Minireview

Unfashionable crop species flourish in the 21st century

Wayne Powell* and Peter Langridge†

Addresses: *Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK. †The Australian Centre for Plant Functional Genomics, School of Agriculture and Wine, Waite Campus, University of Adelaide, SA 5064, Australia.

Correspondence: Wayne Powell. E-mail: Wayne.Powell@adelaide.edu.au

Published: 14 June 2004

Genome ***Biology*** 2004, **5:**233

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2004/5/7/233

Abstract

Genome-level studies are contributing to a major renaissance in crop science. In wheat, there are now more than 500,000 expressed sequence tags, and these are being used in conjunction with specially designed deletion stocks to unravel patterns of genome evolution, recombination and polyploid genome behavior.

The genomic era was founded on the study of a limited number of model organisms [1] that were chosen for their small genome size and experimental tractability. The use of model organisms can be powerful because a community of scientists can work collectively on a single organism, but it also encourages a reductionist approach. Intriguingly, the study of diversity and organism complexity is now gaining more prominence, often at the expense of research on model organisms. The mapping of a large number of wheat expressed sequence tags (ESTs) [2], physical mapping of the wheat genome [3], studies of synteny between related parts of the wheat genome [4,5] and between wheat and other cereals [6], and studies of the organization of sequence polymorphism into haplotypes [7] are big steps forward. These developments in crop genomics vividly illustrate how, although model organisms provide good starting points, their significance may decline as accessibility to genome technologies improves and the social and biological relevance of crop science to the public continues to gain prominence.

The complex wheat genome

Before the emergence of molecular biology, crop plants such as bread wheat (*Triticum aestivum* L.) were considered to be good models for cytogenetic investigations and research into polyploidy. Wheat has one of the largest and most complex genomes known: it is an allopolyploid, containing three different ancestral genomes (designated A, B and D), each of which contains seven pairs of homologous chromosomes. The number of chromosomes in the diploid genome (2n) is therefore 42; this number is also referred to as 6x, as each of the six ancestral genomes has seven chromosomes. The homologous chromosomes and genes in different ancestral genomes are referred to as 'homoeologous'. Although the ancestral genomes are very similar in gene content and gene order, chromosome pairing at meiosis is under genetic control and is restricted to homologous chromosomes. This results in disomic inheritance, as if there were only two sets of 21 chromosomes, which greatly simplifies the pattern and interpretation of genetic segregation data. The size of the wheat genome - 16,000 megabases, approximately three times the size of the human genome - was initially viewed as an impediment to genomic research, and most attention in plants was focused on the small genomes of model plants such as *Arabidopsis* and the smaller-genome crop plant rice. One advantage of polyploidy, however, is that it provides a huge capacity for 'buffering' mutations, as homoeologous genes can make up for the loss of any deleted genes. This has allowed the creation of an unparalleled array of aneuploid stocks that have been used as a resource to locate genes controlling agronomic and biochemical traits and also to find

genes responsible for chromosome pairing. The work of coordinated, global genomic initiatives is dramatically changing the knowledge base for wheat research, and is leading to a renaissance in crop science. This is particularly true in studies of the fascinating relationship between recombination, synteny and genome evolution and of the regulation of gene expression in polyploid organisms.

Recombination and genome evolution

The significance of polyploidy as a basis for chromosome engineering (using aneuploid stocks) has long been recognized, but the use of aneuploid and deletion lines to elucidate the location of genes has been revitalized in the US by the fact that the National Science Foundation (NSF) has funded creation of EST libraries for gene discovery and the physical mapping of these ESTs using aneuploid and deletion lines [2]. As of March 2004, the National Center for Biotechnology Information (NCBI) dbEST database contained 554,289 wheat ESTs from more than 60 different tissues, representing the most extensive EST database available for any plant species. The power of this resource becomes apparent when it is coupled with the use of deletion lines that have been assembled over the past 70 years. In total, 101 deletion lines representing 119 deletions - including deletions within chromosome arms, missing chromosomes (nullisomic-tetrasomic stocks) and missing chromosome arms (ditelosomic stocks) - have been assembled into a panel providing an average of 13 deletions per chromosome [8]. Using this panel, ESTs can be mapped cytologically and physically to one of 159 deletion 'bins' (a bin is a region defined by two adjacent deletion breakpoints in the same chromosome arm) or to one of 21 centromeres.

The mapped ESTs are now being used to study patterns of genome evolution and to initiate cross-genome comparative studies. It has been known for some time that recombination in wheat chromosomes is focused in the telomeric regions: the position of a gene along the chromosome affects its exposure to recombination activity. Genes subject to rapid change - for example, the majority of race-specific disease-resistance genes - are located in the recombinogenic telomeric regions, whereas more highly conserved genes tend to be positioned closer to the centromere. The large size of wheat chromosomes appears to provide a mechanism for developing and maintaining a strong recombination gradient along the chromosomes. Akhunov *et al.* [4] established that synteny between homoeologous wheat chromosomes is inversely proportional to the recombination rate at each relative position along the chromosome. The clear and important result is thus that synteny levels decrease with distance along the centromere-telomere axis. The authors conclude that regions of homoeologous chromosomes with high recombination rates lose synteny faster than do regions of low recombination. Thus, recombination has been a central factor in the evolution of wheat genome organization. The restricted opportunities for recombination because of the self-pollinating nature of wheat reinforces this phenomenon.

A related paper by the same group [5] addresses the question arising from the results of the first study [4]: is recombination a causative agent for genome evolution? This question has not yet been addressed fully in plants. The distal regions of wheat chromosomes have previously been suggested to be gene-rich [9]; this conclusion should be regarded with caution, however, because the selection of markers used may have been biased towards those originating from the distal, high-recombination region of wheat chromosomes. Akhunov *et al.* [5] confirmed that the recombination rate increases along the centromere-telomere axis and found a weak but statistically significant correlation between relative gene density and bin position along the centromere-telomere axis, supporting the observation that gene density increases with distance from the centromere. Further analyses [5] revealed that single-gene loci predominate in the proximal, low-recombination regions of the genome, whereas multi-gene loci consisting of tandemly-duplicated genes were more frequent in distal, high-recombination regions. Two clear messages emerge from these studies [4,5]. Firstly, recombination has influenced the evolution of the wheat genome, with more rapid rates of evolution being observed in the distal regions of wheat chromosomes. This will help with making predictions for the best positional cloning strategies for wheat. Secondly, the studies conducted on wheat [4,5] reveal an evolutionary mechanism that would have been difficult to detect and validate in model organisms.

As well as providing insights into the evolution of wheat, recent studies are also shedding light on the relationship between wheat and its close relatives. The 'unified grass genome' model proposes that different grass genomes have undergone sufficiently little rearrangement for them to be studied effectively as a single syntenic genome; this is a topic of considerable controversy [10]. Recently, Sorrells *et al.* [6] provided much-needed quantitative information on colinearity between cereal genomes (a subset of the domesticated grasses) at the sequence level. Approximately 4,485 ESTs that had been physically mapped into bins along wheat chromosomes were compared using the NCBI BLASTN algorithm [11] to the first draft of the publicly available rice (*Oryza sativa* L.) genome sequence. The resolution of this study was higher than that of previous studies comparing rice-wheat synteny, and it shows significant discontinuity in gene order between rice and wheat as well as the plasticity of cereal genomes. As outlined by Delseny [10], the prior reliance on ancestral shared synteny as a tool to isolate genes from complex genomes therefore now needs to be reconsidered, reinforcing the conclusions of Sorrells *et al.* [6], who emphasized the need to build and establish genomic resources in the species of interest.

Wheat genomic resources

Because the level of synteny between cereal genomes is lower than anticipated, genomics platforms need to be established for each species of interest. In the case of wheat, the extensive EST collection is complemented by bacterial artificial chromosome (BAC) libraries for *Triticum monococcum*, the donor of the ancestral A genome [12], *Aegilops tauschii*, the donor of the D genome [13], a durum wheat (which is tetraploid and has the A and B genomes of wheat), the cultivar Langdon [14], and the hexaploid cultivars Chinese Spring and Renan. Physical maps are an invaluable resource for the positional cloning of genes identified using forward genetics: physical map construction is at an advanced stage for the D genome of wheat, with more than 447,000 clones assembled with an average 17-fold coverage of the D genome [3]. Figure 1 illustrates how the cereal genetics and genomics community is assembling and integrating different technologies in order to make connections between phenotypes, genomes, genes and functional alleles. Recent examples of successful approaches to positional cloning of genes in wheat include the isolation of the leaf-rust resistance gene *Lr10* [15] and of the genes *VRN1* [16] and *VRN2* [17] that are important for vernalization (the induction of seedling growth after a period of cold). Significantly, these studies reveal that *Arabidopsis* and the temperate grasses developed different vernalization pathways that include different genes and regulatory profiles.

As more sequence information becomes available for wheat, more emphasis is being placed on discovering and analyzing intraspecific sequence polymorphism [18]. Wheat ESTs have been exploited as a source of new markers such as simple-sequence repeats [19-21]. Given that various genotypes are represented in the EST database, comparisons between ESTs can identify potential polymorphisms between accessions (plants of different genotype). The electronic discovery of single nucleotide polymorphisms (SNPs) in wheat is complicated, however, by the triplication of genetic information in the hexaploid genome, resulting in the need to distinguish inter-genome polymorphisms (between the A, B and D genomes) from intervarietal polymorphisms. Experimental validation is therefore necessary and requires the generation of genome-specific amplicons that are tested in an aneuploid genetic background provided by the nullisomic-tetrasomic

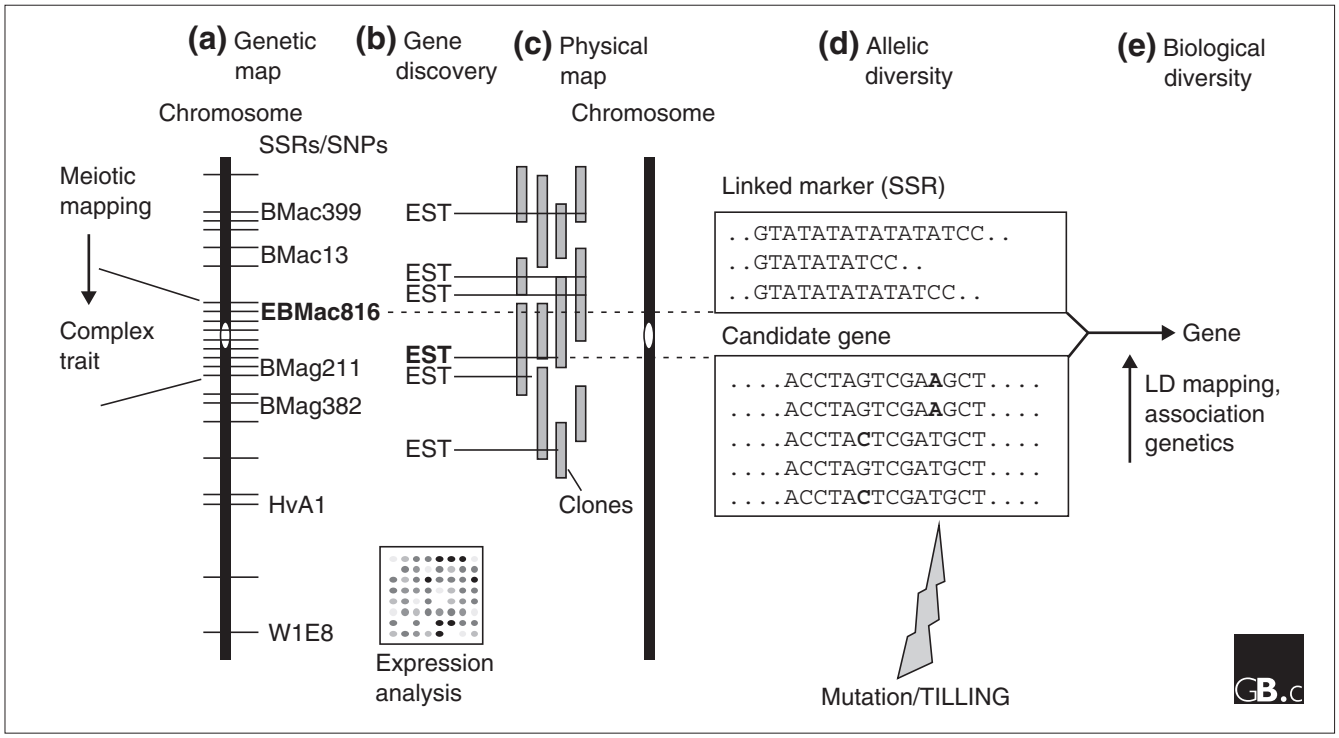

Figure 1

The various kinds of analysis that are being applied to the wheat genome. **(a)** Markers such as single nucleotide polymorphisms (SNPs) and simple-sequence repeats (SSRs) are used in meiotic mapping to narrow down a complex trait to a region of a chromosome. **(b)** ESTs are used to discover new candidate genes within the chromosomal region of interest, and their expression is analyzed using microarrays and other techniques. **(c)** The ESTs are mapped onto the clones that make up a physical map of the genome. **(d)** Allelic diversity, such as the variable-length repeat markers linked to a gene (upper box) and/or SNPs or point mutations inside or outside the gene itself (lower box), can be used for mapping; mutations can be produced using mutagenesis, including using 'target-induced local lesions in genomes' (TILLING, a technique that creates point mutations through chemical mutagenesis and then screens for lesions using high-throughput genotyping methods). **(e)** Linkage disequilibrium (LD) mapping and mapping of the association of markers with the phenotype or quantitative trait of interest can then be used to identify the gene responsible for the trait.

lines of wheat. An example of SNPs detected at the inter-genomic level is in the gene encoding granule-bound starch synthase (GBSS; shown in Figure 2). Somers *et al.* [22] have reported the identification of SNPs by mining the wheat EST database. The overall frequency of sequence variants was one SNP per 24 base-pairs (bp) for homoeologous sequence variants and one SNP per 540 bp between cultivars.

The organization of sequence polymorphism into haplotypes provides an opportunity to unravel the evolutionary history of crop plants. Caldwell *et al.* [7] have recently generated haplotype information specific to the D genome and used it to establish that cultivated wheat originated recurrently, with at least two genetically distinct progenitors contributing to the D genome. A large program funded by the NSF in the US recently commenced with the aim of identifying and mapping 1,800 SNPs across the wheat genome. The information generated from this program will provide a powerful tool for analysis of the genome structure in wheat in far greater detail than has been possible to date.

Gene expression studies in polyploid organisms

Wheat is also emerging as a model for research into the behavior of polyploid genomes, as illustrated by the use of two methods for investigation of gene expression: microarrays and ESTs generated from diverse tissues. Polyploidy is often associated with rapid genetic and epigenetic changes [23]. DNA microarrays have been used to study the effect of autopolypoidy on gene expression in yeast [24], and such an approach may be useful for investigating patterns of gene expression for homoeologous wheat genes. Novel patterns of gene expression occur in polyploids that are not observed in diploid progenitor species [23]. The expression patterns of homoeologous genes in wheat can alternatively be studied using ESTs generated from diverse tissues; one EST study has shown that among sets of homoeologous genes, the gene from one ancestral genome can be expressed while the homoeologs from one or both of the remaining ancestral genomes are silent [25]. More surprisingly, the tissue-specificity was also found to differ between homoeologous genes; for example, a gene in one ancestral genome may be

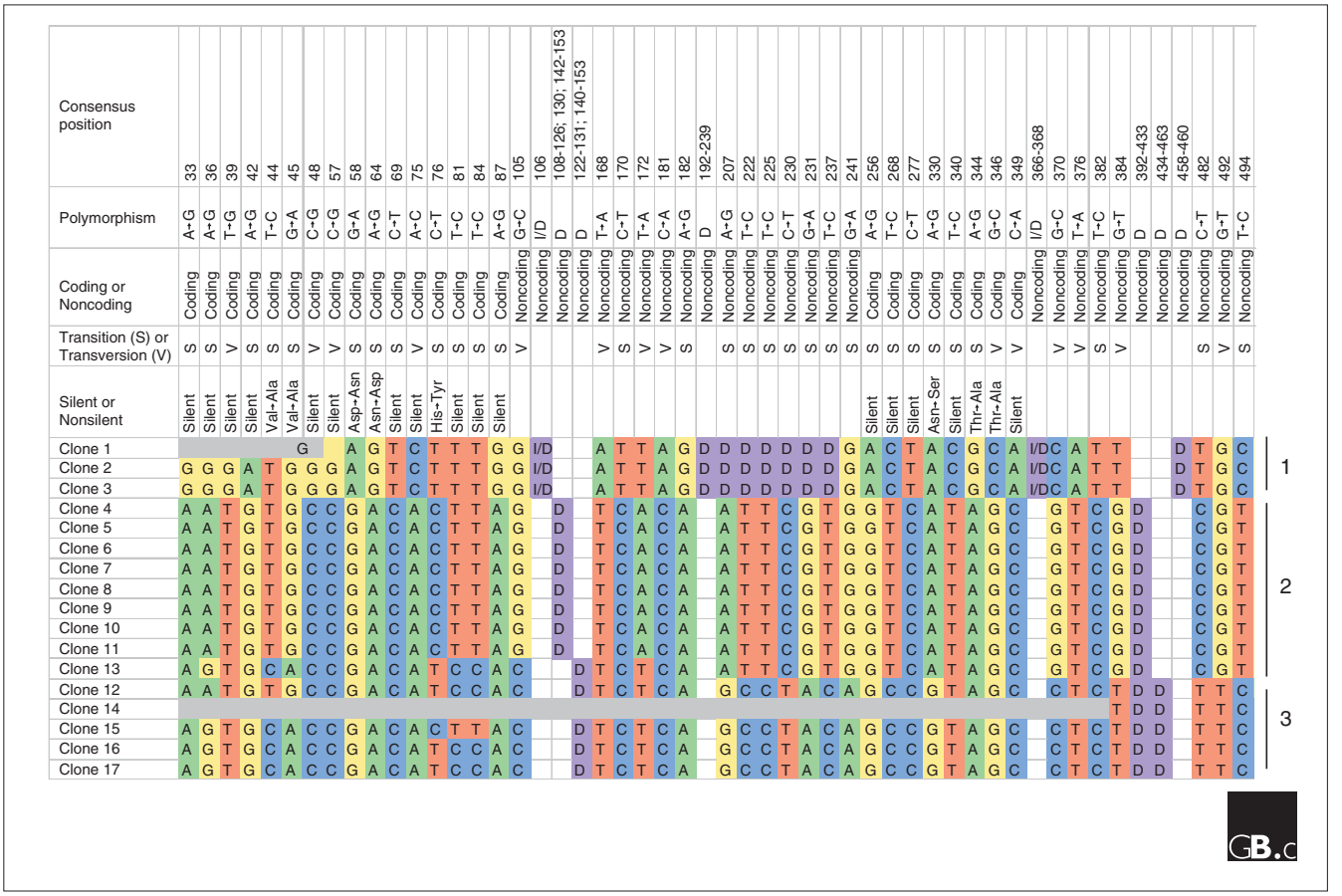

Figure 2
Sample sequencing of 17 clones with primers for the granule-bound starch synthase (GBSS) gene in a single hexaploid wheat accession resulted in the identification of three distinct haplotypes (numbered on the right). These haplotypes must represent inter-genome polymorphism (between the ancestral A, B and D genomes) rather than inter-varietal polymorphism, as they come from a single accession. D, deletion; I, insertion.

expressed only during early grain development whereas the homoeologs are expressed exclusively in leaf tissue [25].

The mechanisms that control chromosome pairing in polyploids are particularly advanced in wheat; several loci have been shown to control pairing and to allow the diploid-like behavior of wheat chromosomes. Genes with the strongest effects on pairing are *Ph1* on the long arm of chromosome 5B and *Ph2* on the short arm of chromosome 3D (both are suppressors of pairing). The *Ph1* locus has been delineated to a region containing fewer than seven genes [26], comparative and functional genomics based approaches are being used to further resolve both the *Ph1* and *Ph2* regions [27], and wheat may prove to be the first plant species for which the genetic basis of chromosome pairing in polyploids can be fully resolved.

It is becoming clear that the distinction between model and crop plants is likely to become blurred as the benefits of public investment in crop genomics becomes more evident. The reality is, however, that opportunities will continue to exist at the interface between model and crop species, where perceived boundaries are rapidly disappearing. Wheat and other crop plants offer notable advantages when compared with model organisms, including the extensive monitoring and archiving of genotypes and associated phenotype data that has already been done and the fact that selective breeding has created unique populations adapted to various environmental conditions. These advantages will become more evident as we enter the post-genomic era. The challenge, therefore, is to synchronize and integrate basic plant science with crop-orientated research to enhance synergy and maximize opportunities for improving crop productivity.

References

1. Davis RH: **The age of model organisms.** *Nat Rev Genet* 2004, **5:**69-76.
2. **The structure and function of the expressed portion of the wheat genomes** [http://wheat.pw.usda.gov/NSF/]
3. Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J: **High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labelling kit and sizing of restriction fragments by capillary electrophoresis.** *Genomics* 2003, **82:**378-389.
4. Akhunov ED, Akhunova AR, Linkiewicz AM, Dubcovsky J, Hummel D, Lazo G, Chao S, Anderson OD, David J, Qi L, *et al.*: **Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates.** *Proc Nat Acad Sci USA* 2003, **100:**10836-10841.
5. Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S, *et al.*: **The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms.** *Genome Res* 2003, **13:**753-763.
6. Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, *et al.*: **Comparative DNA sequence analysis of wheat and rice genomes.** *Genome Res* 2003, **13:**1818-1827.
7. Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo M-C, Wolters P, Powell W: **Haplotype based sequence variation at starch biosynthesis genes provides evidence for recurrent origin of wheat and its relative *Aegilops cylindrica*.** *Genetics* 2004, in press.
8. **The collection of deletion and duplication stocks maintained by the WGRC** [http://www.k-state.edu/wgrc/Germplasm/Stocks/deletion.html]
9. Gill KS, Gill BS, Endo TR: **A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat.** *Chromosoma* 1993, **102:**374-381.
10. Delseny M: **Re-evaluating the relevance of ancestral shared synteny as a tool for crop improvement.** *Curr Opin Plant Biol* 2004, **7:**126-131.
11. **NCBI BLAST** [http://www.ncbi.nlm.nih.gov/blast/]
12. Lukaszewski AJ, Curtis CA: **Physical distribution of recombination in B-genome chromosomes of tetraploid wheat.** *Theor Appl Genet* 1993, **84:**121-127.
13. Moullet O, Zhang, HB, Lagudah ES: **Construction and characterisation of a large DNA insert library from the D genome of wheat.** *Theor Appl Genet* 1999, **99:**305-313.
14. Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J: **Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp *durum*).** *Theor Appl Genet* 2003, **107:**931-939.
15. Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B: **Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome.** *Proc Nat Acad Sci USA* 2003, **100:**15253-15258.
16. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J: **Positional cloning of the wheat vernalization gene *VRN1*.** *Proc Nat Acad Sci USA* 2003, **100:**6263-6268.
17. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J: **The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization.** *Science* 2004, **303:**1640-1644.
18. Rafalski A: **Applications of single nucleotide polymorphisms in crop genetics**. *Curr Opin Plant Biol* 2002, **5:**94-100.
19. Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with the non-repetitive DNA in plant genomes.** *Nat Genet* 2002, **30:**194-200.
20. Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W: **Isolation of EST-derived microsatellites markers for genotyping the A and B genomes of wheat.** *Theor Appl Genet* 2002, **104:**399-407.
21. Leigh F, Lea V, Law J, Wolters P, Powell W, Donini P: **Assessment of EST and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat.** *Euphytica* 2003, **133:**359-366.
22. Somers DJ, Kirkpatrick R, Moniwa M, Walsh A: **Mining single-nucleotide polymorphisms from hexaploid wheat ESTs.** *Genome* 2003, **46:**431-437.
23. Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, Martienssen RA: **Understanding mechanisms of novel gene expression in polyploids.** *Trends Genet* 2003, **19:**141-147.
24. Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR: **Ploidy regulation of gene expression.** *Science* 1999, **285:**251-254.
25. Mochida K, Yamazaki Y, Ogihara Y: **Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags.** *Mol Genet Genomics* 2003, **270:**371-377.
26. Roberts MA, Reader SM, Dalgliesh C, Miller TE, Foote TN, Fish LJ, Snape JW, Moore G: **Induction and characterisation of *Ph1* wheat mutants.** *Genetics* 1999, **153:**1909-1918.
27. Sutton T, Whitford R, Baumann U, Dong C, Able JA, Langridge P: **The *Ph2* pairing homoeologous locus of wheat (*Triticum aestivum*): identification of candidate meiotic genes using a comparative genetics approach.** *Plant J* 2003, **36:**443-456.