Exploring and Harnessing Haplotype Diversity to Improve Yield Stability in Crops

Lunwen Qian[1,2], Lee T. Hickey[3], Andreas Stahl[2], Christian R. Werner[2], Ben Hayes[3], Rod J. Snowdon[2] and Kai P. Voss-Fels[2,3]*

[1] Collaborative Innovation Center of Grain and Oil Crops in South China, Hunan Agricultural University, Changsha, China, [2] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Giessen, Germany, [3] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia

In order to meet future food, feed, fiber, and bioenergy demands, global yields of all major crops need to be increased significantly. At the same time, the increasing frequency of extreme weather events such as heat and drought necessitates improvements in the environmental resilience of modern crop cultivars. Achieving sustainably increase yields implies rapid improvement of quantitative traits with a very complex genetic architecture and strong environmental interaction. Latest advances in genome analysis technologies today provide molecular information at an ultrahigh resolution, revolutionizing crop genomic research, and paving the way for advanced quantitative genetic approaches. These include highly detailed assessment of population structure and genotypic diversity, facilitating the identification of selective sweeps and signatures of directional selection, dissection of genetic variants that underlie important agronomic traits, and genomic selection (GS) strategies that not only consider major-effect genes. Single-nucleotide polymorphism (SNP) markers today represent the genotyping system of choice for crop genetic studies because they occur abundantly in plant genomes and are easy to detect. SNPs are typically biallelic, however, hence their information content compared to multiallelic markers is low, limiting the resolution at which SNP–trait relationships can be delineated. An efficient way to overcome this limitation is to construct haplotypes based on linkage disequilibrium, one of the most important features influencing genetic analyses of crop genomes. Here, we give an overview of the latest advances in genomics-based haplotype analyses in crops, highlighting their importance in the context of polyploidy and genome evolution, linkage drag, and co-selection. We provide examples of how haplotype analyses can complement well-established quantitative genetics frameworks, such as quantitative trait analysis and GS, ultimately providing an effective tool to equip modern crops with environment-tailored characteristics.

Keywords: crop genomics, genomics-assisted breeding, haplotype analysis, SNP haplotype, climate change

INTRODUCTION

Unstable environments and increasing climatic fluctuations like extreme heat and drought events have a severe impact on global crop production (Lesk et al., 2016). In order to meet the food demand of a rapidly growing world population, yields of important commodity crops need to be increased by almost 40% by the middle of this century (Tester and Langridge, 2010). Breeding

of cultivars that consistently achieve high yields even in inconsistent environments is a highly challenging task for the global plant breeding community. The enormous recent advances in DNA sequencing and genotyping technologies may help to overcome this challenge. Low-cost genotyping tools that can capture sequence variation at ultra-high resolution are now available for all agronomically important plant species (Huang and Han, 2014) and commercial genotyping platforms that can generate thousands or millions of data points per genotyping experiment are now in standard use in genomic research (Bassi et al., 2016). These powerful tools provide an effective means for crop genetic research studies (Ganal et al., 2012), for example enabling the assessment of population structure and genotypic diversity on a genome-wide, subgenome-wide, or chromosome-wide scale, facilitating identification of selective sweeps and signatures of directional selection, or providing a basis for genomic selection (GS) or prediction of hybrid performance.

Targeting genetic variants associated with agronomic traits and identifying important underlying candidate genes have become a key area in crop genetic research (Voss-Fels and Snowdon, 2016). Because the majority of high value traits are quantitatively inherited, in recent decades considerable focus has been placed on mapping and characterization of quantitative trait loci (QTL). Quantitative trait loci mapping remains a powerful method to identify genes with major effects, however, for highly complex traits its usefulness is limited because bi-parental mapping populations lack genetic diversity [possessing only two alleles at any given single-nucleotide polymorphism (SNP) locus] and provide only a low mapping resolution due to a lack of recombination events [Brachi et al., 2011; for further methodical details of QTL mapping see Collard et al. (2005)]. Along with the advances in genotyping technologies, huge progress has been achieved in the development of statistical solutions to handle vast amounts of genomic data. Although high-density data provide little additional information in a bi-parental context, genome-wide association studies (GWASs), also known as linkage disequilibrium (LD) mapping using high-density molecular marker information have become a widely used and powerful tool for the genetic dissection of complex traits at an extremely high resolution [see Zhu et al. (2008) for status and prospects of GWAS in plants]. Simultaneously, high-density marker data in genetically diverse plant populations can also provide valuable information about structural diversity of crop genomes.

Typically, high-throughput SNP marker arrays or SNPs detected by reduced-representation DNA sequencing are the genotyping methods of choice for crop genomic investigations. However, in the context of GWAS, SNPs have the major limitation that they normally only provide biallelic information at any individual locus. Further, assuming the rare allele model (Gibson, 2012), a significant fraction of the genetic variance for a given quantitative trait is explained by rare alleles (Maher, 2008) that are often not adequately represented on commercial genotyping arrays and are therefore difficult to detect (Wray et al., 2013; Voss-Fels and Snowdon, 2016). On the other hand, as a consequence of natural and artificial directional selection, most important crop species show abundant levels of LD (Gupta

et al., 2005). Thus, QTL identified by GWAS is unlikely to represent true causal molecular variants, but rather loci that are in LD with a gene or a regulatory element that affects the trait of interest (Mackay et al., 2009; Korte and Farlow, 2013). An effective approach to overcome the biallelic limitations of SNPs and to increase the allelic resolution of candidate genomic regions is to employ haplotypes, the specific combination of jointly inherited nucleotides or markers from polymorphic sites in the same chromosome segment (Stephens et al., 2001; Lu et al., 2010).

Here, we review the latest advances in haplotype analyses in crops, highlighting their potential for the improvement of yield stability in the face of increasing abiotic constraints to crop production. We discuss technical aspects of haplotype description using high-throughput genotyping platforms and how these can be useful for genomics-based crop improvement. We also discuss the importance of haplotypes in the context of polyploidy, co-selection, and linkage drag, and give future perspectives on how they can be incorporated in established and emerging quantitative genetics frameworks, such as quantitative trait analysis and GS.

DEFINITION AND FORMATION OF HAPLOTYPES

Bernardo (2010) describes a haplotype as "two or more SNP alleles that tend to be inherited as a unit." However, the construction of haplotypes based on empirical marker data is not trivial (Korte and Farlow, 2013). Wall and Pritchard (2003) describe key methods to define haplotypes by (i) taking haplotype diversity of a given chromosomal segment as a basis for haplotype assignment or (ii) by using approaches that assign haplotypes based on pairwise LD between markers showing little or no evidence of historical recombination and a joint inheritance in the same a chromosomal block, typically measured as r^2 (Pritchard and Przeworski, 2001). Linkage disequilibrium-based approaches are advantageous because (i) they focus directly on the detection of historical recombination in a given population, via identification of haplotype blocks, (ii) they are easily applicable in diploid data in which the haplotype phase is unknown, and (iii) LD coefficients are easy to visualize. There are numerous factors affecting LD, such as the propagation type and the rate of inbreeding in a given species, the size of the analyzed population, the extent of population stratification and subdivision, genetic drift, recombination frequency, mutation rate, or the strength and type of selection on a given chromosomal fragment (Gupta et al., 2005).

Historically, directional selection for genes or alleles conferring favorable characteristics (Lande and Arnold, 1983) has played a major role in the formation of signatures of selection in all globally important crop species, and numerous examples have been reported for example in rice (*Oryza sativa* ssp. *japonica*, ssp. *indica* and their progenitor *O. rufipogon*; Flowers et al., 2012), maize (*Zea mays* ssp. *mexicana*; Hufford et al., 2013), wheat (*Triticum aestivum* L.; Voss-Fels et al., 2015), sorghum (*Sorghum bicolor* L. ssp. *bicolor*; Mace et al., 2013),

cassava (*Manihot esculenta* L.; Wang et al., 2014), or rapeseed (*Brassica napus* L.; Qian et al., 2014). Signatures of selection, also referred to as selective sweeps or conserved haplotype blocks, typically contain numerous genes and it is likely that their expression is jointly controlled by multiple regulatory alleles, resulting in trait correlations that are more likely to be caused by linked genes rather than pleiotropy (Qian et al., 2016a). Targeting these selection hotspots, unraveling their underlying effects on relevant traits, and increasing recombination by genomics-driven crossing strategies could greatly improve our understanding of quantitative trait complexes and improve modern varieties with tailored agronomical characteristics and better adaptation to extreme climates.

GWAS-BASED HAPLOTYPE ANALYSIS

Since the advent of highly efficient genotyping technologies, GWAS has become the method of choice for the dissection of complex traits in humans, animals, and plants. Numerous examples have demonstrated the power of GWAS for dissection of agronomically relevant quantitative traits in all major crops (Huang and Han, 2014). The most frequently used markers for GWAS are SNPs and commercial genotyping platforms can rapidly provide robust and cheap genome-wide marker data for all important crop species. Single-nucleotide polymorphism markers are typically biallelic and show a much lower mutational rate and lower per-marker information content than other previously used, single-plex marker systems such as simple sequence repeat (SSR) markers (Würschum et al., 2013). Further, it is well documented that significant SNPs in GWAS experiments are unlikely to represent causal molecular variants due to inherent ascertainment biases of SNP arrays and the fact that extreme phenotypes are frequently caused by rare alleles (Wray et al., 2013). Due to genetic heterogeneity, non-causative markers in LD with a true molecular variant can show a more significant statistical marker–trait association than the causal variant itself (Platt et al., 2010; Korte and Farlow, 2013). Nevertheless, such false-positive synthetic associations can still be a valuable predictor of the phenotype if LD with the causal variant is strong. However, in most cases the resolution at which complex relationships between quantitative phenotypes can be depicted with biallelic SNPs is limited. An efficient way to overcome this problem and to increase allelic variation along sites in the genome that are significantly associated with phenotypes is by analysis of haplotype blocks in target regions exhibiting trait associations, as shown in **Figure 1**. Here, candidate regions are identified by GWAS or QTL mapping and local LD surrounding the identified markers is analyzed (**Figure 1A**). Based on LD criteria, marker-based haplotypes consisting of multiple markers are built and their allelic variation is compared to phenotypic measures for the traits of interest. Subsequently, test lines carrying favorable alleles can be screened and selected for crossings using molecular markers in order to accumulate haplotype variants of interest in breeding lines (**Figures 1A,B**).

Stephens et al. (2001) showed that haplotype blocks combining two or more SNPs in strong LD show a highly increased heterozygosity due to their inherent multiallelic nature and are thus much more informative than single biallelic SNPs. That study revealed a higher number of haplotype variants than SNP markers, indicating recombination and recurrent mutation events within and among the genes in the haplotype. In recent years, haplotype analyses on the basis of genome-wide association scans using SNP markers have successfully been applied in a wide range of crop species to identify important candidate genes for various traits. In maize (*Z. mays* L.), for example, Yang et al. (2013) confirmed via GWAS that a CO, CO-LIKE and TOC1 (CCT) domain-containing protein is a main factor for photoperiod response. After locating the candidate gene in a highly conserved selective sweep with pronounced levels of LD, they were able to identify different haplotype variants in the promoter region with highly specific corresponding phenotypes. They further showed that a transposable element in the gene promoter region dramatically reduced flowering time, a phenotype which likely accelerated the worldwide dispersal of maize production into regions characterized by long-day environments. In rice (*O. sativa* L.), Si et al. (2016) mapped a highly significant SNP in a gene for grain size via GWAS. By resequencing the candidate region they subsequently identified two major haplotypes that corresponded to small or long grain phenotypes, respectively. Another recent rice genetics study used ultra-dense SNP marker information from whole-genome resequencing data of 176 rice varieties to identify novel genes with major effects on agronomically important traits (Yano et al., 2016). Deploying information from almost 430,000 polymorphic SNPs and 68,000 insertion–deletion markers, they designed a four-step experimental process for rapid gene identification including (i) significance testing for marker signals using GWAS, (ii) definition of candidate haplotype regions with significant markers on the basis of LD, (iii) extraction of candidate genes based on annotation information, and (iv) validation of functional haplotype alleles through introgressions. In wheat (*T. aestivum* L.), Jiang et al. (2015) analyzed molecular variants of three main cell wall invertase (CWI) genes that were found to have a significant effect on grain yield. By performing haplotype analyses based on SNP and insertion–deletion markers in the gene coding sequences they demonstrated that the haplotype variants with positive effects on grain yield were most frequent in modern cultivars and were strongly positively selected in six major global wheat growing regions. Using haplotype network analyses, Jiang et al. (2015) demonstrated that strong allelic selection in the course of breeding, domestication, and polyploidization has narrowed the genetic diversity of three major wheat CWI genes.

MULTI-TRAIT HAPLOTYPES AND SELECTIVE SWEEPS

Crop breeders have been tremendously successful in combining beneficial loci within germplasm, resulting in an enhancement of yield throughout the history of modern agriculture (Mickelbart et al., 2015). Examples demonstrating the effectiveness of
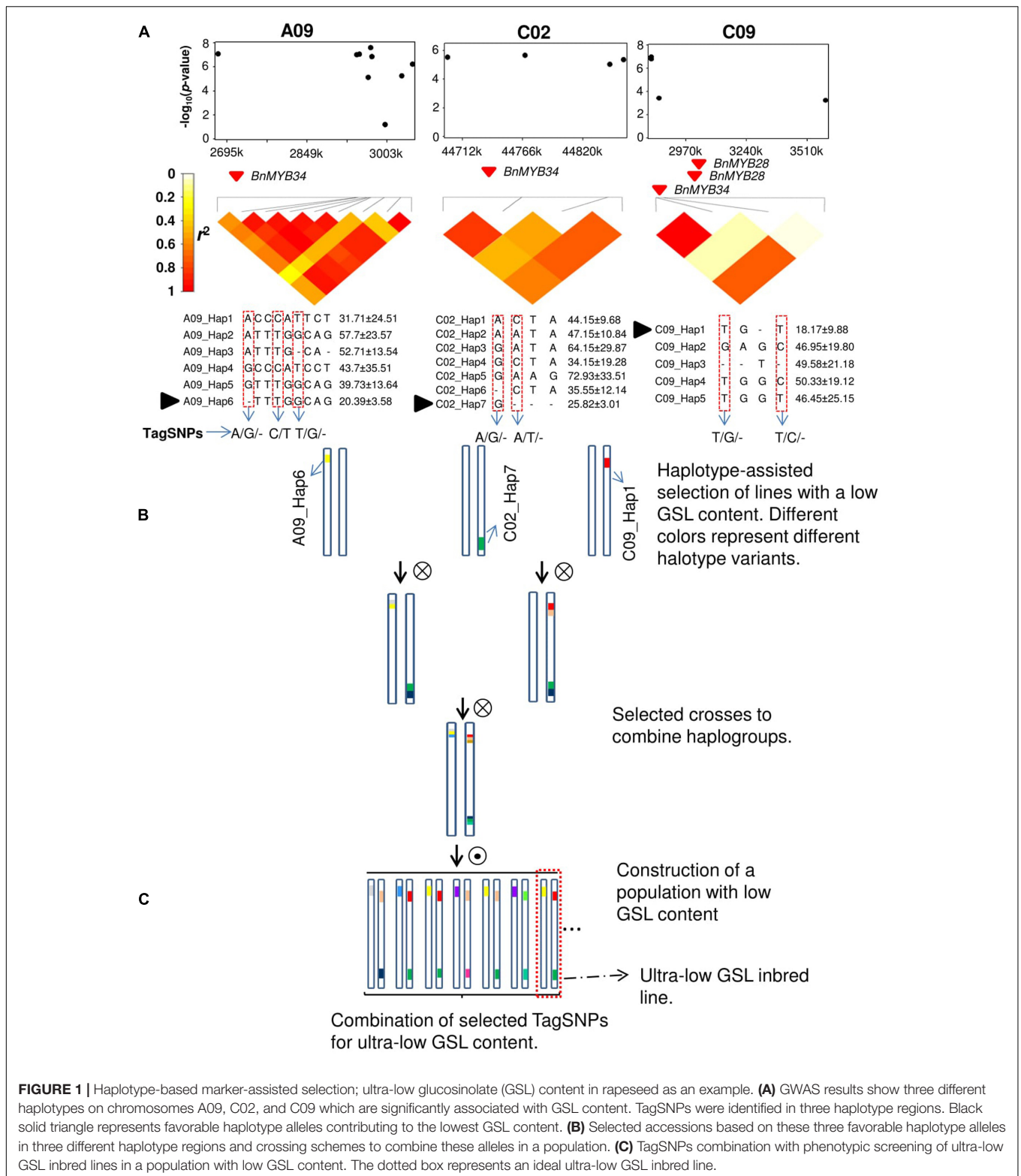
FIGURE 1 | Haplotype-based marker-assisted selection; ultra-low glucosinolate (GSL) content in rapeseed as an example. **(A)** GWAS results show three different haplotypes on chromosomes A09, C02, and C09 which are significantly associated with GSL content. TagSNPs were identified in three haplotype regions. Black solid triangle represents favorable haplotype alleles contributing to the lowest GSL content. **(B)** Selected accessions based on these three favorable haplotype alleles in three different haplotype regions and crossing schemes to combine these alleles in a population. **(C)** TagSNPs combination with phenotypic screening of ultra-low GSL inbred lines in a population with low GSL content. The dotted box represents an ideal ultra-low GSL inbred line.

genetic-driven yield increases have been given in wheat (*T. aestivum* L.) (Austin et al., 1980; Cormier et al., 2013; Laidig et al., 2017a); maize (*Z. mays* L.) (Tollenaar, 1989), rye (*Secale cereale*) (Laidig et al., 2017b), rice (*O. sativa* L.) (Zhu et al., 2016), triticale (× *Triticosecale*) (Losert et al., 2017), and oilseed rape (*B. napus* L.) (Kessel et al., 2012; Stahl et al., 2017). Paradoxically, however, the tremendous success of selective breeding has also caused a drastic reduction of

genetic diversity in many elite crop germplasm pools, limiting the potential for future genetic gain. It is well documented that strong directional selection, either naturally through local adaptation or artificially through breeding, has resulted in pronounced LD patterns across crop genomes. There are numerous examples of selective sweeps and signatures of selection in important crop species, visible as highly conserved haplotype blocks with low levels of recombination (Voss-Fels and Snowdon, 2016).

This is especially critical for future breeding success, as a reduction in genetic diversity ultimately leads to a decrease of long-term breeding progress. To secure continuous breeding success, constant replenishment of genetic diversity is essential (Govindaraj et al., 2015). On the other hand, maintaining a sufficient effective population sizes while minimizing negative effects through linkage drag is challenging (Cowling, 2007; Cowling and Léon, 2013; Griggs et al., 2014). In general, crop gene banks represent a rich diversity of haplotype variants, typically much broader than in commercial breeding programs (Mackay et al., 2016; Riaz et al., 2017). A classic example of the conflict between breeding success and gene pool diversity is seen in oilseed rape/canola (*B. napus* L.), today the second most important oilseed crop in the world. The recent allopolyploid *B. napus* arose only recently from a small number of hybridization events between *B. rapa* and *B. oleracea*, and wild forms do not exist. Today's elite breeding pool was narrowed even further by introduction of the essential seed quality traits "low erucic acid" and "glucosinolate content" from just two trait donors, causing another extreme genetic bottleneck (Hasan et al., 2006; Qian et al., 2014). This resulted in very large, extremely conserved chromosomal blocks in all major ecogeographic gene pools of canola and oilseed rape (Qian et al., 2014), accompanied by low effective population sizes in elite germplasm collections (e.g., Cowling, 2007) and hidden effects of linkage drag. For example, Qian et al. (2016b) discovered a conserved haplotype block in which a gene associated with reduced glucosinolate content caused linkage drag among variants of functionally independent chlorophyll-related genes. This resulted in phenotypic correlations between seed glucosinolate and leaf chlorophyll content that appeared to have negative effects on photosynthesis and oil accumulation in some stress environments. In other species, several mapping studies have shown pleiotropic actions of identified QTL on multiple crop characteristics, for example in rice (*O. sativa* L.) (Bai et al., 2010) and maize (*Z. mays* L.) (Bomblies and Doebley, 2006). More precise identification of gene variants in haplotype blocks based on growing quantities of pangenomic sequence data will provide a basis for more precise selection of environmentally resilient cultivars, for example using high-density molecular markers. The higher allelic resolution of identified candidate regions will sustain a more accurate delineation of complex marker–trait correlations and help to significantly improve resilience of future cultivars to extreme weather conditions, such as heat and drought (Lesk et al., 2016). Whereas adaptation to abiotic constrains was previously mainly achieved by combination of anonymous loci and relatively inaccurate

phenotypic selection, the identification of genetic determinants will become increasingly important in future postgenomic approaches aiming to speed up breeding progress by targeted selection (Mickelbart et al., 2015).

HOMEOLOGOUS GENOME REGIONS AND HAPLOTYPE INTERACTIONS

Many globally important crops are recent or ancestral allopolyploids, yet the direct effect of polyploidy on evolutionary success in recent species is largely unknown. In the formation of many important polyploids, such as wheat (*T. aestivum* L), cotton (*Gossypium arboreum* L.), tobacco (*Nicotiana tabacum* L.), or oat (*Avena sativa* L.), hybridizations between related species merged genomic information from two or more divergent relatives into a single nucleus. Immediate or subsequent genome restructuring and gene loss (Chalhoub et al., 2014; Samans et al., 2017) resulted in divergence from expected genomic additivity with strong consequences for crop evolution and adaptation (Soltis et al., 2009; Mason and Snowdon, 2016). There is growing evidence that duplications and/or deletions of chromosome segments contributed to genetic variation which modulates expression of homoeologous genes. This resulted in phenotypic variation which may sustain adaptation and domestication of polyploid species (Chen, 2007). For example in allotetraploid cotton (*G. arboreum* L.), unequal distribution and expression bias of homoeologous genes as a consequence of asymmetric subgenome evolution were found to be associated to fiber development and a wider environmental adaptation (Zhang T. et al., 2015). In *B. juncea* it was proposed that expression dominance of homoeologous genes facilitated directional selection for genes associated with seed glucosinolate content and lipid metabolism in modern varieties used for food and oil production (Yang et al., 2016). Directional selection for specific homoeologous genetic variants increases the frequency of favorable alleles in a population and affects their standing variation, resulting in the formation of conserved haplotypes with strong surrounding LD. Comparing and dissecting these homoeologous chromosome blocks, while taking into account evolutionary and domestication processes, can improve our understanding of complex genetic mechanisms underlying quantitative traits. For example, a population-scale resequencing approach of a diverse collection of hexaploid wheat (*T. aestivum* L.) lines found that duplicated homoeologous genes are under purifying selection, implying that a fitness benefit could be obtained by a mutation at any one of the homoeologs and that directional selection likely acted on single advantageous mutations in homeologous chromosome regions (Jordan et al., 2015). In rapeseed (*B. napus* L.), a GWAS-based haplotype analysis found that duplicated homoeologous haplotype regions in the two subgenomes additively affect leaf chlorophyll content and specific variants have been co-selected in the same direction (Qian et al., 2016a). These findings indicate that haplotype analyses can shed new light on phenotypic effects of structural variations among homoeologous regions. Use of LD-based haplotypes
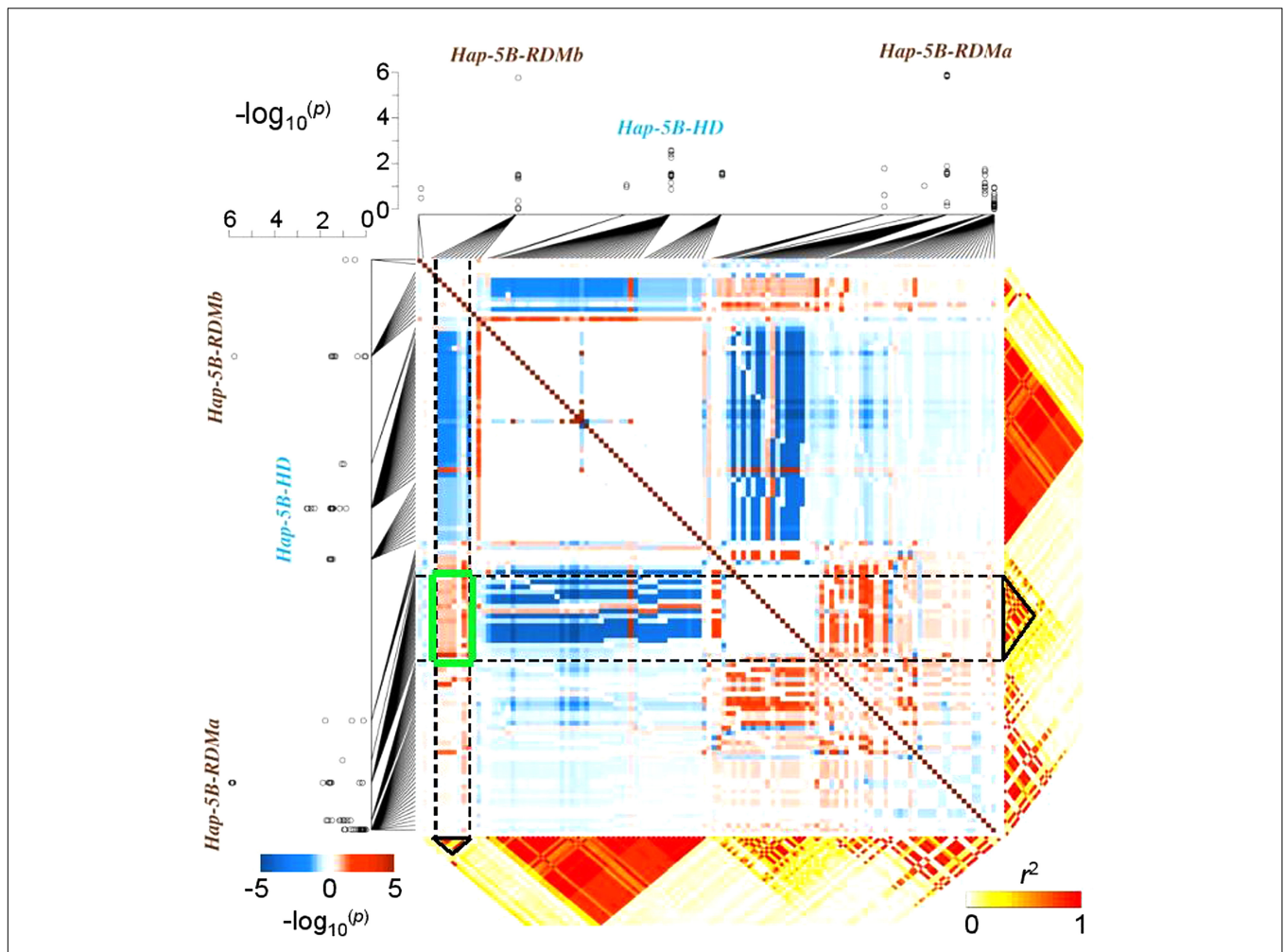
FIGURE 2 | Epistatic interactions among co-selected haplotypes: an example from bread wheat. Genome-wide association mapping results showing the haplotypes *Hap-5B-RDMa* and *Hap-5B-RDMb* which are significantly associated with root dry mass in 215 wheat inbred lines. The squared heat map (blue–red) shows the $-\log_{10}$-(p-values) of the SNP × SNP interaction effects of the SNPs located in haplotype blocks; a negative sign indicates a negative epistatic interaction. SNP × SNP epistatic interactions were confirmed by the epistatic interaction test implemented in the whole-genome association analysis toolset PLINK (v1.07, http://pngu.mgh.harvard.edu/purcell/plink/). The green box highlights a region with strong positive SNP × SNP epistatic between these two haplotypes. Modified from Voss-Fels et al. (2017) and reprinted with permission from John Wiley and Sons © 2017.

to more accurately distinguish between homologous loci can therefore improve our understanding of the consequences of artificial selection on quantitative trait complexes, enhancing the potential for genomics-driven improvements of future cultivars.

Two important genetic features of the variance of complex traits in crops are additive and epistatic effects (Lamkey et al., 1995; Goldringer et al., 1997; Ma et al., 2007). While QTL often additively explain a fraction of the phenotypic variation, different studies have reported epistatic QTL interactions for various important crops and traits, including wheat (*T. aestivum* L.) (Zhang et al., 2008; Li et al., 2014), rice (*O. sativa* L.) (Wang et al., 2012; Chen et al., 2014), maize (*Z. mays* L.) (Ma et al., 2007; Bocianowski, 2013), and rapeseed (*B. napus* L.) (Basunanda et al., 2010; Shi et al., 2011). Since the advent of modern high-throughput genotyping technologies which significantly

increased the genomic resolution and the establishment of GWAS techniques, significant phenotypic effects of epistatic SNP–SNP interactions on various complex traits have been reported in humans (Lindstrom et al., 2011; Ma et al., 2015), animals (Bolormaa et al., 2013; Ali et al., 2015), and crops (Snowdon et al., 2015; Zhang J. et al., 2015). SNP–SNP interactions normally cannot represent intergenic interactions at the single gene level and the exact components for genetic variation of quantitative traits affected by multiple genes in shared biological pathways are elusive. However, selection-driven LD in haplotype blocks and the simultaneous consideration of various multi-locus marker alleles allows a more precise and powerful detection of epistatic interactions between unlinked loci at different locations in the genome.

Different molecular genetic studies have attempted to model the effect of epistasis on quantitative traits. For example,
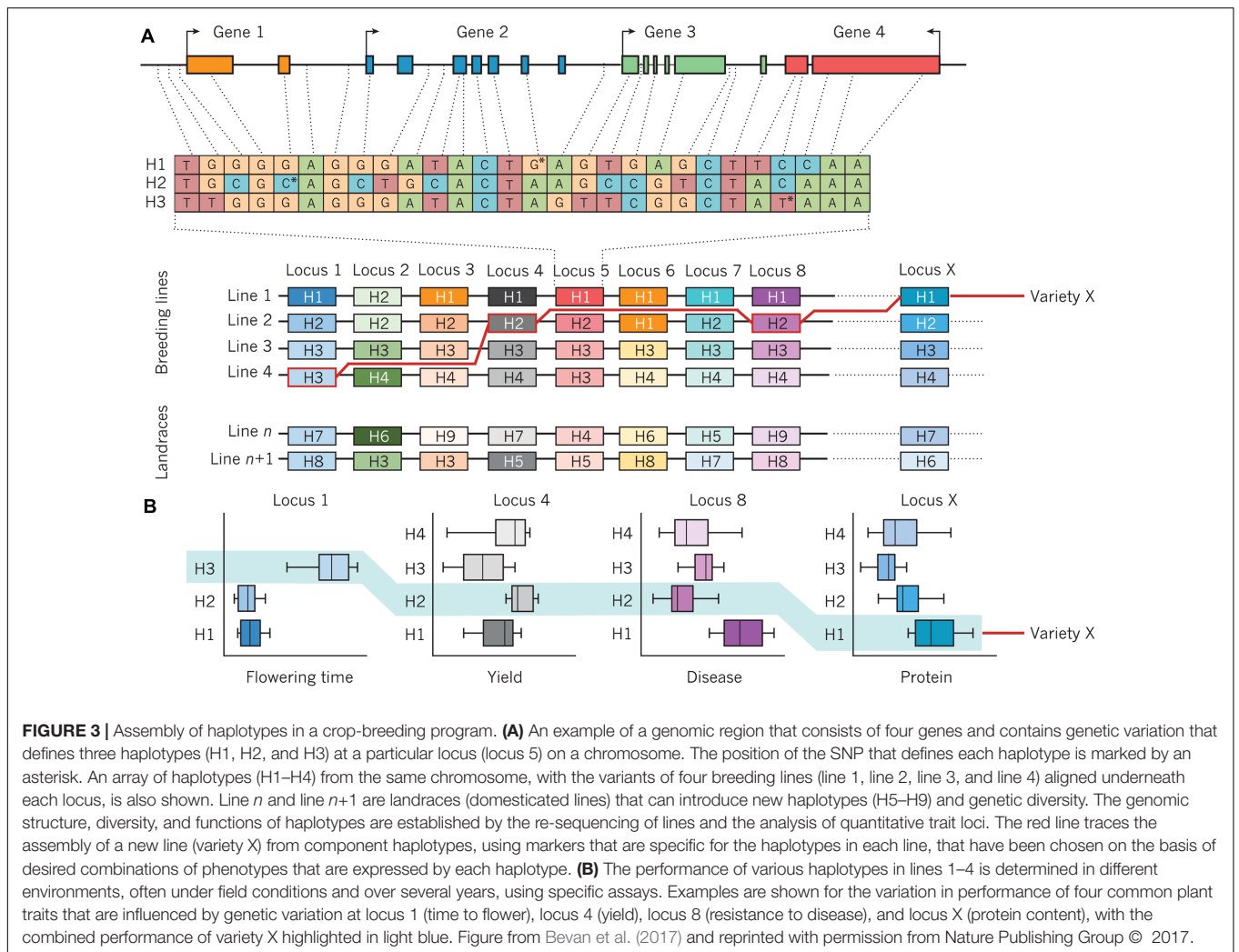
FIGURE 3 | Assembly of haplotypes in a crop-breeding program. **(A)** An example of a genomic region that consists of four genes and contains genetic variation that defines three haplotypes (H1, H2, and H3) at a particular locus (locus 5) on a chromosome. The position of the SNP that defines each haplotype is marked by an asterisk. An array of haplotypes (H1–H4) from the same chromosome, with the variants of four breeding lines (line 1, line 2, line 3, and line 4) aligned underneath each locus, is also shown. Line n and line n+1 are landraces (domesticated lines) that can introduce new haplotypes (H5–H9) and genetic diversity. The genomic structure, diversity, and functions of haplotypes are established by the re-sequencing of lines and the analysis of quantitative trait loci. The red line traces the assembly of a new line (variety X) from component haplotypes, using markers that are specific for the haplotypes in each line, that have been chosen on the basis of desired combinations of phenotypes that are expressed by each haplotype. **(B)** The performance of various haplotypes in lines 1–4 is determined in different environments, often under field conditions and over several years, using specific assays. Examples are shown for the variation in performance of four common plant traits that are influenced by genetic variation at locus 1 (time to flower), locus 4 (yield), locus 8 (resistance to disease), and locus X (protein content), with the combined performance of variety X highlighted in light blue. Figure from Bevan et al. (2017) and reprinted with permission from Nature Publishing Group © 2017.

interactions among molecular variants (often rare alleles) play a major role in determining the susceptibility of humans to a particular disease. Li et al. (2010) proposed a model to map haplotype × haplotype interactions responsible for human diseases. Similar approaches have been adapted to crops, demonstrating that interactions between haplotypes can significantly influence agronomic traits. Prominent examples are the vernalization response in barley (*Hordeum vulgare* L.) (Cockram et al., 2007), yield in wheat (*T. aestivum* L.) (Qin et al., 2014), or chlorophyll content in rapeseed (*B. napus* L.) (Qian et al., 2016a). A recent study showed that haplotype-based approaches are also useful to disclose relationships between crop characteristics that appear to be unrelated, for example inflorescence development and root growth in wheat. By mapping two highly significant haplotypes for root biomass in close proximity to a major locus known to affect spike development, Voss-Fels et al. (2017) uncovered selection-driven linkage drag and strong LD causing inadvertent co-selection of haplotype variants associated with low root biomass in European elite wheat germplasm (**Figure 2**). Using a linear model to estimate pairwise SNP–SNP interaction effects on root biomass,

strong positive epistatic interactions were detected among root-associated haplotypes and negative effects were detected for interactions between LD blocks associated to spike and root traits. High-resolution SNP and LD data are an ideal basis for breeders to disrupt linkage in repulsion among valuable traits, particularly (as for root traits) where phenotypic selection is challenging.

HAPLOTYPE-BASED IMPROVEMENT OF MOLECULAR BREEDING APPROACHES

Improving the resilience of modern crop cultivars in times of increasing climatic fluctuations is a major challenge for breeders (Massawe et al., 2016). Although marker-assisted selection (MAS) could effectively be deployed for traits with a mono- or oligogenic inheritance, selection of highly quantitative traits with a low heritability due to strong environmental influence remains challenging. In this context, MAS strategies have largely failed due to statistical overestimation of QTL-linked markers or complex genetic architectures for most

important traits (Bernardo, 2008). To overcome this problem, the concept of GS with densely spaced genome-wide markers is presently being adopted for many crop breeding programs. Genomic selection methods, initially conceived for animal breeding (Meuwissen et al., 2001), apply the concept that the breeding value of an individual which has not been phenotyped can be estimated purely on the basis of its genome-wide marker profile. Details about methodical concepts of GS and examples of the application of GS in crop breeding are given by Jannink et al. (2010). Given the strong genome structure present in the breeding pools of most crops, the deployment of haplotypes could be a powerful complementary tool to improve accuracy and efficiency of both MAS and GS. Varshney et al. (2005) recognized the high potential of haplotypes for whole-genome selection, based on comprehensive haplotype maps to identify and utilize genome regions are associated with trait of interests in populations with pronounced LD structures. Bevan et al. (2017) refer to retrospective and prospective approaches to implement haplotypes in genomics-assisted breeding (**Figure 3**). In the retrospective approach, favorable haplotypes that have been selected by breeders for a given crop species are identified by first resequencing the genomes of key genotypes with high importance and related, derived breeding lines which have been evaluated across multiple environments and years, as shown in **Figure 3A**. This provides retrospective information about selection decisions made by breeders, identifying genomic haplotypes associated with previous breeding success. These can be used to identify causal gene candidates and underlying networks, as well as desired and deleterious genetic variants that cause specific phenotypes. Subsequently, as shown in **Figure 3B**, molecular markers that define favorable haplotype blocks can be jointly selected to create novel combinations of haplotypes with well-defined trait effects. Haplotype-related markers can also be used to identify lines with novel recombination in chromosomal blocks of interest in order to separate favorable and unfavorable genetic variation. The definition of genome-wide haplotypes by specific functional and genetic relationships differs from classical GS, where genome-wide markers represent anonymous components that are statistically related to phenotypes in a subset of the total breeding population. Instead, Bevan et al. (2017) suggest prospective approaches for haplotype utilization in genomics-based crop improvement through resequencing of large ancestral and wild relative populations of a given crop species, in order to identify haplotypes with a broader range of genetic variation than is present in elite breeding pools. The aim is to define new haplotypes with fewer deleterious genetic components, and subsequently use molecular markers to define haplotype blocks which can be incorporated into breeding programs. A recent study in rice (*O. sativa* L.) showed that the prediction accuracies for different traits using classical GS models like ridge-regression best linear unbiased prediction (RR-BLUP) could significantly be improved by a model extension based on upstream GWAS analyses (Spindel et al., 2015, 2016). Using a collection of 369 elite rice breeding lines, Spindel et al. (2016) mapped significant SNP markers for various traits (e.g., flowering time and plant height, recorded

in multiple environments) and fit these markers as fixed effects in a RR-BLUP model (referred to as a GS + *de novo* GWAS model). They showed that extended GS + *de novo* GWAS models outperformed six other prediction models for various traits. Interestingly, the prediction accuracies of extended GS models outperformed classical models using phenotype data from dry seasons, implying that this approach is particularly suitable for the improvement of drought-resistance in future crop cultivars. Because of their increased information content compared to biallelic SNP markers, fitting haplotypes with statistically significant trait associations to phenotypes as fixed effects in GS models could further improve prediction accuracies. The use of haplotype-assisted GS should more accurately depict the complex relationships between genotypic information and phenotypes than single SNPs alone are able to do; hence, this approach could ultimately help to further increase selection gain per unit of time. Also in horticultural crops, the deployment of haplotype-based approaches holds a great potential for a genetic improvement of future varieties. In cucumber (*Holothuria edulis* L.) for example, Soliman et al. (2016) used PCR-based markers to study sequence diversity in a mitochondrial cytochrome oxidase subunit I (COI), 16S ribosomal RNA, and a nuclear histone. Applying marker-based haplotype analyses they were able to gain insights into population structure, providing valuable information for the management of sea cucumber populations in Okinawa. In grapevine (*Vitis vinifera* L.), Fernandez et al. (2014) analyzed haplotype variation in a gene homologous to *TERMINAL FLOWER 1* from *Arabidopsis thaliana* and found a relationship between haplotype diversity and phenotypic variation for reproductive and inflorescence traits. Gavrilenko et al. (2013) used SSR markers developed from plastid genome sequences to investigate the genetic relationships between cultivated and wild potato (*Solanum tuberosum* Andigenum and *S. tuberosum* ssp. *andigenum*) species. By analyzing haplotype diversity they could gain further insights into the origin of cultivated potatoes. Combining advanced statistical genetics approaches with ultra-fast crop propagation techniques that allow multiple plant generations per year under greenhouse conditions (Riaz et al., 2016) could help to rapidly accelerate environmental resilience of modern crop cultivars to secure future food supply.

AUTHOR CONTRIBUTIONS

LQ, KV-F, and RS conceived the review; LQ, KV-F, AS, and RS conducted the literature survey and drafted the article; LQ and KV-F wrote the manuscript; and LH, CW, and BH contributed additional concepts and carried out critical revision of the manuscript.

FUNDING

REFERENCES

Ali, A. A., Khatkar, M. S., Kadarmideen, H. N., and Thomson, P. C. (2015). Additive and epistatic genome-wide association for growth and ultrasound scan measures of carcass-related traits in Brahman cattle. *J. Anim. Breed. Genet.* 132, 187–197. doi: 10.1111/jbg.12147

Austin, R. B., Bingham, J., Blackwell, R. D., Evans, L. T., Ford, M. A., Morgan, C. L., et al. (1980). Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *J. Agric. Sci.* 94, 675–689. doi: 10.1017/S0021859600028665

Bai, X., Luo, L., Yan, W., Kovi, M. R., Zhan, W., and Xing, Y. (2010). Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus qGL7. *BMC Genet.* 11:16. doi: 10.1186/1471-2156-11-16

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

Basunanda, P., Radoev, M., Ecke, W., Friedt, W., Becker, H. C., and Snowdon, R. J. (2010). Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 120, 271–281. doi: 10.1007/s00122-009-1133-z

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48:1649. doi: 10.2135/cropsci2008.03.0131

Bernardo, R. N. (2010). *Breeding for Quantitative Traits in Plants*. Woodbury, MN: Stemma Press.

Bevan, M. W., Uauy, C., Wulff, B. B. H., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature* 543, 346–354. doi: 10.1038/nature22011

Bocianowski, J. (2013). Epistasis interaction of QTL effects as a genetic parameter influencing estimation of the genetic additive effect. *Genet. Mol. Biol.* 36, 93–100. doi: 10.1590/S1415-47572013000100013

Bolormaa, S., Pryce, J. E., Kemper, K. E., Hayes, B. J., Zhang, Y., Tier, B., et al. (2013). Detection of quantitative trait loci in *Bos indicus* and *Bos taurus* cattle using genome-wide association studies. *Genet. Sel. Evol.* 45:43. doi: 10.1186/1297-9686-45-43

Bomblies, K., and Doebley, J. F. (2006). Pleiotropic effects of the duplicate maize FLORICAULA/LEAFY genes zfl1 and zfl2 on traits under selection during maize domestication. *Genetics* 172, 519–531. doi: 10.1534/genetics.105.048595

Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12:232. doi: 10.1186/gb-2011-12-10-232

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chen, J., Li, X., Cheng, C., Wang, Y., Qin, M., Zhu, H., et al. (2014). Characterization of epistatic interaction of QTLs LH8 and EH3 controlling heading date in rice. *Sci. Rep.* 4:4263. doi: 10.1038/srep04263

Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 58, 377–406. doi: 10.1146/annurev.arplant.58.032806.103835

Cockram, J., Chiapparino, E., Taylor, S. A., Stamati, K., Donini, P., Laurie, D. A., et al. (2007). Haplotype analysis of vernalization loci in European barley germplasm reveals novel VRN-H1 alleles and a predominant winter VRN-H1/VRN-H2 multi-locus haplotype. *Theor. Appl. Genet.* 115, 993–1001. doi: 10.1007/s00122-007-0626-x

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196. doi: 10.1007/s10681-005-1681-5

Cormier, F., Faure, S., Dubreuil, P., Heumez, E., Beauchêne, K., Lafarge, S., et al. (2013). A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 126, 3035–3048. doi: 10.1007/s00122-013-2191-9

Cowling, W. A. (2007). Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crops Res.* 104, 103–111. doi: 10.1016/j.fcr.2006.12.014

Cowling, W. A., and Léon, J. (2013). Sustainable plant breeding. *Plant Breed.* 132, 1–9. doi: 10.1111/pbr.12026

Fernandez, L., Le Cunff, L., Tello, J., Lacombe, T., Boursiquot, J. M., Fournier-Level, A., et al. (2014). Haplotype diversity of VvTFL1A gene and association with cluster traits in grapevine (*V. vinifera*). *BMC Plant Biol.* 14, 209. doi: 10.1186/s12870-014-0209-3

Flowers, J. M., Molina, J., Rubinstein, S., Huang, P., Schaal, B. A., and Purugganan, M. D. (2012). Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. *Mol. Biol. Evol.* 29, 675–687. doi: 10.1093/molbev/msr225

Ganal, M. W., Polley, A., Graner, E. -M., Plieske, J., Wieseke, R., Luerssen, H., et al. (2012). Large SNP arrays for genotyping in crop plants. *J. Biosci.* 37, 821–828. doi: 10.1007/s12038-012-9225-3

Gavrilenko, T., Antonova, O., Shuvalova, A., Krylova, E., Alpatyeva, N., Spooner, D. M., et al. (2013). Genetic diversity and origin of cultivated potatoes based on plastid microsatellite polymorphism. *Genet. Resour. Crop Evol.* 60, 1997–2015. doi: 10.1007/s10722-013-9968-1

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nat. Rev. Genet.* 13, 135–145. doi: 10.1038/nrg3118

Goldringer, I., Brabant, P., and Gallais, A. (1997). Estimation of additive and epistatic genetic variances for agronomic traits in a population of doubled-haploid lines of wheat. *Heredity* 79, 60–71. doi: 10.1038/hdy.1997.123

Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015:431487. doi: 10.1155/2015/431487

Griggs, D., Stafford Smith, M., Rockström, J., Öhman, M. C., Gaffney, O., Glaser, G., et al. (2014). An integrated framework for sustainable development goals. *Ecol. Soc.* 19. 49. doi: 10.5751/ES-07082-190449

Gupta, P. K., Rustgi, S., and Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol.* 57, 461–485. doi: 10.1007/s11103-005-0257-z

Hasan, M., Seyis, F., Badani, A. G., Pons-Kühnemann, J., Friedt, W., Lühs, W., et al. (2006). Analysis of genetic diversity in the *Brassica napus* L. *Genet. Resour. Crop Evol.* 53, 793–802. doi: 10.1007/s10722-004-5541-2

Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715

Hufford, M. B., Lubinksy, P., Pyhäjärvi, T., Devengenzo, M. T., Ellstrand, N. C., and Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genet.* 9:e1003477. doi: 10.1371/journal.pgen.1003477

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, Y., Jiang, Q., Hao, C., Hou, J., Wang, L., Zhang, H., et al. (2015). A yield-associated gene TaCWI, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis. *Theor. Appl. Genet.* 128, 131–143. doi: 10.1007/s00122-014-2417-5

Jordan, K. W., Wang, S., Lun, Y., Gardiner, L.-J., MacLachlan, R., Hucl, P., et al. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.* 16:48. doi: 10.1186/s13059-015-0606-4

Kessel, B., Schierholt, A., and Becker, H. C. (2012). Nitrogen use efficiency in a genetically diverse set of winter oilseed rape (*Brassica napus* L.). *Crop Sci.* 52:2546. doi: 10.2135/cropsci2012.02.0134

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. doi: 10.1186/1746-4811-9-29

Laidig, F., Piepho, H.-P., Rentel, D., Drobek, T., Meyer, U., and Huesken, A. (2017a). Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983-2014. *Theor. Appl. Genet.* 130, 223–245. doi: 10.1007/s00122-016-2810-3

Laidig, F., Piepho, H.-P., Rentel, D., Drobek, T., Meyer, U., and Huesken, A. (2017b). Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. *Theor. Appl. Genet.* 130, 981–998. doi: 10.1007/s00122-017-2865-9

Lamkey, K. R., Schnicker, B. J., and Melchinger, A. E. (1995). Epistasis in an elite maize hybrid and choice of generation for inbred line development. *Crop Sci.* 35:1272. doi: 10.2135/cropsci1995.0011183X003500050004x

Lande, R., and Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution* 37, 1210–1226. doi: 10.1111/j.1558-5646.1983.tb00236.x

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature* 529, 84–87. doi: 10.1038/nature16467

Li, M., Romero, R., Fu, W. J., and Cui, Y. (2010). Mapping haplotype-haplotype interactions with adaptive LASSO. *BMC Genet.* 11:79. doi: 10.1186/1471-2156-11-79

Li, Z. K., Jiang, X. L., Peng, T., Shi, C. L., Han, S. X., Tian, B., et al. (2014). Mapping quantitative trait loci with additive effects and additive x additive epistatic interactions for biomass yield, grain yield, and straw yield using a doubled haploid population of wheat (*Triticum aestivum* L.). *Genet. Mol. Res.* 13, 1412–1424. doi: 10.4238/2014.February.28.14

Lindstrom, S., Schumacher, F., Siddiq, A., Travis, R. C., Campa, D., Berndt, S. I., et al. (2011). Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers–results from BPC3. *PLoS ONE* 6:e17142. doi: 10.1371/journal.pone.0017142

Losert, D., Maurer, H. P., Marulanda, J. J., Würschum, T., and Friedt, W. (2017). Phenotypic and genotypic analyses of diversity and breeding progress in European triticale (x Triticosecale Wittmack). *Plant Breed.* 136, 18–27. doi: 10.1111/pbr.12433

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., et al. (2010). Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19585–19590. doi: 10.1073/pnas.1006105107

Ma, L., Keinan, A., and Clark, A. G. (2015). Biological knowledge-driven analysis of epistasis in human GWAS with application to lipid traits. *Methods Mol. Biol.* 1253, 35–45. doi: 10.1007/978-1-4939-2155-3_3

Ma, X. Q., Tang, J. H., Teng, W. T., Yan, J. B., Meng, Y. J., and Li, J. S. (2007). Epistatic interaction is an important genetic basis of grain yield and its components in maize. *Mol. Breed.* 20, 41–51. doi: 10.1007/s11032-006-9071-9

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* 4:2320. doi: 10.1038/ncomms3320

Mackay, M., Street, K., and Hickey, L. (2016). "Towards more effective discovery and deployment of novel plant genetic variation: reflection and future directions," in *Applied Mathematics and Omics to Assess Crop Genetic Resources for Climate Change Adaptive Traits*," eds A. Bari, A. Damania, M. Mackay, and S. Dayanandan (Boca Raton, FL: CRC Press), 139–150.

Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612

Maher, B. (2008). Personal genomes: the case of the missing heritability. *Nature* 456, 18–21. doi: 10.1038/456018a

Mason, A. S., and Snowdon, R. J. (2016). Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. *Plant Biol.* 18, 883–892. doi: 10.1111/plb.12462

Massawe, F., Mayes, S., and Cheng, A. (2016). Crop diversity: an unexploited treasure trove for food security. *Trends Plant Sci.* 21, 365–368. doi: 10.1016/j.tplants.2016.02.006

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

Mickelbart, M. V., Hasegawa, P. M., and Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Rev. Genet.* 16, 237–251. doi: 10.1038/nrg3901

Platt, A., Vilhjálmsson, B. J., and Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. *Genetics* 186, 1045–1052. doi: 10.1534/genetics.110.121665

Pritchard, J. K., and Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 69, 1–14. doi: 10.1086/321275

Qian, L., Qian, W., and Snowdon, R. J. (2014). Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genomics* 15:1170. doi: 10.1186/1471-2164-15-1170

Qian, L., Qian, W., and Snowdon, R. J. (2016a). Haplotype hitchhiking promotes trait coselection in *Brassica napus*. *Plant Biotechnol. J.* 14, 1578–1588. doi: 10.1111/pbi.12521

Qin, L., Hao, C., Hou, J., Wang, Y., Li, T., Wang, L., et al. (2014). Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene TaGW2. *BMC Plant Biol.* 14:107. doi: 10.1186/1471-2229-14-107

Qian, L., Voss-Fels, K., Cui, Y., Jan, H. U., Samans, B., Obermeier, C., et al. (2016b). Deletion of a stay-green gene associates with adaptive selection in *Brassica napus*. *Mol. Plant* 9, 1559–1569. doi: 10.1016/j.molp.2016.10.017

Riaz, A., Hathorn, A., Dinglasan, E., Ziems, L., Richard, C., Singh, D., et al. (2017). Into the vault of the Vavilov wheats: old diversity for new alleles. *Genet. Resour. Crop Evol.* 64, 531–544. doi: 10.1007/s10722-016-0380-5

Riaz, A., Periyannan, S., Aitken, E., and Hickey, L. (2016). A rapid phenotyping method for adult plant resistance to leaf rust in wheat. *Plant Methods* 12:17. doi: 10.1186/s13007-016-0117-7

Samans, B., Chalhoub, B., and Snowdon, R. J. (2017). Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species *Brassica napus*. *Plant Genome* 10. doi: 10.3835/plantgenome2017.02.0013

Shi, J., Li, R., Zou, J., Long, Y., and Meng, J. (2011). A dynamic and complex network regulates the heterosis of yield-correlated traits in rapeseed (*Brassica napus* L.). *PLoS ONE* 6:e21645. doi: 10.1371/journal.pone.0021645

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Snowdon, R. J., Abbadi, A., Kox, T., Schmutzer, T., and Leckband, G. (2015). Heterotic haplotype capture: precision breeding for hybrid performance. *Trends Plant Sci.* 20, 410–413. doi: 10.1016/j.tplants.2015.04.013

Soliman, T., Fernandez-Silva, I., and Reimer, J. D. (2016). Genetic population structure and low genetic diversity in the over-exploited sea cucumber *Holothuria edulis* Lesson, 1830 (Echinodermata: Holothuroidea) in Okinawa Island. *Conserv. Genet.* 17, 811–821. doi: 10.1007/s10592-016-0823-8

Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079

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

Stahl, A., Pfeifer, M., Frisch, M., Wittkop, B., and Snowdon, R. J. (2017). Recent genetic gains in nitrogen use efficiency in oilseed rape. *Front. Plant Sci.* 8:963. doi: 10.3389/fpls.2017.00963

Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., et al. (2001). Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293, 489–493. doi: 10.1126/science.1059431

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700

Tollenaar, M. (1989). Response of dry matter accumulation in maize to temperature: I. Dry matter partitioning. *Crop Sci.* 29, 1239–1246. doi: 10.2135/cropsci1989.0011183X002900050030x

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

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S., et al. (2015). Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome* 8. doi: 10.3835/plantgenome2015.03.0013

Voss-Fels, K., and Snowdon, R. J. (2016). Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnol. J.* 14, 1086–1094. doi: 10.1111/pbi.12456

Voss-Fels, K. P., Qian, L., Parra-Londono, S., Uptmoor, R., Frisch, M., Keeble-Gagnère, G., et al. (2017). Linkage drag constrains the roots of modern wheat. *Plant Cell Environ.* 40, 717–725. doi: 10.1111/pce.12888

Wall, J. D., and Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Rev. Genet.* 4, 587–597. doi: 10.1038/nrg1123

Wang, W., Feng, B., Xiao, J., Xia, Z., Zhou, X., Li, P., et al. (2014). Cassava genome from a wild ancestor to cultivated varieties. *Nat. Commun.* 5:5110. doi: 10.1038/ncomms6110

Wang, Z., Cheng, J., Chen, Z., Huang, J., Bao, Y., Wang, J., et al. (2012). Identification of QTLs with main, epistatic and QTL × environment interaction effects for salt tolerance in rice seedlings under different salinity conditions. *Theor. Appl. Genet.* 125, 807–815. doi: 10.1007/s00122-012-1873-z

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., and Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Rev. Genet.* 14, 507–515. doi: 10.1038/nrg3457

Würschum, T., Langer, S. M., Longin, C. F. H., Korzun, V., Akhunov, E., Ebmeyer, E., et al. (2013). Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor. Appl. Genet.* 126, 1477–1486. doi: 10.1007/s00122-013-2065-1

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., et al. (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nature Genet.* 48, 1225–1232. doi: 10.1038/ng.3657

Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., et al. (2013). CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16969–16974. doi: 10.1073/pnas.1310949110

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.-C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genet.* 48, 927–934. doi: 10.1038/ng.3596

Zhang, J., Singh, A., Mueller, D. S., and Singh, A. K. (2015). Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *Plant J.* 84, 1124–1136. doi: 10.1111/tpj.13069

Zhang, K., Tian, J., Zhao, L., and Wang, S. (2008). Mapping QTLs with epistatic effects and QTL × environment interactions for plant height using a doubled haploid population in cultivated wheat. *J. Genet. Genomics* 35, 119–127. doi: 10.1016/S1673-8527(08)60017-X

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33, 531–537. doi: 10.1038/nbt.3207

Zhu, C., Gore, M., Buckler, E. S., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1. doi: 10.3835/plantgenome2008.02.0089

Zhu, G., Peng, S., Huang, J., Cui, K., Nie, L., and Wang, F. (2016). Genetic improvements in rice yield and concomitant increases in radiation- and nitrogen-use efficiency in middle reaches of yangtze river. *Sci. Rep.* 6:21049. doi: 10.1038/srep21049