

Accessing complex crop genomes with next-generation sequencing

David Edwards · Jacqueline Batley ·
Rod J. Snowdon

Received: 31 January 2012 / Accepted: 8 August 2012 / Published online: 5 September 2012
© Springer-Verlag 2012

Abstract Many important crop species have genomes originating from ancestral or recent polyploidisation events. Multiple homoeologous gene copies, chromosomal rearrangements and amplification of repetitive DNA within large and complex crop genomes can considerably complicate genome analysis and gene discovery by conventional, forward genetics approaches. On the other hand, ongoing technological advances in molecular genetics and genomics today offer unprecedented opportunities to analyse and access even more recalcitrant genomes. In this review, we describe next-generation sequencing and data analysis techniques that vastly improve our ability to dissect and mine genomes for causal genes underlying key traits and allelic variation of interest to breeders. We focus primarily on wheat and oilseed rape, two leading examples of major polyploid crop genomes whose size or complexity present different, significant challenges. In both cases, the latest DNA sequencing technologies, applied using quite different approaches, have enabled considerable progress

towards unravelling the respective genomes. Our ability to discover the extent and distribution of genetic diversity in crop gene pools, and its relationship to yield and quality-related traits, is swiftly gathering momentum as DNA sequencing and the bioinformatic tools to deal with growing quantities of genomic data continue to develop. In the coming decade, genomic and transcriptomic sequencing, discovery and high-throughput screening of single nucleotide polymorphisms, presence–absence variations and other structural chromosomal variants in diverse germplasm collections will give detailed insight into the origins, domestication and available trait-relevant variation of polyploid crops, in the process facilitating novel approaches and possibilities for genomics-assisted breeding.

Introduction

Polyploidy is a key evolutionary mechanism that has been a driving factor in the success of some of the most important domesticated plants (Leitch and Leitch 2008). Hexaploid wheat (*Triticum aestivum*) and oilseed rape/canola (*Brassica napus*) are prime examples of globally important cereal and oilseed plants with large, complex genomes. The bread wheat genome is allohexaploid, consisting of six sets of chromosomes originating from three distinct diploid genomes (termed A, B and D). The three diploid donor species each contributed seven pairs of chromosomes, resulting in 21 chromosome pairs. The donor species diverged between 2.5 and 6 MYA, and two distinct interspecies hybridisation events occurred which serially gave rise to the current polyploid bread wheat varieties. The first event, between 0.5 and 3 MYA, was a hybridisation between *T. urartu* (A^uA^u) and an unknown

Communicated by R. Varshney.

D. Edwards
Australian Centre for Plant Functional Genomics, School
of Agriculture and Food Sciences, University of Queensland,
Brisbane, QLD 4072, Australia

J. Batley (✉)
Centre of Excellence for Integrative Legume Research, School
of Agriculture and Food Sciences, University of Queensland,
Brisbane, QLD 4072, Australia
e-mail: j.batley@uq.edu.au

R. J. Snowdon
Department of Plant Breeding, IFZ Research Centre for
Biosystems, Land Use and Nutrition, Justus Liebig University,
Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

relative of *Aegilops speltoides* (BB), which produced tetraploid wild emmer wheat, *T. turgidum* (A^uA^uBB). The second hybridisation event produced allohexaploid *T. aestivum* (A^uA^uBBDD) through the hybridisation of *T. turgidum* and *Ae. tauschii* (DD) (Eckardt 2001; Huang et al. 2010; Chantret et al. 2005).

Modern bread wheat (AABBDD, $2n = 6x = 42$) was domesticated some 10,000 years ago, its huge success being largely attributable to an increase in grain size early in the domestication process (Gegas et al. 2010). The amphidiploid crucifer *B. napus* (genome AACC, $2n = 4x = 38$) on the other hand is one of the youngest major crop species, originating from a spontaneous hybridisation between the two closely related diploid progenitors *B. rapa* (AA, $2n = 20$) and *B. oleracea* (CC, $2n = 18$) during mediaeval times (Iñiguez Luy and Federico 2011). The three wheat genomes display a high level of conserved synteny to the chromosomes of other grasses (International Brachypodium Initiative 2010), but display significant amplification of repetitive sequence elements and transposons that have expanded the genome to some 40 times that of rice. In contrast, the genome of *B. napus* is only around 10 times as large as that of the closely related crucifer *Arabidopsis thaliana*, but has a highly complex segmental structure in comparison to the model genome (Parkin et al. 2005; Wang et al. 2011a). Ancestral polyploidy (Lysak et al. 2005) and complex patterns of local gene losses and insertions within the diploid *Brassica* genomes (Town et al. 2006; Wang et al. 2011b), along with homoeologous non-reciprocal translocations between A and C genome chromosomes during polyploidisation (Udall et al. 2005; Gaeta and Pires 2010; Szadkowski et al. 2010), have given rise to a patchwork genome with considerable variation in gene copy number and syntenic conservation.

The structural complexity and size of the polyploid *B. napus* and *T. aestivum* genomes represent considerable obstacles to targeted implementation of germplasm in breeding. In particular the use of map-based cloning to identify genes underlying important quantitative trait loci (QTL) is extremely challenging in complex genomes, and genome-wide association approaches have had only limited success in these crops to date due to insufficient resolution of available genome-wide markers. Candidate gene approaches, implementing knowledge from extremely well-characterised, closely related model species, are also not trivial in wheat and oilseed rape, where multiple homologues and paralogues of any given gene of interest can impede a direct interpretation of their role in the crop genome. The combined pitfalls contributed by polyploidy, large gene families and widespread, abundant repetitive sequences complicate the detection and mapping of single-nucleotide polymorphisms (SNPs), an essential first step in genomics-based crop improvement (McKay and Leach 2011).

Furthermore, although transcriptome profiling arrays are available for both *Brassica* and wheat, a general inability to distinguish between different homologous copies of specific expressed genes using oligonucleotide-based hybridisation methods can make interpretation of microarray expression data difficult. In this article, we review recent advances in next-generation DNA sequencing that may provide solutions to many of these problems, enabling considerably more targeted access to complex crop genomes and more efficient use of their underlying variation in breeding.

Next-generation sequencing of complex crop genomes

Efforts to unravel and understand crop genomes have been boosted in recent years by unprecedented developments in so-called “next-generation” DNA sequencing (NGS) technologies that present exciting new opportunities for crop genetics and breeding (Varshney et al. 2009; Duran et al. 2010; Edwards and Batley 2010). Although semi-conductor DNA sequencing technologies (e.g. Rothberg et al. 2011) are already heralding another new generation of DNA sequencing progress, most of the recent genome sequencing efforts in crop plants have implemented second-generation optical sequencing platforms (Edwards and Batley 2010; Imelfort et al. 2009a; Imelfort and Edwards 2009; Metzker 2010). Dramatic cost reductions accompanying the rapid increase in scale of these systems (Muers 2011) today make DNA sequencing an accessible and powerful option for high-throughput analysis even of more recalcitrant crop genomes. Since the introduction of advanced NGS technology, the bioinformatics analysis of this data has been a challenge (Batley and Edwards 2009), however an increasing number of tools are available to interrogate and analyse this data (Lai et al. 2012b; Lee et al. 2012; Marshall et al. 2010). Figure 1 gives a general overview of different areas in which NGS techniques can contribute to our understanding of complex crop genomes, their origins and structure, and can help identify and utilise useful genetic diversity for genetic analysis, trait dissection and breeding.

Establishing workable reference genome assemblies is the key prerequisite for discovery of DNA sequence variation by resequencing using powerful NGS platforms. The 2.3 Gbp genome of maize, a paleopolyploid with extensive proliferation of long terminal repeat retrotransposons, was the first large crop genome to be published (Schnable et al. 2011). By next-generation resequencing of the gene fraction of genetically diverse maize lines, Gore et al. (2009) were able to generate a haplotype map as a foundation for linking DNA sequence variants to phenotypic variation in agronomically important complex traits. Resequencing in maize also demonstrated for the first time the huge diversity in the DNA sequence and genome structure among varieties of the

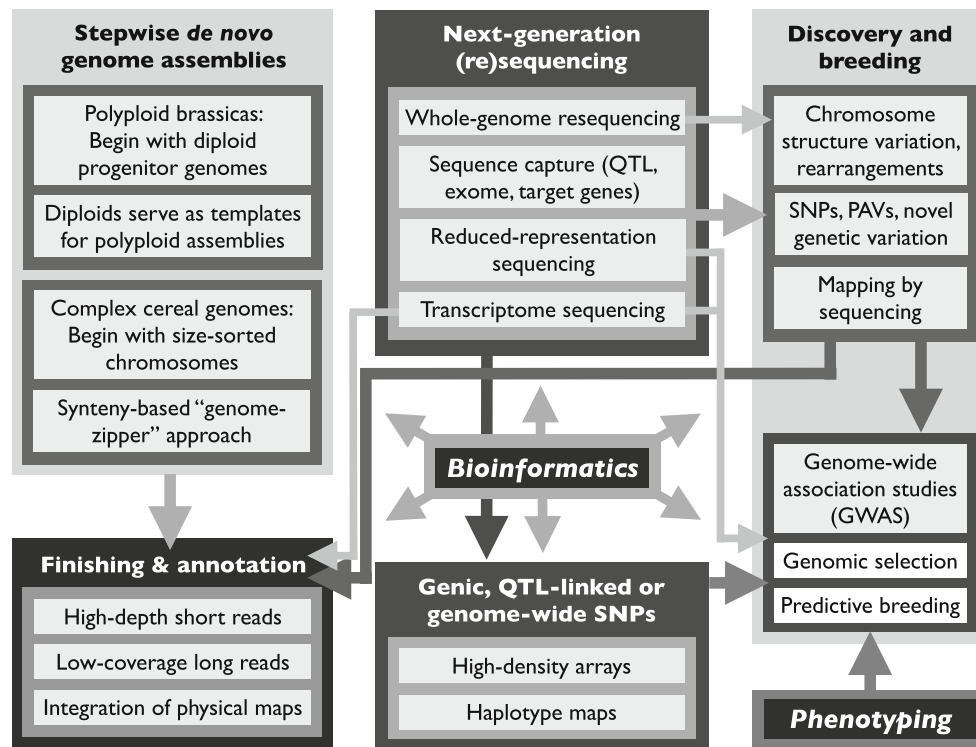


Fig. 1 Next-generation sequencing (NGS) applications have vastly improved our ability to overcome the complexity of large or polyploid crop genomes. The approach of first unraveling the diploid progenitor genomes or individual chromosomes, and implementing synteny or annotation knowledge from less complex relatives, is key to the assembly of large crop genomes like those of oilseed rape or wheat. Whole-genome or reduced-representation resequencing are becoming routine tools for discovery of SNPs, PAVs and chromosome

rearrangements that together contribute to variation. Our ability to relate sequence variation to complex traits using GWAS or genomic selection approaches depends increasingly on appropriate bioinformatic and statistical tools to deal with huge quantities of genotype data. On the other hand, the technologies to generate genome-wide DNA sequence data are quickly outpacing our ability to obtain high-quality phenotype data that accounts for strong environmental effects on complex traits

same crop species (e.g. Springer et al. 2009), with inter-genomic differences between different inbred lines being found on a scale seen only between different species in mammals. These examples from maize demonstrate the enormous power of next-generation sequencing to elucidate the structure of complex, highly repetitive genomes and associate DNA sequence and structural genome variation to important agronomic traits.

Other somewhat less complex crop genomes that have recently been sequenced include the 1.1 Gbp soybean genome (Schmutz et al. 2010) and the 844 Mbp autotetraploid genome of potato (The Potato Genome Sequencing Consortium 2011). The soybean genome, which derived from two ancient genome duplication events followed by extensive gene diversification, gene loss and chromosome rearrangement, was sequenced using a whole-genome shotgun approach, while the relatively small potato genome was deciphered by sequencing a homozygous doubled-monoploid potato clone. The availability of genome sequences has enabled tremendous progress in trait dissection and sequence mining in these and other crops (Hayward et al. 2012b; Hayward et al. 2012c; Raman et al. 2012; Tollenaere et al. 2012). On the other hand, the

huge size of the *T. aestivum* genome and the elevated complexity of the *B. napus* genome, coupled with a particularly high prevalence of gene duplication due to their more recent polyploidy events, represent difficult obstacles that have hindered the development of reliable reference sequences and genome-wide SNP resources for wheat and oilseed rape. Whole-genome DNA sequences for both wheat and oilseed rape were reported in the media during the past two years, however in neither case was a public, whole-genome reference assembly available at the time these articles were written. Numerous international initiatives are currently involved in individual and/or joint efforts to generate reference genome sequences for *B. napus* (for details see <http://www.brassica.info/resource/sequencing.php>). For wheat, the sheer size of the genome necessitates alternative approaches to whole-genome de novo sequencing (<http://www.wheatgenome.info/wiki/>).

Genome sequencing in *Brassica* crops

When the diploid progenitors of a polyploid species are known, a logical approach to simplify de novo genome

sequencing is to begin by assembling the respective diploid genome sequences. In the case of *B. napus* and its progenitors, the A genome of *B. rapa* was chosen as the first *Brassica* genome to be sequenced, based on its smaller size and lower quantity of repetitive DNA elements compared to the C genome of *B. oleracea*. The *B. rapa* v1.1 genome sequence (<http://www.brassicadb.org/brad/>) was published in 2011 by the Multinational *Brassica rapa* Genome Sequencing Project Consortium (Wang et al. 2011b). Completion of the published assembly used a combination of sequential sequencing of physically annotated BAC clones spanning the 10 *B. rapa* chromosomes, supplemented after the origin of next-generation sequencing by high-depth, short-read shotgun sequences. The *B. rapa* genome sequence is a valuable resource to investigate polyploid genome evolution, and also represents a major step towards truly genomics-based breeding of *Brassica* oil and vegetable crops. Whereas the triplication of diploid *Brassica* genomes relative to *Arabidopsis* (Lysak et al. 2005) was confirmed by the completed genome sequence, *B. rapa* was found to contain only 41,174 protein-coding genes instead of the approximately 90,000 that would have been expected in an ancestral crucifer hexaploid. Interestingly, different ancestral subgenomes appear to have been preferentially retained during the evolutionary gene loss process that led to fractionation of the ancestral polyploid genome to what is today *B. rapa* (Wang et al. 2011b). Furthermore, despite their close relationship, a total of 7 % of *B. rapa* genes were found to be not shared by *A. thaliana*. This underlines the enormous value of genome analysis in crop species themselves, rather than our previous reliance on genome information from more or less closely related model species. Completion of the *B. oleracea* C genome is expected to unveil a similar pattern of differential genome fractionation (Town et al. 2006) on a genome-wide scale. The consequence in *B. napus* is that genomic resequencing is highly likely to uncover widespread presence–absence polymorphism caused by removal or retention of different copies of homoeologous genes. In contrast to maize, where fixation of presence–absence variation in different heterotic breeding pools (Springer et al. 2009) results in strong additive effects on heterosis, the different *B. napus* subgenomes appear to have a higher level of plasticity with less structured patterns of gene loss and retention. This is presumably an important driving force determining the degree of “intersubgenomic” or “fixed” heterosis (Basunanda et al. 2010; Zou et al. 2010) in homozygous oilseed rape/canola genotypes. In addition to the biased gene fractionation among subgenomes, subgenome dominance has been observed in both the duplicated maize genome (Schnable et al. 2011) and the triplicated *Brassica* A genome (Cheng et al. 2012). An ability to comprehensively survey such variation by DNA

sequencing will allow us to much more effectively harness additive–additive gene interactions and thereby predict and manage heterosis.

Sequencing in the genome of wheat

The challenges of sequencing the *T. aestivum* genome arise from its large size, polyploidy and abundance of repetitive elements. The genome consists of around 17 billion nucleotides (Paux et al. 2008) and current estimates of gene abundance vary from 77,000 to 295,900 genes (Rabinowicz et al. 2005; Paux et al. 2006; Berkman et al. 2011a, b). The presence of multiple genomes with sequence identity between homoeologous chromosomes often confounds genetic and genomic analysis and complicates the assembly of a reference genome. In addition to the presence of multiple genomes, it is estimated that between 75 and 90 % of the bread wheat genome comprises repetitive sequences (Flavell et al. 1977; Wanjugi et al. 2009), predominantly consisting of nested transposable elements (Li et al. 2004). The abundance of repetitive sequences and the associated increase in genome size further complicates genome sequencing, as the genic and low copy regions are interspersed with long stretches of repeats which are difficult to deconvolute during genome assembly. Together, these factors make wheat one of the more challenging genomes to sequence and analyse.

The establishment of an international wheat genome sequencing project was suggested at a wheat genomics workshop held in November 2003. Initially a BAC-by-BAC approach was adopted, starting with a 5-year pilot project to produce a high-resolution physical map with anchored BACs, followed by the sequencing of a BAC minimum tiling path (Gill et al. 2004). The International Wheat Genome Sequencing Consortium (IWGSC) was established the following year with the goal of producing a complete and annotated genome sequence for *T. aestivum* cv. Chinese Spring (<http://www.wheatgenome.org>). In the 7 years since its establishment, the consortium has produced BAC libraries for all individual wheat chromosome arms (Šafář et al. 2010), physical maps of chromosomes 3B (Paux et al. 2008) and 3DS (Fleury et al. 2010), data analysis pipelines (Sabot et al. 2005) and subgenome-specific molecular markers (You et al. 2011; Nie et al. 2012).

The key to facilitate the assembly of the extremely large hexaploid wheat genome is the ability to physically isolate individual chromosomes (Vrána et al. 2000; Kubaláková et al. 2002), which can be size-separated using flow cytometry (Laat and Blaas 1984; Kubaláková et al. 2000). This method of chromosome isolation has been successfully applied to rice (Lee and Arumuganathan 1999), chickpea (Vlácilova et al. 2002), rye (Kubaláková et al. 2003), barley (Lysak et al. 2005) and wheat (Vrána et al.

2000). Initially only wheat chromosome 3B could be isolated due to its relatively large size, however using cytogenetic stocks exhibiting structural chromosome changes Kubaláková et al. (2002) demonstrated that it is possible to isolate individual chromosome arms from potentially all chromosomes. This approach enables a physical partitioning of the genome into more manageable segments, consequently reducing the complexity of the sequence assembly process.

Whole chromosome shotgun (WCS) sequencing was initially suggested as an alternative to the BAC by BAC approach for wheat genome sequencing (Gill et al. 2004), however the BAC by BAC approach was adopted as the most advanced technology at the time to achieve a thorough and robust reference genome sequence. With the advent of next-generation sequencing technologies, however, several groups have started to apply WCS to large and complex crop genomes. Mayer et al. (2009) assembled Roche 454 sequencing data from DNA from the isolated barley chromosome 1H. To produce a gene-rich scaffold of ordered genes, a “genome zipper” approach was developed based on the syntenic relationship between barley and the sequenced genomes of rice and sorghum (Mayer et al. 2009). This approach has since been extended to the entire barley genome, utilising syntenic comparisons with rice, sorghum and *Brachypodium* to determine the order of ~ 86 % of the predicted barley genes (Mayer et al. 2011). When applied to wheat, the isolation and sequencing of DNA from individual chromosome arms reduces both effective genome size and complexity. This in turn assists downstream bioinformatic assembly and analysis, which is broadly understood to be the current bottleneck in genome sequence assembly. The effectiveness of the WCS approach has been demonstrated for wheat chromosomes 7A, 7B and 7D (Berkman et al. 2011a, b; Lai et al. 2011) as well as chromosome arm 4AL (Hernandez et al. 2012).

Sequence-based transcriptome profiling

Whole-transcriptome profiling via total or reduced-representation mRNA sequencing represents a powerful technique to interrogate gene expression on a global scale (Wang et al. 2010). In comparison to array-based expression analyses, digital gene expression (DGE) methods offer the advantage that expression levels can be quantified even for low-abundance transcripts or previously unknown genes by sequencing at sufficient depth. Sequencing-based whole-transcriptome profiling can also enable large-scale characterisation of transcription start sites and alternative splicing.

Deep mRNA sequencing was demonstrated by Trick et al. (2009) to be an extremely powerful reduced-representation method to distinguish SNPs within expressed

genes from A and C genome *B. napus* transcripts, based on their sequence divergence. In the same way, Bancroft et al. (2009) profiled leaf transcriptomes of doubled-haploid (DH) lines from a genetic mapping population to simultaneously discover and map over 20,000 sequence-annotated SNPs in *B. napus*, generating valuable knowledge on genetic linkage that helped improve previous draft genome assemblies of *B. rapa* and *B. oleracea*. In the process, they also estimated transcript abundance by counting aligned reads per kbp along the lengths of chromosomes, demonstrating that even relatively low-coverage mRNA sequencing was able to provide semi-quantitative data on relative gene expression levels. Upscaling of this method to more recent sequencing platforms today enables a considerably higher sequencing depth and accordingly greater accuracy of transcript quantification. Current NGS-based transcriptome analysis techniques can demonstrably overcome the limitations of microarray techniques with regard to previously unknown genes, and in distinguishing between expression of homoeologous and paralogous gene copies (Parkin et al. 2005). With increasing sensitivity of NGS platforms it is becoming increasingly feasible to quantify even very low-abundance transcripts, or to profile global gene expression even with very small quantities of RNA down to the single-cell level (Ozsolak et al. 2010).

Despite the demonstrated value of NGS technology in transcriptome analysis in other plant species, these have not yet been widely applied to transcriptome analysis in wheat. However, a number of research groups are currently working to analyse wheat transcriptome NGS data. The combination of wheat transcriptome data and the availability of isolated wheat chromosome arm data assemblies will likely lead to a greater understanding of the structure, expression and evolution of the wheat genome, at the same time providing valuable data for annotation of emerging genome sequences.

Applying NGS data for genome-wide association mapping

The application of genome-wide association studies (GWAS) to plants (e.g. Brescghello and Sorrells 2006; Nordborg and Weigel 2008; Gore et al. 2009; Atwell et al. 2010; Huang et al. 2010) has greatly increased the resolution of QTL detection, particularly for species like Arabidopsis, rice and maize where a very high density of SNP markers covering the entire genome was available. In the meantime, NGS platforms have made SNP discovery a more or less routine task for most major crops and genotyping-by-sequencing techniques are now commonly applied in crop species (e.g. Miller et al. 2007; Huang et al. 2010; Chutimanitsakun et al. 2011; Elshire et al. 2011;

Pfender et al. 2011). As sequencing costs continue to drop, it may be only a matter of time until whole-genome or reduced-representation resequencing techniques replace array-based SNP screening as the method of choice for GWAS in crop plants (Chia and Ware 2011). This might be particularly true for polyploid crops in which assignment of genic SNPs to specific homologous loci can often be more reliable using haplotype-oriented sequencing data than on array platforms. Furthermore, because they can theoretically access novel allelic variants in genetically diverse association panels, NGS-based genotyping methods (see Davey et al. 2011 for a comprehensive review) have a potential advantage over array-based SNP screening methods that only survey known alleles in a given panel of SNPs (Chia and Ware 2011). Use of SNPs discovered by genomic resequencing for GWAS was shown by Huang et al. (2010) to be highly effective for uncovering genes involved in important agronomic traits in rice.

Standard GWAS approaches still have the drawback of a lower statistical power of QTL detection, however, meaning that they are generally less suitable for complex traits like seed yield to which numerous minor-effect QTL might contribute. To overcome this drawback, Yu et al. (2008) introduced the concept of nested-association mapping (NAM), an approach which improves the power of genome-wide association studies for the detection of genes underlying complex traits. The success of the NAM technique in maize (McMullen et al. 2009) has prompted efforts in many other major crops to begin developing large, immortal NAM populations for genomic dissection of complex traits. In crops with complex genomes, however, the ability to accurately identify causal genes by GWAS and NAM will ultimately also depend on high-density SNP markers that are able to efficiently distinguish homologous loci.

A major challenge facing plant breeders in coming years is the so-called “phenotyping bottleneck”. Whereas we are now able to generate huge quantities of genotype and sequence data in a short space of time for relatively small amounts of money, phenotyping of complex traits for which such genotyping capacities might give tangible solutions remains a costly and difficult challenge. Increasing abiotic stress tolerance, nutrient efficiency, heterosis, oil content and seed yield in the face of changing environments and emerging production constraints remain challenging goals in this respect. Innovative new approaches are required to analyse and dissect complex traits and associate their components to the underlying genetic mechanisms.

SNP discovery/allele mining in wheat and canola

SNP discovery from NGS data is challenging due to high error rates and short reads (Duran et al. 2009c; Imelfort

et al. 2009b; Azam et al. 2012). In one of the first examples of cereal SNP discovery from next-generation genome sequence data, Barbazuk et al. (2007) identified more than 7,000 candidate SNPs between maize lines B73 and Mo17, with a validation rate of over 85 % using 454 data. Large numbers of wheat expressed-gene SNPs have also been identified from Roche 454 data using autoSNPdb (<http://www.autosnpdb.appliedbioinformatics.com.au/>), software which was originally developed for rice, barley and *Brassica* Sanger sequence data (Duran et al. 2009a, b).

The large data volumes from the Illumina sequencing platform provide the potential to discover very large numbers of genome-wide SNPs (Imelfort et al. 2009a). More than 1 million SNPs have been identified between six inbred maize lines (Lai et al. 2011). This study also identified a large number of presence/absence variations (PAVs), which may be associated with heterosis in this species. More recently, Allen et al. (2011) identified 14,078 putative SNPs in 6,255 distinct wheat reference sequences from Illumina GAIIX data. The validation rate from a subset of 1,659 was 67 % (data accessible at <http://www.cerealsdb.uk.net/NGSdata/AllenSupplement>). Around 3.6 million SNPs were identified by sequencing 517 rice landraces (Huang et al. 2010). This study allowed the association of genome variation with complex traits in rice and is a model for future studies in more complex species, including wheat and canola. A pipeline package called AGSNP has been applied to identify SNPs between two accessions of the diploid wheat, *Ae. tauschii* (Luo et al. 2009). Roche 454 sequencing of *Ae. tauschii* accession AL8/78 has since been combined with Applied Biosystems SOLiD sequencing of genomic DNA and cDNA from *Ae. tauschii* accession AS75 using AGSNP to identify a total of 497,118 candidate SNPs (You et al. 2011). SGSautoSNP (Second-Generation Sequencing autoSNP) was designed specifically to predict SNPs from whole-genome Illumina shotgun sequence data and has been successfully applied to identify more than 1.5 million SNPs in canola, with accuracy greater than 95 % (D. Edwards, unpublished data), as well as more than 900,000 SNPs across the wheat group 7 chromosomes (Edwards et al. 2012; Lai et al. 2012a; Lorenc et al. 2012) (<http://www.wheatgenome.info>). The application of this approach to identify large numbers of genome-wide SNPs even in complex genomes has the potential to be a significant driver in genomics-assisted improvement of wheat, canola and other polyploid crops in coming years (Hayward et al. 2012a).

Several complexity-reduction resequencing approaches have been successfully applied for SNP discovery in complex crop genomes, including *B. napus* and wheat (Bundock et al. 2012; Gholami et al. 2012; Liu et al. 2012; Winfield et al. 2012). These range from sequencing the transcriptome using Illumina technology (Allen et al. 2011;

Bancroft et al. 2009; Trick et al. 2009); anchored EST sequencing (Parkin et al. 2005); sequence capture using oligonucleotide probes (Fu et al. 2010; Pichon et al. 2010; Saintenac et al. 2011), or restriction-based complexity reduction (Elshire et al. 2011). While these approaches can reduce the cost of SNP discovery and genotyping by sequencing, the continued increase in data volumes at an ever-reducing cost may make whole-genome sequencing more cost-effective in the future.

Sequence capture is a common method utilised for SNP discovery from polyploid species. Pichon et al. (2010) demonstrated the usefulness of sequence capture in *B. napus* for SNP discovery between the distinct cultivars Aviso and Montego. Over 7,000 SNPs were identified, of which 266/398 validated SNPs (66.8 %) were suitable for an Illumina GoldenGate™ assay, genotyped across 480 lines. This approach was also utilised in wheat (Saintenac et al. 2011) where the re-sequencing of a 3.5 Mb exonic target covering 3,497 genes in allotetraploid wheat allowed the identification of over 4,000 SNPs. Snowdon and Iniguez Luy (2012), describe the use of different sequence capture techniques in *B. napus*, on the one hand to capture meta-QTL for yield-related traits and on the other hand for targeted resequencing of regulatory genes potentially involved in expression of developmental and yield-related traits.

Transcriptome sequencing has been utilised widely in *B. napus* for SNP discovery (Trick et al. 2009; Parkin et al. 2005). The first of these studies used Illumina RNAseq data in doubled haploid (DH) lines from the Tapidor × Ningyou 7 mapping population. Depending on the stringency applied a total of 23,000 to 41,000 SNPs were identified between the cultivars, using 20 million ESTs from each cultivar. One disadvantage of using transcriptome sequencing for SNP discovery in a polyploid species was demonstrated in this research, however, with over 87 % of SNPs identified as being between homeologous genes.

A public, high-density *B. napus* SNP array

High-density SNP arrays can be a cost-effective alternative for genome-wide polymorphism screens not only for genome-wide association studies, but also by breeders as a tool for comprehensive genome-wide screens of elite germplasm and breeding pools (Batley and Edwards 2007). A major problem in the development of effective SNP arrays for polyploid species is the fact that inter-homoeologue SNPs within one genotype can mask polymorphisms at the same locus between individuals. Large-scale next-generation re-sequencing efforts have facilitated the identification of vast numbers of true, locus-specific SNPs even for complex polyploids, so that development of high-density genome-wide SNP screening arrays has recently become a

viable option for oilseed rape and wheat. In 2011, an international *Brassica* SNP consortium was established in cooperation with Illumina Inc. (San Diego, CA, USA) to generate a public, high-density *B. napus* Infinium SNP array. The International Brassica SNP Consortium array, containing over 50,000 SNPs derived from sequence data contributed by 16 academic and commercial partners from Australia, China, Europe, North and South America, was released in 2012 with a per-sample cost of under US\$70. A similar array is also currently under development for wheat (<http://www.tinyurl.com/wheatSNP>). These resources will open the way for high-throughput germplasm screening and genome-wide association studies in established diversity collections, new “nested” association mapping (NAM) populations and breeding pools. Ultimately, cheap genome-wide SNP screening tools, in tandem with knowledge of global gene expression patterns and relevant expression networks, can be expected to lay the foundation for predictive breeding (Bancroft et al. 2009; Stokes et al. 2010) and genomic selection (Heffner et al. 2009; Jannink et al. 2010; Duran et al. 2010).

Genomic selection and predictive breeding

Since DNA markers were first implemented into practical breeding programs over two decades ago it has been expected that they would significantly improve the efficiency and speed of cultivar development (Beckmann and Soller 1986; Tanksley et al. 1989), by improving the breeder’s ability to select for traits that are difficult to assess phenotypically. Although many breeders indeed implement marker-assisted selection for simple monogenic traits or major-effect QTL, there are still very few examples of successful application of molecular markers for polygenic trait selection (Bernardo 2008; Xu and Crouch 2008). The primary reason for this is that classical QTL analyses tend to identify only small numbers of markers associated with large phenotypic effects, and restrictive costs of single-marker analyses in large breeding populations meant that marker-assisted selection strategies were only cost-effective for traits with high economic return. Recently, this situation has changed dramatically. Using high-throughput SNP array technologies or genotyping-by-sequencing (see above) it is today possible to survey tens of thousands or even hundreds of thousands of markers at a cost that is already not considerably greater than that which a breeder may have previously spent on a small number of major-gene selection markers. By estimating the individual phenotypic effects of large numbers of genome-wide SNPs in a large and diverse “training” population, it becomes possible using appropriate statistical models to calculate genomic estimated breeding values (GEBVs) that can

predict the expected performance of a non-phenotyped individual (Meuwissen et al. 2001). Even without knowing anything about the functional contribution of the SNPs to any given trait of interest, the GEBV represents a fundamentally ideal selection tool because it can theoretically eliminate the need for expensive and time-consuming phenotyping of new breeding lines. This concept, known as “genomic selection”, has revolutionised livestock breeding (Hayes et al. 2009; Bagnato and Rosati 2012). As genotyping and sequencing costs continue to fall there is little doubt that genomic selection will also play an ever-increasing role in plant breeding (Jannink et al. 2010).

Predictive breeding strategies based on the genome-wide SNP markers might also play an important role in selection of optimal crossing partners for the development of high-performing F_1 combinations. In crops like maize and canola, F_1 hybrid cultivars play a major role, however neither the parental line performance nor genetic distance estimates are good predictors of hybrid performance. Instead breeders are faced with time-consuming and expensive evaluation of general combining abilities (GCA) in large numbers of homozygous inbred parental lines from their breeding pools. In a recent study in maize, Riedelsheimer et al. (2012) reported that accurate statistical prediction of GCA could be achieved using genome-wide SNP markers, raising the possibility of selecting cross parents based on genomic predictors rather than phenotypic or population genetic parameters. The use of NGS techniques to survey sequence variation on a whole-genome scale, or to additionally include transcriptome sequence data into prediction models (Bancroft et al. 2009; Stokes et al. 2010), will potentially add further power to genomic selection techniques for breeding of major crops.

Conclusions

The use of NGS technologies has enormous power and potential to access the complex polyploid genomes of many major crops. Discovery and screening of genome-wide SNPs is no longer a bottleneck even in large polyploid genomes like those of oilseed rape and wheat, opening the way for the application of high-resolution association genetics and genomic selection approaches for genetic analysis and improvement of these important crops. As sequencing costs continue to fall, it is feasible that reduced-representation or even whole-genome resequencing could become the method of choice for genotyping in germplasm diversity collections and breeding populations alike. A greater focus on sequence-based global gene expression profiling can also be expected, adding another systemic level of “phenotype” data for genotype-trait associations and predictive breeding strategies. As such

developments progress a major bottleneck in our understanding of complex traits in complex crop genomes will not be our ability to characterise their genetic makeup, but rather a lack of comprehensive and detailed phenotypic characterisation under variable environments.

Acknowledgments DE and JB acknowledge funding support from the Grains Research and Development Corporation (Project DAN00117) and the Australian Research Council (Projects LP0882095, LP0883462, LP110100200 and DP0985953), along with further support from the Australian Genome Research Facility (AGRF) and the Queensland Cyber Infrastructure Foundation (QCIF). RS was supported by a fellowship from the OECD Cooperative Research Programme “Biological Resource Management for Sustainable Agricultural Systems” and by DFG grants SN14/11-1 and SN14/12-1.

References

- Allen AM, Barker GLA, Berry ST, Coghill JA, Gwilliam R, Kirby S, Robinson P, Brenchley RC, D’Amore R, McKenzie N, Waite D, Hall A, Bevan M, Hall N, Edwards KJ (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotech J* (in press, 2012)
- Associated DNA (RAD) linkage map in barley. *BMC Genomics* 12:4
- Atwell S, Huang YS, Vilhjalmsón BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu T et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631
- Azam S, Thakur V, Ruperao P, Shah T, Balaji J, Amindala B, Farmer AD, Studholme DJ, May GD, Edwards D, Jones JDG, Varshney RK (2012) Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. *Am J Bot* 99:186–192
- Bagnato A, Rosati A (2012) From the editors—animal selection: the genomics revolution. *Anim Front* 12:1–2
- Bancroft I, Stokes DR, Morgan CL, Fraser FP, O’Neill C (2009) Prediction of heterosis and other traits by transcriptome analysis. Patent application number 20090300781
- Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. *Plant J* 51:910–918
- Basunanda P, Radoev M, Ecke W, Friedt W, Becker HC, Snowdon RJ (2010) Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theor Appl Genet* 120:271–281
- Batley J, Edwards D (2007) SNP applications in plants. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN (eds) Association mapping in plants. Springer, Berlin, pp 95–102
- Batley J, Edwards D (2009) Genome sequence data: management, storage, and visualization. *Biotechniques* 46:333–336
- Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica* 35:111–124
- Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M, Kubalaková M, Simkova H, Batley J, Fleury D, Dolezel J, Edwards D (2011a) Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. *Plant Biotech J* 9:768–775
- Berkman PJ, Skarshewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai K, Campbell E, Kubalaková M, Simkova H, Batley J,

- Dolezel J, Hernandez P, Edwards D (2011b) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theor Appl Genet* (in press)
- Berkman PJ, Lai K, Lorenc MT, Edwards D (2012) Next generation sequencing applications for wheat crop improvement. *Am J Bot* (in press)
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48:1649–1664
- Bresegghello F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci* 46:1323–1330
- Bundock PC, Casu RE, Henry RJ (2012) Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploid crop plant. *Plant Biotech J* 10:657–667
- Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdilille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *aegilops*). *Plant Cell* 17:1033–1045
- Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X (2012) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *Plos One* 7
- Chia J-M, Ware D (2011) Sequencing for the cream of the crop. *Nat Biotech* 29:138–139
- Chutimanitsakun Y, Nipper R, Cuesta-Marcos A, Cistue L, Corey A, Filichkina T, Johnson E, Hayes P (2011) Construction and application for QTL analysis of a restriction site
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12:499–510
- Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J, Edwards D (2009a) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. *Nucl Acids Res* 37:D951–D953
- Duran C, Appleby N, Vardy M, Imelfort M, Edwards D, Batley J (2009b) Single nucleotide polymorphism discovery in barley using autoSNPdb. *Plant Biotech J* 7:326–333
- Duran C, Appleby N, Edwards D, Batley J (2009c) Molecular genetic markers: discovery, applications, data storage and visualisation. *Curr Bioinform* 4:16–27
- Duran C, Eales D, Marshall D, Imelfort M, Stiller J, Berkman PJ, Clark T, McKenzie M, Appleby N, Batley J, Basford K, Edwards D (2010) Future tools for association mapping in crop plants. *Genome* 53:1017–1023
- Eckardt NA (2001) A sense of self: the role of DNA sequence elimination in allopolyploidization. *Plant Cell* 13:1699–1704
- Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotech J* 7:1–8
- Edwards D, Wilcox S, Barrero RA, Fleury D, Cavanagh CR, Forrest KL, Hayden MJ, Moolhuijzen P, Keeble-Gagnère G, Bellgard MI, Lorenc MT, Shang CA, Baumann U, Taylor JM, Morell MK, Langridge P, Appels R, Fitzgerald A (2012) Bread matters: a national initiative to profile the genetic diversity of Australian wheat. *Plant Biotech J* 10:703–708
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379
- Flavell RB, Rimpau J, Smith DB (1977) Repeated sequence DNA relationships in four cereal genomes. *Chromosoma* 63:205–222
- Fleury D, Luo M-C, Dvorak J, Ramsay L, Gill BS, Anderson OD, You FM, Shoaie Z, Deal KR, Langridge P (2010) Physical mapping of a large plant genome using global high-information-content-fingerprinting: the distal region of the wheat ancestor *Aegilops tauschii* chromosome 3DS. *Bmc Genomics* 11
- Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh CT, Wu W, Swanson-Wagner R, D’Ascenzo M, Millard T, Freeberg L, Aoyama N, Kitzman J, Burgess D, Richmond T, Albert TJ, Barbazuk WB, Jeddelloh JA, Schnable PS (2010) Repeat subtraction-mediated sequence capture from a complex genome. *Plant J* 62:898–909
- Gaeta RT, Pires JC (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. *New Phytol* 186:18–28
- Gegas VC, Nazari A, Griffiths S, Simmonds J, Fish L, Orford S, Sayers L, Doonan JH, Snape JW (2010) A genetic framework for grain size and shape variation in wheat. *Plant Cell* 22:1046–1056
- Gholami M, Bekele WA, Schondelmaier J, Snowdon RJ (2012) A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants. *Plant Biotech J* 10:635–645
- Gill BS, Appels R, Botha-Oberholster A-M, Buell CR, Bennetzen JL, Chalhoub B, Chumley F, Dvorak J, Iwanaga M, Keller B, Li W, McCombie WR, Ogihara Y, Quetier F, Sasaki T (2004) A workshop report on wheat genome sequencing: international genome research on wheat consortium. *Genetics* 168:1087–1096
- Gore MQ, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JL, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. *Science* 326:1115–1117
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92:433–443
- Hayward A, Dalton-Morgan J, Mason A, Zander M, Edwards D, Batley J (2012a) SNP discovery and applications in *Brassica napus*. *J Plant Biotech* 39:49–61
- Hayward A, McLanders J, Campbell E, Edwards D, Batley J (2012b) Genomic advances will herald new insights into the *Brassica:Leptosphaeria maculans* pathosystem. *Plant Biol* 14:1–10
- Hayward A, Vighnesh G, Delay C, Samian MR, Manoli S, Stiller J, McKenzie M, Edwards D, Batley J (2012c) Second-generation sequencing for gene discovery in the Brassicaceae. *Plant Biotech J* 10:750–759
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12
- Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J, Mayer KFX (2012) Next generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J* 69:377–386
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Imelfort M, Edwards D (2009) De novo sequencing of plant genomes using second-generation technologies. *Brief Bioinform* 10:609–618
- Imelfort M, Batley J, Grimmond S, Edwards D (2009a) Genome sequencing approaches and successes. In: Somers D, Langridge P, Gustafson J (eds) *Plant genomics*. Humana Press, New Jersey, pp 345–358
- Imelfort M, Duran C, Batley J, Edwards D (2009b) Discovering genetic polymorphisms in next generation sequencing data. *Plant Biotech J* 7(4):312–317
- Iñiguez Luy FL, Federico ML (2011) The genetics of *Brassica napus* L. In: Bancroft I, Schmidt R (eds) *Genetics and genomics of the Brassicaceae*. Springer, Dordrecht, pp 291–322
- International Brachypodium Initiative (2010) The genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463:763–768
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177

- Kubalaková M, Lysak MA, Vrana J, Simkova H, Cihalikova J, Dolezel J (2000) Rapid identification and determination of purity of flow-sorted plant chromosomes using C-PRINS. *Cytometry* 41:102–108
- Kubalaková M, Vrána J, Čihalíková J, Šimková H, Doležel J (2002) Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 104:1362–1372
- Kubalaková M, Valarik M, Bartos J, Vrana J, Cihalikova J, Molnar-Lang M, Dolezel J (2003) Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow cytometry. *Genome* 46:893–905
- Laat AMMD, Blaas J (1984) Flow-cytometric characterization and sorting of plant chromosomes. *Theor Appl Genet* 67:463–467
- Lai K, Berkman PJ, Lorenc MT, Duran C, Smits L, Manoli S, Stiller J, Edwards D (2011) WheatGenome.info: an integrated database and portal for wheat genome information. *Plant Cell Physiol* 53:1–7
- Lai K, Duran C, Berkman PJ, Lorenc MT, Stiller J, Manoli S, Hayden MJ, Forrest KL, Fleury D, Baumann U, Zander M, Mason AS, Batley J, Edwards D (2012a) Single nucleotide polymorphism discovery from wheat next-generation sequence data. *Plant Biotech J* 10:743–749
- Lai K, Lorenc MT, Edwards D (2012b) Genomic databases for crop improvement. *Agronomy* 2:62–73
- Lee JH, Arumuganathan K (1999) Metaphase chromosome accumulation and flow karyotypes in rice (*Oryza sativa* L.) root tip meristem cells. *Mol Cells* 9:436–439
- Lee H, Lai K, Lorenc MT, Imelfort M, Duran C, Edwards D (2012) Bioinformatics tools and databases for analysis of next generation sequence data. *Brief Funct Genomics* (in press)
- Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. *Science* 320:481–483
- Li W, Zhang P, Fellers J, Friebe B, Gill B (2004) Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J* 40:500–511
- Liu LZ, Wittkop B, Stein A, Sarvari P, Li JN, Dreyer F, Frauen M, Friedt W, Snowdon RJ (2012) A knockout mutation in the lignin biosynthesis gene *CCR1* explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. *Theor Appl Genet* 124:1573–1586
- Lorenc MT, Hayashi S, Stiller J, Lee H, Manoli S, Ruperio P, Visendi P, Berkman PJ, Lai K, Batley J, Edwards D (2012) Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. *Biology* (in press)
- Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo HY, Huo N, Lazo G, Ma Y, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc Natl Acad Sci USA* 106:15780–15785
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. *Genome Res* 15:516–525
- Marshall DJ, Hayward A, Eales D, Imelfort M, Stiller J, Berkman PJ, Clark T, McKenzie M, Lai K, Duran C, Batley J, Edwards D (2010) Targeted identification of genomic regions using TAGdb. *Plant Methods* 6:19
- Mayer KFX, Taudien S, Martis M, Simkova H, Suchankova P, Gundlach H, Wicker T, Petzold A, Felder M, Steuernagel B, Scholz U, Graner A, Platzer M, Dolezel J, Stein N (2009) Gene content and virtual gene order of barley chromosome 1H. *Plant Physiol* 151:496–505
- Mayer KFX, Martis M, Hedley PE, Simkova H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubalaková M, Suchankova P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Dolezel J, Waugh R, Stein N (2011) Unlocking the Barley Genome by Chromosomal and Comparative Genomics. *Plant Cell* 23:1249–1263
- McKay JK, Leach JE (2011) Linkage illuminates a complex genome. *Nat Biotech* 29:717–718
- McMullen MD, Kresovich S, Villeda HS, Bradbury PJ, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C et al (2009) Genetic properties of the maize nested association mapping population. *Science* 325:737–740
- Metzker M (2010) Sequencing technologies—the next generation. *Nat Rev Genet* 11:31–46
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Miller M, Dunham J, Amores A, Cresko W, Johnson E (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17:240–248
- Muers M (2011) Technology: getting Moore from DNA sequencing. *Nat Rev Genet* 12:586–587
- Nie X, Li B, Wang L, Liu P, Biradar SS, Li T, Dolezel J, Edwards D, Luo MC, Weining S (2012) Development of chromosome-arm-specific microsatellite markers in *Triticum aestivum* (Poaceae) using NGS technology. *Am J Bot* (in press)
- Nordborg M, Weigel D (2008) Next-generation genetics in plants. *Nature* 456:720–723
- Ozsolak F, Goren A, Gymrek MA, Guttman M, Regev A, Bernstein BE, Milos PM (2010) Digital transcriptome profiling from attomole-level RNA samples. *Genome Res* 20:519–525
- Parkin IA, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiat DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 171:765–781
- Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J* 48:463–474
- Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeier W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322:101–104
- Pfender WF, Saha MC, Johnson EA, Slabaugh EA (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor Appl Genet* 122:1467–1480
- Pichon J-PD, Riviere N, Duarte J, Dugas O, Wilmer JA, Gerhardt DJ, Richmond T, Albert TJ, Jeddeloh JA (2010) Rapeseed (*B. napus*) SNP discovery using a dedicated sequence capture protocol and 454 sequencing. *Plant and Animal Genome Conference 2010*, San Diego, USA
- Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA (2005) Differential methylation of genes and repeats in land plants. *Genome Res* 15:1431–1440
- Raman H, Harsh R, Raman R, Eckermann P, Coombes N, Manoli S, Zou X, Edwards D, Meng J, Prangnell R, Stiller J, Batley J, Luckett D, Wratten N, Dennis E (2012) Genetic and physical mapping of flowering time loci in oilseed rape (*Brassica napus* L.). *Theor Appl Genet* (in press)
- Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisek J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE

- (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat Genet*. doi:10.1038/ng.1033
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475:348–352
- Sabot F, Guyot R, Wicker T, Chantret N, Laubin B, Chalhoub B, Leroy P, Sourdille P, Bernard M (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. *Mol Genetics Genomics* 274:119–130
- Šafář J, Šimková H, Kubaláková M, Číhalíková J, Suchánková P, Bartoš J, Doležel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet Genome Res* 129:211–223
- Saintenac C, Jiang D, Akhunov ED (2011) Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol* 12:R88
- Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
- Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Nat Acad Sci USA* 108:4069–4074
- Snowdon R, Iniguez Luy (2012) Breeding oilseed rape and canola in the genomics era. *Plant Breed* 131:351–360
- Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddelloh JA, Nettleton D, Schnable PS (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* 5:e1000734
- Stokes D, Fraser F, Morgan C, O'Neill CM, Dreos R, Magusin A, Szalma S, Bancroft I (2010) An association transcriptomics approach to the prediction of hybrid performance. *Mol Breed* 26:91–96
- Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chèvre AM (2010) The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol* 186:102–112
- Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. *Nat Biotech* 7:257–264
- The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195
- Tollenaere R, Hayward A, Dalton-Morgan J, Campbell E, Lee JRM, Lorenc MT, Manoli S, Stiller J, Raman R, Raman H, Edwards D, Batley J (2012) Identification and characterization of candidate Rlm4 blackleg resistance genes in *Brassica napus* using next-generation sequencing. *Plant Biotech J* 10:709–715
- Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, Vigouroux M, Trick M, Bancroft I (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* 18:1348–1359
- Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using solexa transcriptome sequencing. *Plant Biotech J* 7:334–346
- Udall JA, Quijada PA, Osborn TC (2005) Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. *Genetics* 169:967–979
- Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol* 27:522–530
- Vlácilova K, Ohri D, Vrana J, Čihalikova J, Kubalaková M, Kahl G, Doležel J (2002) Development of flow cytogenetics and physical genome mapping in chickpea (*Cicer arietinum* L.). *Chromosom Res* 10:695–706
- Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA, Doležel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). *Genetics* 156:2033–2041
- Wang L, Li PH, Brutnell TP (2010) Exploring plant transcriptomes using ultra high-throughput sequencing. *Brief Funct Genomics* 9:118–128
- Wang J, Lydiate IA, Parkin IAP, Falentin C, Delourme R, Carion PWC, King GJ (2011a) Integration of linkage maps for the amphidiploid *Brassica napus* and comparative mapping with *Arabidopsis* and *Brassica rapa*. *BMC Genomics* 12:101
- Wang XW, Wang HZ, Wang J, Sun RF, Wu J, Liu SL, Bai YQ, Mun JH, Bancroft I, Cheng F, Huang SW, Li XX et al (2011b) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43:1035–1040
- Wanjugi H, Coleman-Derr D, Huo NX, Kianian SF, Luo MC, Wu JJ, Anderson O, Gu YQ (2009) Rapid development of PCR-based genome-specific repetitive DNA junction markers in wheat. *Genome* 52:576–587
- Winfield MO, Wilkinson PA, Allen AM, Barker GLA, Coghill JA, BurrIDGE A, Hall A, Brenchley RC, D'Amore R, Hall N, Bevan MW, Richmond T, Gerhardt DJ, Jeddelloh JA, Edwards KJ (2012) Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotech J* 10:733–742
- Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci* 48:391–407
- You FM, Huo N, Deal KR, Gu YQ, Luo M-C, McGuire PE, Dvorak J, Anderson OD (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* 12:59
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551
- Zou J, Zhu J, Huang S, Tian E, Xiao Y, Fu D, Tu J, Fu T, Meng J (2010) Broadening the avenue of intersubgenomic heterosis in oilseed *Brassica*. *Theor Appl Genet* 120:283–290