New Phytologist

Review

Tansley review

What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits?

Author for correspondence:
Emily B. Josephs
Tel: +1 530 752 2999
Email: emjo@ucdavis.edu

Received: *30 September 2016*
Accepted: *14 November 2016*

Emily B. Josephs[1], John R. Stinchcombe[2]* and Stephen I. Wright[2]*

[1]Department of Evolution and Ecology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA; [2]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, ON, M5S 3B2, Canada

Contents

	Summary	21
I.	Introduction	21
II.	Theoretical predictions	22
III.	What GWAS have told us about selection on quantitative traits	24
IV.	Challenges of connecting GWAS to selection	25
V.	Potential ways forward	29
	Acknowledgements	30
	References	30

New Phytologist (2017) **214:** 21–33
doi: 10.1111/nph.14410

Key words: balancing selection, F_{ST}, genetic architecture, genome-wide association mapping studies (GWAS), mutation–selection balance, selective sweep.

Summary

Understanding the evolutionary forces that shape genetic variation within species has long been a goal of evolutionary biology. Integrating data for the genetic architecture of traits from genome-wide association mapping studies (GWAS) along with the development of new population genetic methods for identifying selection in sequence data may allow us to evaluate the roles of mutation–selection balance and balancing selection in shaping genetic variation at various scales. Here, we review the theoretical predictions for genetic architecture and additional signals of selection on genomic sequence for the loci that affect traits. Next, we review how plant GWAS have tested for the signatures of various selective scenarios. Limited evidence to date suggests that within-population variation is maintained primarily by mutation–selection balance while variation across the landscape is the result of local adaptation. However, there are a number of inherent biases in these interpretations. We highlight these challenges and suggest ways forward to further understanding of the maintenance of variation.

I. Introduction

Genetic variation for quantitative traits is ubiquitous in nature yet, despite widespread interest, we lack empirical data about the evolutionary forces that shape this variation (Johnson & Barton, 2005; Mitchell-Olds *et al.*, 2007). One hypothesis suggests that variation exists as a balance between its creation by mutation and its removal by selection, so that much of this variation is deleterious (Barton & Keightley, 2002). Alternatively, some variation may be maintained directly by balancing selection, either through within-population processes like frequency-dependent selection or through between-population processes, such as local adaptation (Hedrick, 2006). A large body of theory has developed around these alternative hypotheses but empirical tests of their relative importance have been limited.

A number of experimental approaches have attempted to investigate the maintenance of genetic variation for quantitative traits within species and we will briefly discuss two. First, Kelly

*These authors contributed equally to this work.

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

(1999) proposed that the contribution of rare, recessive alleles, which is expected to be large under mutation–selection balance, could be quantified through investigating the change in mean phenotype through both inbreeding and directional selection. Applications of this approach have been limited, but have shown that rare, recessive alleles cannot fully explain the variation in flower size in *Mimulus guttatus* (Kelly & Willis, 2001). Second, reciprocal transplant experiments have played a crucial role in demonstrating the importance of local adaptation (Hereford, 2009). However, while reciprocal transplant experiments can tell us about the fitness of certain genotypes in specific conditions, we need to estimate the selective pressures acting on traits across longer timescales to understand how selection has shaped the variation we see in the present. Incorporating genomic information can provide crucial, complementary information about the selective forces acting on genetic variation (Harrison *et al.*, 2016).

A direct approach to understanding the maintenance of variation will involve generating a sample of loci that control quantitative trait variation and describing how selection acts on these loci (Stinchcombe & Hoekstra, 2008). Over the last decade, genome-wide association studies (GWAS) have identified many loci controlling variation in quantitative traits in a variety of systems. In turn, increasingly sophisticated methods are being applied to population genomics datasets to infer the action of selection across the genome. Is it now possible to integrate population genetics and GWAS to evaluate which selective models best describe the genetic variation observed (Y. W. Lee *et al.*, 2014)? Or are we still unable to both detect a representative sample of causal variants (Rockman, 2012) and identify the selective forces acting on these variants (Tiffin & Ross-Ibarra, 2014), preventing us from making general conclusions about the forces maintaining variation?

Here, we evaluate the potential of GWAS in plants to provide insight into the maintenance of variation in plant species. We first outline specific theoretical predictions for patterns of variation under different selective scenarios. Second, we summarize the studies to date that have used GWAS to investigate the selective forces maintaining variation. Third, we explore the continuing challenges with using GWAS to understand the maintenance of variation and offer potential solutions to these challenges.

II. Theoretical predictions

Linking patterns of genomic variation to various selective scenarios is a key step to understanding how selection maintains trait variation. A number of predictions have been made for what variation in genetic loci under various selective regimes should look like. These predictions encompass the effect size and population-level frequency of alleles, along with the partitioning of variance within and between populations (F_{ST}) and patterns of linked neutral variation around selected loci. We review key predictions later and relate them to what we expect for single nucleotide polymorphisms (SNPs) associated with phenotypic traits under various types of selection.

1. Negative selection against new mutations

Negative selection against new mutations can be a powerful force shaping sequence variation (Hough *et al.*, 2013). Negative selection affects the frequencies of alleles segregating in a population, so the allele frequency spectrum is often used to describe the frequencies of segregating alleles and investigate the selective forces acting on a group of loci. The allele frequency spectrum (also called the site frequency spectrum) summarizes the proportion of alleles present at various bins of allele frequency. Often the allele frequency spectrum is folded, so that it tallies the frequency of the minor (or rarer) allele while being agnostic as to which allele is derived – this information is sufficient for estimating the strength of negative selection (Keightley & Eyre-Walker, 2010).

Negative selection is expected to reduce the frequency of deleterious alleles, leading to the prediction that alleles affecting phenotypic variation should be rarer on average than neutral alleles that do not affect phenotypes or fitness. The allele frequency spectrum of a group of loci will indicate the strength of negative selection acting on these loci in aggregate, and can be used to test for negative selection acting on these loci (Fig. 1a). Because demography can also affect the allele frequency spectrum of both selected and unselected loci, methods for measuring the strength of negative selection acting on a group of loci generally compare the allele frequency spectrum to a neutral allele frequency spectrum, often taken from synonymous sites (Keightley & Eyre-Walker, 2010).

2. Stabilizing selection on traits

Although the allele frequency spectrum can provide clues about the strength of selection acting on individual loci, incorporating information about the relationship between these loci and trait variation can provide important information about how selection acts directly on traits. Stabilizing selection on a quantitative trait should lead to a negative correlation between the effect size and frequency of quantitative trait loci (QTLs) for that trait (Fig. 1b). Specifically, alleles with large effects will be found at lower allele frequencies than alleles with small effects, because negative selection acts more strongly to reduce the allele frequency of large-effect alleles compared with small-effect alleles. Another consequence of this process is that much of the variance for traits under stabilizing selection will be explained by rare large-effect alleles (Turelli, 1984; Barton & Turelli, 1989; de Vladar & Barton, 2014).

3. Balancing selection within populations

The predictions for how balancing selection will shape the frequencies of QTLs are more varied than those for negative and stabilizing selection, in part because balancing selection encompasses a number of distinct properties, such as temporally variable selection and negative frequency-dependent selection. At the sequence level, we generally expect that loci subject to balancing selection should be found at more intermediate allele frequencies than neutral loci (Gillespie & Turelli, 1989; Turelli & Barton,

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

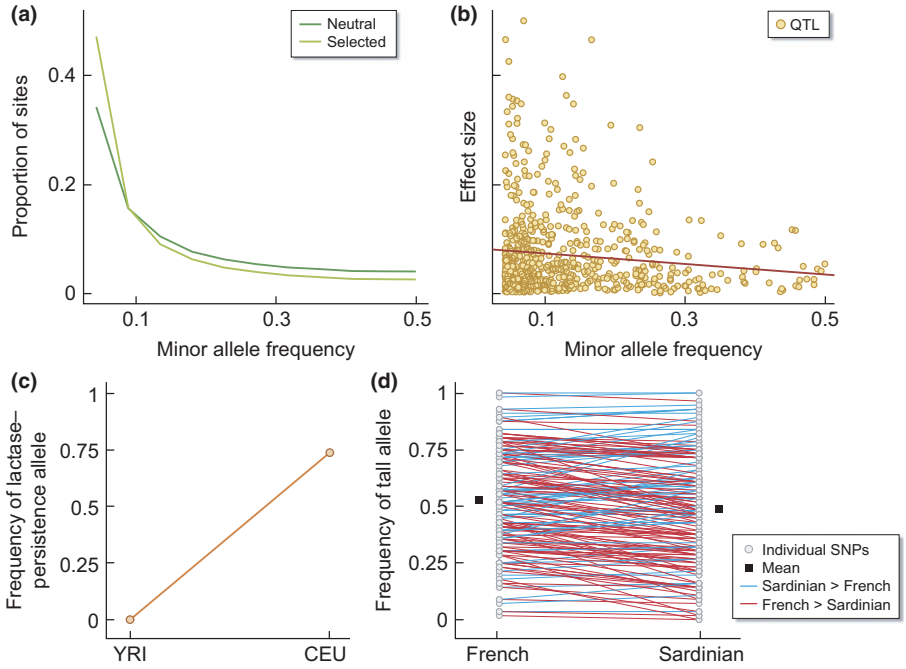

Fig. 1 Patterns of genome-wide association study (GWAS) single nucleotide polymorphisms (SNPs) expected under various selective scenarios. (a) Negative selection is expected to reduce allele frequencies, as observed from this allele frequency spectrum. Data from Williamson *et al.* (2014). (b) Under mutation–selection balance, we expect a negative correlation between effect size and allele frequency. Data from Josephs *et al.* (2015). (c) A classic example of local adaptation, human lactase persistence, is largely controlled by variation at one locus: rs4988235. This locus is at much higher frequency in European populations (CEU, Utah residents with Northern and Western European Ancestry) than in African populations (YRI, Yorubans). Data from The 1000 Genomes Project Consortium *et al.* (2015). (d) The frequency of the tall allele at 163 human height-associated SNPs in French and Sardinian populations. For 104 of 163 alleles, the tall allele is more common in the French population than the Sardinian population. Data from Berg & Coop (2014).

2004; Charlesworth, 2006) although balancing selection can also maintain alleles at low or high allele frequencies (Gillespie, 2010). Additionally, balancing selection can leave characteristic signatures on patterns of linked neutral diversity, namely elevated levels of polymorphism, and long-term balancing selection can also result in trans-specific polymorphism (Charlesworth, 2006).

However, our ability to detect balancing selection in the genome is limited for multiple reasons. First, although coalescent models of balanced polymorphism are well-characterized (Kaplan *et al.*, 1988) the development of explicit model-based approaches that identify balancing selection genome-wide has been limited until recently (Leffler *et al.*, 2013; DeGiorgio *et al.*, 2014). Second, the signature of long-term balancing selection on linked neutral polymorphism can be difficult to detect, because recombination and gene conversion can weaken the association between neutral variants and the site under balancing selection (Andolfatto & Nordborg, 1998). The effects of recombination and gene conversion weakening the signature of balancing selection is one reason why the clearest cases of balanced polymorphisms arise when recombination is suppressed in a genomic region, such as at the self-incompatability locus or on sex chromosomes (Charlesworth, 2006) or in inversions (Lee *et al.*, 2016). However, low rates of recombination in highly selfing plants mean that balancing selection may in fact play a significant role in structuring patterns of shared and unique variation across the genome of selfing species. Overall, while we can detect some forms of balancing selection using sequence data, we may miss many others using current techniques.

4. Balancing selection across populations

A particular type of local adaptation, when mediated by antagonistic pleiotropy at a locus, can lead to between-population balancing selection, although the distinction between balancing selection acting at within- and between-population levels is not always straightforward (Delph & Kelly, 2014). A few broad approaches have been developed to identify local adaptation in sequence data. First, because the loci involved in local adaptation are likely to show increased between-population differentiation, F_{ST}-based scans for between population differentiation are commonly used to identify candidate local adaptation SNPs (Savolainen *et al.*, 2013; Whitlock & Lotterhos, 2015). Specifically, F_{ST} measures the partitioning of variation between and within populations, and signals of reduced within-population variation relative to between population divergence is consistent with local adaptation.

A second commonly sought-after signature of local adaptation is the presence of local selective sweeps, the signature of reduced neutral diversity left by the fixation of a positively-selected allele (Smith & Haigh, 1974; Berry *et al.*, 1991). A number of approaches have been used to detect local selective sweeps in progress, and thus, local adaptation, by specifically looking for excess homozygosity between haplotypes (Toomajian *et al.*, 2006) or allele frequency shifts (Nielsen *et al.*, 2005). The strength of selection, the number of mutations fixed, and the previous evolutionary history of beneficial mutations will affect the shape of selective sweeps (Hermisson & Pennings, 2005; Pennings &

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

Hermisson, 2006; Messer & Petrov, 2013), but differentiating between different types of selective sweeps with sequence data is not always straightforward (Schrider *et al.*, 2015). F_{ST} and selective sweep analyses test slightly different processes related to adaptation but are commonly used concurrently.

5. Polygenic local adaptation

Local adaptation in traits with a polygenic basis can occur through a process called polygenic adaptation: subtle shifts in allele frequency at many loci. The allele frequency shifts involved in polygenic adaptation may not leave clear signatures in sequence data. For example, F_{ST} at the QTLs for a locally adaptive trait is less likely to differ from background F_{ST} as the number of loci involved in a trait increases or as gene flow between adaptively-diverging populations increases (Latta, 1998; Le Corre & Kremer, 2003, 2012; Kremer & Le Corre, 2012). This decoupling between trait divergence and F_{ST} at QTLs for the trait occurs because there are many potential allelic combinations that correspond to a given optimal trait value (Kremer & Le Corre, 2012). Polygenic adaptation is also expected to be difficult to detect using selective sweep tests. Signatures of selective sweeps are not expected around small-effect loci involved in polygenic adaptation because these alleles will take longer to reach fixation, if they even fix at all (Chevin & Hospital, 2008; Pritchard *et al.*, 2010).

Nevertheless, polygenic adaptation can be detectable through covariance between QTL allele frequency and the direction of QTL effects (Latta, 1998; Le Corre & Kremer, 2003). Concretely, this means that the alleles that act in a certain direction will be more common in certain populations if there has been local adaptation for that trait. To illustrate the differences between different modes of local adaptation, we show two examples of adaptation that have occurred in humans (Fig. 1c,d). First, lactase persistence has evolved through a large allele frequency change at one locus (Fig. 1c). Local adaptation of the lactase persistence allele is detectable due to elevated F_{ST} and selective sweep signatures (Tishkoff *et al.*, 2007). By contrast, evolution of human height has proceeded through subtle allele frequency shifts at many loci (Fig. 1d) and local adaptation has been detected through a test for covariance between allelic effects and frequency (Berg & Coop, 2014). Specifically, alleles that increase height are, on average, at higher frequency in the French population sample than the Sardinian population sample. Although this strategy has proven useful in human studies (Berg & Coop, 2014; Robinson *et al.*, 2015) it has not yet been widely applied in plant systems.

III. What GWAS have told us about selection on quantitative traits

Association mapping studies have been conducted in a number of plant species, across various traits and locations, in both population samples and the offspring of multiple controlled crosses, and with genetic information that ranges from a few hundred markers to whole-genome sequence. We will not exhaustively review all plant GWAS, but we instead highlight the findings from selected studies that have attempted to link GWAS results with selection (see Ogura & Busch (2015) for a more comprehensive review). We include GWAS for morphological traits, molecular traits such as expression and methylation, and environmental traits, because all of these traits may either be under selection directly or correlate with traits under selection. It is important to note that many of the plant GWAS conducted to date were not developed to explicitly test evolutionary questions, so the sampling schemes and other elements of experimental design may not be ideal for this purpose. We will discuss the consequences of this problem in a later section.

Only one study, to date, has used a GWAS to investigate selection within a population of plants. Josephs *et al.* (2015) mapped loci associated with *cis* regulatory variation (*cis*-eQTLs) in a single population of the outcrossing plant *Capsella grandiflora*. This study found evidence of negative selection predominating on *cis*-eQTLs, showing that these loci are at lower minor allele frequencies than expected (Josephs *et al.*, 2015). There was also a negative relationship between minor allele frequency and the effect size of *cis*-eQTLs, consistent with stabilizing selection acting on gene expression levels (Josephs *et al.*, 2015). Studies of molecular traits, like gene expression, could be more likely to detect the effects of negative selection than GWAS on physical traits. For example, *cis*-eQTLs may have relatively large effects on the expression of individual genes, allowing them to be mapped, but smaller effects on traits directly under selection, so that they are maintained at detectable allele frequencies.

A few GWAS have tested for the negative correlation between effect size and allele frequency at the range-wide level expected under stabilizing selection. For example, the effect size and frequency of GWAS SNPs associated with multiple maize traits are negatively correlated (Brown *et al.*, 2011; Peiffer *et al.*, 2014; Wallace *et al.*, 2014). In addition, Stanton-Geddes *et al.* (2013) found that observed correlations of effect size and frequency were more negative than those generated from trials on permuted data for height and lower root nodule count in *Medicago truncatula*, consistent with stabilizing selection. However, beyond the studies described earlier, evidence of negative selection and stabilizing selection acting within populations and across species ranges is limited.

There is so far little evidence of within-population balancing selection on GWAS alleles in plants, in part because there have been few GWAS conducted within plant populations. However, a number of GWAS have been conducted on range-wide samples and have uncovered extensive evidence of between-population balancing selection, or local adaptation. An exemplary study showing local adaptation using GWAS comes from Brachi *et al.* (2015), who linked SNPs associated with glucosinolate composition in *Arabidopsis thaliana* leaves to signatures of local adaptation (F_{ST}) and fitness in field conditions, showing that strong divergent selection maintains variation in glucosinolate composition across European *A. thaliana* populations.

A number of additional studies have also demonstrated local adaptation for GWAS candidates in plant populations. One strategy commonly used involves investigating the distribution of alleles associated with a trait of interest across the landscape. An example of this approach comes from Filiault & Maloof (2012), who identified an SNP strongly associated with shade

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

avoidance and found that alternate alleles at this SNP often occurred in neighboring populations. This pattern suggests that selection on shade-response phenotypes is population-specific, possibly due to differences in shade-environment between populations. Additional examples of local adaptation inferred from the distribution of alleles at GWAS SNPs in *A. thaliana* include observations that the distributions of alleles associated with flowering time and fitness follow a latitudinal cline (Li *et al.*, 2010), that an allele associated with lower saline adaptation is more common in populations located farther from the ocean (Baxter *et al.*, 2010), and that genotypes of loci affecting gene-body methylation in *trans* are correlated with latitude (Dubin *et al.*, 2015). An extension of these approaches was used by Hancock *et al.* (2011), who explicitly plotted the geographical distribution of SNPs associated with climate variables. Going further, Fournier-Level *et al.* (2011) combined GWAS results with the logic of a reciprocal transplant experiment, observing that alleles associated with fitness in a particular common garden were more likely to be found near that common garden than genomic controls, consistent with local adaptation. Collectively, these studies show that much of the trait variation mapped through range-wide GWAS in plants appears to be consistent with local adaptation.

Consistent with a view of prevalent local adaption in GWAS loci, a number of studies have also detected population genetic signatures of adaptation around GWAS loci. For example, SNPs associated with flowering time, glucosinolate level and other defense phenotypes in *A. thaliana* are enriched for F_{ST} outliers (Horton *et al.*, 2012; Brachi *et al.*, 2015), as are some SNPs associated with salt tolerance in African rice (Meyer *et al.*, 2016). Selective sweep signatures of various kinds are enriched near SNPs associated with plant development, climate, and ionomic phenotypes in *A. thaliana* (Hancock *et al.*, 2011; Horton *et al.*, 2012), agronomic traits in Foxtail millet (Jia *et al.*, 2013), life history traits and height in *Populus trichocarpa* (Evans *et al.*, 2014), environmental variables in *M. truncatula* (Yoder *et al.*, 2014), and salt tolerance in rice (Meyer *et al.*, 2016). Overall, there are a number of examples of elevated signals of adaptation surrounding GWAS loci, supporting the hypothesis that local adaptation shapes genetic variation, although it is important to keep in mind that this pattern is not always observed.

IV. Challenges of connecting GWAS to selection

The research program of finding loci associated with traits through GWAS and testing for signatures on these loci seems relatively straightforward, at least in principle. Can we now conclude that local adaptation is largely responsible for maintaining plant variation at the range-wide scale while negative selection likely predominates within populations? We argue that although this approach is clearly promising, there are a number of pitfalls that need to be avoided. In addition to problems stemming from the limited number of studies to date, a number of biases in the ascertainment of GWAS SNPs and the types of samples used for current GWAS color the conclusions we can currently make. We outline some of these many challenges later.

1. Biases in the allele frequencies and effect sizes detected by GWAS

The allele frequencies and effect sizes of GWAS loci are unlikely to clearly reflect the underlying distributions of all of the alleles controlling trait variation. Association mapping is likely inherently biased towards finding intermediate frequency alleles and alleles with large effects (Sham & Purcell, 2014; Myles *et al.*, 2009; Box 1) although this pattern is not always straightforward. Additionally, statistical biases such as the winner's curse can cause overestimation of the effect sizes of all alleles, but this overestimation will be worse for rare alleles (Capen *et al.*, 1971; Box 1). Thus, GWAS studies are generally biased towards the identification of common alleles while the rare alleles identified may have larger estimates of effect sizes, even in the absence of selection. These biases can affect not only direct comparisons of allele frequencies and effect sizes but also F_{ST} distributions, and diversity patterns.

One way to get around ascertainment bias is to use permutations to construct a null distribution of the allele frequency spectrum under neutrality. Two studies have compared allele frequencies found through GWAS to a null expectation generated with permuted data (Stanton-Geddes *et al.*, 2013; Josephs *et al.*, 2015). Permutation-based approaches have also been used to control for winner's curse's ability to inflate the allele frequencies of rare alleles when testing for a negative correlation between effect size and allele frequency (Stanton-Geddes *et al.*, 2013). Another useful approach to evaluating the correlation between allele frequency and effect size is subsampling the individuals used in a GWAS in a way to ensure that the associations and effect sizes are estimated from equal numbers of individuals with each genotype, regardless of the population level allele frequency (i.e. downsampling until the association mapping for common alleles is conducted on the same sample size as it is for rare alleles) (Josephs *et al.*, 2015).

Allele frequency biases can also affect F_{ST}, which has implications for using F_{ST} to evaluate evidence for local adaptation at GWAS SNPs. The upper bound of F_{ST} is determined by allele frequency and this bound is higher at moderate allele frequencies (Jakobsson *et al.*, 2013), so any skew towards intermediate allele frequencies in GWAS SNPs may also increase F_{ST}. An alternative may be methods that more explicitly test for increased spatial differentiation in allele frequencies of GWAS SNPs relative to appropriate genomic controls.

2. Controlling for genomic context

Many of the population genetic tests used to identify patterns of selection are sensitive to genomic context. For example, F_{ST} measures the extent of between-population divergence relative to total variation across populations and, as a consequence, F_{ST} is sensitive to forces that reduce total variation, like background selection against neutral mutations linked to deleterious alleles (Charlesworth, 1998; Cruickshank & Hahn, 2014). An alternative may be using absolute measures of divergence that are not scaled by total divergence and so are not sensitive to variation in background selection (as suggested by Charlesworth, 1998). Tests for selective sweeps may also be subject to similar types of biases, since variation

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

Box 1 Statistical and genetic biases in mapping

Mapping the genetic basis of ecologically important traits, with either QTL mapping or GWAS, entails several statistical challenges. It is very important to be aware of these before making general conclusions about the genetic architecture of traits (Rockman, 2012). In addition, these biases can complicate subsequent tests of evolutionary and ecological hypotheses about the genetic loci detected, often with subtle but pervasive effects.

Common statistical biases in GWAS

Crucially, observed patterns of genetic architecture may be affected by numerous statistical biases. Given a certain effect size, SNPs with intermediate allele frequencies will explain more phenotypic variance than SNPs with low allele frequencies, increasing the power of a GWAS to detect intermediate-frequency alleles (Myles *et al.*, 2009; Sham & Purcell, 2014). At the same time, GWAS will also have more power to detect alleles with larger phenotypic effects. Increased power to detect common alleles and alleles with large effect sizes will cause an overestimation of the average effect size and frequency of GWAS-associated alleles. In contrast to this, skewed distributions of traits when run through standard parametric GWAS can lead to an excess of false positives from low-frequency alleles (Brachi *et al.*, 2011), complicating intuitive conclusions about how the allele frequencies of GWAS SNPs will reflect the underlying frequency of loci controlling variation at a trait. Strategies to disentangle power and allele frequencies include subsampling GWAS populations to equalize allele frequencies (see Josephs *et al.*, 2015) or explicit modeling that accounts for frequency and power. Ultimately, any conclusions made about the genetic architecture of traits uncovered through GWAS should be mindful of the statistical biases inherent in GWAS.

Winner's curse

The phrase 'winner's curse' has its roots in economics and reflects dynamics of competitive auction bidding scenarios, in which the winner of the auction is often the individual that overestimated the value of the item, and hence all auction winners overpay (Capen *et al.*, 1971). Within the context of GWAS and genetic mapping, for studies of limited power, loci that are identified as statistically significant will often have their effect sizes overestimated because the data that go into detecting a region or locus as significant in the first place are the same as the data used for estimating effect sizes (Göring *et al.*, 2001; Zöllner & Pritchard, 2007). As a consequence, with limited power or sample size, one only detects a locus as significant if the effect size in the analyzed sample is greater than in the general population (otherwise, one would not have detected it in the first place). One potentially pernicious consequence of the winner's curse is that follow up or validation studies may have too low a sample size and therefore insufficient power, and so fail to replicate initial associations (Ioannidis *et al.*, 2001). For testing evolutionary hypotheses that depend on effect sizes (e.g. distribution of effect sizes of beneficial or deleterious alleles, relationships between effect size and allele frequency), distinguishing biological results from the winner's curse represents a substantial challenge.

Beavis effect

The Beavis effect is named after a series of papers by William Beavis focused on QTL mapping (Beavis, 1994, 1998). Beavis showed through a series of simulations that limited sample size in QTL mapping populations had several adverse consequences on the inferences one could make. First, limited sample sizes lead to a serious underestimation of the number of QTLs affecting a trait, and second, they lead to an overestimation of the effect size of any single QTL. Beavis showed that the former issue is related to the heritability of the traits analyzed and the number of progeny evaluated: as heritability of the traits and sample sizes increased, the power to detect QTLs increased. For example, the power to detect 10 QTLs in a population of 500 F_2 individuals was 0.57 for traits that had a heritability of 0.3, and 0.86 for traits with a heritability of 0.65; for 1000 F_2 individuals, these power values were 0.85 and 0.99. As expected, with greater numbers of QTLs affecting a trait, power declines. Unlike winner's curse, the prominent bias in effect size estimation described by Beavis also has a biological, not statistical, explanation: with fewer individuals, and hence fewer recombination events represented in the population, multiple loci that affect a trait will be mis-identified as a single QTL with larger effect. Beavis (1994) described this as a problem of multicollinearity between the QTLs affecting a trait in small sample. Using larger sample sizes, and hence more recombination events, will lead to fewer QTLs being co-inherited, allowing their individual effects of to be distinguished.

Noor effect

Noor *et al.* (2001) also used simulation to examine how the local genomic environment affected the detection, and effect size, of QTLs. They showed convincingly that QTLs tend to be detected in regions of low recombination in the genome (e.g. near centromeres). In their simulations, multiple QTLs in close proximity to centromeric or low recombination regions were often identified as single QTL with larger effects. They found these results to be robust to the effect size of the QTL or the heritability of the trait (although they also found that higher heritability led to more QTLs being detected). Noor *et al.* (2001) suggest that these results are an inevitable consequence of variation in gene density per centimorgan across the genome.

Population structure

The presence of population structure in a GWAS panel can have important consequences, especially if trait variation covaries with structure. Inadequately controlling for structure in this case will lead to a high number of false-positives (Atwell *et al.*, 2010) because the distribution of many neutral SNPs will match trait variation. Current plant GWAS often use mixed models that incorporate a matrix of relatedness between individuals to control for false-positives due to structure (Thornsberry *et al.*, 2001; Yu *et al.*, 2006; Kang *et al.*, 2010; Zhou & Stephens, 2012). However, it may be impossible to detect true associations if structure and trait are entirely confounded, which can be an especially tough problem in predominantly selfing species (Zhao *et al.*, 2007).

Dealing with effect size biases

Several ways forward exist. For example, in human GWAS, it is now common to identify SNPs or loci as involved in phenotypes in one sample, and then estimate their effects in a second sample (see Kraft *et al.*, 2009 for an overview). The two-sample approach eliminates the double-testing problem inherent in identifying and estimating effects sizes with the same data. Using two samples will often be a challenge in ecological and

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

Box 1 (continued)

evolutionary contexts, and as a consequence alternatives are needed. Several alternatives can be pursued. For example, rather than identifying single SNPs at a time, it may be more profitable to model the contribution of all SNPs, simultaneously to the phenotype (Visscher *et al.*, 2010; Yang *et al.*, 2010). These analyses are not subject to the winner's curse, and often explain substantially more variation in phenotypic traits than SNP at a time GWAS style analyses.

The most straightforward approach to dealing with effect size biases is recognize the consequences of these artifacts, plan accordingly, and to alter interpretations and subsequent work in light of them (Albert *et al.*, 2008; Slate, 2013). While some of the effects Beavis analyzed can be addressed through analytical tools (QTL mapping approaches, statistical analysis), the most fundamental issue, as originally emphasized by Beavis, is the experimental design: the number of recombinant progeny that are in a given mapping experiment, and experimental conditions used to measure the traits. Once an experiment is done, caution about the number and distribution of QTL effect sizes is in order, especially if the goal is to compare to theoretical predictions generated using different models.

in linkage disequilibrium (LD) across the genome could both increase GWAS power, by combining the signal of multiple causal loci, and make it easier to detect selective sweeps. The net result of combining the effects of multiple causal loci and the appearance of a selective sweep signal could cause a spurious overlap between GWAS significance and selective sweeps, so tests for an enrichment of selective sweeps in a set of loci should be careful to control for LD variation across the genome. In general, tests for an overlap between various population genetic parameters and GWAS SNPs should be mindful of confounders.

3. Does species choice limit conclusions made from GWAS?

Plant GWAS have generally been able to map loci that explain significant amounts of trait variation (Atwell *et al.*, 2010; Brachi *et al.*, 2011; Peiffer *et al.*, 2014; Sasaki *et al.*, 2015), especially compared with GWAS in humans, where smaller effect sizes are more commonly observed (Visscher *et al.*, 2012). There are a few possible, nonexclusive explanations for the relative success of plant GWAS. First, plant GWAS can be conducted in controlled environments and, often, with replicated lines, so less environmental variation contributes to measurements of plant traits than those for human GWAS. Second, plant GWAS have almost exclusively been conducted on range-wide samples, where we expect local adaptation to explain a large amount of standing variation, while human GWAS may be mainly mapping variation within populations that is likely to be at mutation–selection–drift balance. Third, the focus of many human GWAS on disease risk could also contribute to lower amounts of variation explained in human GWAS, because the alleles that increase disease risk in humans are likely to be difficult to detect with GWAS (Eyre-Walker, 2010; Maher *et al.*, 2012).

Although the genetic architecture of traits revealed by GWAS clearly differs between plants and humans, there are also significant differences between plant species in the types of variants uncovered by GWAS. Many of these can be explained by mating system; specifically, GWAS in selfers tend to find few large-effect alleles explaining trait variation whereas GWAS in outcrossers tend to find a larger number of alleles and these alleles have smaller effect sizes (Tian *et al.*, 2011; Huang *et al.*, 2012). Mating system could affect observed genetic architecture in multiple ways. First, there may be stronger selection against large-effect flowering-time alleles in outcrossers than selfers, because outcrossers with aberrant flowering times are less likely to be able to mate. This prediction is consistent with observations that the majority of the genetic variation for flowering time in the selfing *A. thaliana* seems to be controlled by a relatively small number of loci with large effects (Sasaki *et al.*, 2015) but by many small-effect loci in maize, an outcrosser (Buckler *et al.*, 2009). Second, mating system may affect the types of genetic variation present within populations. Specifically, theory predicts that stabilizing selection will maintain less within-population genetic variation in selfers than outcrossers (Charlesworth & Charlesworth, 1995; Lande & Porcher, 2015). Third, LD generally decays more slowly in selfers than outcrossers, so GWAS in selfers may be more likely to uncover large-effect combinations of small-effect alleles (Huang & Han, 2014). Finally, selfing species generally have stronger population structure, reducing power to detect associations for traits that have values correlated with structure (Atwell *et al.*, 2010; Platt *et al.*, 2010; Box 1).

However, it is important to note that many of the current observations made about genetic architecture uncovered through GWAS in outcrossing plants comes from maize, specifically GWAS conducted on the Nested Association Mapping panel (NAM). The NAM is made up of the offspring of controlled crosses between 25 diverse maize inbred lines and the same reference line (Yu *et al.*, 2008). Essentially, this means that the population is made up of 25 distinct subpopulations of recombinant inbred lines that share a common parent. The resulting lines allow well-powered mapping, because even variants with low population-level frequency are relatively well-replicated within the subpopulations, although variants at frequencies below 1/25 will not be well-represented. For traits that have been mapped in both the NAM and natural populations, such as height, the NAM has found alleles of smaller effect due to higher power (Peiffer *et al.*, 2014). The increased power seen in the NAM makes it difficult to disentangle GWAS sampling strategy from mating system as long as GWAS studies on outcrossing species focus mainly on the NAM in maize. Additional GWAS conducted in outcrossing plant species will be needed to further investigate the role of mating system in shaping the genetic architecture of quantitative traits and, ultimately, whether the selective forces maintaining trait variation differ between selfers and outcrossers.

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

4. Are crop species representative?

Many of the plant GWAS conducted to date use crop systems such as maize (Buckler *et al.*, 2009; Brown *et al.*, 2011; Peiffer *et al.*, 2014; Wallace *et al.*, 2014), rice (Zhao *et al.*, 2011; Huang *et al.*, 2012; Chen *et al.*, 2014; Begum *et al.*, 2015; Meyer *et al.*, 2016; Yano *et al.*, 2016), sorghum (Morris *et al.*, 2013; Lasky *et al.*, 2015), and foxtail millet (Jia *et al.*, 2013). There are some clear benefits of investigating variation in crops while trying to address questions about the maintenance of genetic variation. First, due to breeders' interest in incorporating natural genetic variation, crop GWAS tend to be conducted in large, diverse panels, providing lots of genetic variation to explore. Second, many of the selective agents that have shaped genetic diversity in crop plants are known, providing clear hypotheses to test. For example, maize has many fewer lateral branches than its wild progenitor, teosinte, and this has been linked to selection at a number of loci (Doebley *et al.*, 1997; Clark *et al.*, 2004). Finally, our intuition is that the selection involved in domestication and breeding is quite strong, making it detectable, but at the same time domestication often involves adaptation at many loci (Morrell *et al.*, 2012; Meyer & Purugganan, 2013), making these systems useful for investigating adaptation in quantitative traits.

If the selective forces acting on trait variation in crops differ from those acting on variation in natural populations, conclusions about the forces maintaining variation made only in crops will be misleading. The specific traits chosen for investigation are important here. Adaptive traits in crops can be divided up into those that were selected for during domestication (domestication traits), and those that were selected for during breeding and vary across the range of the domesticated species (diversification traits) (Meyer & Purugganan, 2013). The timescales and strength of selection on domestication and diversification traits likely differ (Meyer & Purugganan, 2013). Diversification traits can include analogs to those that might vary in wild populations, such as photoperiod sensitivity, but will also include characters that were consciously selected for by humans and may not have analogous natural variation, such as the popping phenotype in maize (Meyer & Purugganan, 2013). Because most GWAS panels focus on domesticated lines, diversification traits are the ones most likely to be investigated during plant GWAS, although there is evidence that when variation in domestication-related traits does persist in domesticated species, this variation is mainly due to small-effect alleles that escaped selection during domestication (Xue *et al.*, 2016). It is unclear if conclusions made about variation in domestication traits or those diversification traits without analogs in natural populations will be applicable to questions about what maintains trait variation in nature.

The process of domestication could also have important consequences for genomic variation. For example, domestication often involves complex demographic changes that could alter population structure (Morrell *et al.*, 2012) or other aspects of genomic variation such as allele frequencies and patterns of linkage disequilibrium, affecting association mapping efficacy (Lohmueller, 2014). In addition, there is evidence that domesticated species have more deleterious mutations than their wild progenitors, potentially due to the repeated bottlenecks that occur during domestication (Renaut & Rieseberg, 2015), which will shape the types of variation present within domesticated species. In sum, while investigations of selection in crop species have been and continue to be useful, there are also many factors that suggest that the evolutionary forces maintaining variation in crop plants likely differ from those acting in wild populations.

5. The samples used for GWAS matter

The choice of what populations to include in GWAS matters for evolutionary conclusions made from these GWAS. Most plant GWAS have been conducted on range-wide samples that sometimes even include multiple closely-related species (for e.g. Huang *et al.*, 2012; Chen *et al.*, 2014). The lack of within-population plant GWAS is likely due to both the field's history of interest in crop breeding and local adaptation and a focus on selfing species. Selfing species are especially tractable for GWAS because they allow researchers to develop inbred lines that can be sequenced once but phenotyped multiple times, allowing for replication and the evaluation of multiple traits. However, the focus on selfing species limits our understanding of within-population processes because these species do not form large interbreeding populations.

There are a couple of sampling strategies that can be used to examine the maintenance of variation within populations without the need to generate expensive sequence data from a single population that can be used only once. First, denser sampling of selfing lines in a limited geographical region such as panels of *A. thaliana* collected from Sweden (Long *et al.*, 2013), has the potential to uncover variation maintained at smaller geographic scales, more closely mimicking the types of dynamics seen within populations, while allowing researchers to take advantage of inbred lines that can be used in multiple experiments. Second, crossing schemes that create inbred lines from an outbred population sample will be useful in species that are primarily outcrossing but also self-compatible, such as *M. guttatus* (Wu *et al.*, 2008). However, outcrossing species can maintain many deleterious recessive alleles and care should be taken to consider the effects that the exposure of deleterious recessive alleles in selfed lines will have on conclusions made from these studies.

In addition, choices made during species-wide sampling efforts may also affect conclusions made about local adaptation. For example, relatively sparse samples conducted on a continental or worldwide scale may miss regional patterns of local adaptation (Anderson *et al.*, 2015). Overall, it is clear that, as in many experiments, the choice of what populations, and at what scale, to sample in a GWAS will have important implications for results and these choices should be thought through carefully.

6. Allelic heterogeneity

GWAS analyses typically test for independent associations between phenotype and one of two genotypes at a SNP. Allelic heterogeneity, the presence of more than two alleles with distinct phenotypic effects at a locus, will make this locus difficult to detect using

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

GWAS. We will describe an example of this phenomenon from Yano *et al.* (2016) to illustrate the general problem. Yano *et al.* (2016) conducted a GWAS for days to heading (flowering time) in a population of *japonica* rice (*Oryza sativa*) varieties and found that one of the strongest association peaks was located *c.* 1 megabase (Mb) from *Hd1*, a gene previously shown to be important for heading date variation. Their sample had 11 haplotypes present at *Hd1*, some of which contained previously-reported null and intermediate alleles for *Hd1* function. The presence of multiple haplotypes and causal alleles at *Hd1* meant that tests for the independent effects of individual SNPs in *Hd1* on days to heading were not significant. However, a region with high LD located *c.* 1 Mb from the *Hd1* gene effectively tagged functional variation in *Hd1*, creating the signal of an association in this location (Yano *et al.*, 2016).

It seems clear that understanding the extent of allelic heterogeneity will be important for understanding our ability to map causal loci using GWAS but allelic heterogenity is difficult to detect using GWAS data alone. The use of a common reference parent in the NAM makes it straightforward to detect series of multiple alleles affecting a trait by looking for parental alleles that have opposite effects relative to the reference parent, and allelic heterogeneity has been observed for a number of traits in QTL detected in mapping studies using the NAM (Buckler *et al.*, 2009; Kump *et al.*, 2011; Cook *et al.*, 2012). In addition, haplotype-based GWAS for eQTLs in *Drosophila melanogaster* have shown that allelic heterogeneity is common (King *et al.*, 2014). If genetic variation at a certain trait is more likely to involve allelic heterogeneity than for other traits, this will lead to an underestimate of effect sizes at heterogeneous loci (Thornton *et al.*, 2013), having important consequences for conclusions made about the selective pressures acting on these loci. Overall, studies attempting to detect selection on GWAS loci should be mindful of the potential effects of allelic heterogeneity.

7. Detecting polygenic adaptation

While plant GWAS have been successful at identifying local adaptation involving changes at a few loci, efforts for identifying polygenic adaptation in plant GWAS SNPs have been limited. One approach used successfully in humans tests for covariance in the direction of allelic effects across populations (Turchin *et al.*, 2012; Berg & Coop, 2014; Robinson *et al.*, 2015), as expected if local adaptation occurs polygenically (Le Corre & Kremer, 2012). Specifically, the allele frequencies and effect sizes of GWAS hits are summarized by constructing predicted phenotypes for populations in a separate genotyping panel and an excess of differentiation in the predicted phenotypes is indicative of local adaptation (Berg & Coop, 2014; Robinson *et al.*, 2015). In addition, a new method that detects very recent shifts in allele frequency due to directional selection could be useful for identifying recent local adaptation by testing for coordinated allele frequency shifts in GWAS SNPs (Field *et al.*, 2016). Applying these new approaches for detecting polygenic adaptation will likely need additional genomic sequencing beyond what is normally conducted for plant GWAS, but should already be feasible in *A. thaliana* and many crop species. However, applying methods developed for humans to plant GWAS will require grappling with how population structure in the plant GWAS sampling populations will affect signatures of polygenic adaptation (Berg & Coop, 2014).

V. Potential ways forward

Although GWAS have, to date, provided insight into the maintenance of variation in plant species, their potential is still relatively untapped and many biases limit the conclusions we can make. In this review we have outlined theoretical predictions for the types of variation we expect to be maintained under different types of selection, described our current understanding of the maintenance of variation based on GWAS, and outlined the challenges that remain and highlighted some solutions to these problems. Below we suggest a few additional ways forward that appear promising.

There are still many limits to using GWAS to address the maintenance of variation within populations. In particular, the GWAS conducted to date have limited power to detect rare alleles. Fully understanding the role of mutation–selection balance in maintaining variation within species will require comprehensive surveys of rare variants. Although increasing sample sizes is a clear way to improve power to detect associations with rare alleles, there will always be a lower bound in the allele frequency detectable by GWAS. A few alternative approaches to detecting the presence of rare QTLs and their properties have been developed for mapping disease traits in humans and could be applied to plants. One approach uses targeted sequencing at a few candidate genes in a very large sample to detect associations between rare alleles and a disease or to test for an enrichment of rare variants in the candidate genes. However, these experiments have had mixed success in detecting associations (Rivas *et al.*, 2011; Hunt *et al.*, 2013; Purcell *et al.*, 2014), and failures to see a significant role of rare deleterious variants may occur because either rare variants do not contribute to the disease in question or because these studies have missed causal rare variants in noncandidate genes and noncoding regions. Additional approaches to detecting rare variants involved in disease are discussed in Zuk *et al.* (2014) and S. Lee *et al.* (2014).

One way to cope with low power to detect rare variants is to combine information from multiple traits. A number of rare variant-based approaches that take advantage of information gained from looking at a large number of traits have been applied to human gene expression. Zeng *et al.* (2015) detected individual human cell lines with aberrant gene expression and showed that aberrant expression was associated with an excess of rare variants in and near genes, consistent with the action of negative selection. Similarly, Zhao *et al.* (2016) found an excess of rare variants in promoters in individuals with extremely high or low expression levels, consistent with the hypothesis that rare variants under negative selection affect gene expression. Li *et al.* (2014) identified genes with putative rare eQTLs by looking for genes with eQTLs that had effects in a single family but lacked common, large effect *cis*-eQTLs in a population level sample and also had rare potentially functional variants near the transcription start site. Genes with rare *cis*-eQTLs had lower dN/dS than genes with common *cis*-eQTLs,

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

suggesting that selection has reduced the allele frequencies of the *cis*-eQTLs affecting evolutionarily constrained genes (Li *et al.*, 2014). All of these approaches could be attempted in plants and in traits other than gene expression. In particular, the approach of Li *et al.* (2014) could be adapted to use QTLs identified in controlled crosses that do not segregate in larger population samples as candidate rare QTLs.

By contrast, our ability to detect within-population balancing selection acting on GWAS loci is limited not by power to detect associations but by both the availability of data from within-population GWAS and a restricted ability to detect the signature of this type of balancing selection in sequence data. There is certainly room for the development of improved model-based approaches. In addition, explicit examination of allele frequency change over time (e.g. Bergland *et al.*, 2014) should enhance the power to detect and quantify the role of temporally-variable selection in maintaining variation in the same way that investigations of spatial variation in frequencies of trait-associated SNPs have been so useful in identifying local adaptation.

It is worth noting that it may not be necessary to directly map the SNPs affecting traits at all in order to make inferences about genetic architecture. Some approaches skip the step of identifying SNPs associated with traits and, instead, estimate the amount of genetic variance contributed by all or a subset of SNPs (Yang *et al.*, 2011; Gusev *et al.*, 2014). The benefit of estimating variance directly is that it allows inferences to be made from the small-effect SNPs that cannot be independently detected due to low power in GWAS (Visscher *et al.*, 2010). Estimates of the genetic variation explained by genotype state at all loci have been successful at identifying much of the genetic variation observed in human height and body mass index and showing that, for height, rare alleles explain more variation than common alleles, consistent with the action of selection (Yang *et al.*, 2010, 2015). These types of approaches are already being used in plants to evaluate the functional importance of various types of annotations in maize (Rodgers-Melnick *et al.*, 2016), the factors contributing to different types of methylation variation in *A. thaliana* (Dubin *et al.*, 2015), and the number of loci involved in trait variation in sorghum (Lasky *et al.*, 2015), but could be used to investigate allele frequency biases or other population genetic parameters within plants.

In conclusion, GWAS in plants have uncovered a large sample of loci that contribute to quantitative genetic variation within species and allowed us to begin to evaluate the role of negative selection contributing to within-population genetic variation and of local adaptation in contributing to species-wide genetic variation. However, ascertainment biases and power strongly determine the variants revealed by GWAS and interpretations made from these variants still require great care. There may not be a one-size-fits-all solution or clear roadmap to answering questions about the maintenance of variation using GWAS results, and instead, the best approaches may be system- and question-specific. Future work to both expand the scope of GWAS conducted in plants and to develop methods that carefully test for the effects of various evolutionary scenarios will be needed to explain the evolutionary forces maintaining variation within plant species.

Acknowledgements

We thank Graham Coop, Jeremy Berg, and the Ross-Ibarra lab for helpful comments and suggestions, as well as Lynda Delph and three anonymous reviewers. E.B.J. is supported by an NSF Postdoctoral Research Fellowship in Biology under Grant no. 1523733. J.R.S. and S.I.W. are supported by Discovery Grants from NSERC Canada.

References

Albert AY, Sawaya S, Vines TH, Knecht AK, Miller CT, Summers BR, Balabhadra S, Kingsley DM, Schluter D. 2008. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* **62**: 76–85.

Anderson JT, Perera N, Chowdhury B, Mitchell-Olds T. 2015. Microgeographic patterns of genetic divergence and adaptation across environmental gradients in *Boechera stricta* (brassicaceae). *American Naturalist* **186**: S60–S73.

Andolfatto P, Nordborg M. 1998. The effect of gene conversion on intralocus associations. *Genetics* **148**: 1397–1399.

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al.* 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627–631.

Barton N, Turelli M. 1989. Evolutionary quantitative genetics: how little do we know? *Annual Review of Genetics* **23**: 337–370.

Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**: 11–21.

Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M *et al.* 2010. A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1. *PLoS Genetics* **6**: e1001193.

Beavis W. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proceedings of the forty-ninth annual corn and sorghum industry research conference.* Chicago, IL, USA: American Seed Trade Association, 250–266.

Beavis WD. 1998. QTL analyses: power, precision, and accuracy. In: Paterson AH, ed. *Molecular dissection of complex traits.* Boca Raton: CRC Press, 145–162.

Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, Virk P, Collard B, McCouch SR. 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE* **10**: e0119873.

Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genetics* **10**: e1004412.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genetics* **10**: e1004775.

Berry AJ, Ajioka J, Kreitman M. 1991. Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. *Genetics* **129**: 1111–1117.

Brachi B, Meyer CG, Villoutreix R, Platt A, Morton TC, Roux F, Bergelson J. 2015. Coselected genes determine adaptive variation in herbivore resistance throughout the native range of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **112**: 4032–4037.

Brachi B, Morris GP, Borevitz JO. 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology* **12**: 232.

Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES *et al.* 2011. Distinct genetic architectures for male and female inflorescence traits of maize. *PLoS Genetics* **7**: e1002383.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC *et al.* 2009. The genetic architecture of maize flowering time. *Science* **325**: 714–718.

Capen EC, Clapp RV, Campbell WM. 1971. Competitive bidding in high-risk situations. *Journal of Petroleum Technology* **23**: 641–653.

Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**: 538–543.

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* **2**: e64.

Charlesworth D, Charlesworth B. 1995. Quantitative genetics in plants: the effect of the breeding system on genetic variability. *Evolution* **49**: 911–920.

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H *et al.* 2014. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature Genetics* **46**: 714–721.

Chevin LM, Hospital F. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **180**: 1645–1660.

Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proceedings of the National Academy of Sciences, USA* **101**: 700–707.

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA. 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology* **158**: 824–834.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**: 3133–3157.

DeGiorgio M, Lohmueller KE, Nielsen R. 2014. A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genetics* **10**: e1004561.

Delph LF, Kelly JK. 2014. On the importance of balancing selection in plants. *New Phytologist* **201**: 45–56.

Doebley J, Stect A, Hubbardt L. 1997. The evolution of apical dominance in maize. *Nature* **386**: 485–4888.

Dubin MJ, Zhang P, Meng D, Remigereau MS, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B *et al.* 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *Elife* **4**: e05255.

Evans LM, Slavov GT, Rodgers-Melnick E, Martin J, Ranjan P, Muchero W, Brunner AM, Schackwitz W, Gunter L, Chen JG *et al.* 2014. Population genomics of *Populus trichocarpa* identifies signatures of selection and adaptive trait associations. *Nature Genetics* **46**: 1089–1096.

Eyre-Walker A. 2010. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences, USA* **107**: 1752–1756.

Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, Yengo L, Rocheleau G, Froguel P, McCarthy MI *et al.* 2016. Detection of human adaptation during the past 2,000 years. *bioRxiv*: e052084.

Filiault DL, Maloof JN. 2012. A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. *PLoS Genetics* **8**: e1002589.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* **334**: 86–89.

Gillespie JH. 2010. *Population genetics: a concise guide. 2nd edn.* Baltimore, MD, USA & London, UK: JHU Press.

Gillespie JH, Turelli M. 1989. Genotype–environment interactions and the maintenance of polygenic variation. *Genetics* **121**: 129–138.

Göring HH, Terwilliger JD, Blangero J. 2001. Large upward bias in estimation of locus-specific effects from genome-wide scans. *American Journal of Human Genetics* **69**: 1357–1369.

Gusev A, Lee SH, Trynka G, Finucane H, Vilhjálmsson BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E *et al.* 2014. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *American Journal of Human Genetics* **95**: 535–552.

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**: 83–86.

Harrison TL, Wood CW, Borges IL, Stinchcombe JR. 2016. Combining phenotypic and genomic approaches reveals no evidence for adaptation to the local mutualist in *Medicago lupulina*. *bioRxiv*. doi: 10.1101/078675.

Hedrick PW. 2006. Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology Evolution and Systematics* **37**: 67–93.

Hereford J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *American Naturalist* **173**: 579–588.

Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**: 2335–2352.

Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A, Sperone FG, Vilhjálmsson BJ *et al.* 2012. Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics* **44**: 212–216.

Hough J, Williamson RJ, Wright SI. 2013. Patterns of selection in plant genomes. *Annual Review of Ecology, Evolution, and Systematics* **44**: 31–49.

Huang X, Han B. 2014. Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology* **65**: 531–551.

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C *et al.* 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* **44**: 32–39.

Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O *et al.* 2013. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* **498**: 232–235.

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. 2001. Replication validity of genetic association studies. *Nature Genetics* **29**: 306–309.

Jakobsson M, Edge MD, Rosenberg NA. 2013. The relationship between F_{ST} and the frequency of the most frequent allele. *Genetics* **193**: 515–528.

Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H *et al.* 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature Genetics* **45**: 957–961.

Johnson T, Barton N. 2005. Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society B* **360**: 1411–1425.

Josephs EB, Lee YW, Stinchcombe JR, Wright SI. 2015. Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences, USA* **112**: 15390–15395.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**: 348–354.

Kaplan NL, Darden T, Hudson RR. 1988. The coalescent process in models with selection. *Genetics* **120**: 819–829.

Keightley PD, Eyre-Walker A. 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data?. *Philosophical Transactions of the Royal Society B* **365**: 1187–1193.

Kelly JK. 1999. An experimental method for evaluating the contribution of deleterious mutations to quantitative trait variation. *Genetical Research* **73**: 263–273.

Kelly JK, Willis JH. 2001. Deleterious mutations and genetic variation for flower size in *Mimulus guttatus*. *Evolution* **55**: 937–942.

King EG, Sanderson BJ, McNeil CL, Long AD, Macdonald SJ. 2014. Genetic dissection of the *Drosophila melanogaster* female head transcriptome reveals widespread allelic heterogeneity. *PLoS Genetics* **10**: e1004322.

Kraft P, Zeggini E, Ioannidis JPA. 2009. Replication in genome-wide association studies. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* **24**: 561–573.

Kremer A, Le Corre V. 2012. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* **108**: 375–385.

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D *et al.* 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics* **43**: 163–168.

Lande R, Porcher E. 2015. Maintenance of quantitative genetic variance under partial self-fertilization, with implications for evolution of selfing. *Genetics* **200**: 891–906.

Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, Juenger TE, Hyma K, Acharya C, Mitchell SE *et al.* 2015. Genome–environment associations in sorghum landraces predict adaptive traits. *Science Advances* **1**: e1400218.

Latta RG. 1998. Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *American Naturalist* **151**: 283–292.

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

Le Corre V, Kremer A. 2003. Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* **164**: 1205–1219.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* **21**: 1548–1566.

Lee S, Abecasis GR, Boehnke M, Lin X. 2014. Rare-variant association analysis: study designs and statistical tests. *American Journal of Human Genetics* **95**: 5–23.

Lee YW, Fishman L, Kelly JK, Willis JH. 2016. A segregating inversion generates fitness variation in yellow monkeyflower (*Mimulus guttatus*). *Genetics* **202**: 1473–1484.

Lee YW, Gould BA, Stinchcombe JR. 2014. Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB Plants* **6**: plu004.

Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G *et al.* 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* **339**: 1578–1582.

Li X, Battle A, Karczewski KJ, Zappala Z, Knowles DA, Smith KS, Kukurba KR, Wu E, Simon N, Montgomery SB. 2014. Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. *American Journal of Human Genetics* **95**: 245–256.

Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. 2010. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **107**: 21199–21204.

Lohmueller KE. 2014. The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genetics* **10**: e1004379.

Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, Zhang Q, Vilhjálmsson BJ, Korte A, Nizhynska V *et al.* 2013. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics* **45**: 884–890.

Maher MC, Uricchio LH, Torgerson DG, Hernandez RD. 2012. Population genetics of rare variants and complex diseases. *Human Heredity* **74**: 118–128.

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* **28**: 659–669.

Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, Dorph K, Barretto A, Gross B, Fuller DQ *et al.* 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nature Genetics* **48**: 1083–1088.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nature Reviews Genetics* **14**: 840–852.

Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**: 845–856.

Morrell PL, Buckler ES, Ross-Ibarra J. 2012. Crop genomics: advances and applications. *Nature Reviews Genetics* **13**: 85–96.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE *et al.* 2013. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences, USA* **110**: 453–458.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* **21**: 2194–2202.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Research* **15**: 1566–1575.

Noor MA, Cunningham AL, Larkin JC. 2001. Consequences of recombination rate variation on quantitative trait locus mapping studies. Simulations based on the *Drosophila melanogaster* genome. *Genetics* **159**: 581–588.

Ogura T, Busch W. 2015. From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. *Current Opinion in Plant Biology* **23**: 98–108.

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CAC, McMullen MD, Holland JB, Bradbury PJ *et al.* 2014. The genetic architecture of maize height. *Genetics* **196**: 1337–1356.

Pennings PS, Hermisson J. 2006. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics* **2**: e186.

Platt A, Vilhjálmsson BJ, Nordborg M. 2010. Conditions under which genome-wide association studies will be positively misleading. *Genetics* **186**: 1045–1052.

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology* **20**: R208–R215.

Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, Chambert K, Bergen SE, Kähler A *et al.* 2014. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**: 185–190.

Renaut S, Rieseberg LH. 2015. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. *Molecular Biology and Evolution* **32**: 2273–2283.

Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, Boucher G, Ripke S, Ellinghaus D, Burtt N *et al.* 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics* **43**: 1066–1073.

Robinson MR, Hemani G, Medina-Gomez C, Mezzavilla M, Esko T, Shakhbazov K, Powell JE, Vinkhuyzen A, Berndt SI, Gustafsson S *et al.* 2015. Population genetic differentiation of height and body mass index across Europe. *Nature Genetics* **47**: 1357–1362.

Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**: 1–17.

Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. 2016. Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences, USA* **113**: E3177–E3184.

Sasaki E, Zhang P, Atwell S, Meng D, Nordborg M. 2015. "Missing" G × E variation controls flowering time in *Arabidopsis thaliana*. *PLoS Genetics* **11**: e1005597.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* **14**: 807–820.

Schrider DR, Mendes FK, Hahn MW, Kern AD. 2015. Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics* **200**: 267–284.

Sham PC, Purcell SM. 2014. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics* **15**: 335–346.

Slate J. 2013. From Beavis to beak color: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution* **67**: 1251–1262.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genetical Research* **23**: 23–35.

Stanton-Geddes J, Paape T, Epstein B, Briskine R, Yoder J, Mudge J, Bharti AK, Farmer AD, Zhou P, Denny R *et al.* 2013. Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS ONE* **8**: e65688.

Stinchcombe J, Hoekstra H. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.

The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA *et al.* 2015. A global reference for human genetic variation. *Nature* **526**: 68–74.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics* **28**: 286–289.

Thornton KR, Foran AJ, Long AD. 2013. Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. *PLoS Genetics* **9**: e1003258.

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics* **43**: 159–162.

Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**: 673–680.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M *et al.* 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* **39**: 31–40.

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust

Toomajian C, Hu TT, Aranzana MJ, Lister C, Tang C, Zheng H, Zhao K, Calabrese P, Dean C, Nordborg M. 2006. A nonparametric test reveals selection for rapid flowering in the Arabidopsis genome. *PLoS Biology* **4**: e137.

Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Hirschhorn JN. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* **44**: 1015–1019.

Turelli M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology* **25**: 138–193.

Turelli M, Barton NH. 2004. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G × E interactions. *Genetics* **166**: 1053–1079.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *American Journal of Human Genetics* **90**: 7–24.

Visscher PM, Yang J, Goddard ME. 2010. A commentary on "Common SNPs explain a large proportion of the heritability for human height" by Yang *et al.* (2010). *Twin Research and Human Genetics* **13**: 517–524.

de Vladar HP, Barton N. 2014. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* **197**: 749–767.

Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genetics* **10**: e1004845.

Whitlock MC, Lotterhos KE. 2015. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. *American Naturalist* **186**: S24–S36.

Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for widespread positive and negative selection in coding and conserved noncoding regions of *Capsella grandiflora*. *PLoS Genetics* **10**: e1004622.

Wu C, Lowry D, Cooley A, Wright K, Lee Y, Willis J. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* **100**: 220–230.

Xue S, Bradbury PJ, Casstevens T, Holland JB. 2016. Genetic architecture of domestication-related traits in maize. *Genetics* **204**: 99–113.

Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, Robinson MR, Perry JRB, Nolte IM, van Vliet-Ostaptchouk JV *et al.* 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics* **47**: 1114–1120.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW *et al.* 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**: 565–569.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**: 76–82.

Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K *et al.* 2016. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics* **48**: 927–934.

Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. 2014. Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics* **196**: 1263–1275.

Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**: 539–551.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203–208.

Zeng Y, Wang G, Yang E, Ji G, Brinkmeyer-Langford CL, Cai JJ. 2015. Aberrant gene expression in humans. *PLoS Genetics* **11**: e1004942.

Zhao J, Akinsanmi I, Arafat D, Cradick TJ, Lee CM, Banskota S, Marigorta UM, Bao G, Gibson G. 2016. A burden of rare variants associated with extremes of gene expression in human peripheral blood. *American Journal of Human Genetics* **98**: 299–309.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P *et al.* 2007. An Arabidopsis example of association mapping in structured samples. *PLoS Genetics* **3**: e4.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J *et al.* 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications* **2**: 467.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.

Zöllner S, Pritchard JK. 2007. Overcoming the winners curse: estimating penetrance parameters from case-control data. *American Journal of Human Genetics* **80**: 605–615.

Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. 2014. Searching for missing heritability: designing rare variant association studies. *Proceedings of the National Academy of Sciences, USA* **111**: E455–E464.

© 2017 The Authors
New Phytologist © 2017 New Phytologist Trust