frontiers in Plant Science

REVIEW
published: 06 July 2017
doi: 10.3389/fpls.2017.01182

Searching for an Accurate Marker-Based Prediction of an Individual Quantitative Trait in Molecular Plant Breeding

Yong-Bi Fu[1], Mo-Hua Yang[1,2], Fangqin Zeng[1] and Bill Biligetu[3]*

[1] Plant Gene Resources of Canada, Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, [2] College of Forestry, Central South University of Forestry and Technology, Changsha, China, [3] Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

Molecular plant breeding with the aid of molecular markers has played an important role in modern plant breeding over the last two decades. Many marker-based predictions for quantitative traits have been made to enhance parental selection, but the trait prediction accuracy remains generally low, even with the aid of dense, genome-wide SNP markers. To search for more accurate trait-specific prediction with informative SNP markers, we conducted a literature review on the prediction issues in molecular plant breeding and on the applicability of an RNA-Seq technique for developing function-associated specific trait (FAST) SNP markers. To understand whether and how FAST SNP markers could enhance trait prediction, we also performed a theoretical reasoning on the effectiveness of these markers in a trait-specific prediction, and verified the reasoning through computer simulation. To the end, the search yielded an alternative to regular genomic selection with FAST SNP markers that could be explored to achieve more accurate trait-specific prediction. Continuous search for better alternatives is encouraged to enhance marker-based predictions for an individual quantitative trait in molecular plant breeding.

Keywords: quantitative trait, RNA-Seq, functional marker, breeding, marker-assisted selection, genomic selection, trait-specific marker selection

OPEN ACCESS

Edited by:
Chengdao Li,
Murdoch University, Australia

Reviewed by:
Yongqing Jiao,
Oil Crops Research Institute (CAAS), China
Ryo Fujimoto,
Kobe University, Japan

****Correspondence:***
Yong-Bi Fu
yong-bi.fu@agr.gc.ca

Specialty section:
This article was submitted to Crop Science and Horticulture, a section of the journal Frontiers in Plant Science

Received: *20 March 2017*
Accepted: *20 June 2017*
Published: *06 July 2017*

Citation:
Fu Y-B, Yang M-H, Zeng F and Biligetu B (2017) Searching for an Accurate Marker-Based Prediction of an Individual Quantitative Trait in Molecular Plant Breeding. Front. Plant Sci. 8:1182. doi: 10.3389/fpls.2017.01182

INTRODUCTION

Molecular plant breeding with the aid of molecular markers has played an important role in modern plant breeding over the last two decades (Moose and Mumm, 2008). Many useful markers have been developed and applied to enhance parental selection in breeding programs (e.g., Randhawa et al., 2013; Grover and Sharma, 2016). Recent advances in next-generation sequencing (NGS) technology (Varshney et al., 2009; Metzker, 2010) have helped to generate abundant low-cost molecular markers and make the molecular markers more useful and informative for plant breeding (Varshney et al., 2014). Currently, there are two major approaches applied for molecular breeding: marker-assisted selection (MAS) and genomic selection (GS or Genome-wide selection) (Jiang, 2013). Traditional MAS is based on the selection of statistically significant, marker-trait associations and enhances parental selection for recessive trait and disease resistance in conventional breeding program without observing phenotypic variation in the traits. However,

traditional MAS is not well-suited for complex traits controlled by many genes (Beavis, 1998). GS, introduced first in animal breeding (Meuwissen et al., 2001), estimates genome-wide marker effects and uses the estimates to predict individual genetic potential (i.e., genomic estimated breeding values). Studies have shown that GS outperforms MAS in parental selection, particularly for those complex traits controlled by a large number of genes (e.g., see Bernardo and Yu, 2007; Massman et al., 2013; Sorrells, 2015; Liu et al., 2016). However, GS applications are not lacking of technical issues and usually display low accuracies of predicting trait performances (Jannink et al., 2010; Windhausen et al., 2012; Riedelsheimer et al., 2013; Bassi et al., 2016; Rabier et al., 2016). Thus, improving trait prediction accuracy is one of the active research areas in molecular plant breeding, and the development of genome-wide informative markers through NGS remains a major theme of research (Yang et al., 2015).

RNA-Sequencing (or RNA-Seq) is a recently developed genomic approach for transcriptome profiling, can be applied to study each transcript of genes affecting a trait at a developmental stage, and has opened many avenues to develop informative markers associated with genes controlling genetically complex traits of agronomical importance (Wang et al., 2009; Ozsolak and Milos, 2011). Here we attempt to search for alternatives to GS for more accurate trait prediction through a literature review on the prediction issues in molecular plant breeding and on the applicability of an RNA-Seq technique for developing function-associated specific trait (FAST) SNP markers. We also perform a theoretical reasoning on whether and how FAST SNP markers could enhance individual trait prediction and verify the reasoning through computer simulation. It is our hope that this effort would seed an alternative with specific trait SNP markers that can be explored to achieve more accurate prediction for a quantitative trait in molecular plant breeding.

MOLECULAR PLANT BREEDING AND ITS LIMITATIONS

Molecular plant breeding is generally termed as the application of molecular markers to improve the characters of interest in plants (Xu, 2010; Jiang, 2013), and is one of the modern breeding strategies with the potential to accelerate breeding efficiency (Moose and Mumm, 2008). Conventional plant breeding is largely relied on phenotypic selection through cycles of crossing and selection and requires substantial breeding efforts with more than 10 years to develop an improved variety. The major challenge lies in the low efficiency of phenotypic selection for desirable traits of quantitative nature such as yield and disease resistance that are controlled by many genes of small effects and their interactions with environments. Thus, efficient methods have been searched to improve the selection of individual plants with desired traits, including MAS.

The idea for the use of markers to assist plant selection could date back to the association analysis done by Sax (1923) between seed color (monogenic trait) and seed weight (polygenic, quantitatively inherited trait) in beans (*Phaseolus vulgaris* L.) and the promotion made by Thoday (1961) on the mapping of polygenic traits with the help of monogenic morphological markers. Although allozyme markers were applied in the early 1980s to identify genotypes, the idea of MAS was not flourished until the development of the first DNA-based genetic markers, restriction fragment length polymorphisms (Botstein et al., 1980). Since then, large efforts have been made to develop molecular markers such as random-amplified polymorphic DNAs, amplified fragment length polymorphisms, simple sequence repeats or single nucleotide polymorphisms (Grover and Sharma, 2016). Such advance in molecular markers not only stimulated the theoretical investigation on MAS efficiency (e.g., see Lande and Thompson, 1990), but also made the MAS practically feasible to complement and enhance the conventional plant breeding (Moose and Mumm, 2008). Accordingly, many MAS techniques have been developed, including marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), and GS (Jiang, 2013). With the recent advance in NGS and the development of genome-wide SNP markers, GS will be more efficient, even in MABC and MARS. These technical developments have made the molecular breeding a standard practice complementary to conventional breeding to improve traits with complex genetic bases (Moose and Mumm, 2008).

As expected with the promise of MAS, several reviews have confirmed that the research and use of molecular markers in plant breeding have continued to increase in the public and private sectors, particularly since the 2000s (Beavis, 1998; Holland, 2004; Collard and Mackill, 2008; Xu and Crouch, 2008; Brumlop and Finckh, 2011; Boopathi, 2013). Successful stories for MAS applications are not lacking (e.g., see Collard and Mackill, 2008; Boopathi, 2013; Randhawa et al., 2013). For example, many molecular markers were deployed to assist selection for disease resistance, agronomic and quality traits in several wheat (*Triticum* spp.) cultivars released for commercial cultivation in Canada (Randhawa et al., 2013). However, MAS applications mainly focused on simply inherited traits, such as monogenic or oligogenic resistance to diseases/pests, although quantitative traits were also involved (Collard and Mackill, 2008). Also, these MAS applications have not achieved the results as expected previously in terms of extent and success (e.g., release of commercial cultivars). For example, Collard and Mackill (2008) listed 10 reasons for the low impact of MAS in general and Jiang (2013) highlighted seven issues associated with MAS applications. Among them are (1) not all markers are breeder-friendly, (2) not all markers can be applicable across populations due to lack of marker polymorphism or reliable marker-trait association, (3) false selection may occur due to recombination between the markers and the genes or quantitative trait loci (QTL) of interest, and (4) imprecise estimates of QTL locations and effects result in slower progress than expected. Improvement of most agronomic traits that are of complicated inheritance and economic importance like yield and quality is still a great challenge for MAS including the newly developed GS (Jannink et al., 2010). Jiang (2013) indicated that MAS is not universally or necessarily advantageous, at least from the viewpoint of a plant breeder.

Last several years have seen increased researches directed toward GS applications (e.g., see Spindel et al., 2015; Bassi et al., 2016). With the advances in the development of cost-effective genome wide markers, it is no doubt that some of the old challenges faced with the MAS applications can be addressed (Jannink et al., 2010). Several applications have demonstrated its usefulness in actual plant breeding programs (e.g., see Sorrells, 2015; Bassi et al., 2016). However, some marked features of GS in plant breeding have also started to emerge (Windhausen et al., 2012; Riedelsheimer et al., 2013; Spindel et al., 2015). First, the accuracy of the genome-wide marker prediction on trait performance has a range of estimates, but is generally low, depending on many factors including crop, trait, marker, training population, GS model, and environment (Rabier et al., 2016). To update the current status of prediction accuracy, we selected 31 peer-review journal publications from 2015 to July of 2016 that reported genomic selections in crop and tree species, and obtained 187 genomic predictions of trait performance with a range of 0.05 to 0.83 and a mean of 0.50 (**Figure 1** and Supplementary Table S1). For example, a range of prediction accuracies from 0.31 to 0.63 for several traits were found in rice (*Oryza sativa* L.) (Spindel et al., 2015); 0.14 to 0.58 for spring barley (*Hordeum vulgare* L.) and 0.40 to 0.80 for winter barley in malting quality traits (Schmidt et al., 2016); 0.10 to 0.51 in maize (*Zea mays* L.) root traits (Pace et al., 2015); and 0.39 to 0.61 in Canola (Jan et al., 2016). Second, some studies have shown that more genomic markers evenly distributed across the genome do not always help to increase the prediction accuracy and as low as 1000 genomic markers can achieve the same level of prediction accuracy for some traits (Spindel et al., 2015; Jan et al., 2016). These features help to explain partly some less optimistic views of GS potential (Bernardo, 2016), and suggest that more research are required on the choice and development of informative genomic markers for GS.

Our analysis of the GS applications with respect to prediction accuracy concurs well with the renewed argument that more research efforts are needed to develop functional markers for MAS, taking advantage of the recent advances in NGS application (Lau et al., 2015; Yang et al., 2015). This realization is not surprising, as the idea for developing functional DNA markers for plant breeding is not new (e.g., see Andersen and Lübberstedt, 2003; Varshney et al., 2005). However, large efforts have been made with limited success, even in major crop species (Iyer-Pascuzzi and McCouch, 2007; Liu Y. et al., 2012; Lau et al., 2015; Yang et al., 2015). The searching for functional markers via QTL and expression QTL (eQTL) analyses (Druka et al., 2010) or gene cloning with limited genomic resources is technically challenging, labor extensive and time consuming (Yang et al., 2015). Acquiring a relevant set of functional or useful markers through genome-wide association mapping (GWAS) is technically possible for marker-based prediction of trait performance, but practically depends highly on the genotyping accuracy and linkage disequilibrium (LD), and is largely limited to the assayed populations in given environments (e.g., see Eichler et al., 2010; Desta and Ortiz, 2014; Spindel et al., 2016). Dr. Hong-Bin Zhang at Texas A&M University has promoted the idea of gene-based breeding system since 2014 and demonstrated substantial gains in trait prediction from gene-based markers in both cotton (*Gossypium hirsutum*; Liu et al., 2017) and maize (Zhang et al., 2017). For example, using 474 Gossypium fiber length genes, they were able to predict fiber length with correlation coefficients ranging from 0.67 to 0.85 (Liu et al., 2017). However, inadequate attention has been paid to gene-based breeding system. This dilemma seems to suggest that a paradigm shift is needed to develop functional or function-associated markers, particularly focusing on specific quantitative traits.

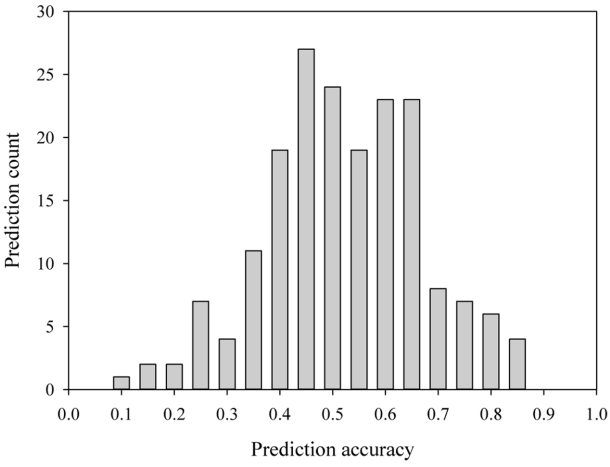

FIGURE 1 | The distribution of 187 prediction accuracies from genomic selection for several traits in crop and tree species, as reported in a selection of 31 peer-review journal publications from 2015 to 2016 (see the Supplementary Table S1).

RNA-Seq AND SNP MARKERS FOR SPECIFIC TRAITS

RNA-Seq is a recently developed genomic technology using NGS to study transcriptome (Marioni et al., 2008; Nagalakshmi et al., 2008; Wilhelm et al., 2008). The transcriptome is usually defined as the set of all RNA molecules transcribed in an organ or tissue at a particular point of time under a given set of environmental conditions. Generally, RNA-Seq has two major components. First, RNA is purified from a sample of interest and converted to a library of cDNA fragments with adaptors attached to one or both ends. Each cDNA fragment, with or without enriched with PCR amplification, is then sequenced using one of high throughput sequencing methods to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). Second, a suite of bioinformatics tools are used to process the raw sequence reads, typically 30–400 bp, map the processed sequences to a reference genome or reference transcripts, or *de novo* assemble without the genomic sequence, and analyze the alternative gene spliced transcripts, post-transcriptional

modifications, gene fusion, mutations/SNPs and changes in gene expression (Garber et al., 2011; Lopez-Maestre et al., 2016).

RNA-Seq has a number of advantages over hybridization-based microarrays. First, it does not require existing genomic sequence information to identify transcripts, making its application to non-model plants more feasible. Second, the detection of differentially expressed genes is more accurate, sensitive and reproducible, with fewer systematic discrepancies among technical replicates (Marioni et al., 2008; Nagalakshmi et al., 2008). Third, RNA-Seq allows for quantification of the abundance or changes of each transcript in a developmental stage or under a specific treatment condition (Mortazavi et al., 2008), making the study of a complex transcriptome possible. The paired-end tag sequencing strategy of RNA-Seq further improves cDNA sequencing efficiency with expanded length of short reads for better understanding of the dynamic transcriptomes (Fullwood et al., 2009).

A variety of RNA-Seq applications have been found, ranging from transcriptome profiling to gene discovery, alternative splicing analysis and molecular marker development (Conesa et al., 2016). RNA-Seq has been successfully applied to study the transcriptomes of different tissues such as root (Postnikova et al., 2013), leaf (Li et al., 2010), flower (Mantegazza et al., 2014), fruit (Kang et al., 2013; Martínez-López et al., 2014), and seed (Jones and Vodkin, 2013). Also, it has been employed to analyze gene expressions for biotic and abiotic responses like diseases resistance (Kong et al., 2015), drought stress (Kakumanu et al., 2012; Bhardwaj et al., 2015), cold stress (Sinha et al., 2015; Hu et al., 2016) and chemical stress (He et al., 2015). Moreover, RNA-Seq applications have been reported in many plant species such as *Arabidopsis*, rice, maize, as well as non-model species such as soybean [*Glycine max* (L). *Merr.*] and wheat (Jiao et al., 2009; Filichkin et al., 2010; Li et al., 2010; Severin et al., 2010; Ramirez-Gonzalez et al., 2015). These applications have demonstrated its tremendous power in characterizing transcriptomes, as it can detect low-expressed transcripts, splice variants, and novel transcripts (Socquet-Juglard et al., 2013). Therefore, RNA-Seq is now regarded as the latest and most powerful tool for sequencing and profiling of transcriptome (Han et al., 2015; Conesa et al., 2016).

Last several years have seen increased efforts toward the development of SSR and/or SNP markers through RNA-Seq in many organisms (e.g., see Wei et al., 2011; Yang et al., 2011; Salem et al., 2012; Ulloa et al., 2015; Zhou et al., 2016). Abundant RNA-Seq SNP markers have been developed to sample polymorphisms within the transcribed region of all genes associated with many traits. Many of these markers may be function-associated with some traits of interest, differing from those selectively neutral markers, but not necessarily are qualified as functional markers for a specific trait. This may reflect the fact that many RNA-Seq analyses were performed with the goals to generate dense, genome-wide function-associated markers for linkage mapping and association mapping, not necessarily for direct GS application. Also, it is challenging to develop truly functional markers for specific complex traits, as it requires specific RNA-Seq designs for specific traits and the resulting SNP markers are required to verify their associations with causal genes influencing the traits.

However, it is practically feasible to develop FAST SNP markers through specific RNA-Seq designs and the developed markers have a high probability of being functionally relevant when compared to randomly selected polymorphisms. Note that FAST markers are not technically new, but termed here to distinguish them from others. For example, Salem et al. (2012) conducted an RNA-Seq whole-transcriptome analysis of pooled cDNA samples from a population of rainbow trout (*Oncorhynchus mykiss*) selected for improved growth versus unselected genetic cohorts and developed many FAST SNP markers for growth traits for fish breeding. Similarly, Ulloa et al. (2015) performed an RNA-Seq analysis of eight low-growth and eight high-growth Zebrafish (*Danio rerio*) and developed 164 SNPs, five of which were associated with genes affecting fish growth. Chopra et al. (2015) applied an RNA-Seq to develop and validate a set of gene-based SNPs in sorghum (*Sorghum bicolor*) genotypes with contrasting responses to cold stress. Ramirez-Gonzalez et al. (2015) implemented an RNA-Seq analysis of bulked pools sampled from a F2 population to identify 175 putative SNP markers associated with *Yr15*, the yellow rust (*Puccinia striiformis f.sp. tritici*) resistance in wheat germplasm. Clearly, the most successful applications in plants were those using RNA-Seq in combination with bulked segregant analysis (Michelmore et al., 1991; Liu S. et al., 2012). Similarly, this approach has also facilitated the development of the high resolution SNP maps for wheat grain protein content (Trick et al., 2012) and for the fertility restorer genes of cytoplasmic male-sterility in radish (*Raphanus sativus* L.) and onion (*Allium cepa* L.) (Lee et al., 2014; Kim et al., 2015). These successful applications are encouraging for developing FAST SNP markers for individual traits of breeding target.

THEORETICAL REASONING AND COMPUTER SIMULATION

Our literature review indicates the importance of using functional markers to increase the trait prediction accuracy for GS. This should not be surprised, as functional markers should be more informative to acquire genetic effects of causal genes for trait prediction than genome-wide neutral markers (Mackay, 2001). Using more random, non-causal SNP markers can inflate individual genomic relationships and decrease trait prediction accuracy (Spindel et al., 2015; Edwards et al., 2016). Also, functional markers can avoid marker validation like those random markers in different breeding populations and could be gene or trait specific (Lau et al., 2015; Yang et al., 2015). Our review also indicates various challenges in the development of ideal functional markers for specific traits, but shows the feasibility of developing FAST SNP markers through RNA-Seq. Thus, we reasoned that FAST SNP markers may not supersede the ideal functional markers, but should be more informative, to predict genetic effects associated with a given trait than those dense, genome-wide neutral markers. This reasoning is based on two expectations that the extent of LD between FAST SNP

markers and causal genes is generally larger than those between genome-wide neutral markers and casual genes, and that the trait prediction accuracy is positively related to LD (Meuwissen et al., 2001; Fernando et al., 2007).

To understand these two expectations, Fernando and his colleagues conducted extensive computer simulations (e.g., Fernando et al., 2007; Kizilkaya et al., 2010) to illustrate the impacts of LD between SNP markers and casual genes on trait predictions of young cattle (*Bos taurus*). In one simulation on an ideal pattern of LD with marker loci either in complete LD or linkage equilibrium with QTL, they found only the prediction method of Bayes-B (Meuwissen et al., 2001) could achieve up to 0.98 prediction accuracy, while the other two methods RR-BLUP and TP-BLUP displayed lower, unstable trait predictions (Fernando et al., 2007). Considering scenarios with more realistic LD patterns for 30 chromosomes with up to 2000 markers each, they found the accuracy of trait prediction by Bayes-B did not increase after 500 markers per chromosome (Fernando et al., 2007). These findings clearly indicate the importance of LD patterns in a trait prediction. To verify their simulated findings, they used actual 50K SNP data of 1,086 purebred (PB) and 924 multibreed (MB) Angus cattle from eight sire breeds, simulated a trait with the heritability of 0.5 controlled by 50, 100, 250, or 500 additive QTL selected randomly from 50K SNPs, and examined five marker panels (mp) with variable levels of LD for genetic evaluation (Kizilkaya et al., 2010). Specifically, for each QTL scenario, mp1 is an ideal case with only QTL genotypes; mp2 is another extreme with both QTL genotypes and equal number of marker loci with the highest linkage disequilibrium (HLD) for each QTL; mp3 reflects the common practice with all genome-wide SNPs, including QTL; mp4 represents a case of markers each having HLD with an QTL; and mp5 reflects a case of markers with all the SNPs minus QTL. The simulated correlations between true and predicted genotypic values by Bayes-B in the PB validation data set are shown in **Table 1**. As expected, the ideal functional markers with QTL genotypes (mp1) displayed the highest prediction accuracies, ranging from 0.72 to 0.95 and increasing with fewer QTL. When there were more than 100 QTL, the highly linked markers (mp4) showed higher prediction accuracies than all genome-wide SNPs including QTL (mp3). Thus, these simulation results are consistent with the two expectations mentioned above for FAST SNPs, as FAST SNPs should approach the behavior of those highly linked markers to QTL (mp4) for a trait predication.

To confirm the simulation results in Angus cattle, particularly with respect to mp4, we also conducted a computer simulation based on existing SoySNP50K data (Song et al., 2015), following exactly the same simulation approach used by Kizilkaya et al. (2010) in Angus cattle. First, we randomly selected 800 soybean plants from the 18,480 domesticated soybean accessions with 42,509 polymorphic SNP markers. After excluding the scaffold SNPs and minor alleles (of frequency less than 0.05) and replacing missing data with common haplotypes, we obtained a final soybean data for this simulation with 800 plants with 36,543 SNP markers and divided them into half, each representing a training or validation set. Second, we simulated the same four QTL scenarios as in cattle with 50, 100, 250, and 500 additive QTL that were randomly selected from 36,543 SNPs, and applied the same five marker panels (mp1 to mp5) as described above and two additional marker panels (mp6 and mp7). Specifically, mp6 consisted of the marker loci in which two markers were randomly selected from loci with the highest 20 LD values for each QTL (HLDr2), and mp7 included both mp6 and a set of random SNP markers (rSNP) each falsely representing an QTL. Third, we also considered two heritabilities 0.5 and 0.2, and applied two extra genomic selection models (Bayes-C, and RR-BLUP), besides Bayes-B. Fourth, for each marker panel with different QTL scenarios, random select 400 soybean marker data representing as the training population and another 400 plants as the validation population for prediction for five times, we generated five more replicates than Kizilkaya et al. (2010) did to get average correlations between true and predicted trait values in each random selected validation set. The simulation was conducted with a custom R script (R Core Team, 2015) that was specifically developed for this confirmation and is available upon request to the first author. Marker effect estimation and genetic value prediction were made using the BGLR statistical package in R (Pérez-Rodriguez and de los Campos, 2014) implemented with three genomic prediction models [RR-BLUP (= GBLUP), Bayes-B and Bayes-C] and confirmed with the *rrBLUP mixed.solve* function (Endelman, 2011).

Our simulation not only confirmed those observed in the Angus cattle, but also revealed some interesting patterns of trait prediction (**Table 1** and Supplementary Table S2). First, the patterns of prediction accuracy by Bayes-B in soybean are the same as in cattle for the QTL scenarios of QTL250 and QTL500. In any QTL scenario, soybean functional markers (mp1) always showed the highest accuracies of trait prediction, followed by highly linked markers (mp4) and all genome-wide SNP markers (mp3) (**Table 1**). Also, the patterns of decreased prediction accuracies by functional markers (mp1) with more QTL were also observed in soybean data (**Table 1**). Second, soybean markers with a little relaxed LD to QTL like mp6 or mp7 still displayed higher prediction accuracies than those genome-wide SNP markers (mp3) in any QTL scenarios assayed. For example, for a trait of heritability 0.5 with 100 QTL, 200 HLDr2 markers (mp6) displayed a correlation of 0.67 while all 36,543 SNP markers (mp3) had only a correlation of 0.61 (**Table 1**). Third, several extra patterns of prediction accuracy were also observed in soybean (**Table 1** and Supplementary Table S2). Three different prediction methods did not show much difference in prediction accuracy. The prediction accuracies became lower for a trait of lower heritability. The prediction accuracies using 36,543 SNPs for a trait of heritability 0.2 ranged from 0.48 to 0.53, while those using highly linked markers (mp6 or mp7) ranged from 0.49 to 0.63. All together, these simulation results demonstrated the potential gain in prediction accuracy from the application of FAST SNP markers in molecular breeding. More evenly distributed markers unlinked to causal genes do not enhance, but rather reduce, trait prediction accuracy.

Our simulation on soybean data had a simple goal to reason the potential of FAST SNP markers and thus

TABLE 1 | Comparative simulation results on the accuracies of predicting a quantitative trait with heritability 0.5 by genomic prediction model Bayes-B based on 50K Angus cattle and 36,543 soybean SNP data with respect to QTL scenario and marker panel.

Angus cattle*		Soybean*	
QTL scenario/marker panel$^{\Phi}$	**Correlation$^{\#}$**	**QTL scenario/marker panel$^{\Phi}$**	**Correlation$^{\#}$**
QTL50		*QTL50*	
mp1: 50 QTL	0.953	mp1: 50 QTL	0.94 (0.01)
mp2: 50 QTL + 50 HLD	0.931	mp2: 50 QTL + 50 HLD	0.93 (0.01)
mp3: 50K SNPs with QTL	0.766	mp3: 36543 SNPs with QTL	0.64 (0.07)
mp4: 50 HLD	0.570	mp4: 50 HLD	0.83 (0.03)
mp5: 50K SNPs − 50 QTL	0.388	mp5: 36543 SNPs − 50 QTL	0.63 (0.07)
		mp6: 100 HLDr2	0.73 (0.06)
		mp7: 100 HLDr2 + 50 rSNP	0.72 (0.06)
QTL100		*QTL100*	
mp1: 100 QTL	0.938	mp1: 100 QTL	0.88 (0.02)
mp2: 100 QTL + 100 HLD	0.914	mp2: 100 QTL + 100 HLD	0.87 (0.02)
mp3: 50K SNPs with QTL	0.585	mp3: 36543 SNPs with QTL	0.61 (0.09)
mp4: 100 HLD	0.513	mp4: 100 HLD	0.77 (0.05)
mp5: 50K SNPs − 100 QTL	0.289	mp5: 36543 SNPs − 100 QTL	0.60 (0.09)
		mp6: 200 HLDr2	0.67 (0.07)
		mp7: 200 HLDr2 + 100 rSNP	0.66 (0.08)
QTL250		*QTL250*	
mp1: 250 QTL	0.840	mp1: 250 QTL	0.78 (0.03)
mp2: 250 QTL + 250 HLD	0.788	mp2: 250 QTL + 250 HLD	0.77 (0.04)
mp3: 50K SNPs with QTL	0.399	mp3: 36543 SNPs with QTL	0.61 (0.07)
mp4: 250 HLD	0.510	mp4: 250 HLD	0.71 (0.05)
mp5: 50K SNPs − 250 QTL	0.247	mp5: 36543 SNPs − 250 QTL	0.61 (0.07)
		mp6: 500 HLDr2	0.63 (0.06)
		mp7: 500 HLDr2 + 250 rSNP	0.62 (0.06)
QTL500		*QTL500*	
mp1: 500 QTL	0.720	mp1: 500 QTL	0.70 (0.06)
mp2: 500 QTL + 500 HLD	0.642	mp2: 500 QTL + 500 HLD	0.70 (0.07)
mp3: 50K SNPs with QTL	0.254	mp3: 36543 SNPs with QTL	0.60 (0.08)
mp4: 500 HLD	0.372	mp4: 500 HLD	0.65 (0.08)
mp5: 50K SNPs − 500 QTL	0.200	mp5: 36543 SNPs − 500 QTL	0.60 (0.08)
		mp6: 1000 HLDr2	0.62 (0.08)
		mp7: 1000 HLDr2 + 500 rSNP	0.61 (0.08)

**The results for Angus cattle were acquired from Table 2 of Kizilkaya et al. (2010) and those for soybean were obtained from this simulation.*

$^{\Phi}$Both cattle and soybean simulations applied the same QTL scenarios and the first five marker panels (mp), but soybean simulation had two extra panels. For each QTL scenario, mp1 is an ideal case with only QTL genotypes; mp2 is another extreme with both QTL genotypes and equal number of marker loci with the highest linkage disequilibrium (HLD) for each QTL; mp3 reflects the common practice with all SNPs, including QTL; mp4 represents a case of markers each having HLD with an QTL; and mp5 reflects a case of markers with all the SNPs minus QTL; mp6 consists of the marker loci in which two markers are randomly selected from loci with the highest 20 LD values for each QTL (HLDr2), and mp7 considers both mp6 and a set of random SNP markers (rSNP) each falsely representing an QTL.

$^{\#}$The correlations between true and predicted genotypic values in the validation data sets and their standard deviations in parentheses. Training and validation sets consisted of 924 multibreed and 1086 purebred cattle, respectively, and for soybean, training and validation populations had 400 plants each.

was not comprehensive. Further detailed simulations are possible to consider all existing marker prediction models, the related parameters associated with QTL genetic model, marker distribution and informativeness, training set and test environment (Heffner et al., 2009, 2010; Zhong et al., 2009; Hickey et al., 2014). However, our simulation results are consistent with several empirical reports from GS analyses that prediction accuracies were higher using only the QTL-linked markers or a subset of informative markers (e.g., Spindel et al., 2015; Thavamanikumar et al., 2015; Arruda et al., 2016; Edwards et al., 2016; Huang et al., 2016; Liu et al., 2016). Thus, the simulations in Angus cattle and soybean, along those empirical reports, provided support for our theoretical reasoning to search for more informative FAST SNP markers through RNA-Seq to improve trait prediction accuracy.

AN ALTERNATIVE FOR INDIVIDUAL TRAIT PREDICTION

Based on the literature review and theoretical reasoning, we synthesized that FAST SNP markers can be developed through RNA-seq for an individual quantitative trait and applied to increase the trait prediction accuracy. To better utilize this

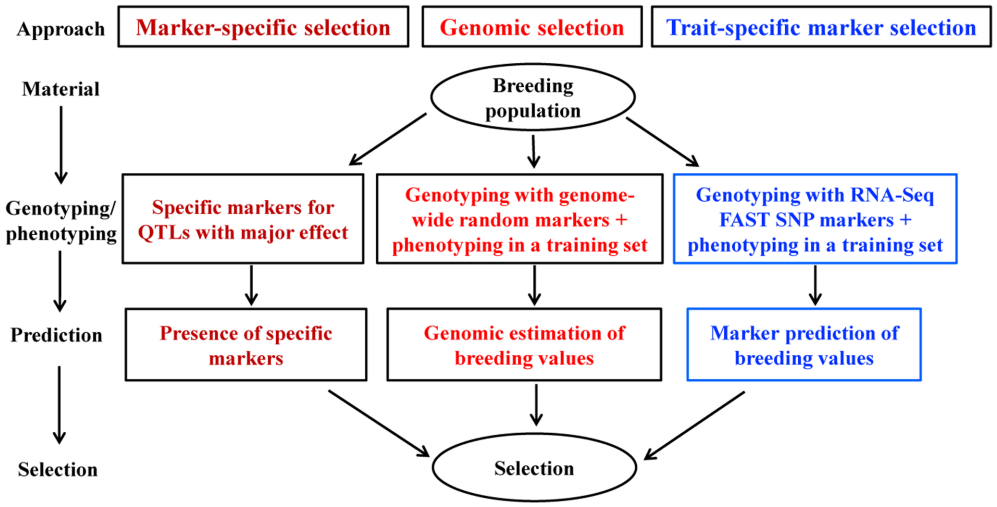

FIGURE 2 | Capturing genes of major and minor effects for a quantitative trait of interest through RNA-Seq function-associated specific trait (FAST) SNP markers in molecular plant breeding. Marker-specific selection uses the specific markers developed from the major QTL of the trait to select desirable parents. Genomic selection applies genome-wide random markers to estimate breeding values in various traits for parental selection. The derived alternative named trait-specific marker selection (TSMS) employs FAST SNP markers developed from RNA-Seq to predict individual performances in a trait of interest for parental selection.

synthesis, we conceived a marker-based and trait-specific strategy as an alternative to regular GS with FAST SNP markers for plant breeders to facilitate parental selection. We termed it as trait-specific marker selection (TSMS) for ease of interpretation and comparison to marker-specific selection and GS. It is our hope that this alternative or its modifications later can provide a useful breeding tool to improve the accuracy of marker-based prediction on individual trait performance.

Trait-specific marker selection represents an added option to GS that can be applied to assist parental selection by predicting specific trait breeding values of individual plants in a breeding population through the separate development and application of RNA-Seq FAST SNP markers for specific traits of interest (**Figure 2**). It requires the development and validation of FAST SNP markers for a given trait in other populations, before the application to genotype a breeding population of interest; estimates the marker "effects" in a training set of the breeding population; and applies the estimated marker "effects" to predict trait performance in the same breeding population. Such an approach differs from traditional MAS with QTL-specific markers in genotyping and prediction, but follows the same idea of GS to predict trait performance with RNA-Seq FAST SNP markers, rather than the genome-wide selectively neutral SNP markers. To make the strategy more understandable, we outline the two major components of TSMS in **Figures 3**, **4** for developing FAST SNP markers through RNA-Seq technology and for performing SNP marker prediction of breeding values, respectively. The proposed RNA-Seq method (**Figure 3**) considers multiple pairs of individual plants with two extreme trait values and collects their sample tissues at given developmental stages for gene expressions associated with the trait of interest. The collected samples will be subjected to RNA-Seq analysis through RNA extraction, cDNA library preparation, multiplexing with barcoding and cDNA sequencing. The collected RNA-Seq data will be analyzed through *de novo* assembly using various bioinformatics tools to identify differential transcripts for each pair and to generate consensus differential transcripts from all the assayed pairs. Identification of a differential transcript in a pair is made based on the presence or absence of a transcript or the difference in abundance of the transcript detected in both plants. Multiple pairs are used to enhance the reliability of identifying differential transcripts for the trait. Based on the consensus differential transcripts, SNP call will be made from all the samples and the detected SNPs will be filtered to generate putative SNPs for the trait based on the differences in allelic frequency between two trait-extreme sets of assayed samples. An empirical validation of putative SNPs in separate population(s) is required to confirm if the acquired SNP markers are truly associated with the trait performance. The validated SNP markers can be applied to genotype all the breeding materials of interest, and some of these genotyped plants will also be assessed with their trait performance as a training set (**Figure 4**). These marker and trait data in the training set can be analyzed using existing marker prediction models for GS such as RR-BLUP or Bayes-B implemented in various R packages (Endelman, 2011; Pérez-Rodriguez and de los Campos, 2014) to estimate marker "effects." The estimated marker "effects" will be utilized to predict these breeding values in the prediction set for genetic ranking of parental lines.

The advantage of this alternative over regular GS mainly lies in the potential gains in marker prediction of individual trait performance through the application of FAST SNP markers. Realizing the gain in trait prediction highly depends on the development of the FAST SNP markers for individual traits, and requires further empirical investigations in breeding programs.

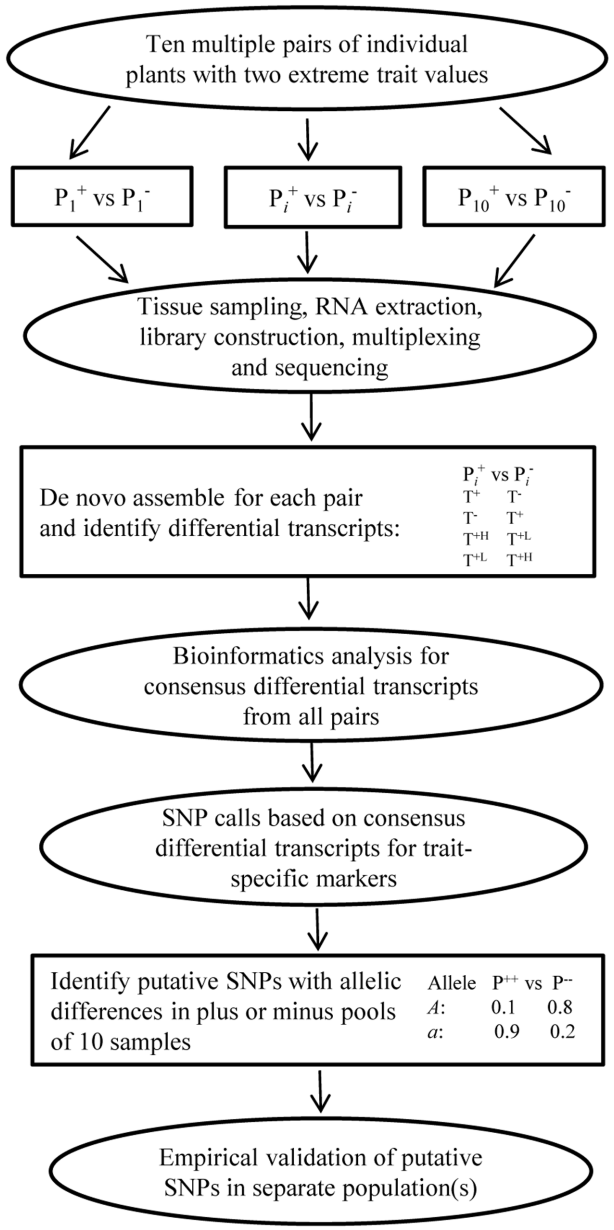

FIGURE 3 | A general procedure for developing plant FAST SNP markers through RNA-Seq. It involves multiple pairs of individual plants with extreme trait values, tissue sampling, RNA-Seq, bioinformatics analysis for differential transcripts for each pair and consensus different transcripts from all the pairs, SNP calls for specific trait markers, putative SNP identification and empirical validation. Identification of a differential transcript in a pair is made based on the presence (T^+) or absence (T^-) of a transcript or the difference in abundance (T^{+H} or T^{+L}) of the transcript detected in both plants. Putative SNPs are identified with a SNP filter based on the differences in allelic frequency between two trait-extreme sample sets.

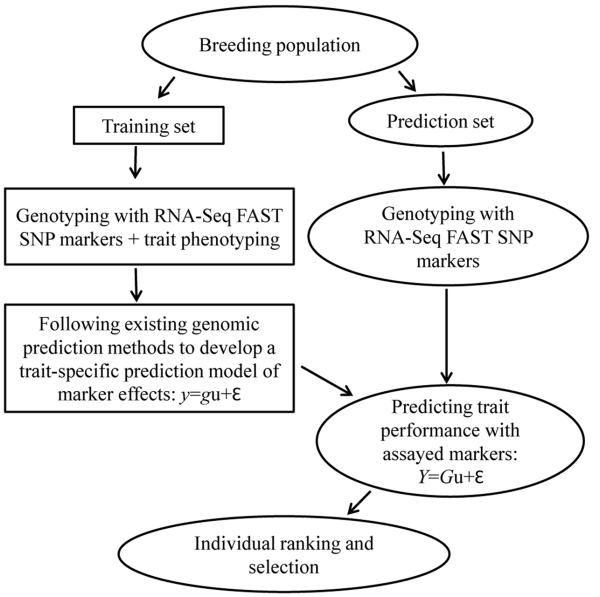

FIGURE 4 | A general procedure for trait performance prediction using RNA-Seq function-associated specific trait (FAST) SNP markers in a breeding population. It follows the same idea of genomic selection to have a training set of breeding materials for both genotyping and phenotyping to develop a trait-specific prediction model of marker effects. The developed prediction model will be applied to predict the trait performance with assayed RNA-Seq FAST SNP markers in the prediction set of breeding population for individual ranking and parental selection. Note that in the prediction equations, *y* or *Y* is the vector of trait values, u is the vector of marker effects, *g* or *G* is the genotype matrix and ε is the random error vector.

To facilitate the development of FAST SNP markers through RNA-Seq (**Figure 3**), we proposed a new procedure, following the principle of BSR-Seq developed by Liu S. et al. (2012) and the methods used by Salem et al. (2012) and Ramirez-Gonzalez et al. (2015). However, it differs from using multiple pairs of individual plants with extreme trait values. We reasoned that the use of multiple pairs should be more powerful than bulking, as it can address not only many issues associated with BSR-Seq such as replication, but also, more importantly, increase the power of identifying consensus differentially expressed transcripts and the accuracy of putative SNP discovery with allelic differential (**Figure 3**). We suggest 10 or more pairs for the effort, but the optimum pairs to be used remain to be empirically determined and they may vary with respect to trait and plant mating system.

However, issues are not lacking in the development and application of FAST SNP markers through RNA-Seq. A complete set of genome wide function-associated SNP markers can be effectively generated through exome capture technology (Mascher et al., 2013; Warr et al., 2015) simultaneously for many traits, but not necessarily specified for a trait of interest. RNA-Seq can produce FAST SNP markers, but these markers may not be comprehensive for the trait, as gene expressions have spatio-temporal specificity. Successful identification of SNP markers associated with casual genes will depend on the gene expression in related tissues over different development stages, so tissue selection and sampling for RNA collection are critical and may vary in effectiveness for different traits. Our proposal (**Figure 3**) did not consider the multiple developmental stages of RNA sampling to capture all expressed genes associated

with the trait performance and may miss some trans-regulatory genes (Druka et al., 2010). Some quantitative traits such as yield, maturity and disease resistance may need more research effort and proper experimental design to sample genes expressed at different developmental stages. Also, research efforts to develop FAST SNP markers for different traits may vary, as the genetic basis of different traits may differ. Uncertainty may also exist in the informativeness of RNA-Seq FAST SNP markers developed in one population for their applicability into other populations. Moreover, developing FAST SNP markers may be more complicated and more consideration may be needed in outcrossing, than selfing, plant species, as the genetic background for a trait in outcrossing plants is more heterogeneous.

In spite of these issues, FAST SNP markers for specific traits can be developed for plant breeding, either following our proposed procedure (**Figure 3**) or using existing methods such as eQTL analysis, GWAS, or those methods used in fish breeding (e.g., Salem et al., 2012). The good examples are the successful developments of FAST SNP markers through RNA-Seq in fish (e.g., see Salem et al., 2012; Ulloa et al., 2015) and 175 putative SNP markers associated with *Yr15*, a major disease resistance gene for wheat yellow rust (Ramirez-Gonzalez et al., 2015). Built upon these leading efforts, our derived alternative will provide an option for plant breeders with new procedures to develop and focus on a set of FAST SNP markers for trait prediction to enhance parental selection. Even with a small number of FAST SNP markers available for a given trait, our alternative is still applicable and may yield more informative parental selection than those with the aid of individual QTL markers in traditional MAS. Also, our synthesis is encouraging, as continuous search for better alternatives based on the other genetic characteristics of a quantitative trait is possible and may be more fruitful to provide much needed accuracy in marker-based prediction of a quantitative trait for molecular plant breeding.

CONCLUDING COMMENTS

Our search for a better marker-based prediction of trait performance through literature review and theoretical reasoning yielded an alternative to regular genome selection for individual trait prediction. More accurate trait predictions can be theoretically achieved through the development of FAST SNP markers with RNA-Seq technique and the application of these markers to genotype plants and to predict breeding values following existing genomic prediction methods in breeding populations. Further empirical investigation is needed to realize how much gain in trait prediction with respect to breeding efficiency could be achieved from the derived alternative in a plant breeding program. The derived alternative may be questioned for its breeding efficiency in multiple-traits breeding, as function-associated SNP markers unspecified for specific traits could be more efficiently developed from exome capture technology than the proposed FAST SNP markers. However, our synthesis is encouraging, as continuous search for better alternatives based on the other genetic characteristics of a quantitative trait is possible and may yield more accurate trait prediction for molecular plant breeding.

ETHICS STATEMENT

The writing process of this manuscript complies with the current laws of Canada.

AUTHOR CONTRIBUTIONS

Y-BF conceived of the research, conducted the literature review, performed the computer simulation and wrote the paper. M-HY conducted the literature review, performed the computer simulation and revised the paper. FZ conducted the literature review and wrote the paper. BB conducted the literature review and revised the paper.

FUNDING

This work was supported by an A-Base research project of Agriculture and Agri-Food Canada to Y-BF and National Science and Technology Program of China (2012BAD21B03), the National Natural Science Foundation of China (31670678) and the China Scholarship Council Postdoctoral Abroad Grant to M-HY.

ACKNOWLEDGMENT

We would like to thank Dr. Nicholas Tinker for his helpful comments on the earlier version of the manuscript and Dr. Qijian Song for his assistance with the acquisition of SoySNP50K data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.01182/full#supplementary-material

REFERENCES

Andersen, J. R., and Lübberstedt, T. (2003). Functional markers in plants. *Trends Plant Sci.* 8, 554–560. doi: 10.1016/j.tplants.2003.09.010

Arruda, M. P., Lipka, A. E., Brown, P. J., Krill, A. M., Thurber, C., Brown-Guedira, G., et al. (2016). Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* 36:84. doi: 10.1007/s11032-016-0508-5

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Sci.* 242, 23–36. doi: 10.1016/j.plantsci.2015.08.021

Beavis, W. D. (1998). "QTL analyses: power, precision, and accuracy," in *Molecular Dissection of Complex Traits*, ed. A. H. Paterson (Boca Raton, FL: CRC press), 145–162.

Bernardo, R. (2016). Bandwagons I, too, have known. *Theor. Appl. Genet.* 129, 2323–2332. doi: 10.1007/s00122-016-2772-5

Bernardo, R., and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690

Bhardwaj, A. R., Joshi, G., Kukreja, B., Malik, V., Arora, P., Pandey, R., et al. (2015). Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop *Brassica juncea*. *BMC Plant Biol.* 15:9. doi: 10.1186/s12870-014-0405-1

Boopathi, N. M. (ed.). (2013). "Success Stories in MAS," in *Genetic Mapping and Marker Assisted Selection: Basics, Practice and Benefits*, (New York, NY: Springer), 187–192. doi: 10.1007/978-81-322-0958-4_9

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

Brumlop, S., and Finckh, M. R. (2011). *Applications and Potentials of Marker Assisted Selection (MAS) in Plant Breeding*. Bonn: Federal Agency for Nature Conservation.

Chopra, R., Burow, G., Hayes, C., Emendack, Y., Xin, Z., and Burke, J. (2015). Transcriptome profiling and validation of gene based single nucleotide polymorphisms (SNPs) in sorghum genotypes with contrasting responses to cold stress. *BMC Genomics* 16:1040. doi: 10.1186/s12864-015-2268-8

Collard, B. C., and Mackill, D. J. (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. doi: 10.1098/rstb.2007.2170

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-Seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19, 592–601. doi: 10.1016/j.tplants.2014.05.006

Druka, A., Potokina, E., Luo, Z., Jiang, N., Chen, X., Kearsey, M., et al. (2010). Expression quantitative trait loci analysis in plants. *Plant Biotechnol. J.* 8, 10–27. doi: 10.1111/j.1467-7652.2009.00460.x

Edwards, S. M., Sørensen, I. F., Sarup, P., Mackay, T. F., and Sørensen, P. (2016). Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *Drosophila melanogaster*. *Genetics* 203, 1871–1883. doi: 10.1534/genetics.116.187161

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., et al. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 11, 446–450. doi: 10.1038/nrg2809

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Fernando, R. L., Habier, D., Stricker, C., Dekkers, J. C. M., and Totir, L. R. (2007). Genomic selection. *Acta Agric. Scand. A* 57, 192–195. doi: 10.1080/09064700801959395

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109

Fullwood, M. J., Wei, C.-L., Liu, E. T., and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* 19, 521–532. doi: 10.1101/gr.074906.107

Garber, M., Grabherr, M. G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-Seq. *Nat. Methods* 8, 469–477. doi: 10.1038/nmeth.1613

Grover, A., and Sharma, P. (2016). Development and use of molecular markers: past and present. *Crit. Rev. Biotechnol.* 36, 290–302. doi: 10.3109/07388551.2014.959891

Han, Y., Gao, S., Muegge, K., Zhang, W., and Zhou, B. (2015). Advanced applications of RNA sequencing and challenges. *Bioinform. Biol. Insights* 9(Suppl. 1), 29–46. doi: 10.4137/BBI.S28991

He, F., Liu, Q., Zheng, L., Cui, Y., Shen, Z., and Zheng, L. (2015). RNA-Seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Front. Plant Sci.* 6:1136. doi: 10.3389/fpls.2015.01136

Heffner, E. L., Lorenz, A. J., Jannink, J.-L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662

Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M., et al. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 54, 1476–1488. doi: 10.2135/cropsci2013.03.0195

Holland, J. B. (2004). "Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities," in *Proceedings for the 4th International Crop Science Congress: New Directions for a Diverse Planet*, eds T. Fischer, N. Turner, J. Angus, L. McIntyre, M. Robertson, A. Borrell, et al. Brisbane, QLD.

Hu, R., Zhu, X., Xiang, S., Zhan, Y., Zhu, M., Yin, H., et al. (2016). Comparative transcriptome analysis revealed the genotype specific cold response mechanism in tobacco. *Biochem. Biophys. Res. Commun.* 469, 535–541. doi: 10.1016/j.bbrc.2015.12.040

Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Van Sanford, D., Costa, J., et al. (2016). Genomic selection for wheat traits and trait stability. *Theor. Appl. Genet.* 129, 1697–1710. doi: 10.1007/s00122-016-2733-z

Iyer-Pascuzzi, A. S., and McCouch, S. R. (2007). Functional markers for xa5-mediated resistance in rice (*Oryza sativa* L.). *Mol. Breed.* 19, 291–296. doi: 10.1007/s11032-006-9055-9

Jan, H. U., Abbadi, A., Lücke, S., Nichols, R. A., and Snowdon, R. J. (2016). Genomic prediction of testcross performance in canola (*Brassica napus*). *PLoS ONE* 11:e0147769. doi: 10.1371/journal.pone.0147769

Jannink, J.-L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9, 166–177. doi: 10.1093/bfgp/elq001

Jiang, G.-L. (2013). "Molecular markers and marker-assisted breeding in plants," in *Plant Breeding from Laboratories to Fields*, ed. S. B. Andersen (Rijeka: InTech), 45–83.

Jiao, Y., Tausta, S. L., Gandotra, N., Sun, N., Liu, T., Clay, N. K., et al. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.* 41, 258–263. doi: 10.1038/ng.282

Jones, S. I., and Vodkin, L. O. (2013). Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS ONE* 8:e59270. doi: 10.1371/journal.pone.0059270

Kakumanu, A., Ambavaram, M. M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., et al. (2012). Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.* 160, 846–867. doi: 10.1104/pp.112.200444

Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N., and Liu, Z. (2013). Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell* 25, 1960–1978. doi: 10.1105/tpc.113.111732

Kim, S., Kim, C.-W., Park, M., and Choi, D. (2015). Identification of candidate genes associated with fertility restoration of cytoplasmic male-sterility in onion (*Allium cepa* L.) using a combination of bulked segregant analysis and RNA-Seq. *Theor. Appl. Genet.* 128, 2289–2299. doi: 10.1007/s00122-015-2584-z

Kizilkaya, K., Fernando, R. L., and Garrick, D. J. (2010). Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J. Anim. Sci.* 88, 544–551. doi: 10.2527/jas.2009-2064

Kong, L. A., Wu, D. Q., Huang, W. K., Peng, H., Wang, G. F., Cui, J. K., et al. (2015). Large-scale identification of wheat genes resistant to cereal cyst nematode *Heterodera avenae* using comparative transcriptomic analysis. *BMC Genomics* 16:801. doi: 10.1186/s12864-015-2037-8

Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.

Lau, W. C., Rafii, M. Y., Ismail, M. R., Puteh, A., Latif, M. A., and Ramli, A. (2015). Review of functional markers for improving cooking, eating, and the nutritional qualities of rice. *Front. Plant Sci.* 6:832. doi: 10.3389/fpls.2015.00832

Lee, Y.-P., Cho, Y., and Kim, S. (2014). A high-resolution linkage map of the Rfd1, a restorer-of-fertility locus for cytoplasmic male sterility in radish (*Raphanus sativus* L.) produced by a combination of bulked segregant analysis and RNA-Seq. *Theor. Appl. Genet.* 127, 2243–2252. doi: 10.1007/s00122-014-2376-x

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S. L., et al. (2010). The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* 42, 1060–1067. doi: 10.1038/ng.703

Liu, G., Zhao, Y., Gowda, M., Longin, C. F. H., Reif, J. C., and Mette, M. F. (2016). Predicting hybrid performances for quality traits through genomic-assisted approaches in central European wheat. *PLoS ONE* 11:e0158635. doi: 10.1371/journal.pone.0158635

Liu, S., Yeh, C.-T., Tang, H. M., Nettleton, D., and Schnable, P. S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS ONE* 7:e36406. doi: 10.1371/journal.pone.0036406

Liu, Y., He, Z., Appels, R., and Xia, X. (2012). Functional markers in wheat: current status and future prospects. *Theor. Appl. Genet.* 125, 1–10. doi: 10.1007/s00122-012-1829-3

Liu, Y.-H., Xu, Y., Zhang, M., Sze, S.-H., Smith, C. W., Xu, S., et al. (2017). "Development of a gene-based breeding system in cotton: a new method powerful and efficient for enhanced fiber quality breeding," in *Proceedings of the Plant and Animal Genome Conference XXV, 14-18 January 2017*, San Diego, CA.

Lopez-Maestre, H., Brinza, L., Marchet, C., Kielbassa, J., Bastien, S., Boutigny, M., et al. (2016). SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Res.* 44:e148. doi: 10.1093/nar/gkw655

Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 33, 303–339. doi: 10.1146/annurev.genet.35.102401.090633

Mantegazza, O., Gregis, V., Chiara, M., Selva, C., Leo, G., Horner, D. S., et al. (2014). Gene coexpression patterns during early development of the native *Arabidopsis* reproductive meristem: novel candidate developmental regulators and patterns of functional redundancy. *Plant J.* 79, 861–877. doi: 10.1111/tpj.12585

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18, 1509–1517. doi: 10.1101/gr.079558.108

Martínez-López, L. A., Ochoa-Alejo, N., and Martínez, O. (2014). Dynamics of the chili pepper transcriptome during fruit development. *BMC Genomics* 15:143. doi: 10.1186/1471-2164-15-143

Mascher, M., Richmond, T. A., Gerhardt, D. J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* 76, 494–505. doi: 10.1111/tpj.12294

Massman, J. M., Jung, H.-J. G., and Bernardo, R. (2013). Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Sci.* 53, 58–66. doi: 10.2135/cropsci2012.02.0112

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626

Meuwissen, T., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease resistance genes by BSA: a rapid method to detect markers in specific genome regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828

Moose, S. P., and Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol.* 147, 969–977. doi: 10.1104/pp.108.118232

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349. doi: 10.1126/science.1158441

Ozsolak, F., and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98. doi: 10.1038/nrg2934

Pace, J., Gardner, C., Romay, C., Ganapathysubramanian, B., and Lübberstedt, T. (2015). Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). *BMC Genomics* 16:47. doi: 10.1186/s12864-015-1226-9

Pérez-Rodriguez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical oackage. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442

Postnikova, O. A., Shao, J., and Nemchinov, L. G. (2013). Analysis of the alfalfa root transcriptome in response to salinity stress. *Plant Cell Physiol.* 54, 1041–1055. doi: 10.1093/pcp/pct056

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rabier, C.-E., Barre, P., Asp, T., Charmet, G., and Mangin, B. (2016). On the accuracy of genomic selection. *PLoS ONE* 11:e0156086. doi: 10.1371/journal.pone.0156086

Ramirez-Gonzalez, R. H., Segovia, V., Bird, N., Fenwick, P., Holdgate, S., Berry, S., et al. (2015). RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnol. J.* 13, 613–624. doi: 10.1111/pbi.12281

Randhawa, H. S., Asif, M., Pozniak, C., Clarke, J. M., Graf, R. J., Fox, S. L., et al. (2013). Application of molecular markers to wheat breeding in Canada. *Plant Breed.* 132, 458–471. doi: 10.1111/pbr.12057

Riedelsheimer, C., Endelman, J. B., Stange, M., Sorrells, M. E., Jannink, J.-L., and Melchinger, A. E. (2013). Genomic predictability of interconnected biparental maize populations. *Genetics* 194, 493–503. doi: 10.1534/genetics.113.150227/-/DC1

Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8, 552–559.

Schmidt, M., Kollers, S., Maasberg-Prelle, A., Großer, J., Schinkel, B., Tomerius, A., et al. (2016). Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. *Theor. Appl. Genet.* 129, 203–213. doi: 10.1007/s00122-015-2639-1

Severin, A. J., Peiffer, G. A., Xu, W. W., Hyten, D. L., Bucciarelli, B., O'Rourke, J. A., et al. (2010). An integrative approach to genomic introgression mapping. *Plant Physiol.* 154, 3–12. doi: 10.1104/pp.110.158949

Sinha, S., Raxwal, V. K., Joshi, B., Jagannath, A., Katiyar-Agarwal, S., Goel, S., et al. (2015). *De novo* transcriptome profiling of cold-stressed siliques during pod filling stages in Indian mustard (*Brassica juncea* L.). *Front. Plant Sci.* 6:932. doi: 10.3389/fpls.2015.00932

Socquet-Juglard, D., Kamber, T., Pothier, J. F., Christen, D., Gessler, C., Duffy, B., et al. (2013). Comparative RNA-Seq analysis of early-infected peach leaves by the invasive phytopathogen *Xanthomonas arboricola* pv. *pruni*. *PLoS ONE* 8:e54196. doi: 10.1371/journal.pone.0054196

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3* 5, 1999–2006. doi: 10.1534/g3.115.019000

Sorrells, M. E. (2015). "Genomic selection in plants: empirical results and implications for wheat breeding," in *Advances in Wheat Genetics: From Genome to Field*, eds Y. Ogihara, S. Takumi, and H. Handa (Dordrecht: Springer), 401–409. doi: 10.1007/978-4-431-55675-6_45

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11:e1004982. doi: 10.1371/journal.pgen.1004982

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J.-L., et al. (2016). Genome-wide prediction models that incorporate *de novo* GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116, 395–408. doi: 10.1038/hdy.2015.113

Thavamanikumar, S., Dolferus, R., and Thumma, B. R. (2015). Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. *G3* 5, 1991–1998. doi: 10.1534/g3.115.019745

Thoday, J. (1961). Location of polygenes. *Nature* 191, 368–370. doi: 10.1038/191368a0

Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C.-C., Febrer, M., and Uauy, C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12:14. doi: 10.1186/1471-2229-12-14

Ulloa, P. E., Rincón, G., Islas-Trejo, A., Araneda, C., Iturra, P., Neira, R., et al. (2015). RNA sequencing to study gene expression and SNP variations associated with growth in Zebrafish fed a plant protein-based diet. *Mar. Biotechnol.* 17, 353–363. doi: 10.1007/s10126-015-9624-1

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630. doi: 10.1016/j.tplants.2005.10.004

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006

Varshney, R. K., Terauchi, R., and McCouch, S. R. (2014). Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12:e1001883. doi: 10.1371/journal.pbio.1001883

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. *G3* 5, 1543–1550. doi: 10.1534/g3.115.018564

Wei, W., Qi, X., Wang, L., Zhang, Y., Hua, W., Li, D., et al. (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* 12:451. doi: 10.1186/1471-2164-12-451

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239–1243. doi: 10.1038/nature07002

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J.-L., Sorrells, M. E., et al. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3* 2, 1427–1436. doi: 10.1534/g3.112.003699

Xu, Y. (2010). *Molecular Plant Breeding*. Wallingford: CAB International. doi: 10.1079/9781845933920.0000

Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48, 391–407. doi: 10.2135/cropsci2007.04.0191

Yang, H., Li, C., Lam, H.-M., Clements, J., Yan, G., and Zhao, S. (2015). Sequencing consolidates molecular markers with plant breeding practice. *Theor. Appl. Genet.* 128, 779–795. doi: 10.1007/s00122-015-2499-8

Yang, S. S., Tu, Z. J., Cheung, F., Xu, W. W., Lamb, J. F., Jung, H. J. G., et al. (2011). Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. *BMC Genomics* 12:199. doi: 10.1186/1471-2164-12-199

Zhang, M., Cui, Y., Liu, Y.-H., Xu, W., Sze, S.-H., Xu, S., et al. (2017). "Gene-based breeding in maize: grain yield breeding by effectively using the genes controlling the targeted trait," in *Proceedings of the Plant and Animal Genome Conference XXV, 14-18 January 2017*, San Diego, CA.

Zhong, S., Dekkers, J. C., Fernando, R. L., and Jannink, J.-L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182, 355–364. doi: 10.1534/genetics.108.098277

Zhou, Q., Luo, D., Ma, L., Xie, W., Wang, Y., Wang, Y., et al. (2016). Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. *Sci. Rep.* 6:20549. doi: 10.1038/srep20549

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.