



Overview

## Cereal genomics: ushering in a brave new world

Roberto Tuberosa<sup>1,\*</sup>, Bikram S. Gill<sup>2</sup> and Steve A. Quarrie<sup>3</sup>

<sup>1</sup>Department of Agroenvironmental Science and Technology, University of Bologna, 40126 Bologna, Italy (\*author for correspondence: e-mail tuberosa@agrsci.unibo.it); <sup>2</sup>Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502, USA; <sup>3</sup>John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

### Introduction

The dawn of the third millennium has witnessed an impressive leap forward in the number of approaches, techniques and materials available to tackle what has seemingly become the number one goal in plant as well as animal and human genetics: assigning functions to genes. The economic consequences, patenting included, and potential benefits ensuing from this exercise are enormous and long-lasting. As trivial and obvious as this may sound, it has become clear to most of those engaged in this exercise that the road to gene function discovery is going to be long and winding and the ride will not be a smooth one, especially for quantitative traits (Beavis, 1994; Flint and Mott, 2001). Nevertheless, map-based cloning of QTLs for agronomically valuable traits has already been reported in rice (Yano *et al.*, 2000; Takahashi *et al.*, 2001) and may soon become a reality also in maize (Salvi *et al.*, this issue).

This special issue presents a number of articles illustrating the type of contributions which genomics can offer to unravel the path from genes to phenotypes and *vice versa* in cereals and, indirectly, also in other crops. In this virtual relay, each article shows how structural and/or functional genomics can improve our capacity to uncover and deploy genetic variation useful to improve food quality and security. A number of articles in this issue were presented at a workshop held in April 2000 at the University of Bologna, Italy under the patronage of the Italian Society of Agricultural Genetics (SIGA) and sponsored by the Biotechnology and Biological Sciences Research Council (BBSRC) and the British Council (BC) in Italy and the CARISBO Foundation. The main objective of the workshop was to bring together a number of scientists in order to debate the state of the art in cereal

genomics with a particular emphasis to Europe and the U.S., and to assess the technological gap existing at that time between Europe and the U.S.

The grouping of the articles into five sections reflects their overall objectives. We are well aware that these articles represent only the tip of the iceberg since many other projects are underway in this fast expanding area, especially following the completion of the sequencing of the *Arabidopsis* genome and the recent release of part of the rice sequence. Although some genomics techniques are still in their infancy (e.g., SNP detection) or remain too expensive for routine applications (e.g., microarray analysis), the healthy status of plant genomics is witnessed by initiatives similar to this one (Chandler and Wessler, 2001) as well as by the growing number of meetings and new specialised journals devoted to genomics.

### The foundations of cereal genomics: molecular markers and genetic maps

During the past 15 years, a vast number of cereal genetic maps of varying resolution have been completed. This painstaking work has provided the foundations for the identification and cloning of genes, the comparative analysis of cereal genomes and the critical interpretation of their evolution. In most cases, the level of resolution of maps remains largely inadequate to allow for even a tentative matching between a putative candidate gene and a related phenotype. A valuable approach to improve the resolution achievable with a mapping population of a given size is presented by Lee *et al.* (this issue): the intermating of the maize (B73 × Mo17) F<sub>2</sub> plants allowed for a considerable expansion of the map distances, thus providing a better resolution for mapping studies.

The intermated B73 × Mo17 population was used by Sharopova *et al.* (this issue) to map over 900 SSRs in order to develop what now represents the most detailed and comprehensive SSR-based map of any plant species. Besides contributing greatly to research in maize genetics, this type of information will be particularly valuable for investigating genetic relatedness, a topic of great relevance also for commercial issues related to essential derivation and breeders' rights. Marker-assisted selection will also be facilitated by the availability of a well-saturated SSR-based map. In sorghum, Menz *et al.* (this issue) have assembled a map of ca. 3000 markers, most of which are AFLPs. The ca. 200 RFLPs representing cDNA and genomic clones mapped in other grass species will provide essential reference markers in comparative mapping. AFLPs filled most of the gaps left by the RFLP/SSR markers demonstrating the effectiveness of AFLP technology in providing excellent genome coverage.

High resolution genetic maps and contigued physical maps will both play critical roles for positional cloning. The availability of contigued BAC/YAC physical maps will allow for the quick mapping of clones monomorphic between parent lines of a mapping population, thus substantially reducing the efforts and time required for positional cloning. A particularly promising avenue for identifying candidate genes is the construction of functional maps based on ESTs: Kantety *et al.* (this issue) indicate how bioinformatics can help us in identifying a wealth of SSRs in coding regions cross-transferable across cereal species. The high informativeness of SSRs has been instrumental for identifying genomic regions influencing adaptation to a range of environmental conditions in wild barley accessions: Ivandic *et al.* (this issue) report associations of SSRs with site-of-origin ecological and geographic data primarily in genomic regions determining plant development.

The next frontier for genome mapping lies in the detection of SNPs and their exploitation through a technological platform allowing for the simultaneous analysis of several hundred SNP loci. Microarrays with series of oligos are being tested for this purpose: Kanazin *et al.* (this issue) report the progress in identifying and exploiting this valuable class of markers in barley. An additional contribution to fine and high-throughput mapping will be provided by the exploitation of indels (insertion-deletions) in the 3' region of maize (Bhatramakki *et al.*, this issue). The detection and utilisation of SNPs and indels is particularly

suitable for the application of non-gel-based techniques in the high-throughput screening required for marker-assisted selection programs and other breeding applications (Salvi *et al.*, 2001).

### Functional genomics

The implementation of new technological platforms also allows us to investigate gene functions and their products at the genome level. Transcriptome profiling will be particularly valuable to investigate the coordinated expression of a suite of genes involved in specific biochemical pathways or in the adaptive response to biotic and abiotic stresses. An example is provided by Ozturk *et al.* (this issue) who have investigated the expression profiles of ca. 1500 drought-related genes in barley subjected to drought and salt stress: a large portion of these transcripts are novel and functionally unknown. Among others, two major shortcomings of expression profiling are that it cannot account for any post-transcriptional modifications and that the expression level detected for a particular gene does not necessarily relate to the abundance and activity of the corresponding protein. Consequently, the integration of transcriptome data with proteome and metabolome data will be necessary to obtain a more realistic view of the role of specific genes. In maize, extensive work has already been carried out using proteome analysis (deVienne *et al.*, 1999). Consoli *et al.* (this issue) review these results and indicate how they can best be integrated with transcriptome analysis for the dissection of the genetic basis of quantitative traits.

### Finding the bricks: QTLs and genes

While forward and reverse genomic approaches now offer unprecedented opportunities for investigating gene functions, the difficulties in producing and evaluating the necessary genetic stocks coupled with gene redundancy and gene interactions have limited the number of success stories in cereals, particularly when compared with *Arabidopsis*. Our capacity to improve steadily cereal production and quality will depend increasingly on our ability to identify and manipulate the genes controlling quantitative traits. Although until now such capacity has been rudimentary, progress in statistical methods allows for a more precise localisation of the QTL peaks. The review presented by Hackett (this issue) summarizes the progress achieved

in the past decade and highlights the main shortcomings still existing in QTL analysis applied to experimental populations derived from inbred lines. The conclusion is that the available methods are adequate for marker-assisted backcrossing but still inadequate for map-based cloning. Progress in QTL cloning in maize is presented by Salvi *et al.* (this issue) who have adopted a positional cloning strategy in order to identify and isolate the gene/s underlying an important QTL (*Vgt1*) for flowering time: the high resolution mapping obtained by testing a large mapping population derived from the cross of two near isogenic lines (NILs) has confined *Vgt1* to an interval amenable to chromosome walking.

An approach complementary to QTL discovery on the basis of mapping populations relies on the exploitation of the residual linkage disequilibrium present at target regions in large germplasm collections, as shown in maize (Thornsberry *et al.*, 2001). The advantage of this method is that it provides a much broader picture of the relevance of allelic variation at specific loci. A drawback to the application of a linkage disequilibrium approach at the whole genome level is the extremely high number of markers required to obtain a meaningful genetic resolution. Although encouraging results have been reported in a study conducted in maize at the genome level (Vuylsteke *et al.*, 2000), a widespread scepticism remains as to the usefulness of a genome-wide approach based on linkage disequilibrium (Remington *et al.*, 2001). The possibility of using microarray platforms, such as the one described in rice by Jaccoud *et al.* (2001), to assess simultaneously allelic composition for thousands of loci will provide new opportunities for investigating linkage disequilibrium at the whole genome level.

As examples of QTL analysis to locate genes of major agronomic importance, Blanco *et al.* (this issue) report QTLs for protein content in durum wheat, showing that most were also closely linked to QTLs for yield, thereby explaining the correlation between these two traits. Otto *et al.* (this issue) describe an important QTL controlling *Fusarium* head blight resistance in wheat using a population of recombinant inbred lines (RILs). An outstanding example of a major gene underlying a QTL is the discovery of puroindoline genes as the functional basis of grain hardness, a trait of immense importance in wheat commerce (Morris, this issue).

Among abiotic stresses restricting cereal production worldwide, a predominant role is played by drought. The strategies required to search for genes

playing an important role in the response to drought stress have been outlined by Cattivelli *et al.* (this issue). This review article compiles all the information available at present on stress-related genes and QTLs in the *Triticeae* and identifies potential candidate genes for stress-tolerance QTLs. Further experimental evidence on the molecular mechanisms underlying the adaptive response of cereals is provided and discussed by Maestri *et al.* (this issue). In rice, Price *et al.* (this issue) describe a number of QTLs influencing more than one trait related to drought, thus accounting for their correlations; in this study, a poor co-location characterized QTLs for drought avoidance traits and QTLs for root growth. Different results are reported in maize (Tuberosa *et al.*, this issue) where QTLs for root traits measured in hydroponics co-localise with QTLs for grain yield under different water regimes in the field. These findings and those available for QTLs influencing root traits and grain yield in other maize crosses warrant the development of NILs for at least one of these QTL regions. NILs and RILs were evaluated by Sanchez *et al.* (this issue) in sorghum to validate the role of QTLs for leaf stay green, an important trait for resistance to premature senescence under soil moisture stress during the post-flowering period. Collectively, these results offer interesting opportunities for the application of marker-assisted selection to improve drought tolerance in cereals, an approach which has already proven its validity in maize (Ribaut *et al.*, 2000).

### Providing the cement: comparative genomics

The accumulation of mapping information (e.g., mutants, ESTs, QTLs, etc.) as well as the availability of the genome sequence for *Arabidopsis* and, at least in part, for rice make comparative genetics particularly attractive. Laurie and Devos (this issue) review the trends in comparative genetics and their potential impact on wheat and barley research; they also analyze the evolution of gene families and our understanding of angiosperm phylogeny. The relative merits of direct map-based cloning in barley and wheat, the utilization of the smaller genome of rice and gene homology methods relying on information from model species such as *Arabidopsis* are briefly discussed. Comparative mapping between model plant species for which genome sequence is available and crop species has been indicated as one of the strategies for the isolation of agronomically valuable genes. Van Buuren

*et al.* (this issue) tested whether comparative mapping between *Arabidopsis* and maize of a small region (754 kb) surrounding the *DREB1A* gene in *Arabidopsis* could lead to the identification of an orthologous region in maize containing the *DREB1A* homologue: the results indicate that the degree of orthology and colinearity between these two species is insufficient to aid gene prediction and cloning in maize. A comparative analysis, but not in terms of syntenic relationships, between *Arabidopsis* and maize was also carried out by Skibbe *et al.* (this issue) to investigate the evolution of the aldehyde dehydrogenase (ALDHs) gene families. These genes are involved in the restoration of male sterility in maize. Although animal, fungal and plant genomes each encode both mitochondrial and cytosolic ALDHs, it appears that either the gene duplications that generated the mitochondrial and the cytosolic ALDHs occurred independently within each lineage or that homogenizing gene conversion-like events have occurred independently within each lineage.

Genome structure was investigated by Boyko *et al.* (this issue) with a high density map (over 700 markers) in *Aegilops tauschii*, the D-genome donor of bread wheat. Because of conserved synteny, this high-density map of the *Aegilops tauschii* genome will be useful for breeding and genetic studies within the *Triticeae*. The retrotransposon marker clusters are common in the pericentromeric regions of the D genome, while resistance and disease resistance gene clusters are more common in distal/telomeric regions. Surprisingly, pericentromeric regions showed negative crossovers interference. Islands with negative, positive and/or no interference were present in interstitial and distal regions. The implications of these findings for the positional cloning of genes are obvious.

Bread wheat is unsurpassed in the inventory of aneuploid stocks and the recently developed deletion lines provide complete physical genome coverage at a resolution of ca. 15 Mb (Endo and Gill, 1996). The deletion maps that have been developed for the 21 chromosomes of wheat have led to the recognition of gene-rich regions which are recombination hot spots. Sandhu *et al.* (this issue) describe in detail the structural and functional organization of one of these gene rich regions of wheat. Wheat species exist at three ploidy levels (2 $\times$ , 4 $\times$ , 6 $\times$ ) and have provided a classic model of genome evolution by allopolyploidy. Huang *et al.* (this issue) report a two-gene system of acetyl-CoA carboxylase (*ACCase*) and 3-phosphoglycerate kinase (*Pgk*) loci for studying grass phylogeny, sys-

tematics and evolution. Both *ACCase* and *Pgk* have a copy in the plastid and cytosol and are useful for clocking the pace of evolution in the grass family. The divergence rates calculated from the gene sequences were consistent with the milestones of grass evolution revealed by other methods.

The grasses show a 40-fold difference amongst them in DNA content although gene content and gene synteny are largely conserved. Bennetzen and Ramakrishna (this issue) show that genome obesity is caused by relative amplification rates of retrotransposons in different cereal species. Furthermore, they document examples of gene rearrangements and orientation at the microsyntenic level. Recent data show that breaks in synteny coincided with genome amplification in the *Triticeae* species (Li and Gill, 2002).

### Building the future: towards 'holistomics'?

In this technology-driven era, it is easy to be lured by the sophisticated platforms available for the profiling of the transcriptome (Lockhart and Winzeler, 2000), proteome (MacBeath and Schreiber, 2000) and metabolome (Fiehn *et al.*, 2000). This wealth of information can provide us with a highly integrated and comprehensive view, hence the term 'holistomics', of the molecular events occurring in different organs and tissues in response to those developmental and environmental cues that ultimately determine cereal yield and its quality. In the next decade, the amount of information generated both *in vivo* and *in silico* will increase exponentially. The ultimate challenge will be to manage such a deluge of information properly (Hess *et al.*, 2001) in order to characterize allelic variation (natural and/or artificially induced in cultivated species as well as their wild relatives) and harness its potential to promote and stabilize productivity in a broader range of environments (Mifflin, 2000). Nonetheless, it is clear that an increased amount of molecular information will not automatically lead to an improved knowledge useful in agricultural terms. An excessive emphasis at the molecular level coupled with an insufficient understanding of the physiology of crops under field conditions will limit the contribution of molecular information on the release of improved cultivars. An example of this is the still scanty evidence available on the transferability to field conditions of the impressive results achieved in the past decade through the overexpression of single

genes conferring tolerance to abiotic stresses under controlled conditions.

Both structural and functional genomics provide excellent opportunities for multidisciplinary approaches. One of such opportunities is offered by the creation and utilization of NILs at target QTLs which can then be used by different categories of scientists for a number of academic and practical purposes. Under this aspect, the remarkable and spectacular progress achieved in genomics is witnessed by the fact that only a decade ago it was argued that the development of NILs was restricted to traits controlled by one or only a few genes (Ludlow and Muchow, 1990).

The articles presented in this issue document how cereal genomics can contribute to usher in the post-genomics era, an era likely to be dominated by bioinformatics and its applications. The availability of the genome sequence of rice will spur a wealth of comparative and functional studies in other cereals. Although the conservation of microsynteny between rice and the other cereals is often incomplete, the possibility to utilise the rice sequence as a source of markers and candidate genes will play a pivotal role in cloning and assigning functions to genes in cereals. The availability of sequence information will also allow for the application of TILLING (Targeting Induced Local Lesion IN Genomes; Colbert *et al.*, 2001) and the other reverse genetics tools necessary to validate the role of putative candidate genes and identify new genetic variation at such loci. As more pieces of the cereal genomic jigsaw puzzle are being patiently put together, a more unifying picture emerges for the crops that feed mankind. Clearly, cereal genomics and its applications have a great and unrivalled potential to shape the future of agriculture and its sustainability.

## References

- Beavis, W.D. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings of the 49th Annual Corn and Sorghum Research Conference American Seed Trade Association, Washington, D.C., pp. 250–266.
- Chandler, V.L. and Wessler, S. 2001. Grasses. A collective model genetic system. *Plant Physiol.* 125: 1155–1156.
- Colbert, T., Till, B.J., Tompa, R., Reynolds, S., Steine, M.N., Yeung, A.T., McCallum, C.M., Comai, L. and Henikoff, S. 2001. High-throughput screening for induced point mutations. *Plant Physiol.* 126: 480–484.
- DeVienne, D., Leonardi, A., Damerval, C. and Zivy, M. 1999. Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. *J. Exp. Bot.* 50: 303–309.
- Endo, T.R. and Gill, B.S. 1996. The deletion stocks of common wheat. *J. Hered.* 87: 295–307.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. 2000. Metabolite profiling for plant functional genomics. *Nature Biotechnol.* 18: 1157–1161.
- Flint, J. and Mott, R. 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Rev. Gen.* 2: 437–445.
- Hess, K.R., Zhang, W., Baggerly, K.A., Stivers, D.N. and Coombes, K.R. 2001. Microarrays: handling the deluge of data and extracting reliable information. *Trends Biotech.* 11: 463–468.
- Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29: e25.
- Li, W.-L. and Gill, B.S. 2002. The colinearity of the *Sh2/A1* homologous regions in rice, sorghum and maize is interrupted and accompanied by genome expansion in the *Triticeae*. *Genetics* (in press).
- Lockhart, D.J. and Winzler, E.A. 2000. Genomics, gene expression and DNA arrays. *Nature* 405: 827–836.
- Ludlow, M.M. and Muchow, R.C. 1990. A critical evaluation of traits for improving crop yields in water-limited environments. *Adv. Agron.* 43: 107–153.
- MacBeath, J. and Schreiber, S.L. 2000. Printing proteins as microarrays for high throughput function determination. *Science* 289: 1760–1763.
- Mifflin, B. 2000. Crop improvement in the 21st century. *J. Exp. Bot.* 51: 1–8.
- Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Joebley, J., Kresovich, S., Goodman, M.M. and Buckler IV, E.S. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. USA* 20: 11479–11484.
- Ribaut, J.M., Edmeades, G., Perotti, E. and Hoisington, D. 2000. QTL analysis, MAS results, and perspectives for drought-tolerance improvement in tropical maize. In: Ribaut, J.M. and Poland, D. (Eds.) *Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments*. CIMMYT, El Batán, Mexico, 21–25 June 1999, pp. 131–136.
- Salvi, S., Tuberosa, R., Phillips, R.L., 2001. Development of PCR-based assays for allelic discrimination in maize by using the 5-nuclease procedure. *Mol. Breed.* 8: 169–176.
- Takahashi, Y., Shomura, A., Sasaki, T. and Yano, M. 2001. *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the  $\alpha$  subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. USA* 98: 7922–7927.
- Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D. and Buckler, E.S. 2001. Dwarf8 polymorphisms associated with variation in flowering time. *Nature Genetics* 28: 286–289.
- Vuylsteke, M., Kuiper, M. and Stam, P. 2000. Chromosomal regions involved in hybrid performance and heterosis: their AFLP®-based identification and practical use in prediction models. *Heredity* 85: 208–218.
- Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y. and Sasaki, T. 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.