale</i> )	2n = 2x = 14	9.4 × 10 <sup>9</sup>	92%	+	+			9195

<sup>a</sup>dbEST release04280 ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)); <sup>b</sup>genome-wide physical map, four draft sequence and complete genome sequence available [29]; <sup>c</sup>physical map construction [52] and <sup>d</sup>genome sequencing project (<http://www.maizegenome.org/>) (in progress); <sup>e</sup>physical map construction and sequencing of the gene-rich regions of genome (in planning phase) (<http://www.wheatgenome.org/>); <sup>f</sup>physical map construction in progress at University of California, Riverside (USA) (<http://phymap.ucdavis.edu:8080/barley/index.jsp>) and in Europe as a part of European Triticeae Genomics Initiative (<http://pgrc.ipk-gatersleben.de/etgi/>); <sup>g</sup>construction of a genome-wide physical map in progress [53]; <sup>h</sup>shotgun sequence data generated by Orion Genomics [34], and DOE JGI (USA) is engaged in sequencing the sorghum genome [33].

markers [10–12]. Although the potential benefits of MAS are substantial, the actual adoption of this technique in breeding programmes among the cereals and other crops has been patchy [13]; it is only relatively recently that MAS has begun to make more than a marginal impact on breeding methodology. The extremes in terms of large-scale MAS deployment in the cereals are represented by maize, where uptake is substantial, and wheat, where its extent is less spectacular. Barley is similar to wheat in terms of breeding system but has enjoyed more progress, possibly because of its simpler (non-polyploid) genetics, whereas rice is particularly relevant because of its global importance both as a crop species in its own right and as a model species for the cereals in general. The remaining small grain cereals (oats, rye, triticale, sorghum, the millets and tef) are largely too minor to have enjoyed any significant investment in marker discovery and commercial deployment [13]. Some notable examples of the successful deployment of MAS in cereals are listed in Table 1. In addition, several programmes and initiatives, such as molecular breeding programmes in wheat and barley in Australia [14] and ‘MASWheat’ (<http://maswheat.ucdavis.edu/index.htm>), are underway to conduct MAS in breeding.

### Association mapping

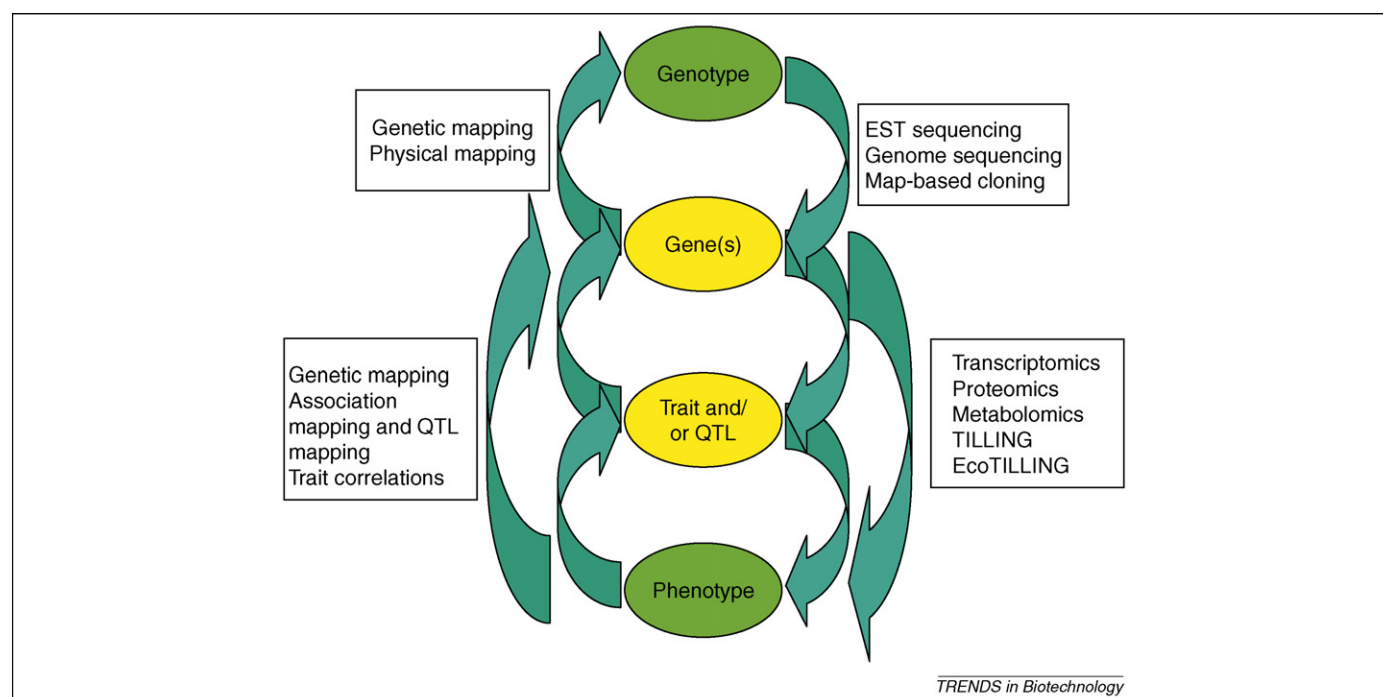
Another approach to identify molecular markers for use in MAS is association mapping, which is based on linkage disequilibrium (LD) (see Glossary). Unlike conventional biparental mapping populations (see Glossary) – such as DH, F<sub>2</sub> or RILs, which have been used in the past for identifying genes or QTLs for the trait of interest, the natural populations are the products of many cycles of recombination and have the potential to show enhanced resolution of QTLs. Association mapping might offer more power than linkage analysis for identifying the genes responsible for the variation in a quantitative trait [15]. The extent of LD around a locus determines the resolution of association analyses and the number of markers that would be required to scan

the entire genome [16]. Because genetic recombination is not evenly distributed throughout the genomes of most species, the linkage distance between markers and candidate genes varies widely.

LD depends on the evolutionary or selection history and, as a result, only genes and/or markers with tight linkage will be detected. Because plant populations are generally structured, Prichard *et al.* [17] proposed a population-based method that enables large-scale assessment of allele and/or trait relationships in structured populations. Under this approach, marker–trait association is only expected when a QTL is tightly linked to the marker because the accumulated recombination events occurring during the development of the lines will prevent the detection of any marker–trait association. By using this approach, association mapping has been demonstrated in: (i) maize, for the *Dwarf8* gene involved in flowering time [18], yellow endosperm colour [19], and sweet taste in maize (E.S. Buckler, personal communication); (ii) barley, for yield and yield stability [20]; and (iii) wheat, for kernel morphology and milling quality [21]. Such high-resolution mapping of traits and/or QTLs to the level of individual genes, in combination with improved statistical methods, will provide new possibilities for studying the molecular and biochemical basis of quantitative trait variation and will help to identify specific targets for crop improvement [22] (Figure 1).

### Map-based cloning (MBC)

The MBC approach involves the use of molecular markers for preparing a high-density genetic map around the region harbouring the gene of interest and, ultimately, the local physical map to isolate the gene [23]. In fact, several MBC projects were started in the mid-1990s, and several genes or QTLs for disease resistance or other traits have been isolated in many cereal species (Table 1). These examples involved long-term efforts (up to 10 years), depending on the availability of resources and the location of gene(s) and/or QTL(s) in the genome. Owing to the availability of



TRENDS in Biotechnology

**Figure 1.** Genomics-based approaches for enhancing the prediction of the phenotype from a genotype. The better prediction of the phenotype that a particular genotype will produce is a primary goal of genomics-based breeding. The figure indicates how the various tools and techniques available today can be used to enhance our understanding of the various components between the genotype and the phenotype. On the left side are those strategies essentially based on classical genetics approaches (often enhanced by modern molecular tools), which can resolve the complexity of a phenotype, trait or QTL at a finer level in the genome. The right side indicates the application of modern genomics approaches to move from a genotype to phenotype. For example, genome and EST sequencing in combination with gene prediction algorithms and map-based cloning lead to the discovery of essentially the full set of genes in a genome. These genes can then be used in transcript profiling (transcriptomics), protein profiling (proteomics) and metabolite profiling (metabolomics) experiments to better understand their role in a trait or phenotype. TILLING and EcoTILLING identify allelic variants of a gene, producing useful germplasm for discovery and confirmation of the role(s) of the particular gene(s) in a final phenotype. These can also be used as sources of superior haplotypes for use in direct enhancement of the phenotype. Traditionally, we have relied on classical genetic approaches such as trait correlation analyses and QTL mapping to dissect a complex phenotype into simpler trait components and/or QTL. QTL mapping has been enhanced in resolution with the availability of high-density molecular marker maps. Association mapping can also be applied to these high-density maps to directly identify gene(s) associated with a particular phenotype and/or trait, without the need for segregating genetic populations. Genetic and physical mapping, using both cytogenetic and large-insert libraries such as bacterial artificial chromosome (BAC) and yeast artificial chromosome (YAC) based approaches, of the genes underlying QTLs provide detailed information regarding the organization of the genes in the genome as well as a basis for understanding the epistatic interactions or epigenetic phenomenon in the genome so that appropriate crop improvement strategies can be devised.

resources and expertise developed recently in cereal genomics (Table 2), it should be easier to isolate genes and/or QTLs in a relatively shorter time frame. For example, by using the genome sequence data of Nipponbare (*japonica* cultivar) and 93-11 (*indica* cultivar), a genome-wide rice DNA polymorphism database has been constructed (<http://shenghuan.shnu.edu.cn/ricemarker/>) that contains 1 703 176 SNPs and 479 406 insertion/deletions (InDels), approximately one SNP every 268 bp and one InDel every 953 bp in the rice genome [24]. Thus, available genomic resources and novel approaches, for example, functional genomics, association mapping or their combination with conventional MBC approach, would certainly facilitate gene isolation in rice and other cereals.

The candidate genes isolated and characterized can be used for allele mining to identify superior haplotypes in addition to producing transgenic plants for the traits. In this context, a strategy based on TILLING (see Glossary) called EcoTILLING was developed for detecting multiple types of polymorphisms in germplasm collections [25]. EcoTILLING enables natural alleles at a locus to be characterized across many germplasm lines, enabling both SNP discovery and haplotyping. This can be done at a fraction of the cost of SNP and/or haplotyping methods, which require large-scale sequencing. EcoTILLING

haplotypes across a range of germplasms can be binned (see Glossary), and confirmatory sequencing done on only the unique haplotypes. EcoTILLING is expected to provide a series of alleles for those genes that are involved in important processes in the plant, although the known variants for these genes have not been observed through genetic studies. For example, Slade *et al.* [26] demonstrated the power of TILLING for practical crop improvement when they identified 196 new alleles in the A and D genome *waxy* genes (granule-bound starch synthase I, *GBSSI*) in only 1 152 individual plants screened in their hexaploid wheat TILLING population, and 50 new alleles in only 768 individuals in their tetraploid pasta wheat TILLING population. Eventually, after identifying all possible alleles that are available in germplasm collections, they must be evaluated for their relative value in adapted genotypes in the target environment. These analyses might help in designing synthetic alleles that are superior to those found in nature.

### Sequence data and genes

#### Whole-genome and/or gene-space sequencing

Because the rice genome is the smallest among the cereal genomes, it was proposed as a target for whole-genome sequencing. Because of international collaborative public



**Table 2. Some important achievements made in the area of cereal genetics and breeding through molecular markers**

Cereal species	Notable examples of MAS	Examples of genes isolated through MBC
Barley	<ul style="list-style-type: none"> <li>•Release of US variety Tango in 2000 that contains two QTL for adult resistance to stripe rust [54]</li> <li>•Advancement of a 'Sloop type' variety with CCN (cereal cyst nematode) resistance for commercial release [55]</li> <li>•Introgression of <i>Yd2</i> gene conferring resistance to barley yellow dwarf virus (BYDV) into a BYDV susceptible background through two cycles of marker-assisted backcrossing [56]</li> </ul>	<ul style="list-style-type: none"> <li>•Powdery mildew resistance genes <i>Mlo</i> [57], <i>Mla</i> [58], <i>Rar1</i> [59]</li> <li>•Stem rust resistance <i>Rpg1</i> [60]</li> <li>•Barley yellow mosaic virus resistance <i>rym3</i> and/or <i>rym4</i> [61]</li> </ul>
Maize	<ul style="list-style-type: none"> <li>•Use of yield-related QTLs for MAS in private sector [13]</li> <li>•Development of quality protein maize (QPM) through marker-aided transfer of <i>opaque2</i> gene in backcross programmes [62]</li> </ul>	<ul style="list-style-type: none"> <li>•Leaf rust resistance <i>Rp1 D</i> [63]</li> <li>•<sup>a</sup>Flowering time QTL <i>Vgt1</i> [64]</li> <li>•<sup>a</sup>Root abscissic acid QTL, <i>ABA1</i> (R. Tuberosa, personal Communication)</li> </ul>
Pearl millet	<ul style="list-style-type: none"> <li>•Release of a Indian pearl millet hybrid cultivar 'HHB 67 Improved' in 2005, which has resistance to downy mildew (C.T. Hash, personal communication)</li> </ul>	
Rice	<ul style="list-style-type: none"> <li>•Release of two Indonesian rice cultivars 'Angke' and 'Conde', in which MAS was used to introduce <i>xa5</i> into a background containing <i>xa4</i> [65]</li> <li>•Pyramiding of disease resistance genes in rice, particularly against blight [66,67], blast [68], and both simultaneously [69]</li> <li>•Pyramiding of insect and blight resistance [67]</li> <li>•The pyramiding of blight resistance with Basmati quality characters [70]</li> </ul>	<ul style="list-style-type: none"> <li>•Bacterial blight-resistance genes <i>Xa1</i> [71], <i>xa5</i> [72], <i>Xa21</i> [73], <i>Xa26</i> [74]</li> <li>•Rice blast-resistance genes <i>PiB</i> [75], <i>Pi ta</i> [76], <i>Pi5 (t)</i> [77], <i>Pi 9</i> [78]</li> <li>•Plant architecture gene <i>Dwarf1</i> [79]</li> <li>•A timekeeper of leaf initiation <i>PLASTOCHRON1</i> [80]</li> <li>•Leaf spotted leaf gene <i>Sp17</i> [81]</li> <li>•Semi-dwarf gene (<i>sd 1</i>) [82]</li> <li>•Seed shattering gene (<i>qSH1</i>) [83,84]</li> <li>•QTLs for heading <i>Hd1</i> [85], <i>Hd3a</i> [86], <i>Hd4</i> and <i>Hd5</i> [87], <i>Hd6</i> [88], <i>Ehd1</i> [89]</li> <li>•QTL for grain production, <i>Gn1a</i> [90]</li> <li>•QTL for salt tolerance [91]</li> </ul>
Sorghum	<ul style="list-style-type: none"> <li>•<sup>a</sup>Pyramiding of stay green QTLs in elite but drought-susceptible sorghum lines (C.T. Hash, personal communication)</li> </ul>	<ul style="list-style-type: none"> <li>•A major aluminum tolerance gene <i>AltSB</i> [92]</li> </ul>
Wheat	<ul style="list-style-type: none"> <li>•&gt;50 000 assays for more than a dozen loci, including tolerance to high soil boron (<i>Bo1</i>), tolerance to late maturity <math>\alpha</math> amylase (LMA) (7BL), barley yellow dwarf virus resistance (BYDV2) (7DL), cereal cyst nematode resistance <i>Cre1</i> (2BL), <i>Cre8</i> (6BL), waxy or granule bound starch synthase (<i>Wx B1</i>) (4A), high molecular weight glutenin subunits (GluD1) (1DL), leaf rust resistances (<i>Lr46</i>) (1BL), (<i>Lr34</i>) (7DS), height or dwarfing genes (<i>Rht1</i>) (4BS), (<i>Rht2</i>) (4DS), (<i>Rht8</i>) (2DS), root lesion nematode resistance (<i>Rlnn1</i>) and yellow flour colour (7AL), stem rust resistances (<i>Sr2</i>) (3BS), (<i>Sr36</i>) (2B) and VPM (Ventricosa x Persicum x Marne), a source for eyespot resistance gene <i>Pch1</i>, are being performed annually in an Australian wheat breeding programme, to implement molecular markers in wheat breeding [14,55]</li> </ul>	<ul style="list-style-type: none"> <li>•Leaf rust resistance genes <i>Lr10</i> [93], <i>Lr21</i> [94]</li> <li>•Powdery mildew resistance gene <i>Pm3b</i> [95]</li> <li>•Major chromosome pairing loci <i>Ph1</i> [96], <i>Ph2</i> [97]</li> <li>•Wheat vernalization genes <i>VRN1</i> [98], <i>VRN2</i> [99]</li> <li>•Wheat domestication gene <i>Q</i> [100]</li> <li>•QTLs (in advanced stage) for resistance to fusarium head blight <i>Qfhs.ndsu 3BS</i> [101], <sup>a</sup>stem rust <i>Sr2</i> [102], <sup>a</sup>leaf rust <i>Lr34</i> [103], <sup>a</sup><i>Stagonospora nodorum</i> [104]</li> </ul>

<sup>a</sup>progress in advanced stage.

and private sector efforts, four drafts [27–29] (<http://rgp.dna.affrc.go.jp/IRGSP/>) and the complete sequence of the whole genome [30] became available for rice: one study was done with *indica* rice and three studies were completed with *japonica* rice [29]. The complete sequence of the rice genome was obtained from 3401 PAC (P1-derived artificial chromosome) and BAC (bacterial artificial chromosome) clones. The total nucleotide sequence of the 12 pseudomolecules (see Glossary) is 370 733 456 bp, excluding the ambiguous nucleotides; therefore, the pseudomolecules so far cover 95.3% of the entire genome and an estimated 98.9% of the euchromatin (<http://rgp.dna.affrc.go.jp/IRGSP/Build3/build3.html>). A total of 55 296 genes,

including those related to transposable elements (TE-related), have been predicted for the finished sequence (or 37 544 genes if you exclude the TE-related genes) [29]. This range of predicted genes is also supported by a full-genome transcription analysis of the *indica* rice subspecies using high-density oligonucleotide tiling microarrays (see Glossary) [31]. This approach involved the hybridization of the custom tiling microarrays: these contain 13 078 888 individual 36mer oligonucleotide probes tiled throughout the nonrepetitive sequence of the rice genome. They are based on the improved whole-genome shotgun (WGS) sequence of the *indica* subspecies [27] and comprise a mixture of cDNA targets derived from

four major tissues to maximize transcript detection. The expression data supported the existence of 35 970 annotated gene models and identified 5464 unique transcribed intergenic regions that share similar compositional properties with the annotated exons and have significant homology to other plant proteins. Furthermore, to characterize fully the rice genome, comparative genomic analysis within the genus *Oryza*, which contains two cultivated and 22 wild species and represents ten distinct genome types, has been recently planned. Furthermore, a comprehensive set of 12 BAC libraries that represent all 10 genome types of *Oryza* has been generated [32].

Compared with rice, the genomes of other cereals are large and complex (Table 2). Despite the magnitude of the task to tackle these genomes, projects to sequence the genome or gene space (see Glossary) of some cereal have been undertaken during the past few years [33], for example, genome-wide sequencing is underway in maize (<http://www.maizegenome.org/>) and sorghum [34] (<http://www.jgi.doe.gov/>). In addition to using traditional methods to obtain genome sequence data, other approaches such as methyl filtration and high Cot analysis strategies [33] are being attempted, to focus in on the gene-rich portion of the genomes. Recently, a consortium called the International Wheat Genome Sequencing Consortium (IWGSC; <http://www.wheatgenome.org/>) has been established to devise the strategies and sequence the gene space of the large and hexaploid wheat genome.

### EST sequencing

Before the start of the genome sequencing projects, large-scale EST-sequencing projects were undertaken in several cereal species, and a large number of ESTs have become available in almost all cereals (Table 2). ESTs provide an alternative means for understanding the genome, or at least the transcriptome, of a given species. For example, analysis of >110 000 barley ESTs revealed that ~41% of barley genes belong to multigene families, and 4% of barley genes undergo alternative splicing [35]. Similarly, after the analysis of 116 232 ESTs generated from 10 wheat tissues, Ogiwara *et al.* [36] studied developmental processes in wheat after analysing correlated expression patterns of these genes across the tissues.

EST resources have been used in applied aspects as well, by exploiting them in the development of molecular markers (FMs) and functional genomics studies by developing cDNA arrays. For example, ESTs have been extensively used for the development of EST-SSR [7], SNP [8] and COS (see Glossary) [37] markers, which are not only used in trait mapping and MAS but also provide information about genome evolution [38]. Similarly, ESTs have been used to develop cDNA arrays [39,40] for identifying the genes involved in seed development processes or unique traits.

### Functional genomics

Functional genomics involves the identification of the function of genes *per se* or those derived from a known allelic difference conferring an improved phenotype. In the latter approach, the objective is to identify the sequence change conferring the improved phenotype; such a

sequence change can then become the basis for a molecular marker that is specific for that allele. Thus, functional genomics in the true sense can be linked or associated with plant breeding for crop improvement programmes (Figure 1).

Several techniques or platforms, such as serial analysis of gene expression (SAGE; see Glossary), massively parallel signature sequencing (MPSS; see Glossary) and micro- and macro-arrays, are available [41] for the estimation of mRNA abundance for large numbers of genes simultaneously. However, because of their advantages compared with other platforms, such as cost and high-throughput outputs, micro- and macro-arrays have been extensively used in cereal species [39,40]. Recently, tiling microarrays based on the whole-genome sequence data have become available for genome-wide transcription analysis in rice [31]. Thus, for rice, GeneChip arrays, full-length cDNA arrays and genome-wide tiling microarrays are available; in the case of wheat and barley, Affymetrix GeneChip arrays have been developed [39]. The macro- and micro-arrays have been successfully used in many cereal species, including maize, rice, wheat, barley and sorghum, for understanding basic physiology, for example, developmental processes, environmental-stress responses and the identification and genotyping of mutations [40–42]. The use of these technologies for applied aspects in plant breeding started in recent years. For example, using ten barley genotypes characterized for six malting-quality parameters and a cDNA array with 1400 unigenes (see Glossary), Potokina *et al.* [40] identified between 17 and 30 candidate genes for each of the six malting parameters. This set of candidate genes contained genes expected to be related to malting quality (e.g. cysteine proteinase 1), genes where the relationship to this trait is unknown (e.g. 70 kDa heat shock protein) and genes of unknown function. Furthermore, the observed linkage of five out of eight mapped candidate genes to known QTLs for malting-quality traits underscores the functional genomics approach. Thus, the functional genomics approach in combination with 'expression genetics' or genetical genomics (when gene expression data are analysed on segregating populations in a quantitative fashion and subjected to expression QTL analysis) provides a candidate set of genes that can be used for understanding the biology of a trait and for the development of perfect (one that predicts the phenotype from the genotype at 100% efficiency rate) or diagnostic marker(s) for use in MAS applied to crop breeding [43].

### Comparative genomics for orphan cereal crops

Besides the major cereal crops (maize, wheat and rice), many cereal crops such as sorghum, pearl millet, small millets and tef (*Eragrostis tef*) are regionally or locally important for nutrition and income, particularly in developing countries [44] (<http://www.cgiar.org/impact/research/millet.html>). Because of relatively low returns in terms of gross economic and welfare impacts, the 'orphan crops' have not received adequate investment in terms of research [45]. Given that significant genomic colinearity has been reported in cereal species [46], the comparative genomics approach using bioinformatics

tools might, therefore, provide an opportunity for efficient transfer of information from model species and major crops to minor and orphan crops. As a result, a relatively small investment in the transfer of genomic information from major or model species to a larger set of orphan crops might potentially result in high payoffs in terms of crop production, yield stability and food security [44,45]. In this context, the model cereal species (i.e. rice as well as major crop plant species such as maize and wheat) possess a great potential that can be exploited in various ways for the improvement of orphan or minor crops. The benefits of transferring genomic information and techniques from model or major crops to minor or orphan crops take one or more of several forms: (i) improved analysis of crop biodiversity and identification of potentially useful variants; (ii) MAS of desired alleles and allele combinations; and (iii) cloning and transfer of desirable alleles among taxa [45].

In addition, there can also be reciprocal benefits of genomics research on minor and orphan crops for improvement of major crops, derived from insights into

the genetic basis of their distinctive attributes. For example, superior alleles for drought tolerance might be found in pearl millet and used in major crops such as wheat and rice [47]. Thus, a shift in the investment perspective from individual (model and major) crops to whole sets of crops (including minor and orphan crops) with common genetic structures could be argued. Thereby, genomics research can fulfil nutritional needs and/or contribute to local incomes and employment in the poorest regions of the world.

### Interdisciplinary genomics approach for crop improvement

Significant progress in the field of cereal genomics has already been made in many cereals. For example, the availability of a variety of molecular markers facilitated the preparation of high-density maps, which proved useful in the identification of molecular markers linked with genes and/or QTLs for a variety of economic traits, including those conferring tolerance to biotic and abiotic stresses. Development of functional molecular markers as a

#### Box 1. Challenges in using genomics for breeding

- **Precision phenotyping.** Appropriate germplasm, accurate phenotyping and high-density genotyping are three crucial components for the successful application of genomics in breeding. Although high-throughput genotyping tools exist, the identification of QTLs or genes through linkage mapping, association mapping or functional genomics approaches requires extensive and precise phenotyping of agronomic traits for breeding materials, mapping populations, natural populations and genebank materials. Sampling of plant material for RNA isolation for functional genomics studies is another important factor that requires detailed knowledge about the correct physiological stage and appropriate plant organs expressing the trait being analyzed. Genetic materials such as near isogenic lines (NILs), mutant stocks and TILLING populations offer the possibilities to validate the identified genes and/or QTLs before their implementation in breeding programmes.
- **Density of genetic maps.** Microsatellite or SSR markers are valuable markers for plant breeding; however, the density of microsatellite markers in genetic maps of crops, such as pearl millet or rye, is not satisfactory. Even for major crops such as barley and wheat, although sufficient numbers of markers are available, the majority of such markers have only mapped in the reference and/or different mapping populations. Therefore, it will be desirable to have all possible markers integrated into a consensus map so that the information available on genetic distances of the markers can help the marker-trait identification. Development of novel marker systems such as DArT, or Illumina-bead technology, based on a SNP genotyping platform, might enhance the density of the maps.
- **Low heritability of traits.** Although marker-trait association can be established for some traits, the low heritability of such traits makes it difficult to use the molecular markers in MAS. Dissecting phenotypes into components can improve heritability and understanding of the biological systems that cause the phenotype.
- **Consideration of epistasis.** Epistasis (see Glossary) has an important role in the phenotype observed from the genetic variation in a genome. However, the epistatic role is often overlooked in contributing to the QTL variation in marker-trait association studies. Simulation and experimental studies have shown the key role of epistasis in the long-term evolution of adaptive traits and in the dynamics of population divergence.
- **Understanding epigenetics.** Epigenetic activities, such as gene silencing and altered chromatin structure, can cause large-scale genomic effects, thus altering transcriptional activity [49]. Epigenetic regulation can be relaxed under stress conditions and can result in the activation of suppressed genes and secondary effects during the re-establishment of genomic order. Selection can then act on the resulting genetic and epigenetic changes in the population. Thus, there is a need to develop the knowledge and tools of epigenetics to control and manipulate it and devise crop improvement strategies accordingly.
- **Contribution of regulatory variation.** Recently, the regulatory variation of gene expression that frequently concerns gene or genomic regions (such as promoters, introns, silencers and other non-coding DNA sequences removed from transcriptional units) has been shown to be more variable than protein-coding DNA sequences. As microarray technology evolves, gene networks and the regulatory factors controlling them need to be considered in gene expression studies dealing with the identification of genes for agronomic traits.
- **Technical difficulties.** High-throughput genotyping platforms can be prone to errors in obtaining precise genotyping data. For functional genomics studies, some studies have shown that different microarray platforms (e.g. Affymetrix, Agilent, Amersham) with the same RNA sample, or analysis of the same microarray gene-expression data with different bioinformatic tools, might not identify the same set of differentially expressed genes for a given trait [105]. Therefore, there is a need for caution and confirmation when analyzing and interpreting functional genomics studies for extracting candidate gene lists. Similarly, in association mapping studies, one of the major concerns is the statistical power of the association testing because, as it stands, there is a trade off between the power and accuracy for reporting associations due to false positives. The major determinant of the levels of false positives and power of associations is the level of population structure in the association population. Therefore, the use of appropriate statistical methods is recommended for control of false positives and to enhance the accuracy of association power [22].
- **Cost investment issues.** The costs for applying genomics strategies and tools for breeding practices frequently exceed the funds that are available in commercial and public breeding programs. This is particularly true for inbreeding crops or crops that are of regional importance only. For example, maize and rice, being important crops in the world and to the private sector, have enjoyed investment of huge amounts of funds in genomics research and, as a result, several agronomic traits for breeding in these crops have been successfully achieved. By contrast, crops such as pearl millet and rye do not have dense genetic maps with a reasonable number of SSR markers to establish the marker-trait associations.



by-product of available sequence data will be useful for marker-trait association studies and examining the functional diversity in breeding germplasm collections or natural populations. Furthermore, genome and/or gene space or EST sequencing provides the sequence data to identify candidate genes for agronomic traits, either through *in silico* approaches, with the help of bioinformatic tools, or 'wet' laboratory experiments such as transcript profiling using micro- or macro-arrays. More interestingly, exploitation of association mapping approaches and expression genetics might provide the best molecular markers (e.g. functional markers) for a trait of interest, which can be used across different genetic backgrounds in MAS. These types of (functional) molecular markers should and/or will ideally co-segregate with the trait of interest. In general, such a marker will often be based on a SNP. The SNPs can be detected in high-throughput systems in such a way that large numbers of plants can be assayed for a particular allele [8].

Integration of the above-mentioned genetic and genomic approaches, together with transcriptomics, proteomics, metabolomics and tools of bioinformatics, as outlined in Figure 1, is essential for the effective use of genomics in breeding: we term this holistic approach as 'genomics-assisted breeding' (GAB) [43]. However, there are several challenges for successful exploitation of GAB that need to be considered or tackled. Some of these challenges include precise phenotyping, low heritability of traits, epistasis (see Glossary), epigenetics, regulatory variation, technical difficulties and cost-investment issues (Box 1). Particularly when using functional genomics and expression genetics approaches for identification of genes and their application in breeding programmes, it is important to consider the phenomenon of gene networks, such as epistatic interactions, and the relationships between gene silencing, DNA methylation, RNA interference (RNAi) and heterochromatic DNA, demonstrating the complexity of RNA regulation operating through small non-coding RNAs [48,49].

Indeed, in the post-genomic era – owing to the availability of high-throughput approaches combined with automation, the rapid increase in sequence data in the public domain and good expertise and tools in the area of bioinformatics – genomics holds great potential to facilitate the prediction of a phenotype more precisely, thereby, increasing the efficiency of breeding. Large-scale application of genomics to breeding will result from new technologies that reduce the costs and increase the throughput of the assays. Although the newly developed genetic and genomics tools will certainly enhance the prediction of phenotype, they will not entirely replace the conventional breeding process.

#### Acknowledgements

The research of authors at ICRISAT (RKV, DAH) is supported by grants from Generation Challenge Programme (GCP) of Consultative Group on International Agricultural Research (CGIAR) and Syngenta Foundation for Sustainable Agriculture (SFSA), whereas AKT acknowledges the support of Department of Biotechnology (DBT), Government of India for supporting research activities at ICPG. Thanks are due to Mahendar Thudi and Spurthi Nayak for their help in formatting and editing references for the article.

#### References

- 1 Varshney, R.K. *et al.* (2004) Molecular maps in cereals: methodology and progress. In *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds), pp. 35–82, Kluwer Academic Publishers
- 2 Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163–185
- 3 Kilian, A. *et al.* (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In *In the Wake of Double Helix: From the Green Revolution to the Gene Revolution* (Tuberosa, R. *et al.*, eds), pp. 443–461, Avenue Media
- 4 Andersen, J.R. and Lübberstedt, T. (2003) Functional markers in plants. *Trends Plant Sci.* 8, 554–560
- 5 Causse, M.A. *et al.* (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138, 1251–1274
- 6 Michalek, W. *et al.* (1999) Sequence analysis and gene identification in a set of mapped RFLP markers in barley (*Hordeum vulgare*). *Genome* 42, 849–853
- 7 Varshney, R.K. *et al.* (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotech.* 23, 48–55
- 8 Rafalski, J.A. (2002) Applications of single nucleotide polymorphism in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100
- 9 Ribaut, J.M. and Hoisington, D. (1998) Marker-assisted selection: new tools and strategies. *Trends Plant Sci.* 3, 236–239
- 10 Li, W. and Gill, B.S. (2004) Genomics for cereal improvement. In *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds), pp. 585–634, Kluwer Academic Publishers
- 11 Jahoor, A. *et al.* (2004) QTLs and genes for disease resistance in barley and wheat. In *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds), pp. 199–252, Kluwer Academic Publishers
- 12 Tuberosa, R. and Salvi, S. (2004) QTLs and genes for tolerance to abiotic stress in cereals. In *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds), pp. 253–315, Kluwer Academic Publishers
- 13 Koebner, R.M.D. (2004) Marker-assisted selection in the cereals: the dream and the reality. In *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds), pp. 199–252, Kluwer Academic Publishers
- 14 Langridge, P. (2005) Molecular breeding of wheat and barley. In *In the Wake of Double Helix: From the Green Revolution to the Gene Revolution* (Tuberosa, R. *et al.*, eds), pp. 279–286, Avenue Media
- 15 Flint-Garcia, S.A. *et al.* (2003) Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374
- 16 Rafalski, A. and Morgante, M. (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet.* 20, 103–111
- 17 Pritchard, J.K. *et al.* (2000) Association mapping in structured populations. *Am. Jour. Hum. Genet.* 67, 170–181
- 18 Thornsberry, J.M. *et al.* (2001) *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genet.* 28, 286–289
- 19 Palaisa, K. *et al.* (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9885–9890
- 20 Kraakman, A.T. *et al.* (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168, 435–446
- 21 Breseghello, F. and Sorrells, M.E. (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177
- 22 Yu, J. and Buckler, E.S. (2006) Genetic association mapping and genome organization of maize. *Curr. Opin. Biotech.* 17, 155–160
- 23 Azhaguvel, P. *et al.* (2006) Methodological advancement in molecular markers to delimit the gene(s) for crop improvement. In *Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues* (Vol. I) (Teixeira da Silva, J.A., ed.), pp. 460–469, Global Science Books
- 24 Shen, Y.J. *et al.* (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135, 1198–1205
- 25 Comai, L. *et al.* (2004) Efficient discovery of DNA polymorphisms in natural populations by EcoTILLING. *Plant Jour.* 37, 778–786



- 26 Slade, A.J. *et al.* (2005) A reverse genetic, non-transgenic approach to wheat crop improvement by TILLING. *Nature Biotech.* 23, 75–81
- 27 Yu, X. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92
- 28 Goff, S. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100
- 29 Vij, S. *et al.* (2006) Decoding the rice genome. *BioEssays* 28, 421–432
- 30 International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800
- 31 Li, L. *et al.* (2006) Genome-wide transcription analyses in rice using tiling microarrays. *Nature Genet.* 38, 124–129
- 32 Ammiraju, J.S. *et al.* (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep coverage large insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res.* 16, 140–147
- 33 Paterson, A.H. (2006) Leafing through genomes of our major crop plants: strategies for capturing unique information. *Nature Rev. Genet.* 7, 174–184
- 34 Bedell, J.A. *et al.* (2005) Sorghum genome sequencing by methylation filtration. *PLoS Biol.* 3, 103–115
- 35 Zhang, H. *et al.* (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant Jour.* 40, 276–290
- 36 Ogihara, Y. *et al.* (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life-cycle by large-scale statistical analyses expressed sequence tags. *Plant Jour.* 33, 1001–1011
- 37 Rudd, S. *et al.* (2005) Plant Markers: a database of predicted molecular markers from plants. *Nucleic Acids Res.* 33, D628–D632
- 38 Sorrells, M.E. *et al.* (2003) Comparative DNA sequence analysis of wheat and rice genomes. *Genome Res.* 13, 1818–1827
- 39 Close, T. *et al.* (2004) A new resource for cereal genomics: 22k barley genechip comes of age. *Plant Physiol.* 134, 960–968
- 40 Potokina, E. *et al.* (2004) Functional association between malting quality trait components and cDNA array-based expression patterns in barley (*Hordeum vulgare* L.). *Mol. Breed.* 14, 153–170
- 41 Milligan, A.S. *et al.* (2004) Functional genomics studies of seed development in cereals. In *Cereal Genomics* (Gupta, P.K. and Varshney, R.K., eds), pp. 447–482, Kluwer Academic Publishers
- 42 Aharoni, A. and Vorst, O. (2002) DNA microarrays for functional plant genomics. *Plant Mol. Biol.* 48, 99–118
- 43 Varshney, R.K. *et al.* (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630
- 44 Naylor, R.L. *et al.* (2004) Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Policy* 29, 15–44
- 45 Nelson, R.J. *et al.* (2004) The role of genomics research in improvement of 'orphan' crops. *Crop Sci.* 44, 1901–1904
- 46 Devos, K.M. (2005) Updating the 'crop circle'. *Curr. Opin. Plant Biol.* 8, 155–162
- 47 Goodman, R.M. *et al.* (1987) Gene transfer in crop improvement. *Science* 296, 92–100
- 48 Morgante, M. and Salamini, F. (2003) From plant genomics to breeding practice. *Curr. Opin. Biotech.* 14, 214–219
- 49 Jones-Rhoades, M.W. *et al.* (2006) MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* 57, 19–53
- 50 Bennett, M.D. and Leitch, I.J. (1995) Nuclear DNA amounts in angiosperms. *Ann. Bot.* 76, 113–176
- 51 Flavell, R.B. *et al.* (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12, 257–269
- 52 Cone, K.C. *et al.* (2002) Genetic, physical, and informatics resources for maize on the road to an integrated map. *Plant Physiol.* 130, 1598–1605
- 53 Draye, X. *et al.* (2001) Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains using the sorghum genome as a foundation. *Plant Physiol.* 125, 1325–1341
- 54 Toojinda, T. *et al.* (1998) Introgression of quantitative trait loci (QTLs) determining stripe rust resistance in barley: an example of marker-assisted line development. *Theor. Appl. Genet.* 96, 123–131
- 55 Cooperative Research Centre for Molecular Plant Breeding (CRCMPB). *Putting Plant Breeding into Fast-forward*, Annual Report 2001/2002, p. 51 ([www.molecularplantbreeding.com/pdfs/CRCMPB-Annual-Report-2001-02.pdf](http://www.molecularplantbreeding.com/pdfs/CRCMPB-Annual-Report-2001-02.pdf))
- 56 Jefferies, S.P. *et al.* (2003) Marker-assisted backcross introgression of the *yd2* gene conferring resistance to barley yellow dwarf virus in barley. *Plant Breed.* 122, 52–56
- 57 Buschges, R. *et al.* (1997) The barley *mlo* gene: a novel control element of plant pathogen resistance. *Cell* 88, 695–705
- 58 Wei, F. *et al.* (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS–LRR gene families and suppressed recombination within a 240 kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* 153, 1929–1948
- 59 Shirasu, K. *et al.* (1999) A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. *Cell* 99, 355–366
- 60 Brueggeman, R. *et al.* (2002) The barley stem rust resistance gene *Rpg1* is a novel disease resistance gene with homology to receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9328–9333
- 61 Stein, N. *et al.* (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive bymovirus resistance in *Hordeum vulgare* (L.). *Plant Jour.* 42, 912–922
- 62 Dreher, K. *et al.* (2000) Is marker assisted selection cost-effective compared to conventional plant breeding methods? The case of quality protein maize. In *Proc 4th Annu Conf Intern Consor on Agricultural Biotechnology Research (ICABR), The Economics of Agricultural Biotechnology*, Ravello, Italy (<http://www.economia.uniroma2.it/Conferenze/icabr00/abstracts/morris.htm>)
- 63 Collins, N. *et al.* (1999) Molecular characterization of the maize *Rp1 D* rust resistance haplotype and its mutants. *Plant Cell* 11, 1365–1376
- 64 Salvi, S. *et al.* (2002) Toward positional cloning of *Vgt1*, a QTL controlling the transition from the vegetative to the reproductive phase in maize. *Plant Mol. Biol.* 48, 601–613
- 65 Toenniessen, G.H. *et al.* (2003) Advances in plant biotechnology and its adoption in developing countries. *Curr. Opin. Plant Biol.* 6, 191–198
- 66 Sanchez, A.C. *et al.* (2000) Sequence-tagged site marker-assisted selection for three bacterial blight resistance genes in rice. *Crop Sci.* 40, 792–797
- 67 Singh, S. *et al.* (2001) Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker assisted selection into *indica* rice cultivar PR106. *Theor. Appl. Genet.* 102, 1011–1015
- 68 He, Y. *et al.* (2004) Gene pyramiding to improve hybrid rice by molecular-marker techniques. In *New Directions for a Diverse Planet: Proc. 4th Intern. Crop Sci. Cong.* Brisbane, Australia ([http://www.cropscience.org.au/icsc2004/](http://www.cropsscience.org.au/icsc2004/))
- 69 Narayanan, N.N. *et al.* (2002) Molecular breeding for the development of blast and bacterial blight resistance in rice cv. IR50. *Crop Sci.* 42, 2072–2079
- 70 Joseph, M. *et al.* (2004) Combining bacterial blight resistance and Basmati quality characteristics by phenotypic and molecular-marker assisted selection in rice. *Mol. Breed.* 13, 377–387
- 71 Yoshimura, S. *et al.* (1998) Expression of *Xa1*, a bacterial blight resistance gene in rice is induced by bacterial inoculation. *Proc. Natl. Acad. Sci. U.S.A.* 95, 1663–1668
- 72 Iyer, A.S. and McCouch, S.R. (2004) The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Mol. Plant Microbe Inter.* 17, 1348–1354
- 73 Song, W.Y. *et al.* (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270, 1804–1806
- 74 Sun, X. *et al.* (2004) *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase like protein. *Plant Jour.* 37, 517–527
- 75 Wang, Z.X. *et al.* (1999) The *Pib* gene for rice blast resistance belongs to the nucleotide-binding and leucine-rich repeat class of plant disease resistance genes. *Plant Jour.* 19, 55–64
- 76 Bryan, G.T. *et al.* (2000) A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene *Pi ta*. *Plant Cell* 12, 2033–2046
- 77 Jeon, J.S. *et al.* (2003) Genetic and physical mapping of *Pi5(t)*, a locus associated with broad spectrum resistance to rice blast. *Mol. Genet. Genomics* 269, 280–289
- 78 Qu, S.H. *et al.* (2006) The broad spectrum blast resistance gene *Pi9* encodes a nucleotide-binding site leucine-rich repeat protein and is a member of a multigene family in rice. *Genetics* 172, 1901–1914

- 79 Ashikari, M. *et al.* (1999) A rice gibberellin-insensitive dwarf mutant gene *Dwarf1* encodes the alpha subunit of GTP binding protein. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10284–10289
- 80 Miyoshi, K. *et al.* (2004) *PLASTOCHRON1*, a timekeeper of leaf initiation in rice, encodes cytochrome P450. *Proc. Natl. Acad. Sci. U.S.A.* 101, 875–880
- 81 Yamanouchi, U. *et al.* (2002) A rice spotted leaf gene, *Spl7*, encodes a heat stress transcription factor protein. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7530–7535
- 82 Spielmeier, W. *et al.* (2002) Semidwarf (*sd 1*), 'green revolution' rice, contains a defective gibberellin 20 oxidase gene. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9043–9048
- 83 Konishi, S. *et al.* (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396
- 84 Li, C. *et al.* (2006) Rice domestication by reducing shattering. *Science* 311, 1936–1939
- 85 Yano, M. *et al.* (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* 12, 2473–2484
- 86 Kojima, S. *et al.* (2002) *Hd3a*, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of *Hd1* under short day conditions. *Plant Cell Physiol.* 43, 1096–1105
- 87 Lin, H. *et al.* (2003) Fine mapping and characterization of quantitative trait loci *Hd4* and *Hd5* controlling heading date in rice. *Breed. Sci.* 53, 51–59
- 88 Takahashi, Y. *et al.* (2001) *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7922–7927
- 89 Doi, K. *et al.* (2004) *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT like gene expression independently of *Hd1*. *Genes Dev.* 18, 926–936
- 90 Ashikari, M. *et al.* (2005) Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745
- 91 Ren, Z.H. *et al.* (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nature Genet.* 37, 1141–1146
- 92 Magalhaes, J.V. *et al.* (2006) High-resolution mapping and cloning of *AltSB*, a major aluminium-tolerance gene in sorghum. In *Plant and Animal Genome Conference XIV*, p. 61 ([http://www.intl-pag.org/14/abstracts/PAG14\\_P61.html](http://www.intl-pag.org/14/abstracts/PAG14_P61.html))
- 93 Feuillet, C. *et al.* (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15253–15258
- 94 Huang, L. *et al.* (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. *Genetics* 164, 655–664
- 95 Yahiaoui, N. *et al.* (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. *Plant Jour.* 37, 528–538
- 96 Griffiths, S. *et al.* (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. *Nature* 439, 749752
- 97 Sutton, T. *et al.* (2003) The *Ph2* pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach. *Plant Jour.* 36, 443–456
- 98 Yan, L. *et al.* (2003) Positional cloning of the wheat vernalization gene *VRN1*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 6263–6268
- 99 Yan, L. *et al.* (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303, 1640–1644
- 100 Simons, K.J. *et al.* (2006) Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172, 547–555
- 101 Liu, S. and Anderson, J.A. (2003) Targeted molecular mapping of a major wheat QTL for *Fusarium* head blight resistance using wheat ESTs and synteny with rice. *Genome* 46, 817–823
- 102 Kota, R. *et al.* (2006) Fine genetic mapping fails to dissociate durable stem rust resistance gene *Sr2* from pseudo black chaff in common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 112, 492–499
- 103 Spielmeier, W. *et al.* (2005) Powdery mildew resistance and *Lr34/Yr18* genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat. *Theor. Appl. Genet.* 111, 731–735
- 104 Schnurbusch, T. *et al.* (2004) Dissection of quantitative and durable leaf rust resistance in Swiss winter wheat reveals a major resistance QTL in the *Lr34* chromosomal region. *Theor. Appl. Genet.* 108, 477–484
- 105 Larkin, J.E. *et al.* (2005) Independence and reproducibility across microarray platforms. *Nature Meth.* 2, 337–343

## Free journals for developing countries

In 2002, the WHO and six medical journal publishers launched the Health InterNetwork Access to Research Initiative, which enabled nearly 70 of the world's poorest countries to gain free or reduced-cost access to biomedical literature through the internet. Currently more than 70 publishers are participating in the program, providing access to over 2000 journals.

Gro Harlem Brundtland, former director-general for the WHO, said that this initiative was "perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries".

**For more information, visit [www.who.int/hinari](http://www.who.int/hinari)**