

# Applications and challenges of next-generation sequencing in *Brassica* species

Lijuan Wei · Meili Xiao · Alice Hayward · Donghui Fu

Received: 7 September 2013 / Accepted: 12 September 2013 / Published online: 24 September 2013  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** Next-generation sequencing (NGS) produces numerous (often millions) short DNA sequence reads, typically varying between 25 and 400 bp in length, at a relatively low cost and in a short time. This revolutionary technology is being increasingly applied in whole-genome, transcriptome, epigenome and small RNA sequencing, molecular marker and gene discovery, comparative and evolutionary genomics, and association studies. The *Brassica* genus comprises some of the most agro-economically important crops, providing abundant vegetables, condiments, fodder, oil and medicinal products. Many *Brassica* species have undergone the process of polyploidization, which makes their genomes exceptionally complex and can create difficulties in genomics research. NGS injects new vigor into *Brassica* research, yet also faces specific challenges in the analysis of complex crop genomes and traits. In this article, we review the advantages and limitations of different NGS technologies and their applications and challenges, using *Brassica* as an advanced model system for agronomically important, polyploid crops. Specifically, we focus on the use of NGS for genome resequencing,

transcriptome sequencing, development of single-nucleotide polymorphism markers, and identification of novel microRNAs and their targets. We present trends and advances in NGS technology in relation to *Brassica* crop improvement, with wide application for sophisticated genomics research into agronomically important polyploid crops.

**Keywords** Polyploidy · Next-generation sequencing · Brassica · Single-nucleotide polymorphism (SNP) · Transcriptome · Genome sequencing

## Introduction

Next-generation sequencing (NGS) platforms rapidly sequence entire genomes in the form of millions of short DNA fragments in a cost-effective and high-throughput manner. NGS has been widely applied in plant and animal genomic de novo sequencing (Al-Dous et al. 2011; Li et al. 2010; Xu et al. 2011), resequencing (Ashelford et al. 2011; Lam et al. 2010; Xia et al. 2009), methylation sequencing (Cokus et al. 2008; Lister et al. 2008, 2009; Baranzini et al. 2010; Downen et al. 2012) and transcriptome and small RNA sequencing (Buggs et al. 2012; Strickler et al. 2012) for a plethora of functional and comparative genomics analyses. In complex polyploid species, including many of the world's most agronomically important crops, genomics may be hampered by poor discrimination between homologous fragments during sequence assembly. However, advances in NGS technologies and the associated computational algorithms required to analyze massive, complex data sets present huge potential for complex genome analysis and development of improved genomics-based breeding strategies (Aversano et al. 2012).

---

L. Wei · D. Fu (✉)  
Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Agronomy College, Jiangxi Agricultural University, Nanchang 330045, China  
e-mail: fudhui@163.com

L. Wei · M. Xiao  
Chongqing Engineering Research Center for Rapeseed, College of Agronomy and Biotechnology, Southwest University, Chongqing 400716, China

A. Hayward  
Centre for Integrative Legume Research, School of Agriculture and Food Sciences, The University of Queensland, St Lucia 4072, Australia

The *Brassica* genus contains the most diverse collection of agronomically important plant species and is a relative of the model plant *Arabidopsis thaliana*, from which it diverged ~20 Mya. The six most agro-economically important *Brassica* species include the three diploid species, *Brassica rapa* (AA,  $2n = 20$ ), *Brassica oleracea* (CC,  $2n = 18$ ) and *Brassica nigra* (BB,  $2n = 16$ ), and the three allotetraploid species, *Brassica juncea* (AABB,  $2n = 34$ ), *Brassica napus* (AACC,  $2n = 38$ ) and *Brassica carinata* (BBCC,  $2n = 36$ ), which were formed through the hybridization of their diploid genome counterparts (U N 1935). World-wide, approximately 12 % of edible vegetable oil is provided by *B. napus*, *B. rapa*, *B. juncea* and *B. carinata*, and the global production of *Brassica* crops has doubled in the last 15 years. Given that many *Brassica* crop species are polyploids, the well-studied relationships between the diploid progenitor species and their corresponding tetraploids, combined with relatively small genomes, make *Brassica* species highly useful models for investigating the genetic and evolutionary mechanisms of polyploidisation (Edwards et al. 2013).

The genome of diploid and tetraploid *Brassica* species (1.2 Gbp) are 3–5 times and 10 times greater than that of *A. thaliana*, respectively (Arumuganathan and Earle 1991). Genome duplication combined with gene loss and insertion (Town et al. 2006; Mun et al. 2009), chromosomal rearrangements, and rapid divergence of repeat sequences (Koo et al. 2011) are frequent effects of polyploidization, which has complicated *Brassica* genome structure. As such, the complex structure of polyploid *Brassica* genomes has made genome sequencing and assembly challenging. Fortunately, with advances in technologies and software, dissecting polyploid genomes has become possible. The current trend towards the application of NGS platforms in numerous plant and animal species will profoundly affect genomic research, improving our understanding of molecular and evolutionary mechanisms underlying variations in a species' phenotype, development and responses to environmental stressors.

This article will compare various NGS platforms and discuss advances and trends in their use with emphasis on their application and challenges in *Brassica* research and breeding. In so doing, this review will highlight the impacts of NGS in *Brassica* research as a model system for understanding the molecular basis of polyploidy in important crop species.

### Characteristics and comparative analysis of various generations of sequencing technologies

#### First-generation sequencing

First-generation sequencing is based on the chain termination method invented by Frederick Sanger (Sanger et al.

1977). Today, this method has been automated and commercialized using sequencing machines available through companies including Applied Biosystems (USA) and Beckman Coulter (USA). Sanger sequencing has dominated the DNA sequencing industry for almost three decades, providing sequence reads between 450 and 1,000 bp in length at high accuracies (99.999 %). The human genome (International Human Genome Consortium 2004) and the genome of model plant species, *A. thaliana*, followed by two rice varieties: *Oryza sativa* L. ssp. *japonica* and *Oryza sativa* L. ssp. *indica* have been completed (Goff et al. 2002; Yu et al. 2002). More recently, the genomes of *Brachypodium distachyon* (The International Brachypodium Initiative 2010), *Populus trichocarpa* (poplar; Tuskan et al. 2006), *Prunus persica* (peach; <http://www.rosaceae.org/peach/genome>), *Sorghum bicolor* (sorghum; Paterson et al. 2009), and *Glycine max* (soybean; Schmutz et al. 2010) were sequenced (Table 1).

For *Brassica* species, the first physical map of *Brassica A* genome of *B. rapa* was constructed based on 67,468 BAC clones, spanning 717 Mb in physical length, fingerprinted by Sanger sequencing (Mun et al. 2008). The *B. rapa* A3 chromosome (31.9 Mb) was also obtained using traditional Sanger sequencing methods, incorporating 348 overlapping BAC clones (Mun et al. 2010).

Although first-generation sequencing enabled initial whole-genome sequencing efforts, significant limitations in Sanger sequencing technology exist. Firstly, conventional Sanger sequencing is still relatively expensive, currently 0.5 per kilobase, and while this cost has declined over the last few years (Snowdon and Luy 2012), it would cost \$5,000,000 to sequence a 1 Gb genome to 10× coverage. Secondly, Sanger technology is difficult to improve upon since it depends on capillary electrophoresis separation of fluorescently labeled fragments (Varshney et al. 2009). Thirdly, the throughput (~70 Kbp of data per run) is exceedingly low, making it time-intensive to obtain genetic information, particularly for large numbers of samples in parallel. Thus, Sanger sequencing technology continues to be beneficial and widely applied for low throughput projects, for example, sequence validation of PCR products and BAC end sequences, but is not efficacious for mainstream genome and transcriptome sequencing, particularly of complex genomes. Such larger-scale studies require higher throughput technologies. As such, sequencing technologies able to multiplex numerous samples in parallel to obtain vast quantities of genetic information were more recently developed, collectively called NGS.

#### Next-generation sequencing

With advancements in molecular research and the increasing demand for large quantities of nucleotide sequences

**Table 1** Plant species that have been sequenced using sequencing technologies

Species	Sequencing accession	Sequencing methods	Genome size	Reference
<i>Arabidopsis</i>	Columbia	Sanger sequencing	125 Mb	Kaul et al. (2000)
Rice	<i>Oryza sativa</i> L. ssp. <i>Indica</i> “93-11”/ <i>japonica</i> “Nipponbare”	Sanger sequencing (4.2×)	466 Mb	Goff et al. (2002), Yu et al. (2002)
<i>B. distachyon</i>	Diploid lines “Bd21”	Sanger sequencing (9.4×)	272 Mb	The International Brachypodium Initiative (2010)
<i>P. trichocarpa</i>	A single female genotype “Nisqually 1,”	Sanger sequencing (7.5×)	370 Mb	Tuskan et al. (2006)
Papaya	Tansgenic fruit tree “SunUp”	Sanger sequencing (3×)	372 Mb	Ming et al. (2008)
Peach	Doubled haploid cultivar ‘Lovell’	Sanger sequencing (7.7×)	227 Mb	International Peach Genome Initiative (2010)
Sorghum	Homozygous genotype “BT × 623”	Sanger sequencing (8.5×)	~730 Mb	Paterson et al. (2009)
Soybean	Cultivar “Williams 82”	Sanger sequencing	1.1 Gbp	Schmutz et al. (2010)
Cucumber	Inbred line ‘Chinese long’ 9930	Sanger (3.9×) and Illumina GA (68.3×)	243.5 Mb	Huang et al. (2009)
Watermelon	Chinese elite inbred line 97103	Illumina GAI (108.6×)	353.5 Mb	Guo et al. (2012a)
Barley	Cultivar “Morex”	Illumina GAI and Roche 454	4.98 Gb	Mayer et al. (2012)
Sweet orange	Dihaploid	Illumina GAI	367 Mb	Xu et al. (2012a)
Potato	Double monoploid “Phureja DM1-3 516 R44”	Illumina GA, Roche 454 and Sanger platforms	844 Mb	Xu et al. (2011)
Cotton	<i>G. raimondii</i>	Illumina HiSeq2000 (103.6×)	775.2 Mb	Wang et al. (2012b)
Bread wheat	Chinese Spring “CS42”	454 pyrosequencing (5×)	17 Gb	Brenchley et al. (2012)

came the development of fast, high-throughput and cost-effective NGS technology. NGS technology has been widely applied in recent years and includes a few major products with different chemistries, including 454 sequencing (Roche Applied Science), Solexa (now Illumina) technology (Illumina inc.), SOLiD (Applied Biosystems), and Polonator G. 007 (<http://www.polonator.org/>) (Table 2).

The first NGS sequencing instrument on the market was the GS20, developed by Roche Applied Science (USA). Recently, the Roche 454 GS FLX+ system was commercialized, yielding up to 700 Mb of sequence per run comprising 1,000 bp sequence reads with an accuracy of 99.997 %. This system includes the software required for assembly and mapping of these sequences reads into contiguous sequence, as well as amplicon variant analysis (Klein et al. 2011; Mundry et al. 2012). The Roche 454 GS has proven efficacy in *Brassica* genome sequencing (Wang et al. 2011b), and improves the accuracy of mapping homeologous fragments because the relatively long length of sequence reads increases specificity. Nonetheless, the Roche 454 GS remains costly as compared to other NGS technologies.

Solexa sequencing, now provided by Illumina (including the HiSeq and MiSeq instruments), can generate up to 200 bp per read and up to 600 Gbp of data per run. Error rates for this system are approximately 0.1 %, primarily resulting from substitution errors, rather than insertions or deletions in the sequencing and detection process

(Minoche et al. 2011). Illumina sequencing is cost-effective and highly suited to the sequencing of short *Brassica* target sequences such as transcriptome sequences and small RNAs. However, its application as a sole NGS technology for whole-genome sequencing is limited by its relatively short-read lengths.

The ABI SOLiD platform (version 4) clonally amplifies template fragments with emulsion PCR, using DNA ligase rather than a polymerase and fluorescently labeled oligonucleotides. Another feature is the use of two-base encoding, which identifies each base twice to reduce the error rate. However, the speed of sequencing is relatively slow, and the read length is only around 85 bp (Edwards and Batley 2010). The Polonator G 007 is similar to the SOLiD system, and the technology platform is open for manipulation and improvement by the user. However, the current read length is less than 30 bp (Morey et al. 2013) and while vast quantities of data can be obtained from clonal amplification sequencing, read length is much shorter than that of other NGS platforms, reducing the applicability to sequence assembly and subsequent analysis in polyploid species including the amphidiploid *Brassicac*s.

### The third-generation sequencing

Single-molecule sequencing technology directly analyses light signals generated by cellular nucleic acids without a requirement for clonal amplification or ligation in template

**Table 2** Comparison of different sequencing technologies

	Company	Current platform	Sequencing products	Sequencing approach	Read length	Time per run	Reads per run	Machine cost (\$)	Accuracy	Advantages	Limitations
Sanger sequencing	Applied Biosystems Inc.	ABI 3730xl	Chain termination reaction	Capillary electrophoresis to detect four-color fluorescent signals	450–1,000 bp	1 day	0.6 Mb	150,000	99.99 %	Long reads length; high accuracy	Requiring capillary electrophoresis
Next-generation sequencing											
Clonally amplification											
454 sequencing	Roche Applied Science	454 GS FLX+ system	Emulsion PCR	Pyrosequencing	1,000 bp	23 h	0.7 Gbp	500,000	99.99 %	Long reads length	Error type: insertions and deletions; requiring more enzyme; high cost
Solexa	Illumina inc.	HiSeq or MiSeq	Bridge PCR and solid-phase amplification	Reversible terminators	200 bp	10 days	600 Gbp	540,000	99.9 %	High coverage; most common technology	Short-read length; error type: substitution
SOLiD	Applied Biosystems	5500 Series Genetic Analysis Systems	Emulsion PCR	Cleavable oligonucleotide probe and sequencing by ligation	85 bp	~7 days	70–200 Gbp	595,000	99.99 %	High accuracy	Short-read length; low sequencing speed; longest run time
Polonator G. 007	Dover, in collaboration with Harvard Medical School	May move on to the H.008	Emulsion PCR	Not cleavable oligonucleotide probe and sequencing by ligation	26 bp	4 days	8–10 Gbp	170,000	>98 %	High accuracy; open source	Shortest read length
The third-generation sequencing											
Single-molecule sequencing											
Helicos biosciences Helicore	Helicos Biosciences	Helicore™ Single Molecule Sequencer	Single molecule	Reversible terminators	55 bp	8 days	37 Gbp	999,000	99.999 %	High accuracy	Short-read length

Table 2 continued

Company	Current platform	Sequencing products	Sequencing approach	Read length	Time per run	Reads per run	Machine cost (\$)	Accuracy	Advantages	Limitations
Pacific Biosciences	PacBio RS	Single molecule	Single-molecule real-time zero-mode waveguide structure	3,000–20,000 bp	30 min	200–250 Mb	700,000	99 %	Long read length; real time	Highest error rates
Ion Torrent/Life Technologies	Ion Proton™	Single molecule	Ion current	35–400 bp	90 min	10 Mb–1 Gb	–	99.99 %	Low cost; long read length	Fast translocation speed; no appropriate products

preparation. Deletions are the main cause of error using this sequencing technology. The HeliScope Single Molecule Sequencer is the first commercial product using this technology, which has been successfully applied to resequencing individual human genomes (Pushkarev et al. 2009). A method of single-molecule direct RNA sequencing without cDNA synthesis has also been established for transcriptome analysis in diagnostics (Ozsolak and Milos 2011); however, the read lengths (on average 30 bp) are relatively short.

Single-molecule real-time sequencing marketed by Pacific Biosciences in 2011 reads the sequence of DNA immobilized on a zero-mode waveguide (ZMW) reaction cell in real-time during polymerization (Eid et al. 2009). This yields approximately 3,000–20,000 bp read lengths, but is more prone to error than second-generation NGS technologies. Due to its high speed and long reads, this technique is highly promising for assembly of large, complicated genomes such as from *Brassica* species, provided that the error rates can be reduced.

Nanopore sequencing is a new technology to detect single molecules rapidly and directly as they pass through a nanoscale pore in a membrane, driven by an ion current. This can produce long read lengths of around 25 kb, or up to 5.4 kb in solid-state nanopores (Branton et al. 2008), and does not require pre-processing of templates, the use of polymerases or ligases or biochemical tags. Furthermore, if applied globally and successfully the cost of this technology is likely to be significantly reduced. The main challenge of this technology is the requirement for fast translocation speeds through the protein nanopore for accurate base reads. Oxford Nanopore Technologies has claimed that the first cost-effective nanopore sequencer will come to market later this year, which can sequence the human genome in 15 min.

Ion Torrent using semiconductor sequencing technology is launched by Life Technologies Corporation, which includes two platforms: Ion Personal Genome Machine (PGM) and Ion Proton. Current PGM 318 platform can produce about 1 Gb data in 2 h, and the length is up to 400 bp. It does not depend on the chemi-luminescence and optics. Undoubtedly continued advances in these third-generation technologies will occur rapidly allowing such technologies to be widely applied.

The choice of sequencing technology depends on the downstream application and is determined by the quantity and quality of the data output (read length, error rate, Gbp output), sequencing cost and time per run. The length of reads is particularly critical for organisms such as *Brassicac*s, where the abundance of homeologous sequences hinders accurate genome assembly. Certainly, if the sequenced sequences are short, e.g., small RNA, many sequencing technologies could be selected, but specificity is also a major problem for RNA sequencing, otherwise non-specific



short reads from different loci can map to a single locus and falsify expression quantities.

### Complete genome sequencing projects using NGS

NGS technologies have contributed to the sequencing of complex genomes in the last few years, paving the way for future genomic and genetic research and crop improvement. The combination of traditional and NGS methods can currently provide a method for rapid and cost-effective sequencing of important plant species. For example, the cucumber genome (*Cucumis sativus*) was sequenced in 2009 with traditional Sanger (3.9×) and Illumina Genome Analyser (GA) NGS reads (68.3×). In this way, a total of 243.5 Mb was assembled and 26,682 genes predicted (Huang et al. 2009). The 353.5 Mb watermelon (*Citrullus lanatus*) genome was obtained at 108.6× coverage using the Illumina GAI system, and annotated with 23,440 genes (Guo et al. 2012a). Similarly, the barley (*Hordeum vulgare*; haploid content of 4.98 Gb) (Mayer et al. 2012) and sweet orange (*Citrus sinensis*; 367 Mb generated) genomes were sequenced using Illumina GAIIx technology (Xu et al. 2012b). The Illumina GA, Roche 454 and Sanger platforms were used to generate 844 Mb of the complex autotetraploid potato genome using a homozygous doubled-monoploid potato (Xu et al. 2011). A diploid cotton genome was sequenced on the Illumina HiSeq2000 platform at 103.6× resolution (about 775.2 Mb) (Wang et al. 2012b), however, given that most cotton cultivars are tetraploid, novel experimental and computational methods need to be applied for characterization of natural polyploids. Wheat has a complex and huge allohexaploid genome that includes numerous repetitive elements. To simplify genome sequencing and assembly in this species, isolated chromosome arms can be sequenced through NGS technologies, reducing the confounding effects of multiple homologous chromosomes (Berkman et al. 2012). Brenchley et al. (2012) has generated 17 Gb of the Chinese Spring (CS43) *Triticum aestivum* (bread wheat), genome with 454 pyrosequencing (5×), containing approximately 95,000 genes.

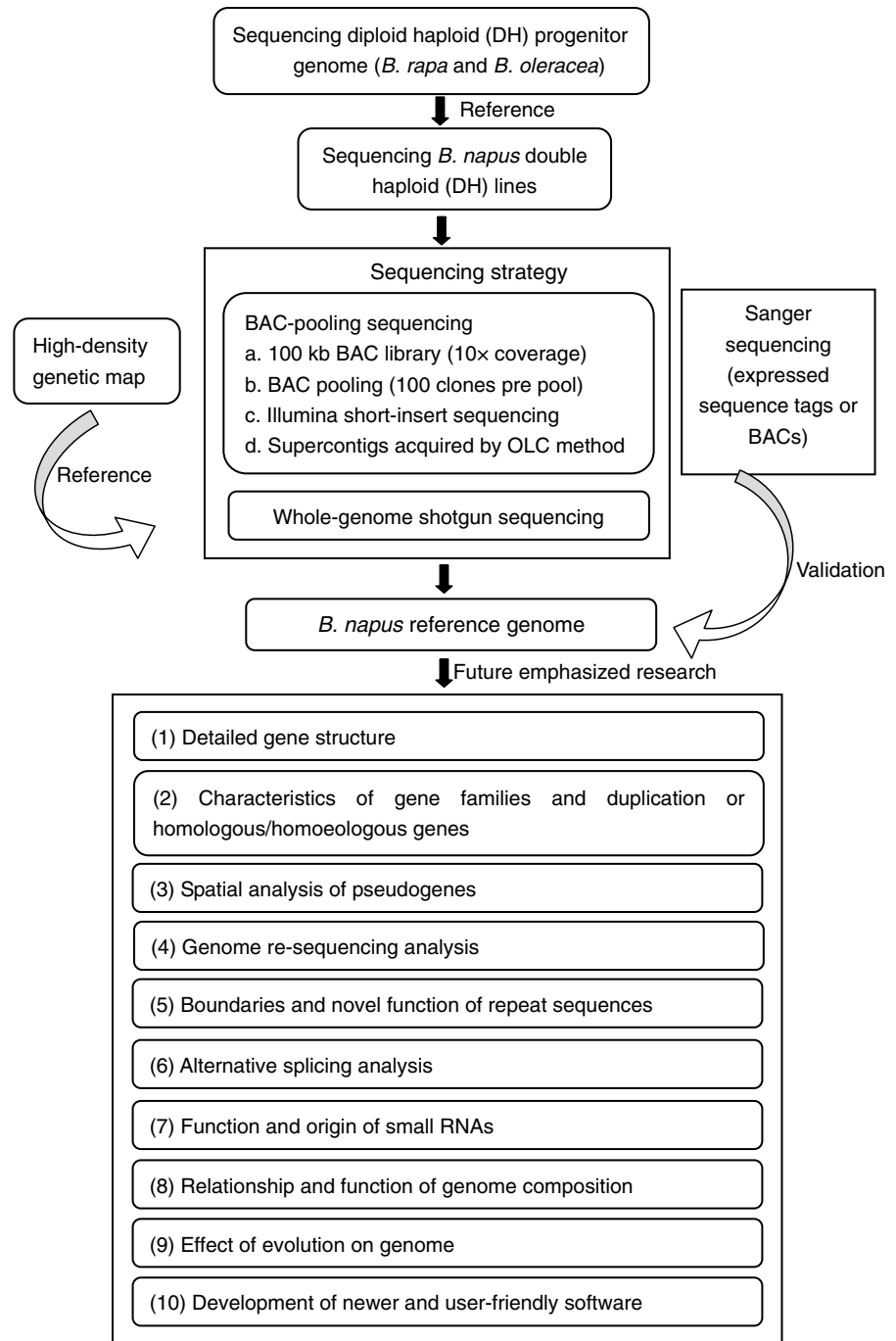
High-coverage and high-quality reference genome sequences are a base or a core of the foundation of omics investigations in of a species, especially polyploidy species. A multinational consortium for *Brassica* genome sequencing was initiated in 2003, with the initial aim to sequence the diploid *B. rapa* A genome using a BAC tiling path and Sanger technology. Following the development of NGS technology, the *B. rapa* (Chinese cabbage) v1.1 genome, including 41,174 protein coding genes, was released in 2011 by the international *B. rapa* Genome Sequencing Project Consortium (Wang et al. 2011b). The completion of *B. rapa* genome provides new insights into genome evolution

in *Brassicaceae* and, importantly, the first *Brassica* reference genome. The *B. oleracea* diploid C genome is the second *Brassica* crop to undergo genome sequencing, using a combined Illumina and Roche 454 sequencing approach, and is expected to be released later this year.

*B. napus* is an allotetraploid formed from hybridization of *B. rapa* and *B. oleracea* and thus possesses a large and complex genome:  $AC 2n = 19$ . Nowadays, genomics investigation of tetraploid *B. napus* is faced with two critical problems: (1) homeologous regions of high sequence identity between the A and C genomes, which interfere with sequence read mapping and thus accurate assembly and correct assignment of homeologous A and C chromosomes; (2) Numerous repeat sequences, including simple repeat sequences, minisatellites, satellites and different categories of transposons, which are enriched in *B. napus* and also hinder mapping and accuracy of assembly. Hence, strategies optimized to study the genome of tetraploid *B. napus* are required, as follows (Fig. 1):

1. The use of double haploid (DH) *B. napus* lines, benefiting from advanced microspore culture techniques in many *Brassica* labs. In these lines, each locus is homozygous, avoiding the interference of allelic variants on sequence assembly. In non-DH heterozygous lines, it is difficult to discriminate the allelic variants within the A or C genomes from the homeologous variants. Recently, an RFAptools pipeline was developed, which was divided into three steps: firstly, a pseudo-reference sequence was assembled; secondly, single-nucleotide polymorphism (SNP) was discovered and genotyped, and finally allelic SNP was discriminated from homeologous loci. It combined with a double digestion RADseq (ddRADseq) approach in a *B. napus* DH population which was developed successfully to discriminate allelic variants from homeologous sequences (Chen et al. 2013).
2. The use of the concomitant genome sequences of *B. rapa* and *B. oleracea*, as the two parental species of *B. napus*, as reference genomes for *B. napus* genome assembly, followed by careful correction for any large genome rearrangements and variations between these progenitors and *B. napus*. Indeed, this is currently being applied in the *B. napus* genome sequencing project (Snowdon and Luy 2012).
3. The use of a high-density genetic map, for the accurate mapping of homeologous regions and genome rearrangements relative to reference sequences. High-density genetic maps enable the correct ordering of sequence contigs to accurately link physical and genetic maps. In *Brassica*, the availability of large mapping populations and high-throughput, genome-

**Fig. 1** Strategies for *B. napus* whole-genome sequencing and future emphasized research on *B. napus* genome. In the sequencing of *B. napus*, double haploid (DH) lines are necessary to avoid the interference of allelic variants on sequence assembling. Diploid progenitor genome and high-density genetic map are also critical for the accurate mapping of homeologous regions and genome rearrangements relative to reference sequences. A combination of BAC-pooling sequencing and whole-genome short-read sequencing will be used to produce a reference sequence of *B. napus*. Expressed sequence tags or BAC sequences sequenced by the Sanger sequencing technique will be used to verify the reference genome. With the production of high-quality reference genomes for *Brassica* species, future numerous works need be launched



specific genotyping technologies make this highly applicable. For example, Illumina’s GoldenGate and Infinium SNP genotyping platforms are currently being developed and applied in *Brassica* species (Durstewitz et al. 2010), which can detect from 384 to 60,000 genome-specific SNPs in parallel.

4. The use of a combination of BAC-pool sequencing and whole-genome short-read sequencing for accurate production of a *B. napus* reference genome. In this method, for example, a 100 kb insert BAC library of

~10,000 clones (10× coverage of *B. napus*) will be evenly and randomly divided into 1,100 pools (100 clones per pool). For each pool, three Illumina short-insert paired-end sequencing libraries will be constructed, sequenced with >50× depth and assembled. Supercontigs can be acquired by merging contigs from each pool using the overlap layout consensus (OLC) method. The redundancy in the assembly can be removed by self-to-self whole-genome alignment and sequencing depth information. Whole-genome

Illumina libraries (200-bp to 20-kb inserts) then will be constructed with 100× coverage and used to aid assembly. Overall, BAC supercontigs can be assembled into Scaffolds and then into a reference genome, with the help of high-density genetic map and gap-filling. Finally, expressed sequence tags (EST) or Sanger-sequenced BACs will be used to verify the reference genome. If these sequences are successfully mapped in the reference genome with a high proportion (e.g., >99 %), the reference genome is of high quality.

Sequencing projects for *B. napus* are currently being mediated by 11 research institutes, with the intention to analyze genetic diversity in oilseed rape. Sequencing of the other three cultivated *Brassica* species, *B. nigra*, *B. juncea* and *B. carinata*, are also in the pipeline (See <http://canseq.ca/>).

### Applications of NGS in *Brassica* species

#### SNP discovery

SNP density varies between and within species as well as different genomic regions. In rice, SNP density averages one in 147 bp (Subbaiyan et al. 2012), while soybean (Choi et al. 2007) and *A. thaliana* (Atwell et al. 2010) average 1/438 and 1/500 bp, respectively. Previously, the application of SNPs was limited because of the high cost of development and detection. SNPs were generally discovered through PCR amplification and Sanger sequencing of genomic regions of interest, or using DNA chips, which were laborious and time consuming. SNPs were also detected with computational tools based on existing EST databases. For instance, a SNP discovery pipeline in barley, wheat, rice and *Brassica* was developed through analysis of assembled EST sequences (Duran et al. 2009), whereby SNPs are identified by blast comparison or keyword search using AutoSNPdb (<http://autosnpdb.appliedbioinformatics.com.au/>).

With the advent of NGS technology, the discovery of large numbers of genome-wide SNPs is now highly achievable. Abundant markers can be discovered through amplicon sequencing, transcriptome sequencing, DNA-rich genome sequencing and whole-genome sequencing (Henry et al. 2012). Davey et al. (2011) reviewed genome-wide marker discovery using NGS in both model organisms and non-model species via reduced representation sequencing methods, including reduced representation libraries (RRLs), complexity reduction of polymorphic sequences (CRoPS) and restriction site-associated DNA sequencing (RAD-seq). The development, validation and application of SNPs had been reviewed systematically (Mammadov et al.

2012). Freely available software including SGSAutoSNP (Lorenc et al. 2012) enable rapid, high-throughput, accurate SNP discovery that can be applied to any species with available NGS data. SNP prediction can be complicated by the error rate of NGS and by repetitive or highly homologous regions causing misassembly of short-read lengths. However, the use of paired-end and large-insert NGS sequence reads in genome assembly and the strict quality control parameters in software, such as SGSAutoSNP (Lorenc et al. 2012), can help to minimize non-specific read mapping, and false SNP predictions.

Given that the current public *Brassica* reference genome is limited to the *B. rapa* A genome, several methods of detecting SNPs to reduce the size of target sequences were developed, such as transcriptome sequencing, EST-based sequencing, RAD sequencing and sequence capture using oligonucleotide probes. Trick et al. (2009b) developed a robust method to discover SNPs through transcriptome analysis of the polyploid *B. napus* cultivars, Tapidor and Ningyou7, using the Illumina platform. In this case, *Brassica* unigene sequences were used as a reference and aligned with sequence reads using MAQ software to discover SNPs. In total, 23,330–42,593 putative SNPs with different read depth were detected, and ~90 % of the SNPs detected were termed as hemi-SNPs, which were homozygous in one line but heterozygous in the other line (Mammadov et al. 2012). The hemi-SNPs between these two lines could be used for genetic mapping. In addition, Hu et al. (2012b) discovered 655 putative SNP markers by 454 sequencing of ESTs of two *B. napus* cultivars: ZY036 and 51070. Similarly, Durstewitz et al. (2010) identified 604 SNPs from ESTs in *B. napus* (one SNP per 42 bp), which were then validated using the Illumina GoldenGate SNP genotyping system. However, the primary limitation of SNP discovery from transcriptome sequencing or EST-based sequencing is the restriction to coding regions, and thus this failed to detect diversity in non-coding regions.

For genome-wide SNP discovery, RAD sequencing technology is a simple, alternative approach for detecting polymorphisms in complex crop genomes by reducing the complexity of the genome. More than 20,000 SNPs and 125 insertions and deletions were identified in about 113,000 RAD clusters of the *B. napus* genome sequenced via the Illumina GAIIx system. At the same time, 26 out of 31 SNPs (84 %) in 16 RAD clusters were validated by Sanger sequencing (Bus et al. 2012). This is simple and effective in genetic mapping, but is limited for genome-wide association studies (GWAS) due to the small number of markers (Mammadov et al. 2012).

Sequence capture is also a technique that reduces the size of sequenced fragments and identifies homologies in the genome by rapidly tagging a targeted region for sequencing. It not only identifies meta-QTL regions associated



with traits but also detects the variance of complex traits, like developmental and flowering traits (Snowdon and Luy, 2012). A total of 87 SNPs and 6 Indels were identified based on existing genomic resources in six *B. napus* varieties (Westermeier et al. 2009). Picho et al. (2010) presented the application of sequence capture in the *B. napus* cultivars Aviso and Montego and identified about 7,000 SNPs, which were useful for QTL mapping and genetic association studies. However, this technology requires a reference sequences for the design of capture probes.

Abundant Illumina read sequences of *B. napus* have been obtained via such complexity reduction methods. Despite combining NGS technology with bioinformatics for SNP discovery, in the allotetraploid *B. napus*, SNP discovery is currently complicated by the presence of highly homeologous ancestral A and C genomes and the absence of a reference genome sequence. Polymorphisms are only useful as inter-cultivar SNPs, and must be distinguished from intra-cultivar SNPs between the homeologous A and C chromosome regions. SNPs derived from both different alleles and homeologues are usually mixed together with above methods. Following the completion of *B. oleracea* genome sequence and de novo *B. napus* genome, comparison to the corresponding diploid species could be used to distinguish these two kinds of SNPs. In addition, Chen et al. (2013) has successfully developed ddRADseq approach, with bioinformatics RFApools, to discriminate allelic SNPs from homoeologous sequences in *B. napus*.

#### SNP arrays in *Brassica*

SNP arrays are the main method for large-scale SNP analysis, which can simultaneously detect millions of SNPs in one reaction. At present there are two main types of SNP array commercialized by Affymetrix and Illumina. The GeneChip Rice 44K SNP Genotyping Array was developed by Affymetrix and used to identify rice varieties and their genetic diversity (McCouch et al. 2010). The 135k *Brassica* exon Array representing 135,201 genes has been designed for whole-transcript profiling and mapping and for analysis of genome evolution and adaptation in the *Brassicaceae* family (Love et al. 2010). The Illumina 1M SNP array has the capacity to identify 1 million SNPs. Although the cost of NGS continues to decline, allowing genotyping by sequencing approaches to become more feasible, SNP microarray techniques retain some exclusive advantages. Firstly, they can provide robust analysis techniques using large public reference datasets at reduced when performing the replicated experiments. For *Brassica*, a public and high-density Illumina SNP array was released in 2012, combining efforts of 16 academic and commercial partners with Illumina Inc (Snowdon and Luy 2012). At present, more than 50,000 SNPs were found to function well in the

*Brassica* A or C genomes using this SNP panel. This SNP array will offer advantages for GWAS and high-throughput screening of germplasm pools. At the same time, it is effective to identify unique genes or primary expressed genes in the genome (Parkin et al. 2010). However, the SNPs in this array are not evenly distributed across the genome, which may lead to bias in downstream analyses. Therefore, future SNP arrays should be developed based on evenly distributed, genome-wide A genome-specific or C genome-specific loci. NAM using a SNP array will aim to uncover the basis of yield and stress tolerance in *B. napus* (Edwards et al. 2013).

#### Genetic map construction and gene mapping

Traditional gene mapping methods for most quantitative traits are based on high-density genetic maps, which are constructed using large numbers of molecular markers. NGS followed by the identification of SNPs, and genetically or physically linked groups of SNPs (SNP haplotype) is an ideal tool to perform high-density mapping. Li et al. (2009) constructed a *B. rapa* linkage map with EST-based SNP markers and identified genes associated with flowering time and leaf morphological traits. In addition, the *B. napus* SNP linkage map was constructed based on SNPs discovered by Illumina sequencing (Bancroft et al. 2011). An integrated genetic map containing 5,764 SNPs and 1,603 PCR markers in *B. napus* was made through SNP genotyping, to produce a higher density, more accurate map than those previously available (Delourme et al. 2013). These SNP maps are applicable to researching complex traits and they are also critical to the assembly of scaffolds in whole-genome sequencing.

Sequencing mixed DNA pools from lines with extreme trait variants in a population enable development of novel molecular markers linked to genes of interest. This is more rapid than gene identification and cloning using conventional methods. Mapping-by-sequencing based on resequencing of bulked segregants using SHOREmap software package (<http://1001genomes.org/software/shoremap.html>) can identify candidate genes, but it is usually limited by the requirement of a completed genome as reference. Galvao et al. (2012) developed a synteny-based method to perform mapping-by-sequencing with few markers in species where whole-genome sequences were unavailable but transcriptome assemblies were available. This was validated to be effective for genetic mapping in *A. thaliana* and its distant relative *B. rapa*. Due to the complexity of the tetraploid *Brassica* genome, at present the diploid progenitor *B. rapa* genome can be used as a reference for candidate gene identification (Tollenaere et al. 2012). A total of 70 SNPs associated with rapeseed pod shatter resistance were discovered and a major QTL was found on chromosome A9 through

the combination of NGS and BSA (Hu et al. 2012a). High-density genetic maps constructed by NGS technology is effective for the identification of SNP markers linked to target traits and can narrow the confidence intervals of QTLs of interest into smaller regions.

#### Association mapping

Traditional linkage mapping identifies the relationships between traits and linked markers following recent recombination events in biparental, structured populations. Meanwhile association mapping, also named as linkage disequilibrium mapping, can be used to identify genes linked with natural variation in populations. Although association mapping can be hampered by confounding population structure, leading to false positives and false negatives due to spurious correlations (Zhao et al. 2007), QTL mapping by association analysis can be a valid approach, for phenological, morphological and quality traits, e.g., in winter rapeseed (Honsdorf et al. 2010). Zhao et al. (2010) constructed a *B. rapa* core collection of 239 accessions for association mapping studies. The genetic loci for oil content, identified in association mapping, were also located within QTL intervals of linkage mapping in *B. napus* (Zou et al. 2010).

Candidate gene sequencing (CGS) and whole-genome scanning (WGS) of natural populations are the two main methods of association mapping. NGS offers abundant molecular markers, producing large quantities of genotyping data. In situations where no reference genomic sequence is available, WGS was carried out by applying SNP markers to gene expression variation data generated by RNA sequencing (Stower 2012). Given that the polyploid nature of *B. napus* complicates the assembly of genome sequences, and no reference genome is currently available, associative transcriptomics was proposed as a method to link molecular markers with trait variation identified in *B. napus* (Harper et al. 2012). This study found that QTLs for the glucosinolate content of seeds were located on genomic regions showing presence–absence variation for the gene of interest in the population. This research offers for a model pipeline for association genetics in species with complex genomes.

However, common association mapping has reduced ability to detect minor-effect QTLs. Hence, the approach of NAM was proposed to solve the problem of various minor-effect QTLs. A NAM population is composed of several recombination inbred line (RIL) families that are derived from the cross of diverse inbred lines to a single reference inbred line. This combines the advantages of linkage analysis and association mapping, enabling analysis of recent recombination events from segregation progeny and historic recombination events from parental inbred lines. NAM can also produce high-resolution mapping

with high allele richness for QTL detection, for example, for quantitative resistance traits (Poland et al. 2011) and genetic compositions of complex traits (Cook et al. 2011). The construction of NAM populations of *Brassica* is underway for genome-wide association analysis of complex traits (Cowling and Balazs 2010). But the disadvantage of this method is that the construction of NAM population is time-consuming, e.g., the hybrids between more than one parental line and the reference line were self-fertilized for six generations, which is a long process.

#### Transcriptome analysis in *Brassicac*s

Transcriptome sequencing is an alternative approach to reduce the size of test sequences but obtain almost equal gene information. In particular, NGS provided a new tool for transcriptome sequencing even where genomic sequence information is not available. The first technology used widely in transcriptome sequencing was Roche 454, due to its long sequence reads. Subsequently, the appearance of Illumina and SOLiD technologies, with high-throughput and relatively short-read lengths, dominated transcriptome sequencing in *Brassica* (Table 3). Bancroft et al. (2011) sequenced the leaf transcriptome of *B. napus* as well as its progenitor species, *B. rapa* and *B. oleracea* using Illumina NGS. The SNP linkage map comprising 23,037 markers in *B. napus* was constructed after the analysis of sequence variation in these species. Transcriptome sequencing with NGS can produce much important information in relation to gene discovery (Higgins et al. 2012), causal SNP discovery within genes, and the discovery of genomic structural loci, for instance, alternative splicing determinants (AS). Detailed information about SNP discovery is delineated above.

#### Digital gene expression

In this system, the transcript levels of given genes are quantified in silico by sequence read profiling, also termed digital gene expression. This aims to determine the level of gene expression in particular biological processes, developmental stages and following various treatments, based purely on NGS read abundance for a specific locus. Previously, DNA microarray technology was used for this, whereby the hybridization intensity determined the levels of gene expression. Trick et al. (2009a) developed a public *Brassica* microarray resource, using the assembly of about 800,000 EST sequences, to analyze gene expression in resynthesised *B. napus* lines and their parents. However, important limitations of this method are (1) sequence information is required, (2) the cost is high, (3) the results can be too complex and inconsistent to clearly interpret and (4) the candidate genes are difficult to determine (Table 4).

**Table 3** The main application of NGS in *Brassica*

Application	Details	Species	NGS technology	Reference
SNP discovery				
Transcriptome sequencing	About 40,000 putative SNP were discovered in Tapidor and Ningyou 7	<i>B. napus</i>	Illumina/Solexa	Trick et al. (2009a)
Sequencing of ESTs	655 putative SNP markers were discovered in cultivars ZY036 and 51070	<i>B. napus</i>	454 sequencing	Hu et al. (2012a)
RAD sequencing	604 SNPs were discovered	<i>B. napus</i>	Illumina/Solexa	Durstewitz et al. (2010)
Sequence capture	20,000 SNPs were identified	<i>B. napus</i>	Illumina GAIIX system	Bus et al. (2012)
	87 SNPs and 6 Indels were identified in six rapeseed varieties	<i>B. napus</i>	–	Westemeier et al. (2009)
	7,000 SNPs were identified, respectively, in <i>B. napus</i> cultivars Aviso and Montego	<i>B. napus</i>	–	Edwards et al. (2013)
Genetic mapping and gene location				
SNP haplotypes	The SNP linkage map was constructed	<i>B. napus</i>	Illumina/Solexa	Bancroft et al. (2011)
NGS-BSA	A major QTL associated with rapeseed pod shatter resistance were found	<i>B. napus</i>	Illumina/Solexa	Hu et al. (2012a)
Association mapping				
Associative transcriptomics	Finding deletion genomic regions in QTLs of glucosinolate content of seeds	<i>B. napus</i>	Illumina/Solexa	Harper et al. (2012)
Transcriptome analysis				
Leaf transcriptome analysis	23,037 markers were identified	<i>B. napus</i>	Illumina/Solexa	Bancroft et al. (2011)
Tag profiling	1,092 genes were discovered relating with water deficit	Chinese cabbage	Illumina/Solexa	Yu et al. (2012a)
	Over seven million ESTs were generated in four oilseeds	<i>B. napus</i>	454 pyrosequencing	Troncoso-Ponce et al. (2011)
Genes discovery	Genes related with stem swelling were discovered	<i>B. juncea</i>	Illumina/Solexa	Sun et al. (2012)
	Genes related with chloroplast were discovered	<i>B. oleracea</i>	Illumina/Solexa	Zhou et al. (2011b)
Alternative splicing	AS patterns are common and change rapidly after polyploidization, greatly contributing to the transcriptome shock	<i>B. napus</i>	–	Zhou et al. (2011a)
MicroRNA				
MiRNA discovery	50 conserved miRNAs, 11 new miRNAs and some miRNA targets were identified in different oil content at developmental stage	<i>B. napus</i>	Illumina/Solexa	Zhou et al. (2011a)
	228 novel and 321 conserved miRNAs were found	Chinese cabbage	Illumina/Solexa	Wang et al. (2012a)
	33 conserved and 19 new mRNA targets were found through degradome sequencing	<i>B. napus</i>	Illumina/Solexa	Xu et al. (2012a)
	59 <i>B. napus</i> miRNA families were found and were associated with seed development and energy storage	<i>B. napus</i>	Illumina	Korbes et al. (2012)

**Table 3** continued

Application	Details	Species	NGS technology	Reference
The differential expression of miRNAs	Heavy metal cadmium-regulated miRNAs were identified	<i>B. napus</i>	Illumina/Solexa	Zhou et al. (2012)
	MiRNAs responded to heat stress were identified	<i>B. rapa</i>	Illumina/Solexa	Wang et al. (2011a), Yu et al. (2012b)
	MiRNAs responded to arsenic stress were identified	<i>B. juncea</i>	Phalanx Plant OneArray	Srivastava et al. (2012)

**Table 4** Advantages and disadvantages of four methods of gene expression analysis

	Advantages	Disadvantages
DNA microarray	High throughput	Need to know gene sequences and design probe sequences; high cost; difficult to identify candidate genes
Serial analysis of gene expression (SAGE)	Do not need to know sequences of genes; no hybridizing; identifying new genes	Requiring quantity of RNA; high cost; considerable sequencing reactions
Massively parallel signature sequencing (MPSS)	Do not need to know sequences of genes; creating digital data; producing more sequences	Requiring quantity of RNA; high cost; relatively new technique
Digital gene expression	High throughput; discriminating homologous genes	Producing quantity of data; high cost

Serial analysis of gene expression (SAGE) (Obermeier et al. 2009) and massively parallel signature sequencing (MPSS) are another two traditional approaches for RNA sequencing. SAGE requires considerable sequencing reactions at high cost, while MPSS requires large quantities of mRNA (2.5–5  $\mu\text{g}$ ) to perform transcriptome analysis. Digital gene expression profiling based on NGS is becoming increasingly widely used in *Brassica* transcriptomics. This technique can discriminate homeologous gene expression in polyploids by comparing with the reference unigene sequences from diploid representative genomes (Higgins et al. 2012). Yu et al. (2012a) analyzed gene expression in a drought model of Chinese cabbage using Illumina NGS technology and found 1,092 genes associated with response to water deficit. In addition, over seven million ESTs were generated from four oilseeds including *B. napus* at four stages of development using 454 pyrosequencing, which can assist future functional and comparative genomic researches (Troncoso-Ponce et al. 2011). These suggest that digital gene expression has been of value in large-scale analysis of gene expression.

#### Gene discovery

In cases where genome sequences of species exist, unigenes can be obtained by assembly of RNA sequence reads (Wang et al. 2010; Chen et al. 2011). The transcriptome of tumourous stem mustard (*B. juncea* var. *tumida* Tsen et Lee) was sequenced by Illumina short-read technology, identifying 146,265 unigenes, in which 1,042 significantly expressed genes were associated with stem swelling and

development (Sun et al. 2012). Zhou et al. (2011b) identified 7,155 genes related to chloroplast development in *B. oleracea*, determined the role of regulatory genes by RNA sequencing with the Illumina Genome Analyzer II system, and discovered 1,600 up-regulated genes in light signaling pathways in green curd tissue. mRNA sequencing with NGS is not only a powerful tool for the identification of the genetic basis of the certain traits but will also help to accelerate gene expression and functional analysis. Numerous related genes were discovered, but major genes and their function were not validated in most studies. Major genes can be found through previous QTL mapping results, which had been made extensively, and then can be identified by real-time quantitative PCR or resequencing candidate gene PCR products.

Mutational sites can be identified directly by deep sequencing with NGS technologies. Short-read sequencing has been successfully applied in the identification of frame shift mutations in *A. thaliana*, with high specificity and high sensitivity, without prior gene information (Laitinen et al. 2010). Such mutant screens can be applied to the *Brassica* genome. Mutations in larger targeted regions or whole exomes of mutant populations were detected in *B. napus* using Illumina sequencing (Plant and Animal Genome meeting) (Sidebottom et al. 2012).

#### Alternative splicing (AS)

AS refers to the various ways that splicing introns in one eukaryotic pre-mRNA may result in several different mRNAs and protein products. AS is important for

determining the complexity of genes, and appears to occur in about 33 % of all rice genes (Zhang et al. 2010) with possibility of reaching 60 % in different tissues, ambient conditions or developmental stages (Syed et al. 2012). In *Brassica*, changes in gene characteristics and AS patterns are common after polyploidization, and AS changes greatly contribute to transcriptome shock, whereby extensive changes in gene expression pattern (Zhou et al. 2011a). However, dynamic changes in AS in whole genomes under different conditions or stresses, and the varied role of AS patterns in the evolution of plant species are still unknown. Akhunov et al. (2013) discovered a high level of AS pattern divergence in homeologous genes of wheat. This indicates that the dynamic changes of AS in polyploidization, different developmental stages, different tissues or different treatments are a promising research direction in *Brassica* species. With the advent of *B. napus* genome sequencing, the AS events and the role of AS in polyploidization will be identified.

### MicroRNA

MicroRNAs (miRNA) are a class of small non-coding RNA of about 18–30 nucleotides that regulate gene expression in plant development and response to environmental stress. miRNAs are widely expressed in plants and animals, even in mosses and fungi, because of their conserved functions in developmental processes. These small RNAs are produced from hairpin-shaped precursors (pri-miRNA) with the help of the endonuclease, DCL1 (DICER-LIKE 1). In the past, the discovery of novel miRNAs depended on cloning and sequencing of individual miRNAs, which could not be distinguished from other non-coding RNAs, such as rRNAs or tRNAs. Emerging microarray technology can detect miRNA genotypes on a large scale but is limited in the capacity to detect novel miRNAs. NGS technology opens new opportunities for novel miRNA discovery and profiling, as well as the identification of miRNA targets. By analyzing small RNA profiles of the embryos of *B. napus* with different oil contents at different developmental stages, a total of 50 conserved miRNAs, 11 new miRNAs and some miRNA targets were identified (Zhao et al. 2012). Korbes et al. (2012) detected 59 *B. napus* miRNA families at different seed developmental stages using Illumina sequencing, in which 13 were novel miRNA families. The putative functions of miRNA target genes were associated with seed development and energy storage. In Chinese cabbage, 228 novel and 321 conserved miRNAs were found using Illumina NGS technology, which laid the foundation for further study of miRNA regulation mechanisms (Wang et al. 2012a). However, these studies only focused on the discovery of miRNAs, but the core miRNAs and their role in *B. napus* were less studied. Numerous works

are still required to determine the function of miRNA in the polyploidization of *B. napus*. In addition, long non-coding RNA is a novel field and not reported in *Brassica* crops up to now. It will become a new star like small RNAs.

Noticeably, degradome sequencing is a vital method for identification of miRNA targeted mRNAs. Traditionally, target mRNA identification relies on computational prediction and subsequent experimental verification, which cannot only lead to inaccurate results but is also time-consuming. Degradome sequencing technology acts directly on the 3'-end of mRNA fragments with a poly-A tail, which are complementary with an miRNA, to figure out the target gene. Recently, Xu et al. (2012a) performed whole-mRNA degradome sequencing of *B. napus* and found 33 conserved and 19 new mRNA targets, providing for the first platform for miRNA regulation functional research in *Brassica*. This exploited the mixed mRNA samples from different tissues, and abundant targets were detected. But degradome sequencing can only detect the targets regulated by the transcript cleavage model, not by translational repression.

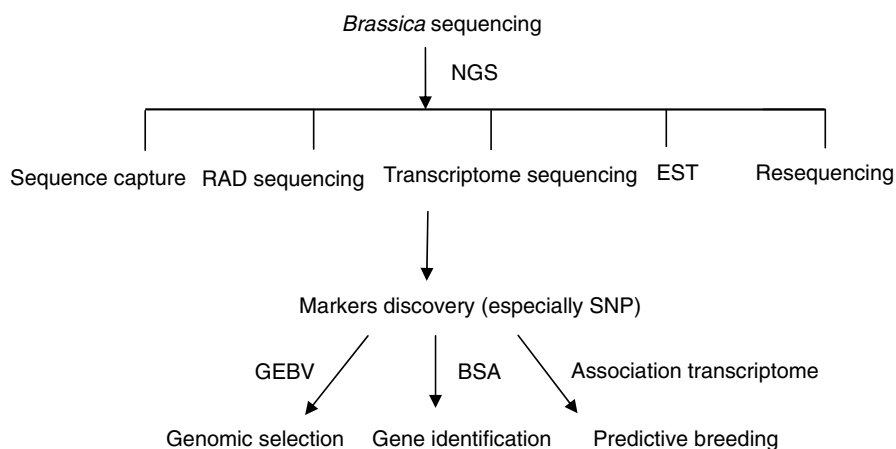
The differential expression of miRNAs in different developmental stages and diverse treatments is a main application of small RNA sequencing. Heavy metal cadmium-regulated miRNAs of *B. napus* were identified by Illumina sequencing technology and novel targets were found to participate in response to cadmium (Zhou et al. 2012). In *B. rapa*, miRNAs responding to heat stress were also found to play a key role in response to heat (Yu et al. 2012b; Wang et al. 2011a). Srivastava et al. (2012) showed via microarray profiling that miRNAs play a great role in response to arsenic stress in *B. juncea*. Indeed, both genetic and epigenetic factors, including heritable DNA methylation profiles and corresponding small RNA activities, contribute to phenotypic traits. Therefore, a combination of genome sequencing, single-base methylation sequencing, transcriptome expression analysis and small RNA analysis will form the basis of complex trait research.

### Marker-assisted selection (MAS) and genome selection in breeding

It is essential to understand the association between phenotypic variation of traits of interest and their intrinsic genetic variation at the DNA sequence level in crop improvement breeding. Molecular markers can be used to accelerate the process of traditional breeding. In backcross breeding, selection toward a genetic background is a critical step in deciding the number of backcrosses with the recurrent parent. MAS can accelerate this process by tracking target genes with linked markers to eliminate linkage drag. Although the idea of MAS has been put forward for many years, successful breeding outcomes are rare and its use is usually blocked by most quantitative



**Fig. 2** The application of NGS in *Brassica* crop breeding



RAD sequencing: restriction-site associated DNA sequencing;

GEBV: genomic estimated breeding value;

BSA: bulked segregant analysis

traits, because only a small number of markers associated with a phenotype are identified and genotyping cost is relatively high. Another method, termed genomic selection, is proposed to solve such problems in breeding, and estimates the values of individual lines with high density markers across the whole genome (Fig. 2). Heffner et al. (2009) made a simulation between the true breeding value and genomic estimated breeding value (GEBV) and the correlation coefficient between them reached 0.85, which indicated that GEBV could be used to estimate the true breeding value. Genomic selection is an ideal tool to select lines of interest with markers alone, circumventing the need for costly and time-consuming phenotyping of breeding lines. The accuracy with genomic selection using ridge regression-best linear unbiased prediction was higher than conventional MAS via composite interval mapping (Guo et al. 2012b). Likewise in *Brassica*, genomic selection will accelerate breeding cycles with the aim to meet the increasing demands of production. However, the lack of valuable markers linked to important agronomic traits has negatively impacted research into genomic selection in *Brassica*. In addition, it is a challenge to develop homeologous gene-specific markers because it is very difficult to select efficient loci which distinguishes different homeologous genes with high similarity.

#### Future emphasized research area of NGS in *Brassica*

With the production of high-quality reference genomes for *Brassica* species, the stage is set for numerous genomics studies. For *Brassica* crops, we have emphasized some fields that require further analysis (Fig. 1). These are:

1. Detailed gene structure. The start and end of the promoter (including transcription factor binding sites (TFBS), untranslated regions (UTR) (including 5-end and 3-end UTRs), coding regions, introns, exons and even enhancers of each gene should be determined.
2. Characteristics of gene families and duplicated genes. Classification, evolution and transcriptional analysis of gene families and duplicated or homologous/homeologous genes should be done.
3. Spatial analysis of pseudogenes. The position, number and role of pseudogenes and their homologous/homeologous genes are worth investigation since they may play a role in species polyploidization and possibly the generation of small RNAs to regulate other genes (Guo et al. 2009).
4. Genome resequencing analysis. Genome resequencing should be done for SNP and Indel discovery and analysis of the position, copy number variation (CNV), and phylogeny of genes among different materials for analysis of species evolution or domestication. A haplotype map or genome variation map should also be produced.
5. Boundaries and novel function of repeat sequences. The boundaries of different kinds of repeat sequences especially transposon sequences should be defined. Though many tools are used to resolve this problem, there are no suitable tools that can adjust the prediction strategies according to the specificity of individual species (Myrick and Gelbart 2007). For *Brassica* crops enriched for repeat sequences, their boundary definition and classification is a big problem. In addition, the novel functions of repeat sequences should be clarified. For example, can some transposons gen-

erate small RNAs that regulate other coding genes? Which transposons are active, or can be induced to become active when environment changes? What roles do they play in maintaining species propagation?

6. Alternative splicing analysis. Alternative splicing events, new genes and fusion genes should be identified and resolved by high depth transcriptome sequencing.
7. Function and origin of small RNAs. The function of numerous small RNAs need be determined. Nowadays, NGS predicts or offers considerable miRNAs in *Brassica* species, but their functions, origins, generation mechanisms and variation/evolution remain unclear.
8. Relationship and function of genome composition. What are the relationship of different genome composition and their functions? Whether the genes, transposons and small RNAs are unevenly distributed (e.g., gene cluster, transposons cluster and small RNA cluster)? Wei et al. (2013) found that nested-LTR retrotransposons were distributed in six *Brassica* BAC clones, and these played great role in the formation of centromeres. If so, how does evolution or domestication affect this phenomenon?
9. Effect of evolution on genome. How does evolution, domestication or artificial selection shape genome structure, distribution and structural variation of different composites in the genome, and alter gene function, transposons and small RNAs?
10. Development of newer and user-friendly software. Although the cost of NGS continues to decline, it can still be prohibitive for analysis of large collections of various accessions. Meanwhile, vast amounts of data are generated from NGS, but errors still exist and intensive, professional computational tools are needed to store and deal with these large amounts of data. Currently, analysis tools are only mastered by a few trained professionals. This can be a limitation for obtaining enough useful sequence information from a lot of sequence reads. Newer, user-friendly software needs to be developed and applied in order to match the fast development of sequencing technology.

## Outlook

Improvements in third-generation sequencing technology are promising for directly acquiring long sequence reads to reduce the complexity and cost of genome assembly. It is worth highlighting that the study of all species can be individualized to reveal the role of gene expression by high-throughput sequencing of respective tissue and organ,

or individual plant response to different treatments and environment. The human genome project was completed in 2003 and had made considerable progress in the application of NGS technology in disease diagnosis and gene functional analyses. In *Brassicaceae*, information gained from the completed genome of *B. rapa*, for instance molecular markers, can be transferred to other *Brassica* crops. Xu et al. (2010) constructed an integrated genetic map of the A genome in *B. napus* using SSR markers originating from *B. rapa* sequenced BACs. In the near future, the release of the *B. oleracea* and *B. napus* genomes will greatly accelerate the large-scale application of NGS in *Brassica* species. This has important implications for downstream functional genomic and epigenetic analyses of the control of important agronomical traits in *Brassicaceae* and other complex crop species.

## References

- Akhunov E, Sehgal S, Liang H, Wang S, Akhunova A, Kaur G, Li W, Forrest K, See D, Simkova H, Ma Y, Hayden M, Luo M, Faris J, Dolezel J, Gill B (2013) Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiol* 161(1):252–265
- Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, Arondel V, Ohlrogge J, Saiee IJ, Suliman-Elmeier KM, Bennetzen JL, Kruegger RR, Malek JA (2011) De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat Biotechnol* 29(6):521–527
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Report* 9(3):208–218
- Ashelford K, Eriksson ME, Allen CM, D'Amore R, Johansson M, Gould P, Kay S, Millar AJ, Hall N, Hall A (2011) Full genome re-sequencing reveals a novel circadian clock mutation in *Arabidopsis*. *Genome Biol* 12(3):R28
- Atwell S, Huang YS, Vilhjalmsdottir BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465(7298):627–631
- Aversano R, Ercolano MR, Caruso I, Fasano C, Rosellini D, Carputo D (2012) Molecular tools for exploring polyploid genomes in plants. *Int J Mol Sci* 13(8):10316–10335
- Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng J, Wang X, Liu S, Trick M (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nat Biotechnol* 29(8):762–766
- Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, May GD, Woodward JE, Caillier SJ, McElroy JP, Gomez R, Pando MJ, Clendenen LE, Ganusova EE, Schilkey FD, Ramaraj T, Khan OA, Huntley JJ, Luo S, Kwok PY, Wu TD, Schroth GP, Oksenberg JR, Hauser SL, Kingsmore SF (2010) Genome,

- epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 464(7293):1351–1356
- Berkman PJ, Lai KT, Lorenc MT, Edwards D (2012) Next-generation sequencing applications for wheat crop improvement. *Am J Bot* 99(2):365–371
- Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA (2008) The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26(10):1146–1153
- Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426):705–710
- Buggs JRA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2012) Next-generation sequencing and genome evolution in allopolyploids. *Am J Bot* 99(2):372–382
- Bus A, Hecht J, Huettel B, Reinhardt R, Stich B (2012) High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *BMC Genomics* 13(1):281
- Chen S, Luo H, Li Y, Sun Y, Wu Q, Niu Y, Song J, Lv A, Zhu Y, Sun C, Steinmetz A, Qian Z (2011) 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Rep* 30(9):1593–1601
- Chen X, Li X, Zhang B, Xu J, Wu Z, Wang B, Li H, Younas M, Huang L, Luo Y, Wu J, Hu S, Liu K (2013) Detection and genotyping of restriction fragment associated polymorphisms in polyploid crops with a pseudo-reference sequence: a case study in allotetraploid *Brassica napus*. *BMC Genomics* 14:346
- Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, van Tassel CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics* 176(1):685–696
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452(7184):215–219
- Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2011) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol* 158(2):824–834
- Cowling WA, Balazs E (2010) Prospects and challenges for genome-wide association and genomic selection in oilseed *Brassica* species. *Genome* 53(11):1024–1028
- 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(7):499–510
- Delourme R, Falentin C, Fomeju BF, Boillot M, Lassalle G, Andre I, Duarte J, Gauthier V, Lucante N, Marty A, Pauchon M, Pichon JP, Ribiere N, Trotoux G, Blanchard P, Riviere N, Martinant JP, Pauquet J (2013) High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L. *BMC Genomics* 14:120
- Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen JM, Nery JR, Dixon JE, Ecker JR (2012) Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci USA* 109(32):E2183–E2191
- Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J, Edwards D (2009) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. *Nucleic Acids Res* 37(Database issue):D951–D953
- Durstewitz G, Polley A, Plieske J, Luerssen H, Graner EM, Wieseke R, Ganai MW (2010) SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*. *Genome* 53(11):948–956
- Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. *Plant Biotechnol J* 8(1):2–9
- Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. *Theor Appl Genet* 126(1):1–11
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910):133–138
- Galvao VC, Nordstrom KJ, Lanz C, Sulz P, Mathieu J, Pose D, Schmid M, Weigel D, Schneeberger K (2012) Synteny-based mapping-by-sequencing enabled by targeted enrichment. *Plant J* 71(3):517–526
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296(5565):92–100
- Guo X, Zhang Z, Gerstein MB, Zheng D (2009) Small RNAs originated from pseudogenes: cis- or trans-acting? *PLoS Comput Biol* 5:e1000449
- Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R, Wu M, Levi A, Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z, Xu Y (2012a) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet* 45(1):51–58
- Guo Z, Tucker DM, Lu J, Kishore V, Gay G (2012b) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. *Theor Appl Genet* 124(2):261–275
- Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat Biotechnol* 30(8):798–802
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic Selection for crop improvement. *Crop Sci* 49(1):1–12

- Henry RJ, Edwards M, Waters DL, Gopala Krishnan S, Bundock P, Sexton TR, Masouleh AK, Nock CJ, Pattermore J (2012) Application of large-scale sequencing to marker discovery in plants. *J Biosci* 37(5):829–841
- Higgins JA, Magusin A, Trick M, Fraser F, Bancroft I (2012) Use of mRNA-Seq to discriminate contributions to the transcriptome from the constituent genomes of the polyploid crop species *Brassica napus*. *BMC Genomics* 13(1):247
- Honsdorf N, Becker HC, Ecker W (2010) Association mapping for phenological, morphological, and quality traits in canola quality winter rapeseed (*Brassica napus* L.). *Genome* 53(11):899–907
- Hu Z, Hua W, Huang S, Yang H, Zhan G, Wang X, Liu G, Wang H (2012a) Discovery of pod shatter-resistant associated SNPs by deep sequencing of a representative library followed by bulk segregant analysis in rapeseed. *PLoS ONE* 7(4):e34253
- Hu ZY, Huang SM, Sun MY, Wang HZ, Hua W (2012b) Development and application of single nucleotide polymorphism markers in the polyploid *Brassica napus* by 454 sequencing of expressed sequence tags. *Plant Breed* 131(2):293–299
- Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan WuZ, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* 41(12):1275–1281
- International Human Genome Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431(7011):931–945
- International Peach Genome Initiative (2010). <http://www.rosaceae.org/peach/genome>
- Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY, Town CD, Venter JC, Fraser CM, Tabata S, Nakamura Y, Kaneko T, Sato S, Asamizu E, Kato T, Kotani H, Sasamoto S, Ecker JR, Theologis A, Federspiel NA, Palm CJ, Osborne BI, Shinn P, Conway AB, Vysotskaia VS, Dewar K, Conn L, Lenz CA, Kim CJ, Hansen NF, Liu SX, Buehler E, Altafi H, Sakano H, Dunn P, Lam B, Pham PK, Chao Q, Nguyen M, Yu GX, Chen HM, Southwick A, Lee JM, Miranda M, Toriumi MJ, Davis RW, Wambutt R, Murphy G, Dusterhoft A, Stiekema W, Pohl T, Entian KD, Terryn N, Volckaert G, Salanoubat M, Choisne N, Rieger M, Ansoerge W, Unseld M, Fartmann B, Valle G, Artiguenave F, Weissenbach J, Quetier F, Wilson RK, de la Bastide M, Sekhon M, Huang E, Spiegel L, Gnoj L, Pepin K, Murray J, Johnson D, Habermann K, Dedhia N, Parnell L, Preston R, Hillier L, Chen E, Marra M, Martienssen R, McCombie WR, Mayer K, White O, Bevan M, Lemcke K, Creasy TH, Bielke C, Haas B, Haase D, Maiti R, Rudd S, Peterson J, Schoof H, Frishman D, Morgenstern B, Zaccaria P, Ermolaeva M, Perlea M, Quackenbush J, Volfovsky N, Wu DY, Lowe TM, Salzberg SL, Mewes HW, Rounsley S, Bush D, Subramanian S, Levin I, Norris S, Schmidt R, Acarkan A, Bancroft I, Quetier F, Brennicke A, Eisen JA, Bureau T, Legault BA, Le QH, Agrawal N, Yu Z, Martienssen R, Copenhaver GP, Luo S, Pikaard CS, Preuss D, Paulsen IT, Sussman M, Britt AB, Selinger DA, Pandey R, Mount DW, Chandler VL, Jorgensen RA, Pikaard C, Juergens G, Meyerowitz EM, Theologis A, Dangl J, Jones JDG, Chen M, Chory J, Somerville MC, In AG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Klein HU, Bartenhagen C, Kohlmann A, Grossmann V, Ruckert C, Haferlach T, Dugas M (2011) R453Plus1Toolbox: an R/Bioconductor package for analyzing Roche 454 Sequencing data. *Bioinformatics* 27(8):1162–1163
- Koo DH, Hong CP, Batley J, Chung YS, Edwards D, Bang JW, Hur Y, Lim YP (2011) Rapid divergence of repetitive DNAs in *Brassica* relatives. *Genomics* 97(3):173–185
- Korbes AP, Machado RD, Guzman F, Almerao MP, de Oliveira LF, Loss-Morais G, Turchetto-Zolet AC, Cagliari A, Dos Santos-Maraschin F, Margis-Pinheiro M, Margis R (2012) Identifying conserved and novel microRNAs in developing seeds of *Brassica napus* using deep sequencing. *PLoS ONE* 7(11):e50666
- Laitinen RA, Schneeberger K, Jelly NS, Ossowski S, Weigel D (2010) Identification of a spontaneous frame shift mutation in a nonreference *Arabidopsis* accession using whole genome sequencing. *Plant Physiol* 153(2):652–654
- Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS, Zhang G (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* 42(12):1053–1059
- Li F, Kitashiba H, Inaba K, Nishio T (2009) A *Brassica rapa* linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. *DNA Res* 16(6):311–323
- Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, Zhang Z, Zhang Y, Wang W, Li J, Wei F, Li H, Jian M, Li J, Zhang Z, Nielsen R, Li D, Gu W, Yang Z, Xuan Z, Ryder OA, Leung FC, Zhou Y, Cao J, Sun X, Fu Y, Fang X, Guo X, Wang B, Hou R, Shen F, Mu B, Ni P, Lin R, Qian W, Wang G, Yu C, Nie W, Wang J, Wu Z, Liang H, Min J, Wu Q, Cheng S, Ruan J, Wang M, Shi Z, Wen M, Liu B, Ren X, Zheng H, Dong D, Cook K, Shan G, Zhang H, Kosiol C, Xie X, Lu Z, Zheng H, Li Y, Steiner CC, Lam TT, Lin S, Zhang Q, Li G, Tian J, Gong T, Liu H, Zhang D, Fang L, Ye C, Zhang J, Hu W, Xu A, Ren Y, Zhang G, Bruford MW, Li Q, Ma L, Guo Y, An N, Hu Y, Zheng Y, Shi Y, Li Z, Liu Q, Chen Y, Zhao J, Qu N, Zhao S, Tian F, Wang X, Wang H, Xu L, Liu X, Vinar T, Wang Y, Lam TW, Yiu SM, Liu S, Zhang H, Li D, Huang Y, Wang X, Yang G, Jiang Z, Wang J, Qin N, Li L, Li J, Bolund L, Kristiansen K, Wong GK, Olson M, Zhang X, Li S, Yang H, Wang J, Wang J (2010) The sequence and de novo assembly of the giant panda genome. *Nature* 463(7279):311–317
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133(3):523–536
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322
- Lorenc M, Boskovic Z, Stiller J, Duran C, Edwards D (2012) Role of bioinformatics as a tool for oilseed *Brassica* species. *Genetics, genomics and breeding of oilseed Brassicas*. Science Publishers Inc, New Hampshire, pp 194–205
- Love CG, Graham NS, Lochlainn SO, Bowen HC, May ST, White PJ, Broadley MR, Hammond JP, King GJ (2010) A *Brassica* exon array for whole-transcript gene expression profiling. *PLoS ONE* 5:e12812
- Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *Int J Plant Genomics* 2012:728398
- Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, Wise RP,

- Stein N (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426):711–716
- McCouch SR, Zhao KY, Wright M, Tung CW, Ebana K, Thomson M, Reynolds A, Wang D, DeClerck G, Ali ML, McClung A, Eizenga G, Bustamante C (2010) Development of genome-wide SNP assays for rice. *Breed Sci* 60(5):524–535
- Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, Schatz M, Nagarajan N, Acob RA, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na JK, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Perez R, Torres MJ, Feltus FA, Porter B, Li Y, Burroughs AM, Luo MC, Liu L, Christopher DA, Mount SM, Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, dePamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya Linnaeus*). *Nature* 452:991–6
- Minoche AE, Dohm JC, Himmelbauer H (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* 12(11):R112
- Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA (2013) A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* 110:3–24
- Mun JH, Kwon SJ, Yang TJ, Kim HS, Choi BS, Baek S, Kim JS, Jin M, Kim JA, Lim MH, Lee SI, Kim HI, Kim H, Lim YP, Park BS (2008) The first generation of a BAC-based physical map of *Brassica rapa*. *BMC Genomics* 9:280
- Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol* 10(10):R111
- Mun JH, Kwon SJ, Seol YJ, Kim JA, Jin M, Kim JS, Lim MH, Lee SI, Hong JK, Park TH, Lee SC, Kim BJ, Seo MS, Baek S, Lee MJ, Shin JY, Hahn JH, Hwang YJ, Lim KB, Park JY, Lee J, Yang TJ, Yu HJ, Choi IY, Choi BS, Choi SR, Ramchiary N, Lim YP, Fraser F, Drou N, Soumpourou E, Trick M, Bancroft I, Sharpe AG, Parkin IA, Batley J, Edwards D, Park BS (2010) Sequence and structure of *Brassica rapa* chromosome A3. *Genome Biol* 11(9):R94
- Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PG (2012) Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. *PLoS ONE* 7(2):e31410
- Myrick KV, Gelbart WM (2007) A modified universal fast walking method for single-tube transposon mapping. *Nat Protoc* 2:1556–1563
- Obermeier C, Hosseini B, Friedt W, Snowdon R (2009) Gene expression profiling via LongSAGE in a non-model plant species: a case study in seeds of *Brassica napus*. *BMC Genomics* 10:295
- Ozsolak F, Milos PM (2011) Single-molecule direct RNA sequencing without cDNA synthesis. *Wiley Interdiscip Rev RNA* 2(4):565–570
- Parkin IA, Clarke WE, Sidebottom C, Zhang W, Robinson SJ, Links MG, Karcz S, Higgins EE, Fobert P, Sharpe AG (2010) Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. *Genome* 53(11):929–938
- Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberger G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Ohtillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551–556
- Picho J-PD, Rivière 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. In: *Plant and Animal Genomes XVIII Conference*, San Diego
- Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc Natl Acad Sci USA* 108(17):6893–6898
- Pushkarev D, Neff NF, Quake SR (2009) Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 27(9):847–850
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74(12):5463–5467
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278):178–183
- Sidebottom CH, koh CS, Gilchrist E, George H, Sharpe A (2012) Mutation detection in *Brassica napus* using Illumina sequencing. In: *International plant and animal genome conference*, San Diego, CA
- Snowdon RJ, Luy FLI (2012) Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breed* 131(3):351–360
- Srivastava S, Srivastava AK, Suprasanna P, D'Souza SF (2012) Identification and profiling of arsenic stress-induced microRNAs in *Brassica juncea*. *J Exp Bot* 64(1):303–315
- Stower H (2012) Plant genomics: associative transcriptomics. *Nat Rev Genet* 13:597
- Strickler SR, Bombarely A, Mueller LA (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species I. *Am J Bot* 99(2):257–266
- Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, Vaddadi S, Henry RJ (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol J* 10(6):623–634
- Sun Q, Zhou G, Cai Y, Fan Y, Zhu X, Liu Y, He X, Shen J, Jiang H, Hu D, Pan Z, Xiang L, He G, Dong D, Yang J (2012) Transcriptome analysis of stem development in the tumorous stem mustard *Brassica juncea* var. *tumida* Tsen et Lee by RNA sequencing. *BMC Plant Biol* 12:53
- Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants—coming of age. *Trends Plant Sci* 17(10):616–623
- The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463(7282):763–768
- Tollenaere R, Hayward A, Dalton-Morgan J, Campbell E, Lee JR, 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 Biotechnol 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(6):1348–1359
- Trick M, Cheung F, Drou N, Fraser F, Lobenhofer EK, Hurban P, Magusin A, Town CD, Bancroft I (2009a) A newly-developed community microarray resource for transcriptome profiling in *Brassica* species enables the confirmation of *Brassica*-specific expressed sequences. *BMC Plant Biol* 9:50. doi10.1186/1471-2229-9-50
- Trick M, Long Y, Meng J, Bancroft I (2009b) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol J* 7(4):334–346
- Troncoso-Ponce MA, Kilaru A, Cao X, Durrett TP, Fan J, Jensen JK, Thrower NA, Pauly M, Wilkerson C, Ohlrogge JB (2011) Comparative deep transcriptional profiling of four developing oilseeds. *Plant J* 68(6):1014–1027
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryabov D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793):1596–1604
- U N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* 7:389–452
- Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol* 27(9):522–530
- Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, Chen X, Li Y (2010) De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11:726
- Wang L, Yu X, Wang H, Lu YZ, de Ruiter M, Prins M, He YK (2011a) A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*). *BMC Genomics* 12:289
- Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park BS, Weissshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IA, Batley J, Kim JS, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Wang J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon SJ, Choi SR, Lee TH, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011b) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43(10):1035–1039
- Wang F, Li L, Liu L, Li H, Zhang Y, Yao Y, Ni Z, Gao J (2012a) High-throughput sequencing discovery of conserved and novel microRNAs in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Mol Genet Genomics* 287(7):555–563
- Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohler RJ, Percy RG, Yu JZ, Zhu YX, Wang J, Yu S (2012b) The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* 44(10):1098–1103
- Wei LJ, Xiao ML, An ZS, Ma B, Mason AS, Qian W, Li JN, Fu DH (2013) New insights into nested long terminal repeat retrotransposons in *Brassica* species. *Mol Plant* 6:470–482
- Westermeier P, Wenzel G, Mohler V (2009) Development and evaluation of single-nucleotide polymorphism markers in allotetraploid rapeseed (*Brassica napus* L.). *Theor Appl Genet* 119:1301–1311
- Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, Cheng T, Jiang T, Becquet C, Xu X, Liu C, Zha X, Fan W, Lin Y, Shen Y, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu H, Yu C, Zhang G, Li J, Cao J, Liu S, He N, Zhou Y, Liu H, Zhao J, Ye C, Du Z, Pan G, Zhao A, Shao H, Zeng W, Wu P, Li C, Pan M, Li J, Yin X, Li D, Wang J, Zheng H, Wang W, Zhang X, Li S, Yang H, Lu C, Nielsen R, Zhou Z, Wang J, Xiang Z, Wang J (2009) Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* 326(5951):433–436
- Xu J, Qian X, Wang X, Li R, Cheng X, Yang Y, Fu J, Zhang S, King GJ, Wu J, Liu K (2010) Construction of an integrated genetic linkage map for the A genome of *Brassica napus* using SSR markers derived from sequenced BACs in *B. rapa*. *BMC Genomics* 11:594
- Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, Orjeda G, Guzman F, Torres M, Lozano R, Ponce O, Martinez D, De la Cruz G, Chakrabarti SK, Patil VU, Skryabin KG, Kuznetsov BB, Ravin NV, Kolganova TV, Beletsky AV, Mardanov AV, Di Genova A, Bolser DM, Martin DM, Li G, Yang Y, Kuang H, Hu Q, Xiong X, Bishop GJ, Sagredo B, Mejia N, Zagorski W, Gromadka R, Gawor J, Szczesny P, Huang S, Zhang Z, Liang C, He J, Li Y, He Y, Xu J, Zhang Y, Xie B, Du Y, Qu D, Bonierbale M, Ghislain M, Herrera Mdel R, Giuliano G, Pietrella M, Perrotta G, Facella P, O'Brien K, Feingold SE, Barreiro LE, Massa GA, Diambra L, Whitty BR, Vaillancourt B, Lin H, Massa AN, Geoffroy M, Lundback S, DellaPenna D, Buell CR, Sharma SK, Marshall DF, Waugh R, Bryan GJ, Destefanis M, Nagy I, Milbourne D, Thomson SJ, Fiers M, Jacobs JM, Nielsen KL, Sonderkaer M, Iovene M, Torres GA, Jiang J, Veilleux RE, Bachem CW, de Boer J, Borm T, Kloosterman B, van Eck H, Datema E, Hekkert BL, Govere A, van Ham RC, Visser RG (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195
- Xu M, Dong Y, Zhang Q, Zhang A, Luo Y, Sun J, Fan Y, Wang L (2012a) Identification of miRNAs and their targets from *Brassica napus* by high-throughput sequencing and degradome analysis. *BMC Genomics* 13:42
- Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao WB, Hao BH, Lyon MP, Chen J, Gao S, Xing F, Lan H, Chang

- JW, Ge X, Lei Y, Hu Q, Miao Y, Wang L, Xiao S, Biswas MK, Zeng W, Guo F, Cao H, Yang X, Xu XW, Cheng YJ, Xu J, Liu JH, Luo OJ, Tang Z, Guo WW, Kuang H, Zhang HY, Roose ML, Nagarajan N, Deng XX, Ruan Y (2012b) The draft genome of sweet orange (*Citrus sinensis*). *Nat Genet* 45:59–66
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. *ssp. indica*). *Science* 296(5565):79–92
- Yu SC, Zhang FL, Yu YJ, Zhang DS, Zhao XY, Wang WH (2012a) Transcriptome profiling of dehydration stress in the Chinese cabbage (*Brassica rapa* L. *ssp. pekinensis*) by tag sequencing. *Plant Mol Biol Report* 30(1):17–28
- Yu X, Wang H, Lu Y, de Ruiter M, Cariaso M, Prins M, van Tunen A, He Y (2012b) Identification of conserved and novel microRNAs that are responsive to heat stress in *Brassica rapa*. *J Exp Bot* 63(2):1025–1038
- Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X, Chen L, Tian W, Tao Y, Kristiansen K, Zhang X, Li S, Yang H, Wang J, Wang J (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res* 20(5):646–654
- Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet* 3(1):e4
- Zhao J, Artemyeva A, Del Carpio DP, Basnet RK, Zhang N, Gao J, Li F, Bucher J, Wang X, Visser RG, Bonnema G (2010) Design of a *Brassica rapa* core collection for association mapping studies. *Genome* 53(11):884–898
- Zhao YT, Wang M, Fu SX, Yang WC, Qi CK, Wang XJ (2012) Small RNA profiling in two *Brassica napus* cultivars identifies microRNAs with oil production- and development-correlated expression and new small RNA classes. *Plant Physiol* 158(2):813–823
- Zhou R, Moshgabadi N, Adams KL (2011a) Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc Natl Acad Sci USA* 108(38):16122–16127
- Zhou X, Fei Z, Thannhauser TW, Li L (2011b) Transcriptome analysis of ectopic chloroplast development in green curd cauliflower (*Brassica oleracea* L. var. *botrytis*). *BMC Plant Biol* 11:169
- Zhou ZS, Song JB, Yang ZM (2012) Genome-wide identification of *Brassica napus* microRNAs and their targets in response to cadmium. *J Exp Bot* 63(12):4597–4613
- Zou J, Jiang C, Cao Z, Li R, Long Y, Chen S, Meng J (2010) Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. *Genome* 53(11):908–916