Investigating Drought Tolerance in Chickpea Using Genome-Wide Association Mapping and Genomic Selection Based on Whole-Genome Resequencing Data

Yongle Li[1]*, Pradeep Ruperao[2], Jacqueline Batley[3,4,5], David Edwards[3,4,5], Tanveer Khan[4,5], Timothy D. Colmer[4,5], Jiayin Pang[4,5], Kadambot H. M. Siddique[4,5] and Tim Sutton[1,6]

[1] School of Agriculture, Food and Wine, The University of Adelaide, Adelaide, SA, Australia, [2] School of Agriculture and Food Sciences, The University of Queensland, Brisbane, QLD, Australia, [3] School of Biological Sciences, The University of Western Australia, Perth, WA, Australia, [4] The UWA Institute of Agriculture, The University of Western Australia, Perth, WA, Australia, [5] UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, [6] South Australian Research and Development Institute, Adelaide, SA, Australia

Drought tolerance is a complex trait that involves numerous genes. Identifying key causal genes or linked molecular markers can facilitate the fast development of drought tolerant varieties. Using a whole-genome resequencing approach, we sequenced 132 chickpea varieties and advanced breeding lines and found more than 144,000 single nucleotide polymorphisms (SNPs). We measured 13 yield and yield-related traits in three drought-prone environments of Western Australia. The genotypic effects were significant for all traits, and many traits showed highly significant correlations, ranging from 0.83 between grain yield and biomass to −0.67 between seed weight and seed emergence rate. To identify candidate genes, the SNP and trait data were incorporated into the SUPER genome-wide association study (GWAS) model, a modified version of the linear mixed model. We found that several SNPs from auxin-related genes, including auxin efflux carrier protein (PIN3), p-glycoprotein, and nodulin MtN21/EamA-like transporter, were significantly associated with yield and yield-related traits under drought-prone environments. We identified four genetic regions containing SNPs significantly associated with several different traits, which was an indication of pleiotropic effects. We also investigated the possibility of incorporating the GWAS results into a genomic selection (GS) model, which is another approach to deal with complex traits. Compared to using all SNPs, application of the GS model using subsets of SNPs significantly associated with the traits under investigation increased the prediction accuracies of three yield and yield-related traits by more than twofold. This has important implication for implementing GS in plant breeding programs.

Keywords: drought tolerance, genome-wide association mapping, genomic selection, chickpea, whole-genome resequencing, auxin

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is ranked second after soybean in terms of global legume production, reaching ~13 million tons in 2014 (FAOSTAT 2017). Consisting of 25% of the total exports worldwide, Australia was the second-largest producer and the largest exporter of chickpea in 2014 (FAOSTAT 2017). Chickpea is an important component of the farming system in Australia, serving as a disease break crop and nitrogen fixer (Knights et al., 2009; Siddique et al., 2013). Chickpea seed is a rich source of protein, essential minerals, and dietary fiber (Bar-El Dadon et al., 2017).

Drought is one of the most important constraints limiting yield potential in cereal and legume crops. Significant differences in terms of drought tolerance, measured by yield reduction, were observed among legume species in a meta-analysis of over 100 studies with chickpea ranking seventh among 13 legume species (Daryanto et al., 2015). There are generally two types of drought: transient drought and terminal drought. Transient drought is a short-term water deficit that can be relieved by precipitation and can occur at any stages of the growing season. Terminal drought is an unrelieved water deficit that terminates the reproductive growth of the plant. Terminal drought is very common in semi-arid tropics (South Asia, north-east Australia) and Mediterranean-type climates such as southern Australia. More than 80% of the world chickpea production is located in South Asia and north-east Australia. Australia has experienced severe drought events from the late 1990s to mid 2000s known as "the Millennium drought." As a consequence, the total production of irrigated rice and cotton fell by 99 and 84% during 2002 and 2009, respectively (van Dijk et al., 2013).

With current climate change projections, extremely hot weather will become more frequent and rainfall will be more erratic in Australia and other regions of the world (Hennesy et al., 2010; Foyer et al., 2016). The reproductive stage of growth is usually the most critical phase influencing grain yield in crops. It is well documented that drought stress during pod filling can lead to pod abortion thus reducing the number of seeds per plant (Leport et al., 1999; Fang et al., 2010; Pang et al., 2017). In a glasshouse experiment, seed yield declined by 85% in chickpea plants exposed to terminal drought at the early podding stage, relative to well-watered plants (Pang et al., 2017). There is an urgent need to develop chickpea varieties that are drought resilient.

Chickpea has a relatively small genome size of 730 Mb, compared to other food legumes such as lentil (*Lens culinaris* L.) and faba bean (*Vicia faba* L.). Thanks to the advance of next-generation sequencing (NGS) technology and a relatively small genome size, chickpea has well-developed genomic resources. Chickpea reference genomes for kabuli, desi, and wild *Cicer* species are available and there are ongoing efforts to improve the assemblies and annotation of the genomes (Jain et al., 2013; Varshney et al., 2013; Ruperao et al., 2014; Parween et al., 2015; Gupta et al., 2017). Whole-genome resequencing (WGRS) has emerged as one of the best methods for genome-wide association studies (GWAS) due to its potential to discover a large amount of sequence variants [single nucleotide polymorphisms (SNPs),

Indel, CNV] in a cost effective manner. A recent study using this method narrowed down a major QTL for ascochyta blight resistance in chickpea (Li et al., 2017). Using WGRS data, several genomic regions were identified under positive selection for plasticity for yield, nitrogen fixation, and δ13C in chickpea under drought and/or heat conditions in the field (Sadras et al., 2016).

To identify QTL/genes associated with drought tolerance in chickpea, different forward-genetic approaches using various molecular markers have been used. A recent study using bi-parental mapping population (ILC588 × ILC 3279) and sparse simple sequence repeat (SSR) markers, identified 15 and 93 QTL associated with different drought-related traits (Rehman et al., 2011; Hamwieh et al., 2013). However, the resulting large QTL intervals have limited practical application in breeding. A "QTL-hotspot" region on chromosome 4 was identified using traditional QTL analysis with SSR and GBS (genotyping-by-sequencing) markers (Varshney et al., 2014b; Jaganathan et al., 2015). This region was further fine mapped to a ~300 kb region, which contains 26 genes using a QTL bin-mapping approach and gene enrichment analysis by adding more SNPs (Kale et al., 2015). The authors also tested 12 genes for differential gene-expression profiling using real-time PCR. Under drought condition, several genes had higher gene-expression levels in the resistant line than in susceptible lines, including E3 ubiquitin ligase, serine/threonine protein kinases, and homocysteine S-methyltransferase. Another study, employing the GWAS approach, discovered over 200 markers associated with drought-related traits using SSR, DArT, and SNP markers (Thudi et al., 2014). The results, albeit needing further validation, are promising for marker-assisted selection. However, the development of these molecular markers is labor-intensive and not cost-effective.

One of the challenges of marker-assisted selection is how to pyramid numerous markers with small effect size, particularly for complex traits such as yield under drought environments (Collins et al., 2008). As such, genomic selection (GS), also known as genome-wide selection, was proposed as an alternative method for marker-assisted selection for complex traits (Meuwissen et al., 2001). GS uses information from all of the markers to estimate the breeding value of plants, thus eliminating the complicated pyramiding process in marker-assisted selection. This approach is more relevant to breeding programs as it can help select the best parents for crossing and reduce the cost and time of a standard breeding cycle; thus it has been adopted rapidly by many livestock and crop breeding programs (Hayes et al., 2009a; Meuwissen and Goddard, 2010; Crossa et al., 2014). Traditionally, animal scientists estimated breeding values by best linear unbiased prediction (BLUP) using the additive-genetic relationship matrix obtained from pedigree information (Henderson, 1975). Thanks to advances in genotyping and NGS technology, a large amount of molecular markers can be obtained at a relatively low cost (Davey et al., 2011; Elshire et al., 2011; Poland and Rife, 2012). The genomic estimated breeding values can be estimated more accurately using the ridge regression BLUP (RR-BLUP) model which replaces the pedigree matrix (**A** matrix) with the genomic relationship matrix (**G** matrix), which is obtained from genome-wide markers (VanRaden, 2008; Hayes et al.,

2009b). A simulation study in barley showed that GS was better than phenotypic selection when the traits had low heritability and the training population was large enough. Ziyomo and Bernardo (2013) demonstrated in a real experiment that GS is superior to indirect phenotypic selection using secondary traits for improving drought tolerance of maize. However, a recent GS study in chickpea showed that prediction accuracies of yield under rainfed environments were much lower than under irrigated environments (Roorkiwal et al., 2016). A similar observation was reported in synthetic wheat, posing a challenge to improving drought tolerance using GS (Jafarzadeh et al., 2016). A new method is needed to increase prediction accuracy, particularly when applied to drought-stressed environments. The objectives of this study were to: (1) identify candidate genes/SNPs significantly associated with yield and yield-related traits under drought stressed environments using GWAS approaches; and (2) investigate whether incorporation of the GWAS result can increase prediction accuracy.

MATERIALS AND METHODS

Plant Materials and Field Experiments

Plant materials included 13 Australian released varieties and 119 Australian and Indian-derived breeding lines, which were selected for yield potential and adaptation to drought-prone environments. The field experiments are described in detail by Pang et al. (2017). Briefly, chickpea accessions were planted in plots (6 × 1.5 m) in Western Australia at one site in 2012 and two sites in 2013. There were three replicates for each site. Rainfall at the three sites during the growing seasons ranged from 196 to 230 mm, and no irrigation was supplied. Twelve traits were measured: grain yield per ha (GY), hundred seed weight (100SW), seed number per plant (SN), empty pod ratio (EPR), harvest index (HI), biomass dry weight (DW), flowering time score (FT), podding time score (PT), maturity score (MA), emergence score (EM), early vigor score (EV), and plant height (PH). Five plants per plot were randomly cut at ground level to measure SN, EPR, HI, and DW. Scores for FT, PT, MA, EM, and EV were on a 1–9 scale.

Phenotypic Analysis

The three sites were first analyzed separately for each trait by fitting a linear mixed model (LMM), which included spatial effects (row and column effects). The resulting best linear unbiased estimator (BLUE) values for each genotype were used to fit a multiple-environment LMM in which environments were treated as random effects. Statistical significance of fixed and random effects were assessed using Wald's test (Wald, 1943) and the likelihood ratio test, respectively (Van Belle et al., 2004). The resulting BLUE values were subsequently used for GWAS analysis. Broad-sense heritability (h^2) was estimated using the following formula:

$$\hat{h}^2 = \hat{\sigma}_g^2 / (\hat{\sigma}_g^2 + \hat{\sigma}_{ge}^2 / t + \hat{\sigma}_e^2 / rt)$$

where $\hat{\sigma}_g^2$, $\hat{\sigma}_{ge}^2$, and $\hat{\sigma}_e^2$ denote genotypic variance, genotype × environment interaction variance, and experimental error variance, respectively. t and r are the numbers of environments and replications within an environment, respectively. All phenotypic analysis was done using GenStat, 16th edition.

WGRS and SNP Discovery

DNA of the 132 genotypes was extracted from young leaf tissues using the Qiagen DNeasy Plant Mini Kit following the manufacturer's instruction. Paired-end sequencing libraries were constructed using the TruSeq library kit for each genotype with an insert size of 500 bp. The procedure was implemented according to the Illumina manufacturer's instruction. Paired-end short reads (150 bp) were generated using the Illumina HiSeq 2000 platform. Sequence data is available from the NCBI Short Read Archive under BioProject accession PRJNA375953. Paired-end reads for each genotype were trimmed, filtered, and mapped to the kabuli reference genome 2.6.3[1] using SOAP2. Homozygous SNPs were called using the SGSautoSNP pipeline (Lorenc et al., 2012).

Population Structure and Linkage Disequilibrium

To correct for confounding effects in the association studies, population structure was estimated based on 144,777 SNPs (MAF > 0.05) using ADMIXTURE (v1.23) software (Alexander and Lange, 2011). Similar to the popular software STRUCTURE, ADMIXTURE uses a model-based algorithm to estimate the ancestry of unrelated individuals. The number of underlying population groups (K) was estimated from 1 to 10 using the maximum likelihood estimation approach with a fast numerical optimization algorithm. Cross-validation method of Alexander and Lange (2011) was used to determine the most likely number of population group (K). Linkage disequilibrium (LD) was measured by the parameter r^2 using SNPs with high confidence (minimum five reads per genotype). An $r^2 = 0.2$ was used as a threshold to determine LD extent. The method to estimate the LD-decay curve under the mutation-drift-equilibrium model was described in detail in Li et al. (2011).

Genome-Wide Association Mapping

Genome-wide association analysis was done using BLUE values of the 132 genotypes with 12 traits and 144,777 SNPs (MAF > 0.05). Adjusting the confounding effects of population structure and kinship, the SUPER GWAS method, implemented in the GAPIT software, was used to estimate each SNP effect (Lipka et al., 2012; Tang et al., 2016). This method can increase statistical power by estimating kinship matrix with a subset of markers which are not in LD with the testing marker (Wang et al., 2014). The kinship matrix was estimated using the VanRaden method and later compressed to its optimum groups using the P3D method to speed up computation time. Default parameters of the SUPER model were used: sangwich.top = "MLM,"

---

[1]http://www.cicer.info/databases.php

sangwich.bottom = "SUPER," LD = 0.1. The significant p-value cut-off was set as $p = 3.45e\text{-}07$, equivalent to the α level of 0.05 after Bonferroni correction. The two genes flanking the significant SNP are reported.

Genomic Selection

Genomic predictions were performed using three different models: RR-BLUP, Bayesian least absolute shrinkage and selection operator (Bayesian LASSO or BL), and Bayesian ridge regression (BRR). The RR-BLUP model is written as:

$$y = \mu 1_n + Zg + e, \quad e \sim N(0, I\sigma^2),$$

where y is the adjusted entry means of phenotypes, μ is the overall mean, $1n$ denotes $n \times 1$ vector of 1s, $\mathbf{Z}$ is an incident matrix for random genotype effect, and g is genotype effect with normal distribution N $(0, \mathbf{G}\sigma_g^2)$, where $\mathbf{G}$ is the genomic relationship matrix obtained from markers (VanRaden, 2008). The markers included all 147,777 markers or a subset of markers selected based on different levels of p-value from GWAS.

The general structure of the two Bayesian linear regressions BL and BRR can be written as,

$$p(\mu, \boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \boldsymbol{\theta}),$$

the posterior probability of unknown parameters includes overall mean μ, marker effect β, and its variance σ^2, given the data y and hyperparameters $\boldsymbol{\theta}$. Estimates of these unknown parameters are obtained by solving the optimization problem and adding a penalty function to $\boldsymbol{\beta}$. For BRR, the same Gaussian prior was assigned to $\boldsymbol{\beta}$, resulting in the same shrinkage for all markers. For BL, a Bayesian version of the least absolute shrinkage and selection operation (Tibshirani, 1996) was introduced in the penalty function of $\boldsymbol{\beta}$, resulting in greater shrinkage of markers with small effects and less shrinkage of markers with large effects. BL has a special feature of both variables selection and shrinkage, whereas BRR only shrinks variables. The detailed similarities and differences between genomic prediction models are reviewed by de Los Campos et al. (2013). The R package synbreed was used to fit the three models (Wimmer et al., 2012).

A fivefold cross-validation was performed to evaluate the prediction performance of the three models. The whole dataset was randomly divided into five mutually exclusive subsets, four of which formed the training set for fitting the model and the fifth was used as a test set. This process was repeated ten times, resulting in 50 cross-validations. Predictive abilities were calculated as Pearson's correlation coefficient between the predicted values and observed phenotypic values of the test set. An average predictive ability of 50 cross-validations was reported.

RESULTS

Yield and Yield-Related Traits

In total, 12 traits including phenology, yield, and yield components were measured. Multiple-environment linear mixed-models were fitted to obtain BLUE values for each genotype (**Table 1** and Supplementary Figure S1). The genotypic

effect of SN was significant at an alpha level of 0.05 while the other 11 traits were highly significant at an alpha level of 0.001 (**Table 1**). Heritabilities (h^2) of the 12 traits ranged from 0.11 for GY to 0.91 for 100SW. Many traits showed highly significant correlations, ranging from 0.83 between GY and DW to −0.67 between 100SW and EM (**Table 2**). GY was positively correlated with 100SW, SN, DW, FT, PT, EM, EV and negatively correlated with EPR, MA, and PH. 100SW was positively correlated with FT, DW, MA and negatively correlated with SN, PT, and EM. GY has a highly positive correlation with DW ($r = 0.83$), which is not surprising, given that a strong and healthy plant with sufficient biomass is advantageous under drought to retain yield (Hamwieh et al., 2013; Kashiwagi et al., 2013). PT and MA were negatively correlated (−0.64), which appears counterintuitive. Due to the low temperatures in early spring in Australian environments, some genotypes originating from India with an early podding trait, aborted their early onset pods which is reflected in the negative correlation between PT and EPR (i.e., the earlier the podding time, the higher the EPR). This supports the observation that chickpea plants need to set pods within a fairly narrow window to optimize yield in Australian environment.

SNP, Linkage Disequilibrium, and Population Structure

A total of 144,777 homozygous SNPs were discovered in 132 genotypes (**Table 3**). The number of SNPs on each chromosome ranged from 25,323 on Ca4 to 4,740 on Ca8, partially reflecting the length of the chromosomes in the Kabuli 2.6.3 reference assembly. The extent of LD on each chromosome ranged from 4,000 kb on Ca3 to 150 kb on Ca6 with an average of 700 kb (**Table 3** and Supplementary Figure S2). The average extent of LD is almost seven times smaller than a previous study, in which mainly Australian-released chickpea varieties were used (Li et al., 2017). The short extent of LD in the 132 genotypes has the potential to enable higher mapping resolution. To avoid false positive results in association analysis, the population structure was investigated using 144,777 SNPs (**Figure 1**). The most likely number of groups (K) in the 132 genotypes was estimated to be two using a cross-validation method from the ADMIXTURE software. The red group in **Figure 1** is mainly the DICC lines (selected from ICRISAT breeding lines) consisting of progenies from the crosses of ICCV98503 × Moti, ICCV96836 × PBG5, ICCV96836 × ICC12004, and ICCV96836 × ICC3996. The green group in **Figure 1** consists of Australian-released varieties and advanced lines. Genotypes with a mixture of red and green have mixed ancestry from ICRISAT, ICARDA, and Australia.

Genome-Wide Association Mapping

In total, 38 SNPs were significantly ($p < 3.45e\text{-}07$) associated with six traits: GY, 100SW, EPR, PT, EM, and EV (Supplementary Table S1 and **Figures 2**, **3**). One SNP, located in Ca3: 18,924,965, was significantly associated with GY (**Figure 2**). The closest gene near this SNP encodes a protein belonging to the ABC transporter B family/p-glycoprotein (PGP). Nine SNPs, located on Ca3, Ca4, Ca5, and Ca6, were significantly associated with 100SW (Supplementary Table S1 and **Figure 2**). Candidate

TABLE 1 | BLUE values (minimum–maximum), genotypic effect, and heritabilities (h^2) of 12 traits obtained from a multi-environment LMM.

Traits	No. of genotypes	Mean	Minimum	Maximum	Wald's test for genotypic effect	H^2
GY (kg)	132	1027.11	623.19	1264.75	$p < 0.001$	0.11
100SW (g)	132	20.00	15.00	37.61	$p < 0.001$	0.91
SN	93	21.96	11.91	32.28	$p = 0.015$	0.32
EPR	59	0.31	0.14	0.46	$p < 0.001$	0.52
HI	93	0.38	0.28	0.46	$p < 0.001$	0.51
DW (g)	93	9.81	6.14	18.02	$p < 0.001$	0.49
FT	132	5.36	1.98	9.56	$p < 0.001$	0.70
PT	132	5.31	0.55	9.27	$p < 0.001$	0.72
MA	132	5.97	4.61	10.04	$p < 0.001$	0.25
EM	132	7.95	5.68	8.88	$p < 0.001$	0.49
EV	132	6.13	3.65	7.93	$p < 0.001$	0.65
PH (cm)	62	53.62	42.88	61.84	$p < 0.001$	0.64

GY, grain yield per ha; 100SW, hundred seed weight; SN, seed number per plant; EPR, empty pod ratio; HI, harvest index; DW, biomass dry weight; FT, flowering time score; PT, podding time score; MA, maturity score; EM, seed-emergence score; EV, early vigor score; PH, plant height.
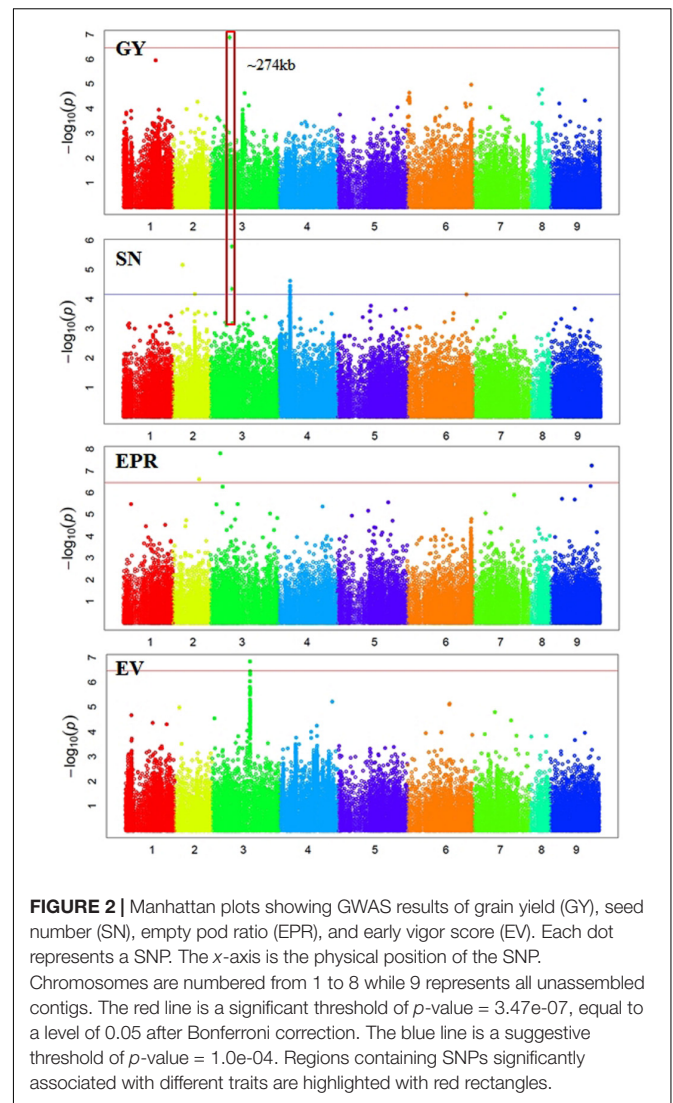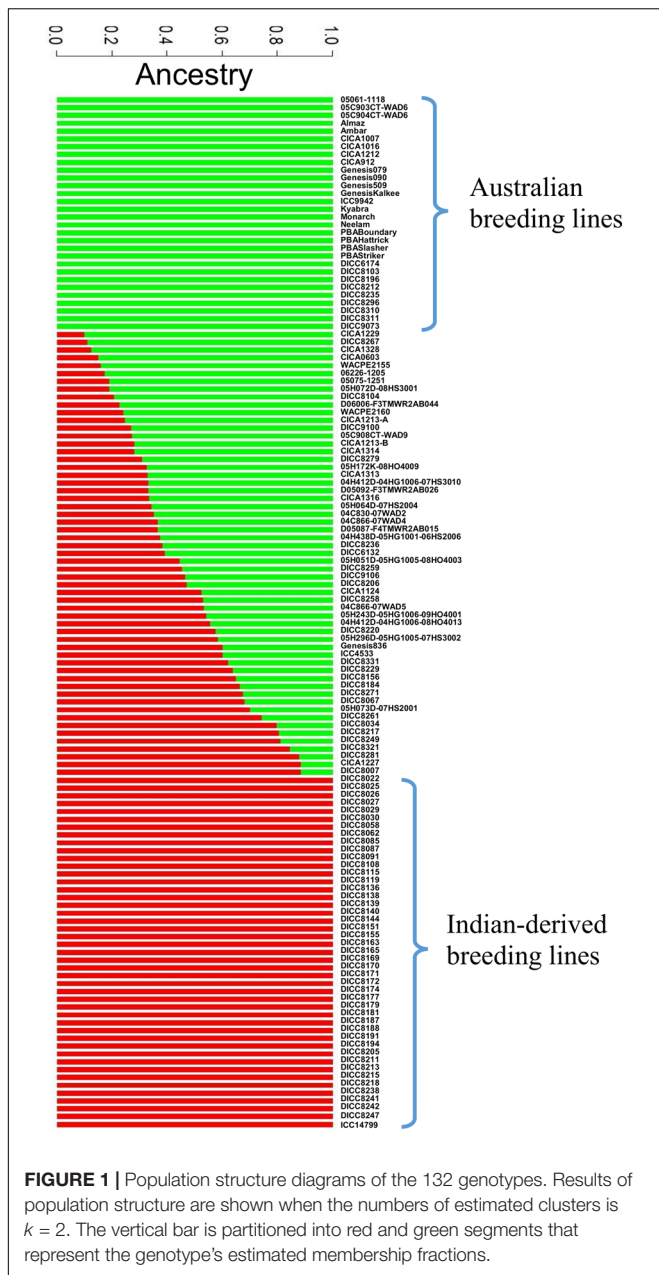
TABLE 2 | Correlation matrix of the 12 traits.

	GY	100SW	SN	EPR	HI	DW	FT	PT	MA	EM	EV	PH
GY	−											
100SW	0.45***	−										
SN	0.47***	−0.36**	−									
EPR	−0.37**	−0.12	−0.22	−								
HI	0.13	−0.27*	0.56***	0.02	−							
DW	0.83***	0.67***	0.15	−0.33*	−0.35**	−						
FT	−0.43***	0.42***	−0.14	−0.55***	−0.28*	0.43***	−					
PT	0.58***	−0.42***	0.27*	0.58***	0.41**	−0.40**	−0.65***	−				
MA	−0.53***	0.70***	−0.17	−0.35**	−0.47***	0.73***	0.54***	−0.64***	−			
EM	0.42***	−0.67***	0.20	0.25	0.27*	−0.59***	−0.29*	0.38**	−0.59***	−		
EV	0.29*	−0.20	−0.03	0.65***	0.03	−0.29*	−0.60***	0.55***	−0.37**	0.25	−	
PH	−0.22	0.25	−0.56***	0.18	−0.58***	0.08	0.11	−0.10	0.28*	−0.27*	0.25	−

*GY, grain yield; 100SW, hundred seed weight; SN, seed number; EPR, empty pod ratio; HI, harvest index; DW, biomass dry weight; FT, flowering time score; PT, podding time score; MA, maturity score; EM, emergence score; EV, early vigor score; PH, plant height. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

TABLE 3 | Summary of LD and SNPs used to estimate LD.

Chromosome	No. of SNPs	Density of SNPs (No. of SNPs/10 kb)	No. of SNPs used to estimate LD[1]	Mean r^2	LD extent (kb)
Ca1	20,837	4.25	2,979	0.06	400
Ca2	11,181	3.01	2,826	0.03	200
Ca3	19,487	2.92	1,674	0.15	4,000
Ca4	25,323	4.30	7,106	0.02	200
Ca5	18,313	2.64	1,428	0.07	200
Ca6	15,620	2.37	2,469	0.05	150
Ca7	13,272	2.36	2,070	0.03	200
Ca8	4,740	2.38	933	0.06	200
Unassembled contigs	16,004	3.25	NA	NA	NA
Total/average	144,777		21,485	0.06	700

[1]*SNPs with high confidence (minimum five reads per genotype).*

FIGURE 1 | Population structure diagrams of the 132 genotypes. Results of population structure are shown when the numbers of estimated clusters is k = 2. The vertical bar is partitioned into red and green segments that represent the genotype's estimated membership fractions.

FIGURE 2 | Manhattan plots showing GWAS results of grain yield (GY), seed number (SN), empty pod ratio (EPR), and early vigor score (EV). Each dot represents a SNP. The x-axis is the physical position of the SNP. Chromosomes are numbered from 1 to 8 while 9 represents all unassembled contigs. The red line is a significant threshold of p-value = 3.47e-07, equal to a level of 0.05 after Bonferroni correction. The blue line is a suggestive threshold of p-value = 1.0e-04. Regions containing SNPs significantly associated with different traits are highlighted with red rectangles.

genes located near these SNPs include two sugar transporters, two nodulin MtN21/EamA-like transporters, one Lateral Organ Boundaries (LOB) domain protein and several uncharacterized genes.

There were five SNPs significantly associated with PT, including two SNPs on Ca1, one SNP on Ca3, and two SNPs on Ca4 (Supplementary Table S1 and **Figure 3**). One was located on Ca4: 35,589,599 near a gene encoding a major latex protein (MLP), which promotes vegetative growth and delays flowering in Arabidopsis (Guo et al., 2011). Another significant SNP was located on Ca3: 38,173,722 close to a transcriptional factor squamosa promoter-binding-like protein 9 (SPL9). It was shown that SPL9 and SPL15 act redundantly in promoting the

juvenile-to-adult phase transition in Arabidopsis (Schwarz et al., 2008).

There were 12 SNPs significantly associated with MA, including one SNP on Ca2, three SNPs on Ca4, six SNPs on Ca5, and two SNPs on Ca6 (Supplementary Table S1 and **Figure 3**). One of the significant SNPs was located on Ca5: 11,580,061, near the nodulin MtN21/EamA-like transporter which has been shown to be involved in auxin homoeostasis (Ranocha et al., 2013). Another significant SNP was located on Ca5: 12,166,907, near a sugar transporter gene with an important role in plant growth (Wobus and Weber, 1999). Six SNPs were significantly associated with EM, including one SNP on Ca3, three SNPs on Ca4, and two SNP on Ca5 (Supplementary Table S1 and **Figure 3**). One of the significant SNPs (Ca4: 35,589,599) was located near a gene encoding MLP, which promotes vegetative growth in Arabidopsis as described above (Guo et al., 2011). Two SNPs were significantly associated with EV (Supplementary Table S1 and **Figure 2**). One of the significant SNPs (Ca3: 38,177,160) was located near a gene encoding the transcriptional

factor SPL9. It has been shown that SPL9 regulates leaf initiation negatively in Arabidopsis, leading to a shorter leaf plastochron, which is the time interval between two successive events of plant growth (Schwarz et al., 2008). There were no SNPs significantly ($p < 3.45e-07$) associated with SN, HI, DW, FT, or PH. This could be attributed to the lack of statistical power due to the small sample size (62–93 genotypes) employed in this study.
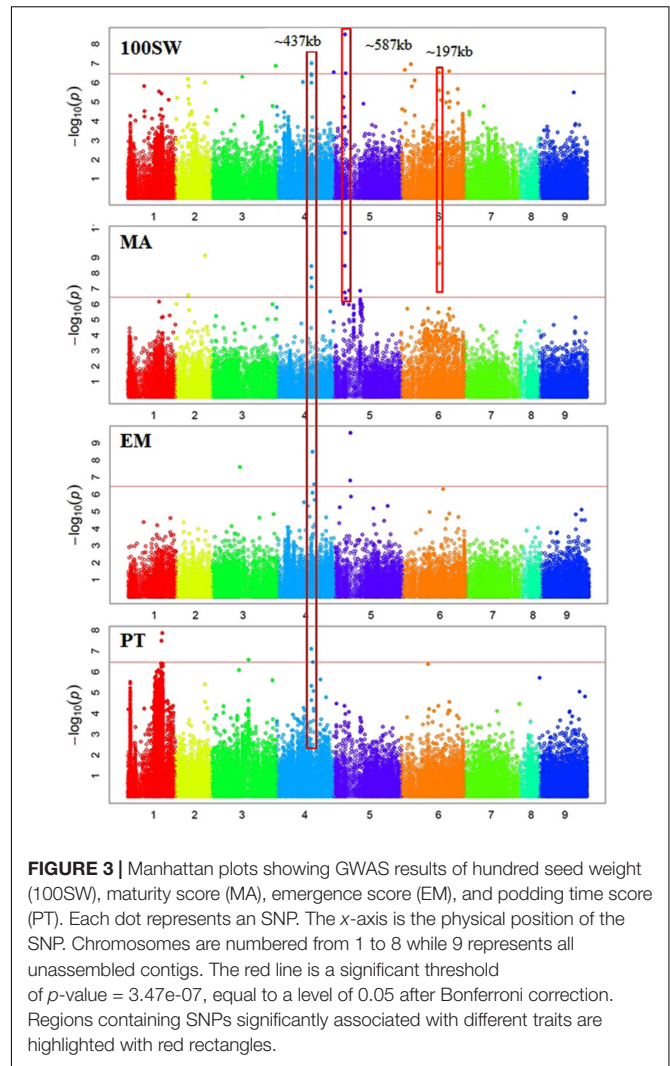
Some regions contain SNPs significantly associated with several different traits, which is an indication of a pleiotropic effect (**Figures 2, 3**). For example, a genomic region of ∼274 kb on Ca3 (18,924,965 to 21,660,191) contains a SNP significantly associated with GY and two SNPs weakly (p-value $= 1.59e-06$ and $4.49e-05$) associated with SN. A ∼437 kb genomic region on Ca4 (35,589,599 to 36,026,910) contains eight SNPs, significantly associated with four traits: EM, maturity, PT, and 100SW. Four auxin-related genes, encoding one auxin efflux carrier protein (PIN3) and three nodulin MtN21/EamA-like transporters, are located in this region on Ca4. Another ∼587 kb genomic region on Ca5 (11,580,061 to 12,166,907) contains two SNPs, which were significantly associated with 100SW and MA. Five auxin-related genes, including one auxin influx transporter (LAX3) and four nodulin MtN21/EamA-like transporters, are located in this region. Two SNPs, significantly associated with both 100SW and maturity, were located in a 197 kb region on Ca6 (39,200,356 to 39,397,897) which contains two sugar transporters.

Genomic Prediction

Prediction accuracies for GY, 100SW, SN, and EV were estimated using RR-BLUP and different subsets of the SNPs based on p-values from GWAS results. Prediction accuracies increased when subsets of the SNPs based on a more stringent level of p-value were used (**Figure 4**). The increments plateaued in all four traits using subsets of SNPs with p-values between 0.05 and 0.01 and dropped dramatically at p-values of $3.45e-07$ (equal to 0.05 after Bonferroni correction). The lowest prediction accuracies in three traits occurred when using all SNPs, which was probably due to noise introduced by non-causal variants as RR-BLUP shrikes each marker effect equally (de Los Campos et al., 2013). We also used BL and BRR to estimate prediction accuracies using subsets of SNPs. The results were similar to the RR-BLUP model.

DISCUSSION

Previous effort on breeding drought-tolerant chickpea has concentrated on accelerating flowering to escape terminal drought (Upadhyaya et al., 2012). This study showed that some India-derived genotypes with early podding trait aborted early onset pods in the Australian environments due to low temperatures in early spring. This suggests that it may be more relevant to focus on breeding for drought tolerance *per se* under Australia environments, traits such as water-use efficiency (Zaman-Allah et al., 2011; Kashiwagi et al., 2013), beneficial

FIGURE 3 | Manhattan plots showing GWAS results of hundred seed weight (100SW), maturity score (MA), emergence score (EM), and podding time score (PT). Each dot represents an SNP. The *x*-axis is the physical position of the SNP. Chromosomes are numbered from 1 to 8 while 9 represents all unassembled contigs. The red line is a significant threshold of *p*-value = 3.47e-07, equal to a level of 0.05 after Bonferroni correction. Regions containing SNPs significantly associated with different traits are highlighted with red rectangles.

root traits (Zaman-Allah et al., 2011), stomatal conductance (Rehman et al., 2011), and osmotic adjustment (Morgan et al., 1991). As pointed out by Berger et al. (2016), the selection pressure for drought escape and drought tolerance *per se* is very different. Excessive use of the drought escape mechanism can compromise yield potential due to shorter life cycles to accumulate water and light resources. Thus a new breeding strategy is warranted, such as the integrated framework proposed by Berger et al. (2016).

Auxin-Related Genes and Sugar Transporters Play an Important Role in Yield-Related Traits under Drought-Prone Environments

Several auxin-related genes, including PIN3, ABC transporter B family/PGP, and nodulin MtN21/EamA-like transporters, were found to be near SNPs significantly associated with GY, 100SW, PT, EM, and MA. Auxin (primarily indole-3-acetic acid) is a well-known phytohormone that plays a pivotal
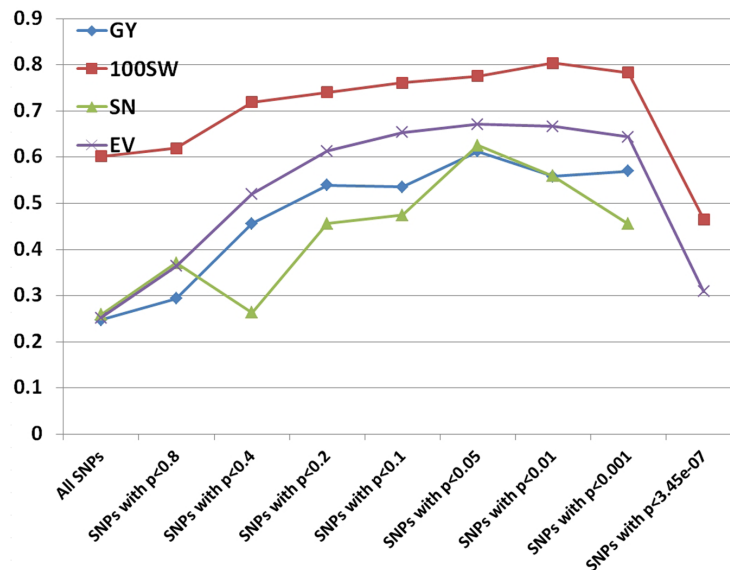
FIGURE 4 | Prediction accuracies for grain yield (GY), hundred seed weight (100SW), seed number (SN), and early vigor score (EV) using different subsets of SNPs based on *p*-values from GWAS results.

role in plant growth, seed development, and abiotic stress response (Zhao, 2010; Kazan, 2013; Locascio et al., 2014). A recent review paper summarized how plants coordinate auxin biosynthesis, transport, perception under osmotic stresses induced be drought, salinity (Naser and Shani, 2016). Auxin was found to enhance drought tolerance via the regulation of root architecture, expression of abiotic stress genes (*DREB2A* and *DREB2B*), ROS metabolism, and metabolic homeostasis in Arabidopsis (Shi et al., 2014). PIN3, belonging to the auxin efflux carrier protein family, have been characterized as important regulators involved in plant growth, phototropic response, and drought stress response (Ding et al., 2011; Zhang et al., 2012, 2013). A study in rice showed that drought, cold and heat stress affected the expression of genes involved in auxin signaling and polar transport, such as the PIN protein family (Du et al., 2013). Several studies have shown that PGP is involved in auxin transport through the plasma membrane and can stabilize the PIN protein family (Geisler and Murphy, 2006; Blakeslee et al., 2007; Titapiwatanakun et al., 2009; Zazimalova et al., 2010). Arabidopsis WAT1, a homolog of the nodulin MtN21/EamA-like transporter, was recently identified as a vacuolar auxin transporter required for auxin homoeostasis, a process that maintains an endogenous steady-state concentration of primary auxin (Ranocha et al., 2013). Several sugar transporters were found to be near SNPs significantly associated with 100SW and MA. Comprising hexose and sucrose transport proteins, the sugar transporters are members of the major facilitator superfamily. Sugar transporters play a key role in plant growth, source–sink partitioning, molecular signaling, and seed development, and are therefore important for optimal plant development and crop yield (Wobus and Weber, 1999; Wingenter et al., 2010; Doidy et al., 2012).

Using a bi-parental QTL mapping population, a "QTL-hotspot" region on Ca4 13,239,546 to 13,547,009 (based on the kabuli reference genome v1.0) was associated with at least seven traits including root traits, 100SW, PH, and days to flowering (Varshney et al., 2014b; Jaganathan et al., 2015; Kale et al., 2015). In this study, we did not identify SNPs from the "QTL-hotspot" region significantly associated with any traits. We identified a ∼437 kb genomic region on Ca4: 35,589,599 to 36,026,910 (Ca4: 37,933,355 to 38,412,853 based on the kabuli reference genome v1.0) containing eight SNPs significantly associated with four traits: EM, maturity, PT, and 100SW. Different 100SW QTL were identified between the two studies, which may be attributed to the different mapping populations examined.

Confounding Effects of Self-Incompatibility with Pod Abortion under Drought

Three SNPs, located on Ca2, Ca3, and an unassembled contig, were significantly associated with EPR (Supplementary Table S1 and **Figure 2**). The kinesin-4 and self-incompatibility (SI) proteins were adjacent to two of the three significant SNPs. The kinesin-4 family plays an important role in cell elongation and has been shown to affect the length of siliques and seeds produced per silique in Arabidopsis (Kong et al., 2015). The SI protein (IPR010264) is highly homologous (with a total score of 141 and E value of 2e-32 using NCBI blastn) to a Medicago gene Medtr1g057250.1, which was well characterized in *Papaver rhoeas* (Foote et al., 1994; Wilkins et al., 2014). SI is a mechanism used by many flowering plants to prevent self-fertilization and inbreeding depression. Pollen from SI plants, carrying the same haplotype as the pistil, was rejected via the program cell death

mechanism (Wilkins et al., 2014). Chickpea is generally a self-pollinating crop with an outcrossing rate of less than 2% (Toker et al., 2006); however, SI plants with empty pods yet viable pollen were observed in an F2 segregating population from a cross (H 82-5 × E100 ym) × Bhim with ∼22% of F2 SI plants (Lather and Dahiya, 1992). Pod abortion has long been thought to be introduced mainly by abiotic stress (Fang et al., 2010; Pang et al., 2017). Our findings, however, indicated that SI might have confounding effects with pod abortion under drought.

Incorporating the Results of GWAS Increased Prediction Accuracy

Prediction models, employing variable selection procedures such as the LASSO, are considered to be better than RR-BLUP theoretically because they remove non-causal variants and variants not in LD with causal variants (Daetwyler et al., 2010; Meuwissen and Goddard, 2010). However, Wimmer et al. (2013) showed that LASSO failed to achieve its superiority and thus suggested to "*preselect markers according to biological prior information.*" Our study supports this assumption by showing that prediction accuracies were significantly improved using a subset of SNPs significantly (between $p < 0.05$ and $p < 0.01$) associated with traits. Several computer simulation studies (Meuwissen and Goddard, 2010; Ober et al., 2012) speculated that using large amounts of markers from WGRS may increase prediction accuracy, particularly in cases where the training population is distantly related to the prediction population. We argue that having more markers alone may not help to increase prediction accuracy, but it may help to identify the causal variants. If the prediction is done subsequent to the identification of causal variants, then the prediction may increase, as demonstrated in the current study, and thus the advantage of employing WGRS in GS can be realized.

Many studies have been published on the effect of marker density on prediction accuracy (Lorenzana and Bernardo, 2009; Vazquez et al., 2010; Weigel et al., 2010; Hoffstetter et al., 2016). Most have selected markers randomly or based on equal space and found that prediction accuracy increased when the number of markers increased, but reached a plateau depending on the extent of LD and the population size (Vazquez et al., 2010). A few studies selected markers based on biological prior information (Hoffstetter et al., 2016; Kooke et al., 2016; Spindel et al., 2016). Compared to using all markers in the RR-BLUP model, prediction accuracies doubled by using a subset of markers with significant association with grain yield in wheat (Hoffstetter et al., 2016). Another GS study in rice also showed that prediction accuracies were 7.0%–29.8% higher based on RR-BLUP with all markers and markers (selected from GWAS) fitted as fixed effects compared to that based on RR-BLUP with all markers alone (Spindel et al., 2016). Using different models, a recent GS study in chickpea indicated that prediction accuracies of yield under rainfed environments ranged from 0.148 to 0.186, which is similar (0.25) to this study using all 144,777 SNPs, but much smaller (0.56–0.61) than when using a subset of SNPs significantly associated with yield. We speculate

that prediction accuracy may increase if an approach described here is adopted.

Training population size is an important factor in GS. Several studies have been conducted to investigate the optimum size for a training population in plants. Generally, the accuracy of estimated marker effects increases as the sample size increases (Albrecht et al., 2011; Endelman et al., 2014). Compared to other prediction models, one simulation study showed that RR-BLUP is robust with a small training population size even as low as $n = 75$, with diminishing benefits between $n = 125$ and $n = 300$ (Lorenz, 2013). Riedelsheimer and Melchinger (2013) reached a similar observation and recommended to allocate more resources to the selection candidates (prediction set) instead of the training population when budget is fix. Compared to a GS study in chickpea conducted by Roorkiwal et al. (2016), the training population size in this study is relatively small. Because the main objective of this study is to test the prediction accuracy based on subsets of significant SNPs. The result of this study should hold since the size of the training population was the same in different subsets of SNPs. For real breeding application, such as selecting candidate genotypes without phenotypic data, larger training populations should be used to increase prediction accuracy. Additionally, the training population needs to be updated regularly to maintain a close relationship with selection candidates (Neyhart et al., 2017).

Grain yield is a complex trait controlled by numerous genes with small effect. We found only one SNP significantly associated with grain yield in this study, probably due to limited statistical power to identify genes which underline complex traits. Even if all yield-related genes could be identified using a larger sample size, pyramiding favorable alleles from all genes into a single genotype using traditional marker-assisted selection or transgenic approaches would be extremely difficult. The superiority of the GS approach is that it can use all marker information simultaneously and thus circumvent the complex process of pyramiding. That is not to say that GWAS and marker-assisted selection do not have a place in molecular breeding; for example these approaches are useful for targeting simple traits (Mendelian traits) such as disease resistance (Varshney et al., 2014a; Li et al., 2017). This study also shows that incorporating the results of GWAS into the prediction model can significantly increase prediction accuracy. However, this gain of prediction accuracy is only examined in a cross-validation scheme. Further study is needed to investigate whether this result holds true when this approach is applied to selection candidates.

AUTHOR CONTRIBUTIONS

YL and TS conceived the study; YL performed the GWAS and GS analysis; JB contributed to sequencing; PR analyzed the sequencing data; DE supervised sequencing-data analysis; KS, JP, TK, and TC contributed to phenotyping; YL wrote the manuscript and all the authors read and approved the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00190/full#supplementary-material

REFERENCES

Albrecht, T., Wimmer, V., Auinger, H. J., Erbe, M., Knaak, C., Ouzunova, M., et al. (2011). Genome-based prediction of testcross values in maize. *Theor. Appl. Genet.* 123, 339–350. doi: 10.1007/s00122-011-1587-7

Alexander, D. H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246. doi: 10.1186/1471-2105-12-246

Bar-El Dadon, S., Abbo, S., and Reifen, R. (2017). Leveraging traditional crops for better nutrition and health - The case of chickpea. *Trends Food Sci. Technol.* 64, 39–47. doi: 10.1016/j.tifs.2017.04.002

Berger, J., Palta, J., and Vadez, V. (2016). Review: an integrated framework for crop adaptation to dry environments: responses to transient and terminal drought. *Plant Sci.* 253, 58–67. doi: 10.1016/j.plantsci.2016.09.007

Blakeslee, J. J., Bandyopadhyay, A., Lee, O. R., Mravec, J., Titapiwatanakun, B., Sauer, M., et al. (2007). Interactions among PIN-FORMED and P-glycoprotein auxin transporters in *Arabidopsis*. *Plant Cell* 19, 131–147. doi: 10.1105/tpc.106.040782

Collins, N. C., Tardieu, F., and Tuberosa, R. (2008). Quantitative trait loci and crop performance under abiotic stress: Where do we stand? *Plant Physiol.* 147, 469–486. doi: 10.1104/pp.108.118117

Crossa, J., Perez, P., Hickey, J., Burgueno, J., Ornella, L., Ceron-Rojas, J., et al. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112, 48–60. doi: 10.1038/hdy.2013.16

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi: 10.1534/genetics.110.116855

Daryanto, S., Wang, L., and Jacinthe, P. A. (2015). Global synthesis of drought effects on food legume production. *PLoS One* 10:e0127401. doi: 10.1371/journal.pone.0127401

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012

de Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313

Ding, Z., Galvan-Ampudia, C. S., Demarsy, E., Langowski, L., Kleine-Vehn, J., Fan, Y., et al. (2011). Light-mediated polarization of the PIN3 auxin transporter for the phototropic response in *Arabidopsis*. *Nat. Cell Biol.* 13, 447–452. doi: 10.1038/ncb2208

Doidy, J., Grace, E., Kuehn, C., Simon-Plas, F., Casieri, L., and Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. *Trends Plant Sci.* 17, 413–422. doi: 10.1016/j.tplants.2012.03.009

Du, H., Liu, H., and Xiong, L. (2013). Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. *Front. Plant Sci.* 4:397. doi: 10.3389/fpls.2013.00397

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X. C., Sorrells, M. E., et al. (2014). Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci.* 54, 48–59. doi: 10.2135/cropsci2013.03.0154

Fang, X. W., Turner, N. C., Yan, G. J., Li, F. M., and Siddique, K. H. M. (2010). Flower numbers, pod production, pollen viability, and pistil function are reduced and flower and pod abortion increased in chickpea (*Cicer arietium* L.) under terminal drought. *J. Exp. Bot.* 61, 335–345. doi: 10.1093/jxb/erp307

Foote, H. C. C., Ride, J. P., Franklin-Tong, V. E., Walker, E. A., Lawrence, M. J., and Franklin, F. C. H. (1994). Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2265–2269. doi: 10.1073/pnas.91.6.2265

Foyer, C. H., Lam, H. M., Nguyen, H. T., Siddique, K. H. M., Varshney, R. K., Colmer, T. D., et al. (2016). Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* 2:16112. doi: 10.1038/nplants.2016.112

Geisler, M., and Murphy, A. S. (2006). The ABC of auxin transport: the role of p-glycoproteins in plant development. *FEBS Lett.* 580, 1094–1102. doi: 10.1016/j.febslet.2005.11.054.

Guo, D., Wong, W. S., Xu, W. Z., Sun, F. F., Qing, D. J., and Li, N. (2011). *Cis-cinnamic acid-enhanced* 1 gene plays a role in regulation of Arabidopsis bolting. *Plant Mol. Biol.* 75, 481–495. doi: 10.1007/s11103-011-9746-4

Gupta, S., Nawaz, K., Parween, S., Roy, R., Sahu, K., Pole, A. K., et al. (2017). Draft genome sequence of *Cicer reticulatum* L., the wild progenitor of chickpea provides a resource for agronomic trait improvement. *DNA Res.* 24, 1–10. doi: 10.1093/dnares/dsw042

Hamwieh, A., Imtiaz, M., and Malhotra, R. S. (2013). Multi-environment QTL analyses for drought-related traits in a recombinant inbred population of chickpea (*Cicer arientium* L.). *Theor. Appl. Genet.* 126, 1025–1038. doi: 10.1007/s00122-012-2034-0

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009a). Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92, 433–443. doi: 10.3168/jds.2008-1646

Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009b). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/S0016672308009981

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31, 423–447. doi: 10.2307/2529430

Hennesy, K., Whetton, P., and Preston, B. (2010). "Climate projection," in *Adapting Agriculture to Climate Change: Preparing Australian Agriculture, Forestry and Fisheries for the Future*, eds C. Stokes, and M. Howden (Canberra: CSIRO), 13–20.

Hoffstetter, A., Cabrera, A., Huang, M., and Sneller, C. (2016). Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. *G3* 6, 2919–2928. doi: 10.1534/g3.116.032532

Jafarzadeh, J., Bonnett, D., Jannink, J. L., Akdemir, D., Dreisigacker, S., and Sorrells, M. E. (2016). Breeding value of primary synthetic wheat genotypes for grain yield. *PLoS One* 11:e0162860. doi: 10.1371/journal.pone.0162860

Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2015). Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. *Mol. Genet. Genomics* 290, 559–571. doi: 10.1007/s00438-014-0932-3

Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tpj.12173

Kale, S. M., Jaganathan, D., Ruperao, P., Chen, C., Punna, R., Kudapa, H., et al. (2015). Prioritization of candidate genes in "QTL-hotspot" region for drought

tolerance in chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:15296. doi: 10.1038/srep15296

Kashiwagi, J., Krishnamurthy, L., Gaur, P. M., Upadhyaya, H. D., Varshney, R. K., and Tobita, S. (2013). Traits of relevance to improve yield under terminal drought stress in chickpea (*C. arietinum* L.). *Field Crops Res* 145, 88–95. doi: 10.1016/j.fcr.2013.02.011

Kazan, K. (2013). Auxin and the integration of environmental signals into plant root development. *Ann. Bot.* 112, 1655–1665. doi: 10.1093/aob/mct229

Knights, E. J., Siddique, K. H. M., Khan, T. N., and Hobson, K. B. (2009). "Development of the Australian chickpea industry: booms and blights," in *Milestones in Food Legumes Research*, eds M. Ali, and M. Kumar (Kanpur: Indian Institute of Pulses Research), 36–57.

Kong, Z., Ioki, M., Braybrook, S., Li, S., Ye, Z. H., Julie Lee, Y. R., et al. (2015). Kinesin-4 functions in vesicular transport on cortical microtubules and regulates cell wall mechanics during cell elongation in plants. *Mol. Plant* 8, 1011–1023. doi: 10.1016/j.molp.2015.01.004

Kooke, R., Kruijer, W., Bours, R., Becker, F., Kuhn, A., van de Geest, H., et al. (2016). Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in Arabidopsis. *Plant Physiol.* 170, 2187–2203. doi: 10.1104/pp.15.00997

Lather, V. S., and Dahiya, B. S. (1992). Self-incompatibility in Chickpea. *Int. Chickpea Pigeonpea Newsl.* 14, 4–5.

Leport, L., Turner, N. C., French, R. J., Barr, M. D., Duda, R., Daves, S. L., et al. (1999). Physiological responses of chickpea genotypes to terminal drought in a Mediterranean-type environment. *Eur. J. Agron.* 11, 279–291. doi: 10.1016/s1161-0301(99)00039-8

Li, Y., Haseneyer, G., Schön, C. C., Ankerst, D., Korzun, V., Wilde, P., et al. (2011). High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol.* 11:6. doi: 10.1186/1471-2229-11-6

Li, Y. L., Ruperao, P., Batley, J., Edwards, D., Davidson, J., Hobson, K., et al. (2017). Genome analysis identified novel candidate genes for *Ascochyta* blight resistance in chickpea using whole genome re-sequencing data. *Front. Plant Sci.* 8:369. doi: 10.3389/fpls.2017.00359

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Locascio, A., Roig-Villanova, I., Bernardi, J., and Varotto, S. (2014). Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. *Front. Plant Sci.* 5:412. doi: 10.3389/fpls.2014.00412

Lorenc, M. T., Hayashi, S., Stiller, J., Lee, H., Manoli, S., Ruperao, P., et al. (2012). Discovery of single nucleotide polymorphisms in complex genomes using SGSautoSNP. *Biology* 1, 370–382. doi: 10.3390/biology1020370

Lorenz, A. J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. *G3* 3, 481–491. doi: 10.1534/g3.112.004911

Lorenzana, R. E., and Bernardo, R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* 120, 151–161. doi: 10.1007/s00122-009-1166-3

Meuwissen, T., and Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185, 623–631. doi: 10.1534/genetics.110.116590

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.

Morgan, J. M., Rodriguezmaribona, B., and Knights, E. J. (1991). Adaptation to water-deficit in chickpea breeding lines by osomoregulation: relationship to grain-yields in field. *Field Crops Res.* 27, 61–70. doi: 10.1016/0378-4290(91)90022-n

Naser, V., and Shani, E. (2016). Auxin response under osmotic stress. *Plant Mol. Biol.* 91, 661–672. doi: 10.1007/s11103-016-0476-5

Neyhart, J. L., Tiede, T., Lorenz, A. J., and Smith, K. P. (2017). Evaluating methods of updating training data in long-term genomewide selection. *G3* 7, 1499–1510. doi: 10.1534/g3.117.040550

Ober, U., Ayroles, J. F., Stone, E. A., Richards, S., Zhu, D. H., Gibbs, R. A., et al. (2012). Using whole-genome sequence data to predict quantitative trait phenotypes in drosophila melanogaster. *PLoS Genet.* 8:e1002685. doi: 10.1371/journal.pgen.1002685

Pang, J., Turner, N. C., Khan, T., Du, Y. L., Xiong, J. L., Colmer, T. D., et al. (2017). Response of chickpea (*Cicer arietinum* L.) to terminal drought: leaf stomatal conductance, pod abscisic acid concentration, and seed set. *J. Exp. Bot.* 68, 1973–1985. doi: 10.1093/jxb/erw153

Parween, S., Nawaz, K., Roy, R., Pole, A. K., Suresh, B. V., Misra, G., et al. (2015). An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:12806. doi: 10.1038/srep12806

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005

Ranocha, P., Dima, O., Nagy, R., Felten, J., Corratge-Faillie, C., Novak, O., et al. (2013). *Arabidopsis* WAT1 is a vacuolar auxin transport facilitator required for auxin homoeostasis. *Nat. Commun.* 4:2625. doi: 10.1038/ncomms3625

Rehman, A. U., Malhotra, R. S., Bett, K., Tar'an, B., Bueckert, R., and Warkentin, T. D. (2011). Mapping QTL associated with traits affecting grain yield in chickpea (*Cicer arietinum* L.) under terminal drought stress. *Crop Sci.* 51, 450–463. doi: 10.2135/cropsci2010.03.0129

Riedelsheimer, C., and Melchinger, A. E. (2013). Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor. Appl. Genet.* 126, 2835–2848. doi: 10.1007/s00122-013-2175-9

Roorkiwal, M., Rathore, A., Das, R. R., Singh, M. K., Jain, A., Srinivasan, S., et al. (2016). Genome-enabled prediction models for yield related traits in chickpea. *Front. Plant Sci.* 7:1666. doi: 10.3389/fpls.2016.01666.

Ruperao, P., Chan, C. K. K., Azam, S., Karafiatova, M., Hayashi, S., Cizkova, J., et al. (2014). A chromosomal genomics approach to assess and validate the desi and kabuli draft chickpea genome assemblies. *Plant Biotechnol. J.* 12, 778–786. doi: 10.1111/pbi.12182

Sadras, V. O., Lake, L., Li, Y., Farquharson, E. A., and Sutton, T. (2016). Phenotypic plasticity and its genetic regulation for yield, nitrogen fixation and delta13C in chickpea crops under varying water regimes. *J. Exp. Bot.* 67, 4339–4351. doi: 10.1093/jxb/erw221

Schwarz, S., Grande, A. V., Bujdoso, N., Saedler, H., and Huijser, P. (2008). The microRNA regulated SBP-box genes *SPL9* and *SPL15* control shoot maturation in Arabidopsis. *Plant Mol. Biol.* 67, 183–195. doi: 10.1007/s11103-008-9310-z

Shi, H., Chen, L., Ye, T., Liu, X., Ding, K., and Chan, Z. (2014). Modulation of auxin content in Arabidopsis confers improved drought stress resistance. *Plant Physiol. Biochem.* 82, 209–217. doi: 10.1016/j.plaphy.2014.06.008

Siddique, K. H. M., Erskine, W., Hobson, K., Knights, E. J., Leonforte, A., Khan, T. N., et al. (2013). Cool-season grain legume improvement in Australia-use of genetic resources. *Crop Pasture Sci.* 64, 347–360. doi: 10.1071/cp13071

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redona, 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

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., et al. (2016). GAPIT version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9, 1–9. doi: 10.3835/plantgenome2015.11.0120

Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *PLoS One* 9:e96758. doi: 10.1371/journal.pone.0096758

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58, 267–288.

Titapiwatanakun, B., Blakeslee, J. J., Bandyopadhyay, A., Yang, H., Mravec, J., Sauer, M., et al. (2009). ABCB19/PGP19 stabilises PIN1 in membrane microdomains in Arabidopsis. *Plant J.* 57, 27–44. doi: 10.1111/j.1365-313X.2008.03668.x

Toker, C., Canci, H., and Ceylan, F. O. (2006). Estimation of outcrossing rate in chickpea (*Cicer arietinum* L.) sown in autumn. *Euphytica* 151, 201–205. doi: 10.1007/s10681-006-9140-5

Upadhyaya, H. D., Kashiwagi, J., Varshney, R. K., Gaur, P. M., Saxena, K. B., Krishnamurthy, L., et al. (2012). Phenotyping chickpeas and pigeonpeas for adaptation to drought. *Front. Physiol.* 3:179. doi: 10.3389/fphys.2012.00179

Van Belle, G., Fisher, L., Heagerty, P., and Lumley, T. (2004). *Biostatistics: A Methodology for the Health Sciences*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/0471602396

van Dijk, A., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., et al. (2013). The millennium drought in southeast Australia (2001-2009):

natural and human causes and implications for water resources, economy, and society. *Water Resour. Res.* 49, 1040–1057. doi: 10.1002/wrcr. 20123

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Varshney, R. K., Mohan, S. M., Gaur, P. M., Chamarthi, S. K., Singh, V. K., Srinivasan, S., et al. (2014a). Marker-assisted backcrossing to introgress resistance to fusarium wilt race 1 and Ascochyta Blight in C 214, an elite cultivar of chickpea. *Plant Genome* 7. doi: 10.3835/plantgenome2013.10.0035.

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491

Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., et al. (2014b). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6

Vazquez, A. I., Rosa, G. J., Weigel, K. A., de los Campos, G., Gianola, D., and Allison, D. B. (2010). Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. *J. Dairy Sci.* 93, 5942–5949. doi: 10.3168/jds.2010-3335

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* 54, 426–482. doi: 10.1093/biostatistics/kxs015

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z. (2014). A SUPER powerful method for genome wide association study. *PLoS One* 9: e107684. doi: 10.1371/journal.pone.0107684

Weigel, K. A., de los Campos, G., Vazquez, A. I., Rosa, G. J. M., Gianola, D., and Van Tassell, C. P. (2010). Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J. Dairy Sci.* 93, 5423–5435. doi: 10.3168/jds.2010-3149

Wilkins, K. A., Poulter, N. S., and Franklin-Tong, V. E. (2014). Taking one for the team: self-recognition and cell suicide in pollen. *J. Exp. Bot.* 65, 1331–1342. doi: 10.1093/jxb/ert468

Wimmer, V., Albrecht, T., Auinger, H. J., and Schon, C. C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28, 2086–2087. doi: 10.1093/bioinformatics/bts335

Wimmer, V., Lehermeier, C., Albrecht, T., Auinger, H. J., Wang, Y., and Schon, C. C. (2013). Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* 195, 573–587. doi: 10.1534/genetics.113.150078

Wingenter, K., Schulz, A., Wormit, A., Wic, S., Trentmann, O., and Hoermiller, II, et al. (2010). Increased activity of the vacuolar monosaccharide transporter TMT1 alters cellular sugar partitioning, sugar signaling, and seed yield in Arabidopsis. *Plant Physiol.* 154, 665–677. doi: 10.1104/pp.110.162040

Wobus, U., and Weber, H. (1999). Sugars as signal molecules in plant seed development. *Biol. Chem.* 380, 937–944. doi: 10.1515/bc.1999.116

Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011). A conservative pattern of water use, rather than deep or profuse rooting, is critical for the terminal drought tolerance of chickpea. *J. Exp. Bot.* 62, 4239–4252. doi: 10.1093/jxb/err139

Zazimalova, E., Murphy, A. S., Yang, H. B., Hoyerova, K., and Hosek, P. (2010). Auxin transporters - Why so many? *Cold Spring Harb. Perspect. Biol.* 2:a001552. doi: 10.1101/cshperspect.a001552

Zhang, K. X., Xu, H. H., Yuan, T. T., Zhang, L., and Lu, Y. T. (2013). Blue-light-induced PIN3 polarization for root negative phototropic response in Arabidopsis. *Plant J.* 76, 308–321. doi: 10.1111/tpj.12298

Zhang, Q., Li, J., Zhang, W., Yan, S., Wang, R., Zhao, J., et al. (2012). The putative auxin efflux carrier OsPIN3t is involved in the drought stress response and drought tolerance. *Plant J.* 72, 805–816. doi: 10.1111/j.1365-313X.2012.05121.x

Zhao, Y. (2010). Auxin biosynthesis and its role in plant development. *Annu. Rev. Plant Biol.* 61, 49–64. doi: 10.1146/annurev-arplant-042809-112308

Ziyomo, C., and Bernardo, R. (2013). Drought tolerance in maize: indirect selection through secondary traits versus genomewide selection. *Crop Sci.* 53, 1269–1275. doi: 10.2135/cropsci2012.11.0651