Theor Appl Genet (2015) 128:779–795
DOI 10.1007/s00122-015-2499-8

REVIEW

Sequencing consolidates molecular markers with plant breeding practice

Huaan Yang[1] · Chengdao Li[1,2] · Hon-Ming Lam[3] · Jonathan Clements[1] · Guijun Yan[4] · Shancen Zhao[3,5]

Received: 18 September 2014 / Accepted: 14 March 2015 / Published online: 28 March 2015

Abstract
Key message **Plenty of molecular markers have been developed by contemporary sequencing technologies, whereas few of them are successfully applied in breeding, thus we present a review on how sequencing can facilitate marker-assisted selection in plant breeding.**
Abstract The growing global population and shrinking arable land area require efficient plant breeding. Novel strategies assisted by certain markers have proven effective for genetic gains. Fortunately, cutting-edge sequencing technologies bring us a deluge of genomes and genetic variations, enlightening the potential of marker development. However, a large gap still exists between the potential of molecular markers and actual plant breeding practices. In this review, we discuss marker-assisted breeding from a historical perspective, describe the road from crop sequencing to breeding, and highlight how sequencing facilitates the application of markers in breeding practice.

Communicated by R. K. Varshney.

✉ Huaan Yang
huaan.yang@agric.wa.gov.au

✉ Shancen Zhao
zhaoshancen@genomics.cn

1 Department of Agriculture and Food Western Australia, 3 Baron-Hay Court, South Perth 6151, Australia

2 State Agricultural Biotechnology Centre, Murdoch University, Murdoch, WA 6150, Australia

3 Center for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

4 Faculty of Sciences and The UWA Institute of Agriculture, School of Plant Biology, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

5 BGI-Shenzhen, Shenzhen 518083, China

Introduction

Plant domestication by human beings commenced some 10,000 years ago. Gregor Mendel's classical pea experiments in the 1850s laid the foundation of plant genetics and traditional plant breeding that was primarily based on bi-parental crossing. Over the last few decades, the emergence of modern plant breeding technologies has further enhanced the power of plant genetic improvement. The development of hybrid rice and maize has demonstrated the potential for dramatic yield increases simply by combining beneficial genes and alleles (Cheng et al. 2004; Moose and Mumm 2008). Double haploid (DH) breeding either by tissue culture or by inducer-based approach to produce homozygous diploid plants from haploid pollen cells has greatly shortened the breeding cycle to expedite the release of new cultivars (Xu 2010). A fast generation system based on single seed decent has also been developed in a number of plants to complement the DH technology for the quick production of recombinant inbreeding lines or near isogenic lines (Zheng et al. 2013; Ma et al. 2011). Recurrent selection employs the power of re-selection from generation to generation with selected plants as the initial population each cycle, ensuring that superior inbred plants can be easily isolated and selected (Vales et al. 2009). Multi-parent advanced generation inter-cross (MAGIC) breeding employs complex crossing that introduces multiple alleles to increase recombination resolutions (Cavanagh et al. 2008). Transgenic plants containing artificially inserted genes also have significant economic benefit to farming communities.

In both classical and modern plant breeding approaches, markers are important to accelerate genetic improvement. Although thousands of articles were published with the term "marker-assisted selection" (MAS) or "quantitative trait loci" (QTLs) or "molecular markers", a large gap still exists between the expectations and actual applications in practical plant breeding (Xu and Crouch 2008; Brumlop and Finckh 2011). This review attempts to identify the major factors limiting the application of molecular markers in plant breeding, and illustrates how sequencing technologies, if used properly, could resolve these limitations and advance molecular plant breeding.

Historical view of marker-assisted plant breeding

Plant breeding essentially includes two major activities: the creation of genetic variation and the selection of beneficial traits. Marker development, in association with agronomic traits, is crucial for the transition from crop genomics to breeding practice. In marker-assisted plant breeding, economic consideration usually comes first. To achieve effective cost, molecular markers for application in plant breeding must satisfy the following criteria: (1) polymorphic and evenly distributed in the crop genome to provide adequate genetic resolution for marker-trait association discovery, (2) cost-effective in genotyping procedure, (3) high tolerance of impurity in DNA samples from rapid and inexpensive extraction methods, and (4) amenable to automation to efficiently process large numbers of crop samples.

Different molecular breeding methods require different number of markers to work with, which in turn determines the overall cost and the choice of genotyping platforms. Marker-based molecular breeding methods include marker-assisted germplasm evaluation (MAGE), marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), and marker-assisted gene pyramiding (Xu 2010; Varshney et al. 2012b). These applications usually involve in a small number of plant samples, while large scale MAS is commonly associated with screening and selection on large scale of F_2 populations (Collard and Mackill 2008).

DNA markers developed in the last 30 years can be roughly classified into two categories according to their suitability for application in plant breeding: (1) markers without sequence information generated from generic DNA fingerprinting, and (2) markers with sequence specificity to alleles in the genome. The first generation DNA marker developed in the mid-1980s was restriction fragment length polymorphism (RFLP), which required a large amount of high purity DNA to run. RFLPs are no longer used nowadays because more efficient genotyping methods emerged and many RFLPs have been replaced by or converted to PCR-based methods. Several prevailing PCR-based marker systems were developed a few years later such as random amplified polymorphic DNA (RAPD) markers. Unfortunately, these markers were later found to be unreliable and unrepeatable. Amplified fragment length polymorphism (AFLP) markers appeared in the mid-1990s with improved reliability and reasonable efficiency. Microsatellite-anchored fragment length polymorphism (MFLP) markers were developed taking advantage of high efficiency of AFLP in capturing SSR markers. In the early 2000s, microarray-based DNA fingerprinting methods (e.g. DArT) were invented to increase the volume of markers for genotyping. However, it is very difficult to apply those markers in MAS practice due to several reasons. For example, the genotyping procedures in AFLP, MFLP and DArT are too complex for automation, let alone their high costs; the lack of sequencing information of these markers makes them difficult to be transferred and interpret in different breeding populations; and the deficiency of suitable markers linked to genes of interest restricts MAS application.

Most of the markers applied in MAS were PCR markers with DNA sequence specificity to loci in genomes (Shahidul et al. 2013), such as sequence-tagged microsatellite site (STMS) or simple sequence repeat (SSR), sequence characterized amplified region (SCAR), sequence-tagged site (STS) markers and allele-specific PCR (AS-PCR) markers. Sequence-specific PCR markers tolerate high levels of impurities in crude DNA extracts, making them cost-effective and robust for marker implementation in breeding programs. Usually, several sequence-specific PCR markers can be multiplexed in genotyping, which increases throughput with reduced costs. Theoretically, non-specific markers from electrophoresis gel-based DNA fingerprinting methods can be converted into sequence-specific PCR markers. A successful conversion requires sequence information beyond the marker fragments to design a pair of sequence-specific primers. Unfortunately, it was often difficult to obtain flanking sequences before the sequencing era.

The experience with narrow-leafed lupin as an example illustrates the importance of marker development to follow the needs of plant breeding. In the last 20 years, several molecular marker technologies have been used in lupin, including RAPD (Wolko and Weeden 1994), AFLP (Brien et al. 1999), RFLP (Francki and Mullan 2004; Nelson et al. 2006, 2010) and DArT (Kroc et al. 2014). Disappointingly, none of these markers has been used for MAS owing to the failure in meeting the basic requirements of MAS, despite some being linked to agronomic genes of interest. One exception is MFLP markers, where DNA polymorphisms can be easily converted into cost-effective and sequence-specific PCR markers (Yang et al. 2002, 2004, 2008, 2010; Boersma et al. 2007a, b, 2009; Li et al. 2010b, 2011, 2012a,

b). With them, about 20,000–30,000 F_2 breeding progenies from approximately 200 crosses were screened annually in the Australian National Lupin Breeding Program. Even with MFLP, however, not all attempts were successful in marker conversion (Yang et al. 2001).

The trends of crop sequencing toward breeding

The advances in parallel sequencing technologies have opened an unprecedented opportunity to systematically understand crop genomics. There has been a flood of crop genomes being sequenced in three waves during recent years, taking advantage of improved high-throughput read lengths and single-base accuracy, reduced costs, as well as matching analytical approaches in genomic sequencing. The earliest wave aimed at models for crop genomics employed bacterial artificial chromosome (BAC) clones from corresponding physical maps such as Arabidopsis (The Arabidopsis Genome Initiative 2000), rice (Goff et al. 2002; Yu et al. 2002; International Rice Genome Sequencing Project 2005), Brachypodium (The International Brachypodium Initiative 2010) and Medicago (Young et al. 2011). These small genomes, generated and assembled by Sanger sequencing, lead to high quality and close-to-complete genome sequences. The second wave primarily focused on staple or major economic crops such as sorghum (Paterson et al. 2009), soybean (Schmutz et al. 2010) and maize (Schnable et al. 2009). The crop genomes decoded during this wave were relatively large and full of repetitive sequences (>60 %), resulting in high quality rather than complete genomes despite the combined application of the Sanger method and the whole genome shotgun (WGS) strategy. Maize was an exception due to its high level (~85 %) of transposable elements. The BAC-by-BAC strategy was adopted to sequence a minimum tiling path (Schnable et al. 2009). During this stage, new sequencing or so-called next-generation sequencing (NGS) platforms emerged and matured, making genome sequencing affordable with the collaboration of a few laboratories. The third and current wave has extended to important crops with large genomes [e.g. barley (The International Barley Genome Sequencing Consortium 2012) and wheat (Brenchley et al. 2013)], orphan crops [e.g. pigeonpea (Varshney et al. 2012a), chickpea (Varshney et al. 2013)] and horticultural plant species, incorporating fruits [e.g. apple (Velasco et al. 2010), peach (International Peach Genome Initiative et al. 2013)], vegetables [e.g. potato (Xu et al. 2011), cabbage (Wang et al. 2011)] and flowers [e.g. plum flower (Chen et al. 2012), Carnation (Yagi et al. 2014)]. Tens of thousands of genomes are in the process of being sequenced. In the frame of the WGS strategy, these genomes were sequenced and assembled using either single or multiple NGS platforms. However, an intrinsic feature of the NGS is their short read length. To circumvent this problem, most modern assemblers such as Velvet (Zerbino and Birney 2008) and SOAPdenovo (Li et al. 2010a) employ the de Bruijn graph-based algorithm and hierarchical pair-end information generated by different insert-size libraries. Nevertheless, these assemblers are to some extent stymied by sequencing errors; as a result, most genome assemblies are draft sequences and need continuous effort to pursue a high quality or complete version.

Despite the progress in crop genomics, some genomes remain challenging. From an evolutionary perspective, plants tend to take an alternative strategy such as whole genome duplications to adapt to changing environments. Factors such as a relatively high level of heterozygosity, large and diverse gene families, ubiquitous variations, active transposable elements and repetitive redundancy, and frequent polyploidy contribute to the size and dynamic complexity of crop genomes and exacerbate challenges to decode them. As shown by *K*-mer analysis, a lower peak, reflecting heterozygosity, will appear before the highest one, which is then followed by a peak caused by polyploidy (Fig. 1). The high head indicates sequencing errors whereas a long tail means abundant repetitive sequences in a large genome. Figure 1 provides some empirical criteria to distinguish simple and complex plant genomes. Unfortunately, cultivated wheat (*Triticum aestivum*) has all the difficulties except heterozygosity, making it the Himalayan Mountain in crop genomics. Controversy remains in the wheat community about the best approach to sequence the wheat genome. The international community has largely concentrated on using sorted chromosome arms to construct physical maps by BAC-by-BAC sequencing of the minimal tiling path (Feuillet et al. 2011; The International Wheat Genome Sequencing Consortium (IWGSC) et al. 2014). Considering time-saving and cost-effectiveness of the method, Ling et al. (2013) and Jia et al. (2013) used the WGS approach to assemble the progenitors of the wheat A sub-genome (*Triticum urartu*) and D sub-genome (*Aegilops tauschii*), respectively. At a technical level, these studies represent a transient accomplishment that improves genomic resources and understanding of polyploidy genomes and breeding efforts in relation to Triticeae crops.

Typically, the road of crop sequencing to breeding practice could be divided into the genome scale, population scale and panel scale (Fig. 2). The genome scale primarily focuses on the quality of the crop genome. Most of the crops stay at this stage except some model and staple crops. Considerable effort is made to obtain a crop genome, to improve the genome from draft to high quality, and eventually to obtain a complete genome. Sometimes, a single genome is insufficient for the studies of a given crop species. Take rice as an example, considering the genome divergence between *Japonica* and *Indica*, at least two

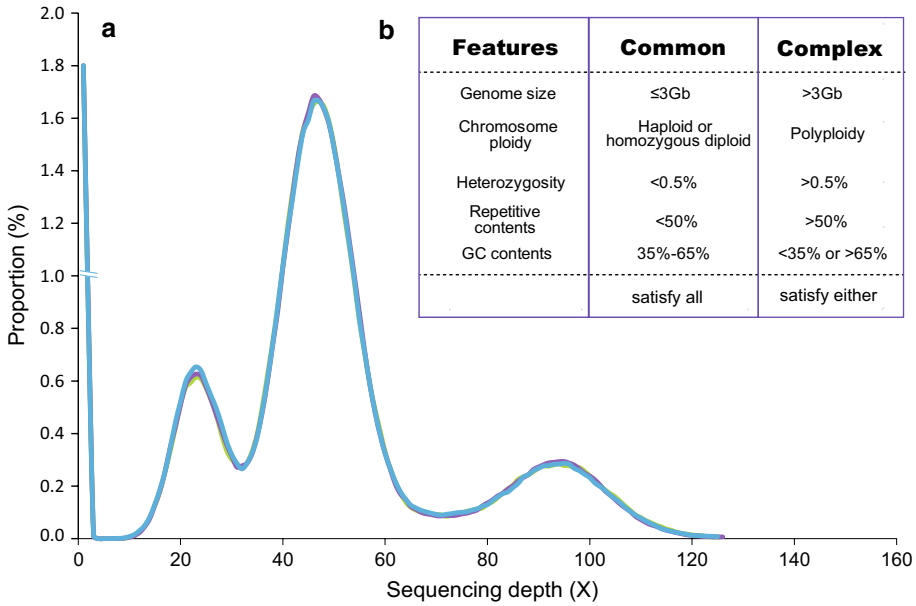

Fig. 1 Genomic features exhibited by *K*-mer analysis and standards fora complex plant genome. The difficulties that affect plant genome assembly can be reflected using *K*-mer analysis: the *lower peaks* before and after the *highest* indicate heterozygosity and polyploidy, respectively. The *long tail* probably represents amount of repetitive sequences and a large genome

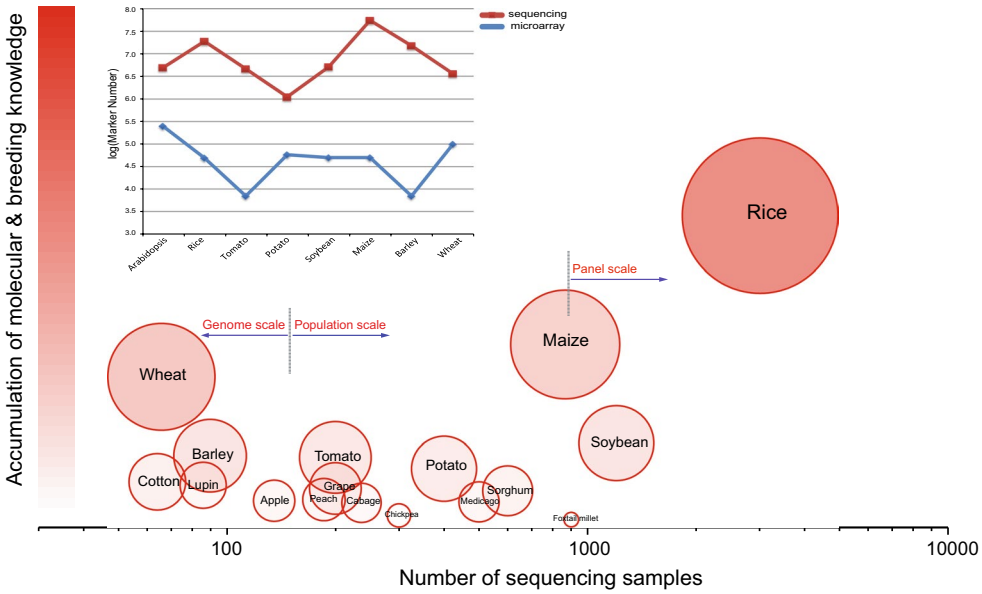

Fig. 2 The trends of crop sequencing toward breeding practice. *X*-axis represents sequencing samples and *Y*-axis represents the accumulation of molecular and breeding knowledge. The *inset* shows the accumulation of genetic variations based on sequencing and molecular markers on microarrays for several crops. The crop sequencing could be divided into the genome scale, population scale and panel scale. At each stage, we need to accumulate different knowledge for a given crop as described

separate reference genomes are needed. A similar scenario occurs in maize. Cultivated crops are differentiated from their wild ancestors by morphological, physiological and genomic changes that have occurred during domestication and subsequent crop improvement. A genome reference for the wild ancestor or close relative of a given crop is often needed to explore its genetic diversity, which has been realized in some crops such as rice (Chen et al. 2013) and soybean (Kim et al. 2010; Qi et al. 2014).

Having the genome for a given crop allows genomic studies to advance to the population scale. At this stage, the most critical step is to mine genetic variations that can be used to infer population stratification and to detect the relationship between genotypes and phenotypes. Generally, a given crop needs both natural and breeding populations to elucidate different features. Parallel sequencing technologies enable us to sequence a specified population at low depth coverage for each individual. A series of statistical methods have been developed to improve the accuracy of variation calling, which is decisive for subsequent analysis (Nielsen et al. 2011, 2012). These variations usually refer to single nucleotide polymorphisms (SNP), small insertion and deletions (InDel), copy number variations and structural variations. Based on the variations in a natural population, genetic diversity can be calculated to infer population structure, haplotypes and linkage disequilibrium (LD).

These form the basis for understanding the demographic process, including the geographic origin, natural selection signals, domestication history, and tracking introgressions between cultivated and wild taxa. High-density genetic variations make it possible to explore causal loci in genome-wide association studies (GWAS). Whereas for breeding populations, abundant markers provided by sequencing improve the resolution of linkage analysis using combination of different genetic mapping strategies. In addition, novel RNAs and splice events are now being identified and quantified by sequencing that target specific aspects of the transcriptome (Mortimer et al. 2014). These studies promote understanding the relationship between genes and traits.

After population-based genomics, crop studies can then move to the panel scale where more attention is paid to precise assessments of crop phenotypes, to further characterize genotypes. The marker-trait association and marker effects need to be evaluated under specific environmental conditions. Thus, modern phenotyping largely depends on how environmental factors contribute to a specific phenotype, which is particularly important for traits such as abiotic stress tolerance (Xu et al. 2012). A panel of populations with large-scale phenotyping and multiple environment typing (e-typing) is crucial at this stage. By detecting transcriptomic regulation and epigenetic modification, the interaction between genes and traits and environments can be better deciphered. By accumulating these large data sets and the corresponding data mining, novel germplasm of higher productivity and better quality will eventually be created through computational modeling and modern breeding technologies such as MARS, advanced backcross quantitative trait loci (QTL) analysis, genomic selection, gene pyramiding, etc.

Sequencing leads to explosion of cost-effective markers

The overall cost in molecular plant breeding consists of marker development and marker application. The rapid advancement of NGS technologies substantially reduces the cost in marker development in plant breeding programs. Serving as high-throughput genotyping platforms, NGS technologies triggered the explosion of cost-effective DNA markers (Fig. 2). More than 55 million SNPs were discovered in maize by genome-wide sequencing of several hundred lines representing wild, landrace and improved genotypes (Jiao et al. 2012; Chia et al. 2012; Hufford et al. 2012). With the same strategy as in maize, about 5.1 million SNPs were identified in 55 soybean accessions (Li et al. 2013). About 3.6 million variations were detected in cucumber germplasm (Qi et al. 2013) and 11.6 million in tomato (Lin et al. 2014). Recently, an international community reported approximately 18.9 million SNPs by re-sequencing a core collection of 3000 rice accessions from 89 countries (Li et al. 2014).

For crops with available genome sequences, known DNA markers can be anchored to corresponding positions in the genomes. New variations can be identified from different cultivars by sequencing or PCR amplification using these sequences as templates. Sequence-specific primers or probes flanking the variations can then be designed to convert original markers into cost-effective markers. For lupin breeding in Australia, marker implementation has recently switched to SNP technologies following the adoption of the Fluidigm SNP-genotyping platform and the LightScanner high resolution melting (HRM) system. The successful and smooth transition was achieved by the NGS-based draft genome sequences, from which corresponding scaffolds for all 16 MFLP markers were identified (Yang et al. 2013b). Using these scaffolds as templates, the MFLP-derived InDel markers were converted into sequence-specific SNP markers, which are now being applied in lupin breeding.

For plant species without genome references, several NGS-based methods of complexity-reduced representation offer a rapid approach to develop cost-effective DNA markers. These methods include reduced-representation libraries (RRLs) (Gore et al. 2009; Hyten et al. 2010), complexity reduction of polymorphic sequences (Mammadov et al. 2010), restriction-site associated DNA sequencing (RAD-seq) (Baird et al. 2008), sequence-based polymorphic marker technology (Sahu et al. 2012), low-coverage multiplexed shotgun genotyping (Andolfatto et al. 2011) and genotyping by sequencing (GBS) (Elshire et al. 2011; Poland et al. 2012b), of which RAD-seq and GBS are the most popular in recent years. It is worth noting that the RAD method was originally developed on a microarray platform (Miller et al. 2007), and then adapted to NGS platforms (Baird et al. 2008). Beginning with a specific restriction enzyme digestion, DNA fragments were ligated to P1 adaptors, pooled, randomly sheared, size selected, then ligated to P2 adaptors for RAD-seq. In this way, almost all of the fragments with enzyme recognition sites could be sequenced (Davey et al. 2011). For GBS, digested fragments were ligated to barcoded and common adaptors before being pooled and amplified. Thus, only short fragments with a barcode adaptor on one side and a common adaptor on the other could be sequenced.

RAD-seq is widely applied to detect SNPs and InDels for QTL mapping (Baird et al. 2008; Bus et al. 2012), in addition to evolutionary analysis (Emerson et al. 2010; Wang et al. 2013). RAD-seq can also be used in de novo assembly to identify missing SNPs by paired-end sequencing (Etter et al. 2011; Willing et al. 2011). However, the cost is still prohibitive for using RAD-seq in crops with

large genomes, whereas GBS is a suitable alternative (Chutimanitsakun et al. 2011; Poland et al. 2012b). Moreover, the cost of GBS can be further reduced when processing with a methylation-sensitive restriction enzyme, *Ape*KI for instance. Considering simple and quick library construction, lower coverage requirement, and more importantly, lower cost, GBS is particularly wellsuited for GWAS and genomic selection (GS) in large sample size genotyping (Poland et al. 2012a; Uitdewilligen et al. 2013; Ward et al. 2013; Morris et al. 2013). In essence, NGS-based complexity-reduced representation sequencing shares the same principles as traditional DNA fingerprinting methods in sampling genome-wide DNA markers (Yang et al. 2012). By this approach, 416,856 markers were obtained in wheat (Saintenac et al. 2013), then 34,000 in barley (Poland et al. 2012b), 25,047 in soybean (Hyten et al. 2010), and over 10,000 markers in eggplant (Barchi et al. 2011).

The most distinct advantage of sequencing generated markers over traditional fingerprinting markers is that all markers have DNA sequence information available. The sequences can be readily applied for primer design or probe design. This enables sequencing-derived markers to be easily converted to cost-effective formats, which are amenable to large samples in plant breeding. Researchers have applied RAD-seq in the absence of lupin genome to develop SNP markers for anthracnose and phomopsis stem blight (PSB) disease resistance (Yang et al. 2012, 2013a). More than 30 SNP markers linked to disease resistance genes were discovered in each case. Although the sequencing reads are short, many of the SNPs were located in the middle, leaving enough flanking sequences for primer design. The SNPs were successfully converted into sequence-specific PCR markers, which have been applied in lupin breeding using the HRM platform.

The cost of marker application in plant breeding is largely determined by genotyping platforms. With the fast accumulation of sequence-defined markers, a number of high-throughput genotyping platforms emerged for marker development and implementation in plant breeding (Table 1). High-density oligonucleotide arrays designed by Affymetrix or Illumina are now commercially available for several crops, allowing parallel genotyping of a few individuals for thousands of markers or, alternatively, thousands of individuals for a few markers (Gupta et al. 2008). These array-based markers are widely used in crop genomic investigation such as population dynamics, origin and evolution, GWAS and QTL interval mapping (Atwell et al. 2010; Molina et al. 2011; Dai et al. 2012). Thousands of SNP markers can be assayed in microarrays in some crops, which potentially can be applied to large-scale plant breeding practices (Fig. 3). For example, sequence-specific markers were applied to genotype a segregating population in soybean using the Illumina GoldenGate genotyping platform (Hyten et al. 2010). Marker-assisted breeding has been routinely used by some large private companies benefit from their long-term basic research and investment on genotyping infrastructures. Several versatile high-throughput genotyping platforms have been developed in recent years (Table 1) including KASPar, Fluidignnanofluidic arrays, HRM, SNaPshot multiplex SNP genotyping and TaqMan SNP genotyping. They provide more choices to meet the cost-effective requirement, which may help small-sized companies, public sector and developing countries to take up molecular plant breeding. A comprehensive review on marker genotyping platforms was published recently (Xu et al. 2013).

Sequencing facilitates development of functional or diagnostic markers

Plant breeding programs usually involve germplasm from diverse genetic backgrounds. Implementation of molecular markers in different breeding populations requires high consistency between marker genotypes and trait phenotypes in a wide range of germplasm, and the markers should be able to screen progeny resulting from complex crosses involving multiple parents. Unfortunately, most molecular markers are not "perfect", for example, distinct from target genes. Owing to genetic recombination, cultivars showing desirable marker alleles may not necessarily possess targeted genes, and vice versa. To deal with this prevalent problem, molecular biologists need to conduct marker validation to figure out suitable markers for a certain cross, which not only increases the overall cost, but also slows down MAS application.

The best solution for this plight is to develop functional markers or non-genic diagnostic markers that can be applied in a wide range of breeding germplasm without marker validation. Functional markers are part of the genes of interest, whose genotypes will always match up with trait phenotypes on all individuals in a breeding program. Hence, marker validation is no longer necessary in marker-assisted plant breeding (Gebhardt 2013). Two famous markers implemented in Australia were "diagnostic" for cereal cyst nematode (CCN) resistance in all breeding crosses (Ogbonnaya et al. 2001).

Gene sequences obtained from genome sequencing greatly facilitate the development of genic markers for plant breeding. In soybean, functional markers were developed for the fragrance gene (Juwattanasomran et al. 2012), for the phosphorus stress genes (Zhang et al. 2014), and for seed oil storage genes (Goettel et al. 2014). Genic markers were developed in rice for the bacterial leaf blight disease resistance genes *xa5* (Iyer-Pascuzzi and McCouch 2007), *xa13* (Antony et al. 2010), *xa21* (Park and Ronald

Table 1 High throughput marker genotyping platforms based on the suitability of application in plant breeding

Genotyping platforms	Simple description	Variations based and origin	General cost	Cost per datapoint	Suitability for MAS
Illumina Goldengate SNP assay	PCR amplicons of SNP nucleotides hybridized on BeadChip detected by fluorescence	PCR amplification by oligo primers with allelic nucleotides for each SNP	Medium	Low	Up to 1536-plex to genotype 147,456 data points on one plate
Affymetrix SNP arrays	DNA probes with specific SNP nucleotides binded on array slide. Test samples are fragmented and hybridized to probes; and attached to biotin	SNP markers detected by DNA probe	Medium	Low	10,000 or more SNPs can be screened on one slide
KASPar genotyping	Competitive allele-specific PCR enables bi-allelic scoring of SNP and indel markers	PCR amplification with oligo primers specific to SNP/indel sites.	Medium	Low	Assays can be done on 96-, 384- and 1526-well plates
Fluidign SNP genotyping	PCR amplification with primers allelic to SNP nucleotides, and resolved on nanofluidic array slide	Detection of SNP by allelic primers with fluorescence signals	Medium	Low	A versatile system where markers and test samples can be changed on every array slide
TaqMan SNP genotyping	Probes with minor groove binder (MGB) hybridized with test DNA for form match and mismatch at SNP sites	Detection by 5′ nuclease cleavage of allele-specific 5′ dye label	Low	Low	Restricted to species with available TaqMan assays
High resolution melting (HRM)	PCR amplification by a pair of sequence-specific primers flanking a SNP or InDel marker	Polymorphic alleles resolved by high resolution melting (HRM)	Low	Low	A cost-effective method for marker implementation without the need for complex equipment
SNaPshot Multiplex SNP genotyping	Primer extension-based method. Products are resolved on capillary array	PCR amplification with oligo primers specific to SNP/indel sites	Low	Low	A versatile system using unlabeled user-defined primers. Up to 10-plex capability
MassARRAY	Locus-specific PCR followed by allele-specific single-base primer extension	Mass resolution by MALDI-TOF of primer extension products	Low	Low	A versatile platform which can resolve multiplexed reactions, which can acquire and interpret data very quickly

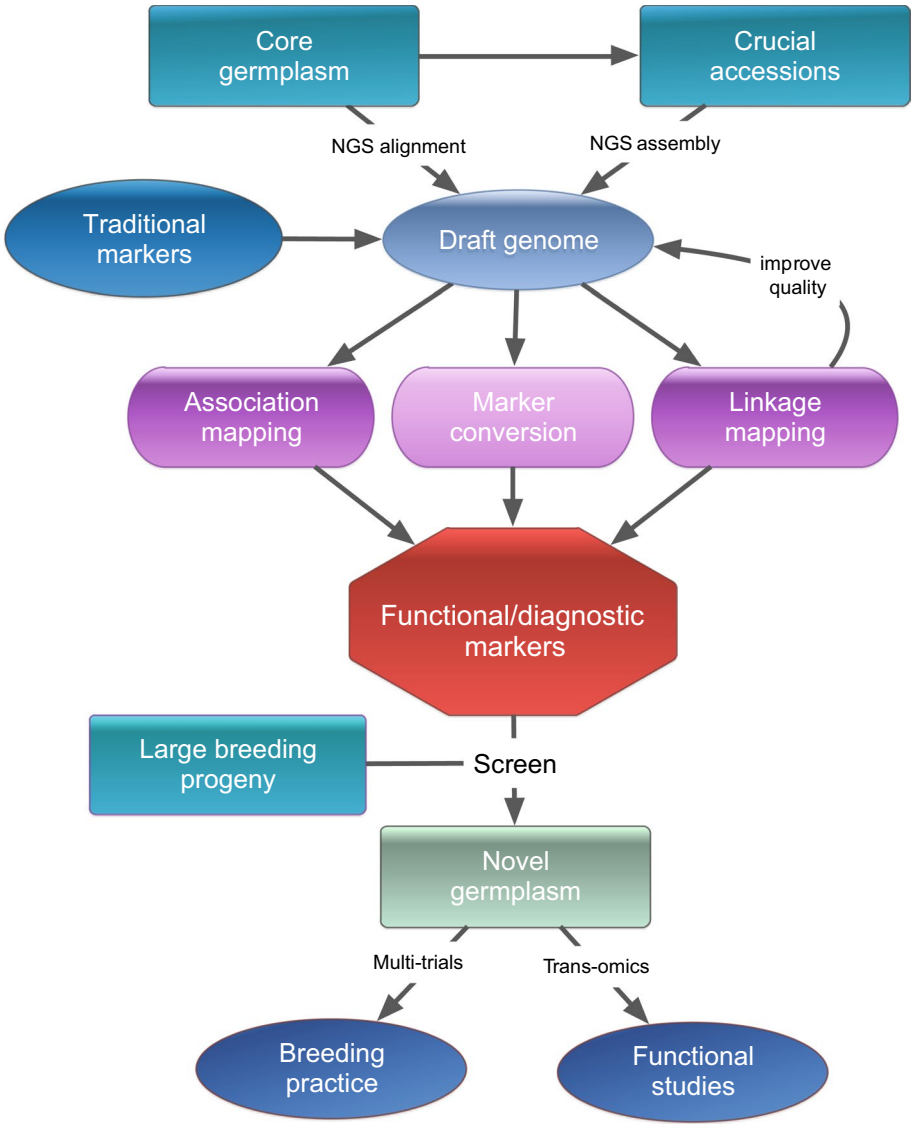

Fig. 3 A schematic overview of sequencing strategies for marker-assisted plant breeding. For a given crop, at least a core germplasm is needed, from which crucial accessions can be picked out to construct a draft genome by sequencing. Traditional markers can be converted into sequence-specific, diagnostic markers using the draft genome. Then natural and segregating populations can be sequenced to develop diagnostic markers through linkage and association mapping. The information obtained from the linkage mapping can also be used to improve the quality of the draft genome for more precise analyses. The diagnostic markers can then be used to screen large breeding progeny, based on which novel germplasm will be constructed both for breeding practice through multi-trials and for functional studies through *trans*-omics strategies

2012), and for plant hopper resistance gene *Bph14* (Zhou et al. 2013). Functional markers were also developed in wheat for the gene *Pm3* conferring resistance against powdery mildew disease (Brunner et al. 2010), and for the polyphenol oxidase gene (PPO) (He et al. 2007). A thorough understanding of the gene structure and function can greatly expedite the development of functional markers. In barley for example, the beta-Amylase (α-1,4-glucan maltohydrolase) is a key enzyme in the hydrolysis of starch in germinating barley grains, which determines the diastatic power during malting and brewing. A non-synonymous substitution of Arg115Cys in the gene determines the binding ability with the enzyme inhibitor. Another substitution of Val233Ala influences the refolding after heating giving a benefit during malt kilning. The third substitution of Leu347Ser confers maintenance of activity at higher mashing temperature. Each of these amino acid substitutions is caused by a SNP variation, and three genic markers were separately designed for these variations. Now, barley breeders can select specific malting and brewing quality traits with different enzyme activity, thermostability and inhibitor binding ability without the time consuming chemical analysis for malting and brewing quality (Evans et al. 2010).

Although genic markers are clearly the most ideal ones for molecular plant breeding, it still remains a difficult task to develop them. Even with available genome sequences, it requires a considerable amount of research work to pinpoint the corresponding genes for agronomic traits of interest. Fortunately, non-genic diagnostic markers and semi-diagnostic markers can be developed without sequence knowledge of causal genes. The SSR marker *Xgwm382* was identified in wheat through bulk segregant analysis, which was diagnostic for yellow rust disease resistance (Akfirat

et al. 2013). Sequence-tagged microsatellite markers were identified to be diagnostic for stem rust disease resistance gene *Sr2* in wheat (Hayden et al. 2004; Ellis et al. 2014). In lupin, the first marker AntjM1, tagging the anthracnose disease resistance (*R*−) gene *Lanr1*, was developed in 2002 (Yang et al. 2004). Unfortunately, the initial excitement quickly subsided when cultivars highly susceptible to the disease were found to have the *R*-marker allele. With a careful marker validation, AntjM1 was used to screen F_2 breeding populations from only 32 and 61 crosses of approximately 400 crosses in 2003 and 2004, respectively. The resistance to anthracnose disease was so important to have a veto power on the release of new lupin cultivars (Yang et al. 2001) that researchers kept on developing more markers for *Lanr1* gene over the following 2 years until the identification of the diagnostic marker AntjM2 (You et al. 2005). Based on this observation, a new protocol was formulated to identify diagnostic or semi-diagnostic markers: (1) multiple candidate makers are generated (2) and then validated on a set of key cultivars; (3) markers with the least false positives are selected and converted into sequence-specific PCR markers for routine breeding (Yang et al. 2008). This protocol was applied to later marker implementation, which to some degree broadens the application of developed markers in lupin breeding (Li et al. 2010b, 2011, 2012a, b).

NGS-based reduced complexity sequencing technologies provide powerful tools to develop non-genic but diagnostic or semi-diagnostic markers. In wheat, diagnostic SNP markers were developed using RNA-seq for the yellow rust disease resistance gene *Yr15* (Ramirez-Gonzalez et al. 2014). When RAD-seq was applied in marker development for PSB disease resistance in lupin, 7241 SNP markers were obtained across 20 plants. Of which, 33 were identified as candidate markers linked to the *R* gene *PhtjR*. Using NGS platforms, the rapid discovery of a large number of molecular markers associated with a target gene provides an excellent source to select diagnostic markers. Linkage analysis confirmed that two SNP markers co-segregated with the *R* gene on a population of 187 recombinant inbred lines (Yang et al. 2013a). Validation tests found that these two SNP markers, which were 1.1 cM away from the *R* gene, had marker genotypes consistent with disease phenotypes on all commercial cultivars and breeding lines, which is "diagnostic" for lupin breeding. In comparison, two markers at 0 cM to the *R* gene had a large number of false positives on commercial cultivars (Yang et al. 2013a). These results further indicated that a diagnostic marker may not necessarily be the one most closely adjacent to the target gene of interest (Sharp et al. 2001; Yang et al. 2008).

Genome sequencing and re-sequencing are a valuable resource for searching diagnostic markers. In lupin, the scaffold2922 (33,979 bp in length) was linked to anthracnose disease resistance by a combination of genome sequencing and genetic mapping from lupin cultivar Tanjil (Yang et al. 2013b). By re-sequencing ten cultivars, researchers developed several InDel and SNP markers linked to anthracnose resistance genes by sequence alignment on scaffold2922. These markers showed significant genotypic variations on commercial cultivars, even though they originated from the same scaffold (Table 2). One InDel marker showing diagnostic genotypes for anthracnose disease phenotypes; the other four markers had false positives on 8–15 out from 27 commercial cultivars (Table 2). In rice, by genome-wide sequencing a segregating population, each recombination bin was identified to contain an average of 525 SNP markers (Huang et al. 2009). Markers within each recombination bin are closely linked, which therefore provide a wide scope for the discovery of diagnostic markers. Some of the markers within a bin could be from the causative gene that determines the phenotypic variations. Such markers will be "perfect" that are not only diagnostic to a wide range of breeding populations, but also provide 100 % accuracy in predicting the phenotype of interest in MAS (Varshney et al. 2005).

Sequencing boosts molecular breeding of neglected crops

At present, highly complete genome sequences are only available for a few plant species, such as Arabidopsis and rice, which serve as model species for plant research. A few major staple crops have attracted large amounts of research investment worldwide. However, plant breeding covers hundreds of plant species, but most of them are "neglected" on molecular research and with little genomic information available. Until recently, examples of marker application in plant breeding have mostly concentrated on a few major crop species. For neglected crops, little available molecular resources make it difficult for marker development, thus few successful marker applications for MAS have been observed on these plant species. Reduced cost of NGS now makes it feasible to sequence the genomes of many neglected crops. Low-cost and medium-depth sequencing can provide valuable genome resources for the breeding of these crops. For example, a draft genome sequence of lupin (~1.1 Gb) was achieved by sequencing 27-fold depth for less than US $20,000. The scaffold N50 length reached 12.5 kb, which is sufficient for sequence-specific primer and probe design in developing allele-specific markers. Researchers have located the previous InDel markers linked to various key genes to the respective scaffolds (Yang et al. 2013b), whose length ranged from 3 to 64 kb. With those scaffolds as templates, sequence-specific and cost-effective SNP markers have been successfully developed for each of

Table 2 Disease resistance phenotypes for the anthracnose disease resistance gene *Lanr1* and the five linked molecular markers derived from Scaffold number 2992[1] of the reference lupin genome assembly showing large marker genotype variations on 27 cultivars of *Lupinus angustifolius* illustrating the potential of identification of diagnostic markers for MAS

	Cultivars	Disease resistance genotype	SNP3299[2] [T/C][3]	SNP10116 [T/C]	SNP22156 [A/G]	Indel 23269 [TTAC][4]	SNP31649 [T/C]
1	Uniwhite	S	C	C	G	-	T
2	Uniharvest	S	C	C	G	-	T
3	Unicrop	S	C	C	G	-	T
4	Marri	S	T[5]	T	A	-	T
5	Illyarrie	S	T	T	A	-	T
6	Yandee	S	T	T	A	-	T
7	Chittick	S	C	C	G	-	T
8	Danja	S	C	C	G	-	T
9	Geebung	S	C	C	G	-	T
10	Gungurru	S	T	T	A	-	C
11	Yorrel	S	T	T	A	-	C
12	Warrah	S	T	T	A	-	C
13	Merrit	S	T	T	A	-	C
14	Myallie	S	C	C	G	-	T
15	Kalya	S	T	T	A	-	T
16	Wonga	R	T	T	A	+	C
17	Belara	S	T	T	A	-	C
18	Tallerack	S	C	C	G	-	T
19	Tanjil	R	T	T	A	+	C
20	Moonah	S	T	T	A	-	T
21	Quilinock	S	T	T	A	-	C
22	Jindalee	S	T	T	A	-	C
23	Mandelup	S	T	T	A	-	T
24	Coromup	S	T	T	A	-	T
25	Jenabillup	S	T	T	A	-	C
26	Gunyidi	R	T	T	A	+	C
27	Barlock	R	T	T	A	+	C
Number of "false positives"			**15**	**15**	**15**	0[6]	8

Plant disease resistance genotypes are presented in blue

Cultivars possessing the *R* gene are highlighted in green

[1] Reference genome sequence assembly derived from cultivar Tanjil was published by Yang et al. (2013b)

[2] Marker positive on scaffold 2992

[3] SNP nucleotide

[4] InDel nucleotides

[5] False positives (marker genotype does not match disease resistance genotypes) are highlighted in red color

[6] Markers showing no false positive are diagnostic markers

those agronomic genes of interest. The immense number of markers developed by whole genome re-sequencing provides ample choice for molecular lupin breeding.

For plant species without a reference genome sequence, NGS-based reduced genome representation sequencing approaches are available to generate a large number of high quality and cost-effective markers. Before publication of the pepper genome, transcriptome analysis by RNA-seq of two cultivars yielded 11,849 reliable SNP and 853 SSR markers on pepper (Nicolai et al. 2012). Sequencing of reduced-representation libraries identified 55,465 SNP markers in eight flax (*Linum usitatissimum*) genotypes (Kumar et al. 2012). The markers generated from these alternative strategies also provide an easy way to construct high-density genetic maps on plants without any preceding molecular knowledge. NGS-based genetic mapping is more rapid and cost-effective than traditional genetic mapping. For example, several research projects have invested millions of dollars over the last 20 years in genetic mapping on the same mapping population of narrow-leafed lupin in Australia (Brien et al. 1999; Nelson et al. 2006, 2010; Kroc et al. 2014). The combined reference maps from these projects contained <1500 markers; and most of these markers were anonymous without sequence information. Thus, they cannot be transferred and interpreted on different breeding germplasm. In contrast, a dense genetic map with 8,246 sequence-defined markers was constructed within a month using RAD-seq in lupin (Yang et al. 2013b). This NGS-based map has several major advantages over previously reported maps. First, the marker number is several times higher than those reported, which provides higher resolution landmarks for the lupin genome. It will provide lupin geneticists and breeders a broader suite of options to choose markers for a wide range of research purposes. Second, all sequence-defined markers can be easily and unambiguously transferred and interpreted in

Table 3 The impact of MAS on the Australian lupin breeding program as shown by efforts for screening resistance to anthracnose disease in response to implementation of DNA markers in the past 12 years

	2002 (before MAS)	2013 (after 12 years of MAS)
Glasshouse disease screening	All year around	Stopped
Size of field disease screening trials	5 bays[a]	1 bays[a]
Breeding lines showing disease resistance in field trials	50 %	98 %
Working time spent by (1) plant pathologist; and (2) technical officer	(1) 20 % (2) 80 %	(1) 5 % (2) 20 %

[a] Each bay is a 90 m × 18 m field plot

any lupin germplasm, and are useful for comparative genomic studies. Third, 7563 of the markers are SNP markers, which are compatible with high-throughput SNP-genotyping platforms for molecular breeding and genetics studies.

Recent work demonstrated that QTL maps could not only be employed for mapping quantitative traits, but could also improve the genome quality of the parents of a segregating population (Gao et al. 2013). A core segregating system here refers to a mapping population together with its parental lines, which could also be expanded to a larger population of breeding. By taking full advantage of the sequencing data of a segregating population, the genome sequence of parents improved by anchoring scaffolds to chromosomes, filling up gaps and correcting single-base errors. It provided an alternative approach to improve the genome sequence of crops. Furthermore, the improvement will be accumulative as more mapping populations derived from a shared parent are sequenced. It is concluded that, for each crop, at least one segregating population and its parents should be sequenced to initiate a modern breeding platform (Fig. 3). Core and larger mapping populations should be developed and sequenced step-by-step depending on funding availability. In conclusion, NGS-based rapid and high-density genetic mapping provides a cost-efficient approach for neglected crop species as well as major crops on marker discovery for MAS and GS in breeding practice (Pérez-de-Castro et al. 2012; Nakaya and Isobe 2012).

Sequencing ties up breeders and molecular biologists

Human society has experienced a surprising population explosion over the past 100 years. Amazingly, plant breeding has been successful in keeping mankind well fed. To some extent, the major health concern in contemporary society is now over-nutrition related diseases, such as diabetes, gout, hyperlipemia and obesity. However, slow progress in producing molecular markers for practical plant breeding, and competition from molecular biologists in research funding, can hamper the enthusiasm of breeders to adopt marker technologies. In Australia, for example, multi-million dollar investments have been made in molecular research on several legume crops in the last 20 years, including field pea, chickpea, lentil and faba bean. To date, no marker has been applied for MAS on F_2 populations in these breeding programs. As such, it is not easy to convince plant breeders to believe in molecular markers. When one author of this review (Huaan, Y.), a plant pathologist at the time, developed the first two InDel markers that linked to PSB disease resistance (Yang et al. 2002), the breeders required thorough validation of the effectiveness of markers before adopting them. Since then, 20 more InDel markers have been developed and linked to various agronomic genes of interest in lupin (Shahidul et al. 2013; Yang et al. 2013b). Nowadays, MAS has become an integral part of the breeding program yielding significant economic and genetic benefit. Take anthracnose resistance for instance; before MAS was applied in lupin breeding, glasshouse testing was a year-around task plus field trials in five disease nurseries (1620 m^2 each). DNA markers linked to the *R* gene *Lanr1* conferring resistance to anthracnose disease were then applied to screen all F_2 plants (Yang et al. 2004; You et al. 2005), which made the glasshouse anthracnose disease test redundant. The field disease nursery is now reduced to one bay annually, which serves as verification for the advanced breeding lines (Table 3). The application of molecular markers not only results in cost saving from glasshouse trials, field trials and manpower, but more importantly, it greatly increases the efficiency of genetic improvement. Without molecular markers, about 50–60 % of advanced breeding lines were thrown away due to anthracnose susceptibility, and the breeder could only use the remaining lines to find new cultivars. After MAS, 98 % of breeding lines are resistant to the disease (Table 3), which means that with the same effort the lupin breeding team has twice as many advanced breeding lines, from which new cultivars can be selected.

Once molecular markers are cost-effective and applicable in plant breeding, it is often straightforward to engage plant breeders to apply them in breeding programs. For example, breeding for resistance to PSB is a key objective in lupin breeding (Yang et al. 2002). Several conventional screening methods for PSB resistance are labor-intensive

and time-consuming (Cowling et al. 1987; Williamson et al. 1991; Shankar et al. 1996). With NGS technology, SNP markers were developed for two major *R* genes conferring PSB resistance. One linked to the *PhtjR* gene was developed by the RAD-seq method (Yang et al. 2013a), and the other linked to the *Phr1* gene was developed by mining the draft genome (Yang et al. 2013b). These SNP markers are now being implemented in large-scale MAS to screen for PSB resistance on F_2 populations in lupin breeding.

Another example of marker-assisted breeding is the selection for the soft-seed gene *mollis*. Seeds of wild lupin are "hard-seeded", which means that the seed coat is impermeable to water to prevent uniform germination as a survival strategy in the wild. During lupin domestication, a soft-seediness gene (*mollis*) was selected, which confers seed coat permeability to water (Boersma et al. 2007a), ensuring a satisfactory germination rate of the crop. Screening for the *mollis* gene by the conventional method of soaking seeds in water to observe water absorption is problematic, since a portion of hard-seeded lupin seeds can naturally absorb water, which can be confused with truly soft-seeded plants. More importantly, the permeability of the seed coat is maternally determined. That is, if a soft-seeded plant is used as a female parent crossed with a hard-seeded wild type, the F_1 (*Mollis*/*mollis*) seed is soft seeded, and the resultant F_2 seeds are hard seeded including homozygous soft-seediness individuals. Therefore, the *mollis* gene is regarded as the most difficult domestication gene to be identified and selected by conventional phenotyping methods. Using the draft genome sequence and previously identified InDel markers (Boersma et al. 2007a; Li et al. 2012a), two co-dominant SNP markers linked to soft-seed gene *mollis* were developed. Marker-assisted selection of the *mollis* gene has removed the difficulties encountered by conventional selection methods.

With breeders and molecular biologists working together, molecular markers can be applied to deal with some tough challenges in plant breeding, such as broadening the genetic-based breeding pools. Genetic gains in plant breeding are proportional to the phenotypic variations present in the original source population, and the phenotypic variation is positively associated with genetic diversity in the breeding pools (Moose and Mumm 2008). The genetic reduction and narrowed gene pools have become major issues in many plant breeding programs, which is particularly evident in lupin. From the release of the first cultivar Unicrop in 1973 to the latest commercial cultivar Barlock in 2013, lupin yield has increased by 85 %, from 0.78 to 1.85 tonnes per hectare. However, all 27 commercially released cultivars in Australia can be traced back to no more than 12 wild lupin accessions by recorded pedigrees. The narrow gene pool makes it difficult for further yield improvement. The major reason for the narrowed genetic pool is that the five wild genes must be replaced with domestication alleles when introduced into modern lupin cultivars, including the high alkaloid gene *Iucundis*, the hard-seeded gene *Mollis*, the late-flowering gene *ku*, and the two pod-shattering genes *Tardus* and *Le* (Clements et al. 2005). The percentage of F_2 individuals containing homozygous domesticated alleles of all five genes is very low in breeding populations derived from W × D crosses. Thus, it is difficult to select such F_2 plants with domestication alleles by conventional methods. With draft genome sequences, SNP markers for each of the five genes were developed and are now applied for MAS in the F_2 populations (Yang et al. 2013b). A major advantage of MAS over conventional selection is the ability to differentiate homozygous individual from heterozygous plants, making it possible to select and retain F_2 plants containing domestication alleles for the five genes simultaneously. If the number of F_2 plants with homozygous genotypes is low, F_2 plants with heterozygous genotype are retained and the selection of homozygous genotypes is delayed until the following generation. In this way, MAS promoted to integrate wild germplasm into the domesticated gene pool of lupin.

Insufficient communication between molecular biologists and plant breeders is another major factor limiting MAS application. For example, we heard molecular biologists wondering "how can a marker happen to have false positives" and "the marker test should be finished in 1 year on all interested germplasm" and "why use the same marker year after year". Weak knowledge on plant genetics and breeding practice among molecular biologists could hinder their ability to develop applicable markers for plant breeding. To date, all molecular markers applied in lupin breeding were developed by the Australian lupin breeding team. Nevertheless, the application of molecular markers to accelerate genetic improvement is a joint mission of plant breeders and molecular biologists. The rapid NGS-based marker development protocol (Yang et al. 2012) and lupin draft genome (Yang et al. 2013b) further convinced lupin breeders to willingly invest in modern genotyping platforms for characterizing breeding germplasm and potential parental lines, which help breeders design and optimize different crosses. The close cooperation between breeders and molecular geneticists succeeds in lupin breeding.

Concluding remarks

Molecular markers have been gradually applied to assist plant breeding in a few major agricultural crops in the last 30 years. A small number of commercial varieties obtained from MAS breeding programs have been released in rice, soybean, maize, barley, wheat and potato (Brumlop and Finckh 2011). Generally speaking, the potential of

molecular markers for crop genetic improvement is yet to be fully realized. To date, most fruitful examples of marker-assisted plant breeding are associated with major genes (Gupta et al. 2008; Brumlop and Finckh 2011). However, many agronomic traits of breeding interest, such as yield and protein content, are controlled by multiple genes with small effect in multiple QTLs. Although thousands of studies have been published on mapping QTLs in crops, very few have been translated into crop genetic improvement. Of the few successful cases, the application of MAS based on QTL mapping was restricted to one or a few breeding populations resulting from bi-parental cross. The scale up of MAS in a breeding program requires a new theory for integrating QTL information across numerous populations, since the marker validation for monogenic traits is difficult to apply to markers linked to quantitative traits. The major difficulties hindering the wide application of molecular markers in plant breeding could be summarized as: prolonged methods for identification of marker-trait association; limited numbers of established molecular markers linked to traits of breeder interest; the lack of functional/diagnostic markers resulting in marker validation; the low reliability and accuracy of QTL mapping across different populations; the high unit cost associated with MAS uptake; the lack of suitable infrastructure and platforms for high-throughput genotyping in public sector and developing countries; and the knowledge gaps between molecular biologists and plant breeders (Collard and Mackill 2008; Ribaut et al. 2010).

Traditional DNA fingerprinting-based marker development methods contributed to some of the above technical difficulties due to their low efficiency in generating DNA markers, and the absence of DNA sequence information. NGS-based strategies have superseded previous gel-based methods in marker development, which eventually would facilitate identification, selection and application of molecular markers for plant breeding. An advantage of sequencing technologies is the extraordinarily high efficiency in generating large numbers of sequence-defined DNA markers without the need for prior genome knowledge. Sequencing-derived markers can be easily and affordably applied to any plant. Sequencing technologies are capable of having a major impact on molecular plant breeding even on neglected plant species, for example lupin.

In recent years, GS has attracted great attention in plant molecular research. It simultaneously estimates all loci or all markers in a genome to calculate the genomic estimated breeding value (GEBV) (Desta and Ortiz 2014). Selection of desirable breeding progeny is based on the joint effect of genome-wide dense markers for the total genetic gains. It is contrary to traditional MAS where selection is based on a set of defined markers with significant genetic effects. GS also provides a new avenue for marker-assisted selection for QTL controlled complex traits (Varshney et al. 2009; Connelly and Akey 2012). Various statistical models for prediction of genetic effects in GS are being proposed (Jannink et al. 2010). In simulation studies, GS produced 43 % more genetic gain than MAS for polygenic traits (Rex and Yu 2007), and has the potential to reduce a 7-year MAS breeding program into a 3-year program (Heffner et al. 2009). Despite the fact that GS has been applied in animal breeding for some time (Schaeffer 2006), it is now still in its infancy in crop breeding (Lorenz et al. 2011; Desta and Ortiz 2014). Example of practical application of GS in plant breeding programs is yet to be seen. The escalating capacity of sequencing technologies in developing massive molecular markers, coupled with parallel genotyping platforms, holds great promise to advance molecular plant breeding in the near future.

Author contribution statement S. Zhao and H. Yang conceived and drafted the main text; C. Li, H. Lam, J. Clements, and G. Yan commented and revised the manuscript.

Acknowledgments We greatly appreciate the two anonymous reviewers for their invaluable comments and suggestions. The research is supported by the Western Australian Government through the Lupin Molecular Marker Strategy Project to H. Y.; and the Hong Kong RGC Collaborative Research Fund (CUHK3/CRF/11G) and the Hong Kong RGC General Research Fund (468610) to H. M. L.

Conflict of interest The authors declare that they have no conflict of interest.

References

Akfirat FS, Ertugrul F, Hasancebi S et al (2013) Chromosomal location of genomic SSR markers associated with yellow rust resistance in Turkish bread wheat (*Triticum aestivum* L.). J Genet 92:233–240

Andolfatto P, Davison D, Erezyilmaz D et al (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res 21:610–617

Antony G, Zhou J, Huang S et al (2010) Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os-11N3*. Plant Cell 22:3864–3876

Atwell S, Huang YS, Vilhjálmsson BJ et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Baird NA, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Barchi L, Lanteri S, Portis E et al (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics 12:304

Boersma JG, Buirchell BJ, Sivasithamparam K, Yang H (2007a) Development of a PCR marker tightly linked to *mollis*, the gene that controls seed dormancy in *Lupinus angustifolius* L. Plant Breed 126:612–616

Boersma JG, Buirchell BJ, Sivasithamparam K, Yang H (2007b) Development of two sequence-specific PCR markers linked to

the le gene that reduces pod shattering in narrow-leafed Lupin (*Lupinus angustifolius* L.). Genet Mol Biol 30:623–629

Boersma J, Nelson M, Sivasithamparam K, Yang H (2009) Development of sequence-specific PCR markers linked to the Tardus gene that reduces pod shattering in narrow-leafed lupin (*Lupinus angustifolius* L.). Mol Breed 23:259–267

Brenchley R, Spannagl M, Pfeifer M et al (2013) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705–710

Brien SJ, A CW, Potter RH et al (1999) A molecular marker for early maturity (Ku) and marker-assisted breeding of *Lupinus angustifolius*. In: Proceedings of the 9th international lupin conference, Klink/Muritz, Germany, pp 115–117

Brumlop S, Finckh MR (2011) Applications and potentials of marker assisted selection (MAS) in plant breeding. BundesamtfürNaturschutz (BfN), Bonn, Germany

Brunner S, Hurni S, Streckeisen P et al (2010) Intragenic allele pyramiding combines different specificities of wheat Pm3 resistance alleles. Plant J 64:433–445

Bus A, Hecht J, Huettel B et al (2012) High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. BMC Genomics 13:281

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221

Chen W, Sun L, Zhao F et al (2012) The genome of *Prunus mume*. Nat Commun 3:1318

Chen J, Huang Q, Gao D et al (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying Oryza genome evolution. Nat Commun 4:1595–1599

Cheng SH, Cao LY, Yang SH, Zhai HQ (2004) Forty years' development of hybrid rice: China's experience. Rice Sci 11:225–230

Chia J-M, Song C, Bradbury PJ et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803–807

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A et al (2011) Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. BMC Genomics 12:4

Clements JC, Dracup M, Buirchell BJ, Smith CG (2005) Variation for seed coat and pod wall percentage and other traits in a germplasm collection and historical cultivars of lupins. Aust J Agric Res 56:75–83

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B: Biol Sci 363:557–572

Connelly CF, Akey JM (2012) On the prospects of whole-genome association mapping in *Saccharomyces cerevisiae*. Genetics 191:1345–1353

Cowling WA, Hamblin J, Wood PM, Gladstones JS (1987) Resistance to Phomopsis stem blight in *Lupinus angustifolius* L. Crop Sci 27:648–652

Dai F, Nevo E, Wu D et al (2012) Tibet is one of the centers of domestication of cultivated barley. Proc Natl Acad Sci USA 109:16969–16973

Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601

Ellis JG, Lagudah ES, Spielmeyer W, Dodds PN (2014) The past, present and future of breeding rust resistant wheat. Front Plant Sci 5:641

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Emerson KJ, Merz CR, Catchen JM et al (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proc Natl Acad Sci USA 107:16196–16200

Etter PD, Preston JL, Bassham S et al (2011) Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. PLoS One 6:e18561

Evans DE, Li C, Eglinton JK (2010) The properties and genetics of barley malt starch degrading enzymes. In: Zhang G, Li C (eds) Advanced topics in science and technology in China. Springer, Berlin, pp 143–189

Feuillet C, Leach JE, Rogers J et al (2011) Crop genome sequencing: lessons and rationales. Trends Plant Sci 16:77–88

Francki MG, Mullan DJ (2004) Application of comparative genomics to narrow-leafed lupin (*Lupinus angustifolius* L.) using sequence information from soybean and Arabidopsis. Génome 47:623–632

Gao Z-Y, Zhao S-C, He W-M et al (2013) Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. Proc Natl Acad Sci USA 110:14492–14497

Gebhardt C (2013) Bridging the gap between genome analysis and precision breeding in potato. Trends Genet 29:248–256

Goettel W, Xia E, Upchurch R et al (2014) Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. BMC Genomics 15:299

Goff SA, Ricke D, Lan T-H et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Gore MA, Chia JM, Elshire RJ et al (2009) A first-generation haplotype map of maize. Science 326:1115–1117

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Hayden MJ, Kuchel H, Chalmers KJ (2004) Sequence tagged microsatellites for the Xgwm533 locus provide new diagnostic markers to select for the presence of stem rust resistance gene Sr2 in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1641–1647

He XY, He ZH, Zhang LP et al (2007) Allelic variation of polyphenol oxidase (PPO) genes located on chromosomes 2A and 2D and development of functional markers for the PPO genes in common wheat. Theor Appl Genet 115:47–58

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1

Huang X, Feng Q, Qian Q et al (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Hufford MB, Xu X, van Heerwaarden J et al (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

Hyten DL, Cannon SB, Song Q et al (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

International Peach Genome Initiative, Verde I, Abbott AG et al (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45:487–494

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Iyer-Pascuzzi A, McCouch S (2007) Functional markers for xa5-mediated resistance in rice (*Oryza sativa* L.). Mol Breed 19:291–296

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Jia J, Zhao S, Kong X et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91–95

Jiao Y, Zhao H, Ren L et al (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44:812–815

Juwattanasomran R, Somta P, Kaga A et al (2012) Identification of a new fragrance allele in soybean and development of its functional marker. Mol Breed 29:13–21

Kim MY, Lee S, Van K et al (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc Natl Acad Sci USA 107:22032–22037

Kroc M, Koczyk G, Swiecicki W et al (2014) New evidence of ancestral polyploidy in the Genistoid legume *Lupinus angustifolius* L. (narrow-leafed lupin). Theor Appl Genet 127:1237–1249

Kumar S, You FM, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genomics 13:684–694

Li R, Zhu H, Ruan J et al (2010a) *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272

Li X, Renshaw D, Yang H, Yan G (2010b) Development of a co-dominant DNA marker tightly linked to gene *tardus* conferring reduced pod shattering in narrow-leafed lupin (*Lupinus angustifolius* L.). Euphytica 176:49–58

Li X, Yang H, Buirchell B, Yan G (2011) Development of a DNA marker tightly linked to low-alkaloid gene *iucundus* in narrow-leafed lupin (*Lupinus angustifolius* L.) for marker-assisted selection. Crop Pasture Sci 62:218–224

Li X, Buirchell B, Yan G, Yang H (2012a) A molecular marker linked to the *mollis* gene conferring soft-seediness for marker-assisted selection applicable to a wide range of crosses in lupin (*Lupinus angustifolius* L.) breeding. Mol Breed 29:361–370

Li X, Yang H, Yan G (2012b) Development of a co-dominant DNA marker linked to the gene *lentus* conferring reduced pod shattering for marker-assisted selection in narrow-leafed lupin (*Lupinus angustifolius*) breeding. Plant Breed 131:540–544

Li Y-H, Zhao S-C, Ma J-X et al (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics 14:579

Li Z, Fu B-Y, Zhang G, McNally KL (2014) The 3,000 rice genomes project. GigaScience 3:1–6

Lin T, Zhu G, Zhang J et al (2014) Genomic analyses provide insights into the history of tomato breeding. Nat Genet 46:1220–1226

Ling H-Q, Zhao S, Liu y et al (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87–90

Lorenz AJ, Chao S, Asoro FG et al (2011) Genomic selection in plant breeding. In: Sparks DL (eds) Advances in agronomy. Elsevier, pp 77–123

Ma J, Yan GJ, Liu CJ (2011) Development of near-isogenic lines for a major QTL on 3BL conferring *Fusarium* crown rot resistance in hexaploid wheat. Euphytica 183:147–152

Mammadov J, Chen W, Ren R et al (2010) Development of highly polymorphic SNP markers from the complexity reduced portion of maize [*Zea mays* L.] genome for use in marker-assisted breeding. Theor Appl Genet 121:577–588

Miller MR, Dunham JP, Amores A et al (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Molina J, Sikora M, Garud N et al (2011) Molecular evidence for a single evolutionary origin of domesticated rice. Proc Natl Acad Sci USA 108:8351–8356

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147:969–977

Morris GP, Ramu P, Deshpande SP et al (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci USA 110:453–458

Mortimer SA, Kidwell MA, Doudna JA (2014) Insights into RNA structure and function from genome-wide studies. Nat Rev Genet 15:469–479

Nakaya A, Isobe SN (2012) Will genomic selection be a practical method for plant breeding? Ann Bot 110:1303–1316

Nelson MN, Phan HTT, Ellwood SR et al (2006) The first gene-based map of *Lupinus angustifolius* L.-location of domestication genes and conserved synteny with *Medicago truncatula*. Theor Appl Genet 113:225–238

Nelson MN, Moolhuijzen PM, Boersma JG et al (2010) Aligning a new reference genetic map of *Lupinus angustifolius* with the genome sequence of the model legume, *Lotus japonicus*. DNA Res 17:73–83

Nicolai M, Pisani C, Bouchet J-P et al (2012) Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). Genet Mol Res 11:2295–2300

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12:443–451

Nielsen R, Korneliussen T, Albrechtsen A et al (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One 7:e37558

Ogbonnaya FC, Subrahmanyam NC, Moullet O et al (2001) Diagnostic DNA markers for cereal cyst nematode resistance in bread wheat. Aust J Agric Res 52:1367–1374

Park C-J, Ronald PC (2012) Cleavage and nuclear localization of the rice XA21 immune receptor. Nat Commun 3:920

Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Pérez-de-Castro AM, Vilanova S, Cañizares J et al (2012) Application of genomic tools in plant breeding. Curr Genomics 13:179–195

Poland J, Endelman J, Dawson J et al (2012a) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012b) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253

Qi J, Liu X, Shen D et al (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet 45:1510–1515

Qi X, Li M-W, Xie M et al (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat Commun 5:4340

Ramirez-Gonzalez RH, Segovia V, Bird N et al (2014) RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnol J. doi:10.1111/pbi.12281

Rex B, Yu J (2007) Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Ribaut J-M, de Vicente MC, Delannay X (2010) Molecular breeding in developing countries: challenges and perspectives. Curr Opin Plant Biol 13:213–218

Sahu BB, Sumit R, Srivastava SK, Bhattacharyya MK (2012) Sequence based polymorphic (SBP) marker technology for targeted genomic regions: its application in generating a molecular map of the *Arabidopsis thaliana* genome. BMC Genomics 13:20

Saintenac C, Jiang D, Wang S, Akhunov E (2013) Sequence-based mapping of the polyploid wheat genome. G3 (Bethesda) 3:1105–1114

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218–223

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Shahidul I, Yang H, Yan G (2013) Molecular markers for genetics and plant breeding: the MFLP marker system and its application in narrow-leafed lupin (*Lupinus angustifolius*). Methods Mol Biol 1069:179–201

Shankar M, Cowling WA, Sweetingham M (1996) The expression of resistance to latent stem infection by *Diaporthe toxica* in narrow-leafed lupin. Phytopathology 86:692–697

Sharp PJ, Johnston S, Brown G et al (2001) Validation of molecular markers for wheat breeding. Aust J Agric Res 52:1357–1366

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

The International Wheat Genome Sequencing Consortium (IWGSC), Mayer KFX, Rogers J et al (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

Uitdewilligen JG, Wolters A-MA, Bjorn B et al (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8:e62355

Vales M, Dossmann J, Delgado D, Duque MC (2009) Parallel and interlaced recurrent selection (PAIRS): demonstration of the feasibility of implementing PAIRS to improve complete and partial resistance to blast (*Magnaporthe grisea*) and some other main traits in rice. Field Crops Res 111:173–178

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Varshney RK, Chen W, Li Y et al (2012a) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Ribaut J-M, Buckler ES et al (2012b) Can genomics boost productivity of orphan crops? Nat Biotechnol 30:1172–1176

Varshney RK, Song C, Saxena RK et al (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31:240–246

Velasco R, Zharkikh A, Affourtit J et al (2010) The genome of the domesticated apple (*Malus* × *domestica Borkh.*). Nat Genet 42:833–839

Wang X, Wang H, Wang J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Wang XQ, Zhao L, Eaton DAR et al (2013) Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. Mol Ecol Resour 13:938–945

Ward JA, Bhangoo J, Fernández-Fernández F et al (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics 14:2–15

Williamson PM, Sivasithamparam K, Cowling WA (1991) Formation of subcuticular hyphae by *Phomopsis leptostromiformis* upon latent infection of narrow-leafed lupins. Plant Dis 75:1023–1025

Willing E-M, Hoffmann M, Klein JD et al (2011) Paired-end RAD-seq for *de novo* assembly and marker design without available reference. Bioinformatics 27:2187–2193

Wolko B, Weeden NF (1994) Linkage map of isozyme and RAPD markers for the *Lupinus angustifolius* L. ISA Press, Lisbon, pp 42–49

Xu Y (2010) Molecular plant breeding. CABI, Wallingford, pp 1–755

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. Crop Sci 48:391–407

Xu X, Pan S, Cheng S et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Xu Y, Lu Y, Xie C et al (2012) Whole-genome strategies for marker-assisted plant breeding. Mol Breed 29:833–854

Xu Y, Xie C, Wan J et al (2013) Marker-assisted selection in cereals: platforms, strategies and examples. In: Gupta PK, Varshney RK (eds) Cereal genomics II. Springer, Netherlands, pp 375–411

Yagi M, Kosugi S, Hirakawa H et al (2014) Sequence analysis of the genome of carnation (*Dianthus caryophyllus* L.). DNA Res 21:231–241

Yang H, Sweetingham MW, Cowling WA, Smith PM (2001) DNA fingerprinting based on microsatellite-anchored fragment length polymorphisms, and isolation of sequence-specific PCR markers in lupin (*Lupinus angustifolius* L.). Mol Breed 7:203–209

Yang H, Shankar M, Buirchell BJ et al (2002) Development of molecular markers using MFLP linked to a gene conferring resistance to *Diaporthe toxica* in narrow-leafed lupin (*Lupinus angustifolius* L.). Theor Appl Genet 105:265–270

Yang H, Boersma JG, You M et al (2004) Development and implementation of a sequence-specific PCR marker linked to a gene conferring resistance to anthracnose disease in narrow-leafed lupin (*Lupinus angustifolius* L.). Mol Breed 14:145–151

Yang H, Renshaw D, Thomas G et al (2008) A strategy to develop molecular markers applicable to a wide range of crosses for marker assisted selection in plant breeding: a case study on anthracnose disease resistance in lupin (*Lupinus angustifolius* L.). Mol Breed 21:473–483

Yang H, Lin R, Renshaw D et al (2010) Development of sequence-specific PCR markers associated with a polygenic controlled trait for marker-assisted selection using a modified selective genotyping strategy: a case study on anthracnose disease resistance in white lupin (*Lupinus albus* L.). Mol Breed 25:239–249

Yang H, Tao Y, Zheng Z et al (2012) Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. BMC Genomics 13:318–328

Yang H, Tao Y, Zheng Z et al (2013a) Rapid development of molecular markers by next-generation sequencing linked to a gene conferring phomopsis stem blight disease resistance for marker-assisted selection in lupin (*Lupinus angustifolius* L.) breeding. Theor Appl Genet 126:511–522

Yang H, Tao Y, Zheng Z et al (2013b) Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. PLoS One 8:e64799

You M, Boersma JG, Buirchell BJ et al (2005) A PCR-based molecular marker applicable for marker-assisted selection for anthracnose disease resistance in lupin breeding. Cell Mol Biol Lett 10:123–134

Young ND, Debellé F, Oldroyd GED et al (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480:520–524

Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhang D, Song H, Cheng H et al (2014) The acid phosphatase-encoding gene GmACP1 contributes to soybean tolerance to low-phosphorus stress. PLoS Genet 10:e1004061

Zheng Z, Wang HB, Chen GD, Yan GJ, Liu CJ (2013) A procedure allowing up to eight generations of wheat and nine generations of barley per annum. Euphytica 191:311–316

Zhou L, Chen Z, Lang X et al (2013) Development and validation of a PCR-based functional marker system for the brown planthopper resistance gene Bph14 in rice. Breed Sci 63:347–352