



OPEN ACCESS

EDITED BY

Nisha Singh,
Gujarat Biotechnology University, India

REVIEWED BY

Anuj Kumar,
Dalhousie University, Canada
Priyanka Jain,
Amity University, India

*CORRESPONDENCE

Pusarla Susmitha

✉ pusarlasushmita94@gmail.com

Te Ming Tseng

✉ tt1024@amsstate.edu

[†]These authors have contributed equally to this work

RECEIVED 14 December 2022

ACCEPTED 08 June 2023

PUBLISHED 14 August 2023

CITATION

Susmitha P, Kumar P, Yadav P, Sahoo S, Kaur G, Pandey MK, Singh V, Tseng TM and Gangurde SS (2023) Genome-wide association study as a powerful tool for dissecting competitive traits in legumes. *Front. Plant Sci.* 14:1123631. doi: 10.3389/fpls.2023.1123631

COPYRIGHT

© 2023 Susmitha, Kumar, Yadav, Sahoo, Kaur, Pandey, Singh, Tseng and Gangurde. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide association study as a powerful tool for dissecting competitive traits in legumes

Pusarla Susmitha^{1*†}, Pawan Kumar^{2†}, Pankaj Yadav³, Smrutishree Sahoo⁴, Gurleen Kaur⁵, Manish K. Pandey⁶, Varsha Singh⁷, Te Ming Tseng^{7*} and Sunil S. Gangurde⁸

¹Regional Agricultural Research Station, Acharya N.G. Ranga Agricultural University, Andhra Pradesh, India, ²Department of Genetics and Plant Breeding, College of Agriculture, Chaudhary Charan Singh (CCS) Haryana Agricultural University, Hisar, India, ³Department of Bioscience and Bioengineering, Indian Institute of Technology, Rajasthan, India, ⁴Department of Genetics and Plant Breeding, School of Agriculture, Gandhi Institute of Engineering and Technology (GIET) University, Odisha, India, ⁵Horticultural Sciences Department, University of Florida, Gainesville, FL, United States, ⁶Department of Genomics, Prebreeding and Bioinformatics, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India, ⁷Department of Plant and Soil Sciences, Mississippi State University, Starkville, MS, United States, ⁸Department of Plant Pathology, University of Georgia, Tifton, GA, United States

Legumes are extremely valuable because of their high protein content and several other nutritional components. The major challenge lies in maintaining the quantity and quality of protein and other nutritional compounds in view of climate change conditions. The global need for plant-based proteins has increased the demand for seeds with a high protein content that includes essential amino acids. Genome-wide association studies (GWAS) have evolved as a standard approach in agricultural genetics for examining such intricate characters. Recent development in machine learning methods shows promising applications for dimensionality reduction, which is a major challenge in GWAS. With the advancement in biotechnology, sequencing, and bioinformatics tools, estimation of linkage disequilibrium (LD) based associations between a genome-wide collection of single-nucleotide polymorphisms (SNPs) and desired phenotypic traits has become accessible. The markers from GWAS could be utilized for genomic selection (GS) to predict superior lines by calculating genomic estimated breeding values (GEBVs). For prediction accuracy, an assortment of statistical models could be utilized, such as ridge regression best linear unbiased prediction (rrBLUP), genomic best linear unbiased predictor (gBLUP), Bayesian, and random forest (RF). Both naturally diverse germplasm panels and family-based breeding populations can be used for association mapping based on the nature of the breeding system (inbred or outbred) in the plant species. MAGIC, MCILs, RIALs, NAM, and ROAM are being used for association mapping in several crops. Several modifications of NAM, such as doubled haploid NAM (DH-NAM), backcross NAM (BC-NAM), and advanced backcross NAM (AB-NAM), have also been used in crops like rice, wheat, maize, barley mustard, etc. for reliable marker-trait associations (MTAs), phenotyping accuracy is equally important as genotyping. Highthroughput

genotyping, phenomics, and computational techniques have advanced during the past few years, making it possible to explore such enormous datasets. Each population has unique virtues and flaws at the genomics and phenomics levels, which will be covered in more detail in this review study. The current investigation includes utilizing elite breeding lines as association mapping population, optimizing the choice of GWAS selection, population size, and hurdles in phenotyping, and statistical methods which will analyze competitive traits in legume breeding.

KEYWORDS

breeding, genomic selection, linkage, mapping, phenotyping, protein

Introduction

The term legume originated from the Latin word “legumen”, which denotes “seeds harvested in pods”. During the Neolithic Revolution, which marked the beginning of human farming methods, farmers were accompanied by legumes that belong to the family Fabaceae. It is acknowledged that inadequate protein-energy intake and micronutrient deficits are two of the primary causes of undernutrition. Legumes play a minor but significant role in our food system. They are the superior economical dietary solutions due to their rich protein content (17%–30%) and relevant nutritional richness compared to expensive food sources containing animal-based protein and dairy products that may be difficult to obtain in situations where there is food insecurity (Marinangeli et al., 2017).

Compared with cereals, legumes provide a substantial quantity of protein throughout the complete plant, notably in grains. The incorporation of leguminous crops in cropping systems enabled an enhancement in soil quality (Hasanuzzaman et al., 2020). Legumes’ ability to fix atmospheric nitrogen in symbiotic relationships with soil bacteria such as *Rhizobium* and *Bradyrhizobium* minimizes the requirement for chemical fertilizers during crop growth and contributes to a reduction in greenhouse gas emissions like nitrous oxide (N₂O) and carbon dioxide (CO₂). In addition, they can help to reduce the utilization of fossil-based energy inputs in the chain of agriculture and food production by infusing high-quality organic matter, facilitating nutrient circulation, and promoting water retention in the soil (Stagnari et al., 2017). Legumes are rich in nutraceuticals, such as vitamin B6, calcium, magnesium, sodium, zinc, copper, and manganese. Thus, it is crucial to expand the genetic background and foster the breeding of legume crops, which will serve the needs of the growing human population under changing climatic conditions. Therefore, it is essential to come up with high-yielding cultivars that have enhanced resistance to diseases, higher nitrogen fixation ability, and tolerance to abiotic and biotic stresses, which can be achieved using biotechnological and genomics-assisted breeding approaches.

Genome-wide association study (GWAS) is an effective technique for determining the genes underlying a particular trait. To accomplish this, it is ideal to assess the genomic regions where

genotypic and phenotypic variations are correlated with each other. In comparison to standard biparental populations, GWAS offers greater mapping precision for detecting interactions among molecular markers and desirable characteristics in a variety of crops (Liu et al., 2016; Cui et al., 2017). It has become a vital tool in agricultural genetics due to its techniques that build upon the mixed linear model (MLM) framework and deliver radically improved computational speed and statistical power.

Furthermore, improvements can be applied in fields like omic-wide association studies, which utilize GWAS techniques to analyze relationships among desirable morphological traits and other kinds of omics data that include transcriptomic or metabolomic. GWAS requires structuring the population of diverse panels to estimate genetic distinction and minimize the detection of spurious connections (Sul et al., 2016). Breeders can develop new varieties owing to recent innovations in NGS applications and technologies that enable advanced tools to characterize genetic variation at a high resolution (Gali et al., 2019). The ultimate objective of this review is to quantify the genetic diversity, GWASs, and other related aspects or techniques that could be used to break the plateau of yield in legume crop production and can be utilized for further crop improvement.

Mapping population in association studies

Association mapping (AM), an alternative to QTL mapping, is dependent on linkage disequilibrium (LD) and uses collections of genotypes with known or unknown ancestry that have a significant degree of genetic variation due to hundreds of recombination cycles. The ultimate goal of association studies is to find a strong correlation between a genome-wide DNA marker and an interesting attribute that can be highly useful in marker-assisted selection for crop development. GWAS and candidate gene (CG)-based analysis are two important approaches to AM.

The creation of a mapping population that will be tested for the marker–trait relationship is a prerequisite for the GWAS. Both broad-based natural populations and narrow-based breeding populations can be utilized as the mapping population for GWAS

(Figure 1). The sort of mapping population needed for the success of GWAS hangs significantly on the mode of pollination (inbreeding or outbreeding) of the plant species. Both natural diverse germplasm panels and family-based breeding populations can be used for this. Among the breeding population, both biparental and multiparental mapping populations such as Multiparent Advanced Generation Inter-Cross ([MAGIC), Multiline Cross Inbred Lines (MCILs), Recombinant Inbred Advanced Intercross Lines (RIAILs), Nested Association Mapping (NAM), and Random Open- parents Association Mapping (ROAM) are being used for AM in several crop plants. Populations such as doubled haploid NAM (DH-NAM), backcross NAM (BC-NAM), and advanced backcross NAM (AB-NAM) that are modifications of NAM have also been used in recent times. The selection of the mapping population should be taken care of enough to avoid the false-positive marker–trait association. Because of the problematic inconsistent phenotyping scores of segregating lines over the years and location, heterozygote segregating individuals should not be included with the inbred lines as one population when creating the AM panel. When significant features like days to bloom and maturity are influencing the target trait, extreme genotypes should be eliminated from the AM panel for proper scoring of trait data (Kulwal and Singh, 2021). Each population has unique virtues and flaws, which will also be discussed further in the review study.

analysis of the target characteristic. These populations, however, are not appropriate for QTL mapping. When a germplasm accession collection represents the natural population, it may be a core collection or a sample that is more resilient to environmental changes. The population is excellent for assessing the QTLs for rare alleles that can help develop elite breeding lines or highly heritable domestic features. QTLs for some agronomically key characteristics have been uncovered in germplasm of several crops using GWAS, such as in 135 pea accessions (Gali et al., 2019), 366 sesame accessions (Cui et al., 2017), and 119 accessions in rice (Pawar et al., 2021).

The cultivars and lines created by a deliberate breeding program are known as elite inbred lines. These lines are unbreakable and can be maintained by numerous researchers in various places to identify QTLs using an AM panel. For instance, two AM panels of maize having 306 dent corn and 292 European flint corn inbred lines were individually assessed using single-nucleotide polymorphism (SNP) markers in the cold and control growth chamber conditions to identify genes related to cold tolerance (Revilla et al., 2016). For GWAS research in sorghum, AM panels of 377 tropical accessions from various geographic and climatic zones, significant U.S. breeding lines, and the wild species have been brought together to be used as AM panels (Casa et al., 2008). GWASs in legumes mostly include the natural populations and elite advanced breeding lines (Table 1), whereas GWAS using artificial mapping populations is more or less a recent phenomenon, and they are still underway.

Natural population and elite breeding lines as association mapping population

Any naturally occurring panmictic population with a significant history of recombination events can undergo AM. Utilizing hundreds of recombination events makes it simple to do an LD

Biparental mapping population and association mapping

Recombinant inbred lines (RILs) and Near Isogenic Lines (NILs) are the most used biparental population, usually used for linkage mapping or Quantitative Trait Locus (QTL) or QTL

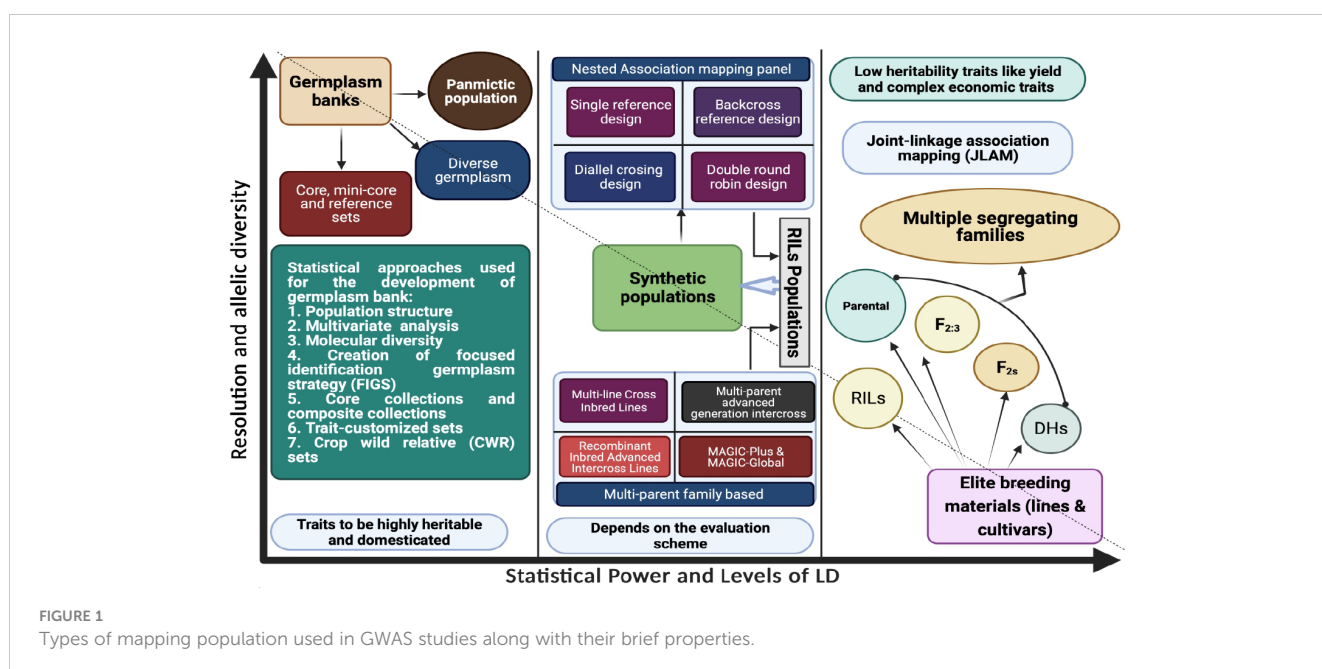


FIGURE 1 Types of mapping population used in GWAS studies along with their brief properties.

TABLE 1 GWAS studies for various traits in different leguminous crops.

Crop	Mapping population	Traits	QTLs/Marker trait associations	References
Gram	132 varieties and Advanced Breeding Lines (ABLs)	Yield traits	38 MTAs (marker trait association)	Li et al. 2018
	192 desi & kabuli accessions	Seed weight	8 MTAs	Bajaj et al., 2016
	182 diverse genotypes	Phenological, physiological and yield traits	14-34 MTAs in different environment condition	Jha et al. 2021
	75 ABLs	Fusarium wilt	3 MTAs	Jha et al. 2021
	165 chickpea genotypes	resistance to <i>Ascochyta rabiei</i>	30 MTAs	Farahani et al. 2022
	280 accessions	Grain Nutrient and Agronomic Traits	20 and 46 MTAs for grain nutrient and agronomic traits, respectively	Srungarapu et al., 2022
Arhar	Diverse collection of 142 pigeonpea lines	Flowering related traits	22MTAs	Kumar et al. 2022
	Pangenome based on 89 accessions	9 agronomic traits	229 MTAs	Zhao et al. 2020
Faba beans	481 elite breeding lines	Agronomic Traits	30 MTAs	Keller et al., 2020
Lentil	188 lines of the USDA Lentil Core Collection	Pea aphid	15 candidate genes	Das et al. 2022
Pea	135 pea accessions	Agronomic and Seed Quality Traits	251 MTAs	Gali et al., 2019
	135 pea accessions	Heat and Drought Adaptive Traits	15 MTAs	Tafesse et al. 2021
Mungbean	127 test genotypes	Mungbean yellow mosaic India virus resistance	15 MTAs	Singh et al. 2020
	95 cultivated mung bean genotypes	Seed Mineral content	43 MTAs	Wu et al. 2020
Blackgram	100 diverse genotypes	Agronomic traits	42 QTLs	Singh et al. 2022
	99 diverse genotypes	Agronomic traits	83 MTAs	Nkhata et al. 2021
Soybean		Protein, Oil, unsaturated fatty acid, oleic acid	SNP	(Hwang et al., 2014; Zhang et al., 2019; Zhao et al., 2019; Liu et al., 2020)
		Nematode resistance, Iron deficiency and Canopy wilt, brown stem rot, Diseases resistance	SNP	(Butenhoff, 2015; Vuong et al., 2015; Chang et al., 2016; Rincker et al., 2016; Zhang et al., 2017a; Zhang et al., 2017b; Zatybekov et al., 2018; Do et al., 2019; Ravelombola et al., 2019; Tran et al., 2019; Che et al., 2020; Lin et al., 2020)

(Continued)

TABLE 1 Continued

Crop	Mapping population	Traits	QTLs/Marker trait associations	References
		Salt tolerance, Flood tolerance, Drought tolerance, Water Use Efficiency	SNP	(Dhanapal et al., 2015; Zeng et al., 2017; Chen et al., 2018; Khan et al., 2018; Yu et al., 2019; Assefa et al., 2020)
		Agronomic trait	SNP	(Wen et al., 2015; Zhang et al., 2015; Contreras-Soto et al., 2017; Yan et al., 2017; Zhang et al., 2021; Zatybekov et al., 2017; Pan et al., 2018; Hu et al., 2019a; Li et al., 2019; Kim et al., 2020)
		Physiological traits	SNP	(Sui et al., 2020; Wang et al., 2020; Yang et al., 2020)
Groundnut	170 genotypes	Quality traits	SNP	(Shaibu et al., 2019a)
	125 ICRISAT groundnut mini core collection	Physiological traits	SNP	(Shaibu et al., 2019b; Shaibu et al., 2020)
	158 peanut accessions; 195 peanut accessions	Agronomic traits	SNP	(Zhang et al., 2017c; Wang et al., 2019)
	120 genotypes	Disease resistance	SNP	(Zhang et al., 2019; Zhang et al., 2020)
	249 peanut accessions	Abiotic stress tolerance	SNP	(Zou et al., 2020)
Chickpea		Agronomic traits	SNP	(Bajaj et al., 2015a; Bajaj et al., 2016; Kujur et al., 2015a; Upadhyaya et al., 2015; Upadhyaya et al., 2017; Basu et al., 2018; Orsak et al., 2019; Fayaz et al., 2022; Srungarapu et al., 2022)
		Abiotic stress tolerance	SNP	(Thudi et al., 2014; Li et al., 2018; Kohli et al., 2020; Ahmed et al., 2021; Kalve et al., 2022)
		Physiological traits	SNP	(Basu et al., 2019)
		Quality traits		(Upadhyaya et al., 2016a; Upadhyaya et al., 2016b; Parida et al., 2017; Samineni et al., 2022)
		Biotic stress resistance		(Li et al., 2017; Agarwal et al., 2019; Agarwal et al., 2022; Farahani et al., 2022; Raman et al., 2022)
Beans		Agronomic traits/ Quality	SNP	(Warsame et al., 2019; Rasool et al., 2022)
		Abiotic stress	SNP	(Ali et al., 2016; Li et al., 2017; Hu et al., 2019b; Breria et al., 2020; Abou-Khater et al., 2022; Maalouf et al., 2022; Sallam et al., 2022)
		Biotic stress	SNP	(Faridi et al., 2021)
Lentils		Agronomic traits	SNP, SSR	(Kumar et al., 2018a; Kumar et al., 2018b; Singh et al., 2019; Karthika et al., 2021; Neupane et al., 2022)
		Quality traits, Seed quality	SNP	(Khazaei et al., 2017; Khazaei et al., 2018; Johnson et al., 2021; Hang, 2022; Puspitasari et al., 2022)
		Biotic stress	SNP	(Banoo et al., 2020; Gela et al., 2021)
		Abiotic stress	SSR	(Singh et al., 2017; Kumar et al., 2019; Ma et al., 2020)

mapping. Whereas the power of QTL identification is higher in linkage mapping as compared to AM, the resolution has a reverse relationship with both mapping schemes. The concept of joint linkage AM (JLAM) was introduced to fully exploit the capabilities of both mapping methods. JLAM uses either a biparental population set or one or more multiparental AM panels, or two sets of genotypes consisting of germplasm and biparental mapping populations, which are genotyped utilizing the same set of markers (Myles et al., 2009; Lu et al., 2010; Reif et al., 2010 and Wurschum et al., 2012). Hence, JLAM is also

recognized as integrated mapping that identifies more significant marker–trait associations and increases the power of AM. Using JLAM (by combining germplasm accessions and full-sib F2 population of a bioenergy crop Shrub willow (*Salix* sp.) identified several major QTLs along with QTL hotspots (Carlson et al., 2019). Several studies using JLAM include QTL identification and CG identification for drought tolerance in maize (Lu et al., 2010), pleiotropic QTLs for silique length and seed weight in rapeseed (Li et al., 2014), and the epistatic QTLs for agronomically important characters in sugarbeet (Reif et al., 2010). Recent studies claim that

regulating population structure and addressing rare alleles can be accomplished through cofactors and a demographic effect accounting for JLAM, which enhances the predictive power of the methods (Wurschum et al., 2012).

Multiparent mapping population for GWAS

The multiparent populations include several founder parents, which reflect wider genetic diversity. Hence, in AM studies, the use of multiparent mapping populations helps limit the demerit of recombination frequency in biparental populations. Multiparent mapping populations provides tools to control population structure and balance allele frequencies. The historical and artificial recombinational events of the multiparent mapping populations such as NAM and MAGIC populations and their derivatives increase the efficiency of QTL identification in AM. Because of the controlled crosses, NAM population has higher power because of maximized population structure and minimal familial relatedness and accumulated frequency of rare alleles. The population facilitates cost-effective GWAS and allows the perpetual sharing of the NAM panel with global researchers.

To generate sets of RILs, NAM populations can be developed using reliable mating strategies such as diallel mating, NCD-II (North Carolina design II), eight-way cross, and single/double round robin. NAM population was first developed in maize using RILs developed from a diverse set of parents. Twenty-five diverse families in maize were used to develop 5,000 RILs that were evaluated for southern leaf blight disease resistance (Kump et al., 2011), and the wide diversity helped in the identification of 32 QTLs for the trait. A NAM population was developed using 23 different inbreds of barley in a twofold round-robin design to identify QTLs and CGs for grain morphology (Shrestha et al., 2022). NAM population has been established in both autogamous and allogamous species such as barley, rice, wheat, sorghum, and maize (Maurer et al., 2015; Bajgain et al., 2016). Several modifications of NAM, such as DH-NAM, BC-NAM, and AB-NAM, have also been used in recent times. An AB-NAM of barley consists of 796 BC₂F_{4,6} lines, which were derived from 25 wild barley accessions by backcrossing to the cultivar Rasmusson (Nice et al., 2016). Using 384 SNP markers, the AB-NAM population with minimal undesirable and unadapted characteristics of the wild barley parents was genotyped and encountered 10 QTLs for grain protein content (Nice et al., 2016).

A MAGIC population is created by a group of RILs from a complex cross or a group of crosses with numerous parents. Multiple rounds of recombination occur as these populations mature, improving the accuracy of desirable recombination and desirable alleles, thereby increasing the resolution of QTL mapping. With the aid of single seed descent (SSD), highly homozygous lines will be developed to establish the MAGIC population. To develop MAGIC populations for wheat and rice that can be deployed for QTL mapping, indica and japonica lines have been adopted. Seven cycles of SSD selfing resulted in 305 F8 lines in cowpea (*Vigna*

unguiculata) (Huynh et al., 2018). In the MAGIC indica rice population, 400 lines from S2 bulk were chosen on the basis of agronomic attributes and evaluated in mega-environment trials to select elite lines (Bandillo et al., 2013).

New high-throughput genotyping technologies for plants

The molecular markers are being progressively used to expedite breeding efforts in the post-genome sequencing era. Modern plant breeding is shifting from classical breeding to molecular breeding, where various genotyping technologies are being used for the discovery of molecular markers. In the last decade, a huge number of molecular markers were used for structural analysis of large germplasm populations to understand the diversity and use in GWAS. The whole-genome sequencing for most of leguminous crops has already been completed. Chromosome-level genomes are completed for most of the leguminous crops (Varshney et al., 2013). In the pre-genome sequencing era, the simple sequence repeat (SSR) markers were very powerful and potentially used for GWAS analysis. SSRs are tandem repeats highly polymorphic, abundant, co-dominant, and distributed throughout the genome. However, SSR markers are very laborious and time-consuming when compared with modern genotyping platforms such as double-digest restriction site-associated DNA sequencing (ddRAD-Seq) or specific locus amplified fragment sequencing (SLAF-Seq), whole-genome resequencing (WGRS), genotyping-by-sequencing (GBS), SNP-chip arrays, diversity array technology (DArT) array technology. With Illumina, gigabases of DNA sequencing data may be generated in a short period and cost-effectively in the NGS era (Bentley et al., 2008), Roche (Rothberg and Leamon, 2008), and AB-SOLiD (Pandey et al., 2008).

Molecular markers have become crucial components in molecular breeding over the past 2 years (Nadeem et al., 2018; Horst and Wenzel, 2007; Eathington et al., 2007). Molecular breeding has gained popularity and has been accepted by plant scientists because of its rapid and precise results for germplasm classification, back cross-breeding, and marker-assisted selection (Kumar et al., 2011; Nair and Pandey, 2021). A plethora of studies has been done using molecular markers (Kujur et al., 2015a; Wu et al., 2010; Song et al., 2013; Deokar et al., 2014; Shao et al., 2022). Different types of markers have been used for genotyping of legume, which includes rapid amplified polymorphic DNA (RAPD) (Doldi et al., 1997; Thompson et al., 1998; Iruela et al., 2002; Talebi et al., 2008), amplified fragment length polymorphism (AFLP) (Nguyen et al., 2004; Singh et al., 2008; Ude et al., 2002), inter-SSR (ISSR) (Yadav et al., 2014; Bhagyawant and Srivastava 2008; Iruela et al., 2002; Souframanien and Gopalakrishna, 2004), and SSR (Saxena et al., 2010; Choudhary et al., 2012; Zavinon et al., 2020).

However, continuous improvement of next-generation sequencing (NGS) technologies in recent years has made it cost-effective and accessible for any crop, including legumes (Poland et al., 2012). Reference genome sequencing has been completed in some legume crops like soybean, pigeon pea, groundnut, cowpea,

chickpea, and common bean (Afzal et al., 2022; Salgotra and Stewart, 2022). The currently available NGS technologies sequence each molecular or base pair of the DNA of any organism and make it feasible for us to identify the number of SNP markers with high precision and in a very short period (Liew et al., 2004). Although SNPs are biallelic and their polymorphism information is much lower compared to SSRs, they cover a significantly large part of the genome, which makes them markers to go for GWASs. In the last decade, a plethora of genotyping studies were carried out using SNPs in chickpea (Kujur et al., 2015b; Gaur et al., 2012; Hiremath et al., 2012; Deokar et al., 2014), pigeon pea (Raju et al., 2010; Singh et al., 2016; Arora et al., 2017), groundnut (Varshney, 2016; Pandey et al., 2017; Abady et al., 2021), soybean (Wu et al., 2010; Song et al., 2013; Shao et al., 2022), and other legume crops (Bohra et al., 2021; Shilpa and Lohithaswa, 2021).

PCR-based genotyping methods

Amplification of DNA segments with PCR leads to the development of multiple genotyping methods. If the primers in a PCR reaction include the variation of interest, then it is called as allele-specific PCR. Allele-specific markers are generally used during foreground selection during marker assisted selection. PCR-Restriction Fragment Length Polymorphism (RFLP) is another method of PCR-based genotyping (Saiki et al., 1985), where the genomic region of interest is PCR-amplified using the markers and then digested with restriction enzymes specifically recognize a DNA sequence, so that the digested product can produce alleles of different size, which can distinguish among the individuals. Microsatellites or short tandem repeat polymorphisms are ideal markers for PCR-based genotyping as the length of the amplified DNA fragment varies based on repeats of microsatellites in the genome (Weber and May, 1989). Before NGS technologies, a variety of DNA-based markers have been developed and used for genotyping, for instance, RAPD, SSRs (Hong et al., 2021), ISSRs, and AFLP. Among them, SSRs were most widely used in genotyping and genetic mapping studies. PCR-based genotyping methods are cheaper as compared to NGS technologies. However, the PCR-based genotyping methods are laborious and not highly efficient as NGS-based genotyping. The NGS-based genotyping includes restriction digestion of DNA and sequencing of libraries.

Double-digest restriction site-associated DNA

Although the SSRs are a potent marker system because of high reproducibility, co-dominance, and polymorphism, it is time, therefore, to generate the thousands of genome-wide SNP markers, restriction-sites associated with DNA sequencing (RADSeq) for large populations to study population genetics and genetic dissection of complex traits (Davey and Blaxter, 2011). However, in RADSeq, ~30%–50% of data were discarded because of

repeated variable sites. The more reliable technique of double-digest restriction site-associated DNA sequencing (ddRAD-Seq) was developed to boost the efficiency (Peterson et al., 2012). The ddRAD-seq simultaneously uses two restriction enzymes to decrease the genome entanglement and library preparation cost by five-folds and can capture the genomic regions in hundreds of thousands for enhanced representation of the genome. It was successfully used in genetic mapping studies in peanut to map the QTLs for late leaf resistance and plant type-related traits (Zhou et al., 2014; Zhou et al., 2016). The ddRAD-seq was further advanced to reduce the repetitive DNA sequences, and the optimized version of ddRAD-Seq was developed called SLAF-Seq. The steps in SLAF-Seq are the same as in ddRAD-Seq. The DNA fragments are optimized for even distribution and to reduce the repetitive sequences. However, both technologies do not cover the whole genome (Sun et al., 2013).

Genotyping-by-sequencing

GBS is a robust genotyping technology used for SNP discovery for a multitude of applications (Elshire et al., 2011). It is a variation of ddRAD-seq, first discovered in maize and barley used for genotyping recombinant inbred line populations. In GBS, methylation-sensitive restriction enzymes play a vital role in DNA digestion that lessens the genome complexity while constructing the sequence libraries. The genomic areas that are difficult to access to contemporary sequencing techniques can be captured by GBS. The GBS was efficiently used in groundnut for trait mapping (Jadhav et al., 2021) and diversity analysis (Khera et al., 2013). Pandey and co-workers (2014) performed GWAS analysis using SSR and GBS-based SNP genotyping data to identify the SNPs associated with aflatoxin contamination and agronomic traits in groundnut. GBS was used for genotyping cultivated and wild accessions of chickpea to discover 82,489 SNPs used for diversity, population structure, and LD analysis (Bajaj et al., 2015a; Kujur et al., 2015a). A total of 3,187 SNPs were used to reveal the genetic cluster associated with black-seeded genotypes of chickpea. GBS was also used for genotyping biparental populations in trait mapping studies to identify the QTLs for sterility mosaic disease (Saxena et al., 2017a), fusarium wilt (Saxena et al., 2017b), and fertility restoration (Saxena et al., 2018) in pigeon pea. In chickpea, drought tolerance-related “*QTL-hotspot*” was discovered with 743 SNP loci (Jaganathan et al., 2015), and 3,228 SNP loci were used for mapping and identification of CGs of seed traits (Verma et al., 2015). The multiplex sequencing strategy by using adapter sequences makes GBS very inexpensive. However, it produces more missing calls, and imputations are highly recommended during quality analysis. However, GBS is also incomplete, as its sequencing covers only a limited genome (~2.5%). GBS has replaced the previous genotyping markers, i.e., RAPD, ISSR, and SSRs, as it requires less time and labor and is highly cost-effective. GBS technology has been done in legumes like chickpea and soybean (Shingote et al., 2022; Torkamaneh et al., 2021; Iquira et al., 2015; Bajaj et al., 2016; Sudheesh et al., 2021; Kujur et al., 2015b; Verma et al., 2015).

Diversity array technology

The polymorphic DNA segments called DArT markers in a genome are recognized through differential hybridization on a diversity genotyping array (Jaccoud et al., 2001). DArT is a very cost-effective whole-genome DNA fingerprinting tool for a variety of genetic analyses. It is a high-throughput sequence-independent technology that combines restricted-based hybridization and PCR. It is a very efficient marker system that can discover thousands of polymorphic sites in a very short time in any crop species. DArT is very popular in terms of high genome coverage, speed, reproducibility, and reliability (Aitken et al., 2014). Furthermore, polymorphic fragment calling does not require the reference genome. The DArT technology can be effectively used for genomic selection (GS) (Varshney et al., 2017) and marker-assisted selection (Stojaowski et al., 2011). However, the DArT markers are redundant due to clones with common sequences. Therefore, the presence of redundancy and markers with low frequencies (~41%) may affect the statistical analysis that is needed to filter out. DArT procedure includes generating a diversity panel followed by genotyping using a diversity panel. The first-ever genetic map of any legume crop was designed using DArT technology by Yang and coworkers (2011) in pigeon pea. A biparental population (F₂) was screened using 554 DArT markers. Olukolu et al. (2012) used the DArT marker technology for genetic diversity assessment of 124 accessions of groundnut representing 25 countries of Africa. Roorkiwal et al. (2014) used the DArT markets to diversify the 10 *Cicer* species, including 94 genotypes. Aldemir and coworkers (2017) used an advanced version of DArT technology, i.e., DArT sequencing (DArTseq), for the identification of QTL for iron content in lentil seeds. DArTseq is also a hybridization-based technology but combines with NGS and provides a much simpler form of sequencing than DArT (Courtois et al., 2013; Aldemir et al., 2017). Ates (2019) estimated the genetic diversity of 94 lentil landraces with DArT-based 19,383 SNPs.

SNP arrays

NGS technologies discovered an ample number of SNP markers because the demand for high-throughput genotyping has increased. The hybridization-based microarray or SNP arrays are very popular in genetic mapping, diversity analysis, and population genomics (You et al., 2018). SNP array or DNA microarray are highly polymorphic and use designed probes hybridized with fragmented DNA, which determines the alleles of all the SNP positions for hybridized DNA samples (LaFramboise, 2009). On the basis of the density, the SNP arrays can be divided into high-density (>50K), mid-density 5–10K), and low-density (>5K) SNP arrays. High-density SNP arrays can be used for high-density genetic mapping, GWAS, and population genomics studies. Mid-density assays can be used in GS because a few thousand SNPs are enough based on the genome size of the individual. However, the low-density SNP arrays can be used for foreground and background selection during marker-assisted selection and several breeding purposes. For

instance, the quality control panel of rice is a low-density SNP array (25 SNPs), highly used for F1 confirmation, hybrid purity testing, and DNA fingerprinting in rice (Ndjiondjop et al., 2018). SNP arrays have been efficiently developed in several crops for genotyping, such as maize (600K SNP array) (Unterseer et al., 2014), apple (480K SNP array) (Bianco et al., 2016), and rice (700K SNP array) (McCouch et al., 2016). In leguminous crops such as peanut, the SNP *Arachis* array with 58K SNPs (Pandey et al., 2017) was very successful for genetic mapping (Pandey et al., 2020) and association analysis (Gangurde et al., 2020) for several traits. In pigeon pea, 56K Axiom *Cajanus* SNP Array and chickpea 11K Axiom *Cicer* SNP Array were developed (Roorkiwal et al., 2018). However, they are fixed and may not capture all recombination or diversity in an association panel, which are the limitations of SNP arrays. For instance, for genotyping a multi-parent population such as MAGIC or NAM, the whole-genome resequencing-based genotyping is helpful to capture maximum recombination regions.

Whole-genome resequencing

Advanced NGS technologies reduced per-sample sequencing cost, and WGRS-based genotyping was used for many populations to identify the presence of absence variations for genome-wide association analysis. WGRS can be carried out at high depth or low depth based on the objective of the study. For instance, in the case of genetic mapping, 0.5–1.0X coverage is sufficient; however, for GWAS, 10–15X coverage can be used. Several NGS platforms can be used for generating WGRS data, such as Illumina Hi-seq (read length of 150–250 bp), PacBio (10–25Kb), and NanoPore (read size of 500 bp to 2.3 Mb). Large LD blocks (several hundred kilo-base pairs) in plants, specially self-pollinating. Large LD blocks include several CGs. Therefore, with dense genotyping, we can have SNP variants in each of the CGs in the block and individual CGs can be identified using GWAS carried out on WGRS genotyping data. A gene for salinity tolerance *Glyma03g32900*, using sequencing data on 106 soybean diversity panels and the SNP-based KASP markers, was developed to improve salinity tolerance in soybean (Patil et al., 2016). Recently, 2,980 chickpea accessions are sequenced to discover 3.94 million SNPs, phenotyping data on 16 traits was used for GWAS analysis and identified 205 SNPs associated with 11 traits, and the associated SNPs were in the genomic regions of 79 CGs playing a role in controlling key traits like seed weight (Varshney et al., 2021).

Alleviating the phenotyping bottleneck

In the era of different omics like genomics, transcriptomics, and proteomics with the help of NGS technologies, genotyping of large germplasm at multiple locations has become feasible for plant scientists. Thus, phenotyping these large germplasms/populations with higher accuracy have become difficult. Thus, high-throughput genotyping technologies have shifted the bottleneck of plant science

from genotyping to phenotyping (Mir et al., 2019). Thus, it has become the need for time to develop high-throughput phenotyping (HTP) approaches (Mir et al., 2019). Several advanced artificial intelligence-based HTP platforms have been developed for crops like rice, maize, and Arabidopsis (Yang et al., 2020). Still, a lot of improvement is required in HTP, which can record multiple phenotypic traits in less time and manpower, which can be associated with large genotypic data of large populations (Mir et al., 2019). The major limitation in phenotyping is recording the multiple traits (agronomic traits, physiological traits, and stress-related scoring) data of large populations at multiple locations in several replications (Furbank and Tester, 2011). There are a lot of chances for error in phenotypic data when recorded manually, and less accuracy leads to false significant associations with molecular markers and wrong interpretation of alleles and genes. HTP is a non-destructive data recording method that allows the plant scientist to increase the size of the experiment by the number of genotypes or replication, or locations (Awlia et al., 2016). PHENOPSIS was one of the first automated imaging and weighing systems developed in Arabidopsis to estimate its response to water deficiency (Granier et al., 2006). However, it has its limitations. HTP platforms are of two types, i.e., HTP platforms for greenhouse or laboratory experiments and open field experiments (Shafiekhani et al., 2017). Although, HTP technologies have been used successfully for genetic dissection of agronomic traits in major field crops like rice, maize, wheat, barley, and brassica (Zhang et al., 2017a; Shi et al., 2013; Yang et al., 2014; Muraya et al., 2017; Topp et al., 2013; Tanabata et al., 2012). The use of these HTP platforms in legume crops is yet to be evaluated at the large fields, population, and multiple location levels (Zhang et al., 2021). A handful of studies has been conducted on legumes such as pea, soybean, and chickpea using a HTP approach for biotic and abiotic stress (Zhang et al., 2012; Friedli et al., 2016; Humplík et al., 2015). Zhang et al. (2021) used the quadcopter unmanned aircraft vehicle multispectral imaging data to predict the yield of chickpea and dry pea with a multivariate regression model. Humplík et al. (2015) used the automatic red blue green image analyzing software in pea to estimate the shoot biomass and photosynthetic activity for cold tolerance. Friedli et al. (2016) used the terrestrial 3D laser scanning system in soybean for canopy-related traits.

Advanced methods and tools for GWAS

GWAS has continuously expanded in the last few decades due to advancements in sequencing technologies and the collective effort of the research community. In addition, HTP technologies have allowed us to measure many plant traits that are now frequently analyzed through GWAS tools. Recent years have seen GWAS methods solving issues of computation complexity or enhancing statistical power. It is utilized to detect new associations with traits of interest and to replicate loci detected by other different approaches. A diverse set of researchers is involved in rare-variant detection, statistical model optimization, synthetic associations, and using GWAS findings to better our knowledge of

disease etiology. These methods can detect genetic variants associated with biochemical or agronomic and molecular phenotypes. In the future, this will enhance the utility of GWAS methods and their implications for plant science.

Naïve methods

In the GWAS, linear or logistic regression models are used to test for associations. The linear model is used for continuous traits such as plant height, whereas logistic regression models are used for binary traits indicating that the disease is present or absent. In addition, some covariates are included to account for confounding effects from demographic factors. However, naïve approaches often suffer from inflated false-positive rates that might be induced due to genetic relatedness among study participants (Oetjens et al., 2016). In GWAS, usually, diverse populations are selected, which often have related individuals, making subpopulations within the population. This might lead to spurious associations between SNPs that are more common in each subpopulation and phenotypes of interest if the phenotype has a higher prevalence in that subpopulation.

Mixed linear model methods

The MLM frameworks used in GWAS have drastically decreased the false-positive rates in comparison with conventional naïve approaches. Among these, the fast GWA tool is an ultra-efficient tool for MLM-based GWAS analysis of biobank-scale data (Jiang et al., 2019). MLM approaches resolve the issue of genetic relatedness among individuals following correction at two levels. These refer to population structure and kinship (Yu et al., 2006). At the first level, the population structure is inferred using genotype data with STRUCTURE tool (Pritchard et al., 2000) or through principal component analysis (Price et al., 2006). The kinship matrix is used at the second level to estimate inter-individual relatedness using the genotype data (Yu et al., 2006). In recent years, many methods have been developed to efficiently solve MLM equations. For instance, a recently available method referred to as EMMA (efficient mixed-model association) provided superior computational speed by eliminating the duplicate matrix operations at each iteration while estimating the likelihood function (Kang et al., 2008). MLM-based methods become computationally intensive for large numbers of samples. The FaST-LMM solves this issue but requires that the number of SNPs be less than the number of samples to derive kinship. The SUPER (Settlement of MLM Under Progressively Exclusive Relationship) method has been developed to extract a subset of SNPs and use them in FaST-LMM to increase the statistical power. Moreover, the compress MLM (CMLM) and enriched CMLM (ECMLM) methods are available for kinship optimization. The modified MLM method called multiple-locus linear mixed model (MLMM) incorporates multiple markers simultaneously as covariates in a stepwise MLM to partially remove the confounding between testing markers and kinship. Furthermore, a new method referred to as fixed and random model circulating

probability unification (FarmCPU) completely removes the confounding by dividing MLM into a fixed-effect model and a random-effect model and using them iteratively. The FarmCPU can analyze the dataset with half million individuals and half million markers within 3 days. However, the random-effect model is computationally intensive in FarmCPU. The new method called Bayesian information and linkage disequilibrium iteratively nested keyway (BLINK) replaces the random-effect model with the fixed-effect model by using Bayesian information criteria. This method also replaces the bin method used in FarmCPU with LD information to eliminate the requirement that quantitative trait nucleotides be uniformly distributed throughout the genome. These all methods are summarized in [Table 2](#).

Machine learning methods

Recent years have seen tremendous growth in the machine learning methods targeted for GWAS. The approaches used by these methods include classification, regression, ensemble-based learning, and neural networks.

Regression

Logistic regression coupled with the least absolute shrinkage and selection operator (LASSO) regularization approach is a famous method for GWAS. The penalized logistic regression method was used for the classification of patients with Crohn's disease using genotyping data at the genome-wide level. The LASSO and ridge regression are among the most frequently utilized penalized regression algorithms ([Tibshirani et al., 1996](#); [Hoerl et al., 1970](#)). Recently, a faster and more powerful algorithm was developed by binning the closely

occurring SNPs based on LD ([An et al., 2020](#)). In addition, the SNPs and phenotypes were mapped using LASSO regression in this method. This method was found to provide a reduced type 1 error rate in comparison with regular MLM and LASSO. To discover variations closely associated with the duloxetine response, some researchers used the standard genome-wide logistic regression ([Maciukiewicz et al., 2018](#)). In addition, they extracted the top predictors using LASSO regression. In another study, a preconditioned random forest regression was used to predict late genitourinary toxicity after radiotherapy. This preconditioning involved usage of logistic regression for making a continuous surrogate outcome from the original binary outcomes, which were followed by random forest regression where the surrogate outcome is utilized as a target for prediction. In this study, five-fold cross-validation was conducted for testing the model stability against existing baseline models ([Lee et al., 2018](#)). The major drawback of regression approaches is the failure to find higher-order non-linear SNP interactions involved in susceptibility to diseases. The process developed by Zhang and coworkers (2012) utilizes prior information from proteomics and biological pathways for SNP groups. To find the top predictive SNP groups, they used linear regression standardized by group sparse constraint. In the end, group LASSO was used for the regularized linear regression ([Yuan et al., 2005](#)). Thus, this approach overcomes the limitations of the regular MLM used in GWAS.

Classification

Support vector machine (SVM)-based classification methods such as COMBI have been developed for unknown phenotype prediction for a given unseen genotype ([Mieth et al., 2016](#)). In this approach, the SNPs having larger SVM weight are chosen, and the remaining SNPs are removed. Next, a chi-squared test is

TABLE 2 Advanced methods and tools for GWAS.

S.No.	Method	Description	Reference
1.	MLM	At the first level, the population structure is inferred using genotype data with STRUCTURE tool or through principal component analysis. The kinship matrix is used at the second level to estimate inter-individual relatedness using the genotype data.	Jiang et al., 2019
2.	CMLM	Clusters the individuals into groups and fits the genetic values of groups as random effects in the model that improves statistical power compared to regular MLM methods.	Zhang et al., 2010
3.	ECMLM	Calculate kinship using several different algorithms and then choose the best combination b/w kinship algorithms and grouping algorithms.	Li et al., 2014
4.	FaST-LMM	An algorithm for genome-wide association studies (GWAS) that scales linearly with cohort size in both run time and memory use. This method requires that the number of SNPs be less than the number of samples to derive kinship.	Lippert et al., 2011
5.	SUPER	Uses the associated genetic markers referred as pseudo quantitative trait nucleotides instead of all the markers, to derive kinship.	Wang et al., 2014
6.	MLMM	Include multiple markers simultaneously as covariates in a stepwise MLM to partially remove the confounding between testing markers and kinship.	Segura et al., 2012
7.	FarmCPU	Uses a bin method under the assumption that quantitative trait nucleotides are evenly distributed throughout the genome. Completely eliminates the confounding by dividing MLM into a fixed effect model and a random effect model and using them iteratively.	Liu et al., 2016
8.	BLINK	Replaces the random effect model with the fixed effect model by using Bayesian information criteria. Uses linkage disequilibrium information to eliminate the requirement that quantitative trait nucleotides be uniformly distributed throughout the genome.	Huang et al., 2019

performed, and SNPs that exhibit a p-value below the significant criterion are taken into consideration for intensive study. The SVM method separates labeled data points into two groups with a large difference between them. Some authors proposed using SVM for genetic risk prediction (Mittag et al., 2012). This method has been used for genome-wide risk profiling for diseases such as type 1 diabetes and Parkinson's disease. In this algorithm, model training is performed using SNP data, which is followed by binary classification of the test dataset. Another researcher used the K-nearest neighbor learning algorithm for the classification of individuals into breast cancer positive and negative groups using their SNPs (Hajiloo et al., 2013). They used a leave-one-out cross validation strategy and external holdout methods for evaluating the performance of their classification algorithm.

Ensemble learning methods

These methods comprise an ensemble of decision trees. For example, random forest is an example of the ensemble learning algorithm. A bootstrapped subsample of the initial training dataset is used to create each decision tree in this instance. Some authors used gradient-boosting and random forest approaches to identify potent SNPs (Dorani et al., 2018). Nguyen et al. (2015) used a random forest method for selecting informative SNPs. They used a two-stage quality-based approach in model learning for the selection of informative SNPs. This method seems quite useful for the high-dimensional GWAS data. They also used five-fold cross-validation for assessing the potential of the model on different GWAS datasets. In addition, gradient boosting of decision trees was used for GWAS datasets. Others proposed using the XGBoost model for SNP selection (Behravan et al., 2018). This model could be used as an alternative to polygenic risk scoring. In addition, SVM classifier is used at the backend for SNP classification. Using principal component analysis and logistic regression, Oh et al. (2017) suggested a preconditioned random forest regression that converts a binary variable into a continuous variable. Later, Lee and a group of researchers (2020) used this preconditioned model for predicting the risk of breast cancer.

Neural network-based methods

Liu et al. (2019) developed a convoluted neural network (CNN) model for phenotype prediction using the SNP dataset. Moreover, they applied a saliency map for the first time to choose significant SNPs from training model. They also compared them with statistical methods such as best linear unbiased prediction and Bayesian ridge regression (BRR). In this study, association analysis was performed for quantitative traits of soybean and SNP datasets. Some authors found that increasing the hidden neuron's number does not affect the performance of the classification model for the case-control settings (Romagnoni et al., 2019). In a different study, authors compared the deep mixed model constituted of CNN and long and short-term memory with standard univariate testing and MLM (Wang et al., 2019).

Transcriptome-wide association study methods

Transcriptome-wide association study (TWAS) methods perform association analysis for gene expression variations and quantitative traits. TWAS is an approach based on genes with the ability to expand GWAS for a better understanding of functional relationships in complex traits. These methods are alternatives to variant-based association methods representing a subgroup of multi-marker association or locus-based methods. The locus-based methods have been so popular due to the larger apprehension and acceptability of the polygenic framework of the complex traits. In principle, locus-based approaches rely on multiple genetic variants to estimate the contribution of a gene or loci. TWAS uses GWAS results and transcriptome-level information to perform association testing at the gene level (Pividori et al., 2020). The ability to separate and assess the analytical procedures in TWAS simultaneously, provides several opportunities for the development of effective statistical models for the study of gene disease connections.

PheWAS methods

PheWAS methods perform unique associations in addition to utilizing known genotype–phenotype associations acquired through GWAS. These established relationships might serve as “positive controls” for additional high-throughput analysis. PheWAS methods suffer from high false-positive rates due to thousands of genotype–phenotype associations being tested in such studies (Bastarache et al., 2022). In addition, sample sizes usually also vary across studies impacting the statistical power and the replication among studies. PLATO tool is used to identify associations in PheWAS (Hall et al., 2017). DNAnexus is another tool for genomic analysis that was hosted on Amazon Web Services. This provides a distributed cluster of computers on the cloud allowing much lesser computation time for such studies. With the assistance of the DNAnexus app for PLATO, scatter-process-gather can be used on the platform to train regression models concurrently. This scatter-gather approach initiated multiple AWS virtual machines to simultaneously fit the regression models. Deep-PheWAS is another platform for PheWAS that intertwines quantitative phenotypes from primary care data, disease progression, longitudinal trajectories of quantitative measures, and drug response phenotypes with the composite phenotypes generated from clinically curated data (Packer et al., 2023). Moreover, several tools are available on this platform for efficiently analyzing the association with genetic data under different genetic models.

GWAS-assisted genomic selection

GS has been utilized as a practical genomic approach for upgrading complex traits in various crops (Thudi et al., 2016; Sandhu et al., 2022). In segregating populations, GS allows identifying lines with higher genomic estimated breeding value

(GEBV) using genome-wide marker data. A training population (TP) is used to estimate GEBV, which consists of elite breeding lines and is evaluated for multi-seasons and locations for the target phenotype. Then, a candidate population (CP) is developed by selecting parents based on the GEBVs. GS utilizes all the available genome-wide marker data irrespective of any significant effects. The GS prediction accuracies depend on several factors, including the genome size, ploidy level, interactions between gene and QTL, sample number, relatedness, number and distribution of markers, and model (Yadav et al., 2020). Several statistical methods are used for GS, including Ridge regression best linear unbiased prediction (rrBLUP) and genomic best linear unbiased predictor (gBLUP); both hypothesize a normal distribution of the SNP effects, whereas Bayesian methods like BayesA, BayesB, BayesC, and BayesR allow different variance distributions considering marker effect sizes (Heslot et al., 2012; de Los Campos et al., 2013). On the other hand, kernel approaches help predict non-additive models along with complex multi-environment/trait data (Gianola and van Kaam, 2008; Bandeira E Sousa et al., 2017).

Zhang and coworkers (2016) showed a GWAS-assisted GS with 309 soybean lines and 31,405 SNPs for seed weight using the rrBLUP approach. They showed GS prediction accuracies of 0.75–0.87, outperforming marker-assisted selection with prediction accuracies of 0.62–0.75. Ravelombola et al. (2020) performed a GS approach for soybean cyst nematode tolerance with biomass reduction using 234 soybean accessions in the greenhouse. They used five methods to compute GEBVs, including gBLUP (Zhang et al., 2007), random forest (RF) (Ogutu et al., 2011), rrBLUP (Meuwissen et al., 2001), SVMs (Maenhout et al., 2007), and Bayesian LASSO (Legarra et al., 2011). They found that the prediction accuracies were dependent on the model used, the marker set, and the size of TP. However, the accuracy of GS was higher than the SNPs from GWAS for all selection models and TP sizes.

In alfalfa, Li et al. (2015) used clonal ramets from 185 to 190 individuals for GS of biomass yield across three locations and recorded prediction accuracies of 0.43 to 0.66 for each location. Another study used 322 individual genotypes from 75 genetically diverse alfalfa populations. They tested three Bayesian models (BayesA, BayesB, and BayesC) for 25 agronomic traits, including forage quality traits, dry matter, and fall dormancy (Jia et al., 2018). They reported prediction accuracies of 0.0021 to 0.6485 with no significant differences in the three Bayesian models.

In chickpeas, Roorkiwal et al. (2016) used 320 breeding lines and six different models, including rrBLUP, RF, Bayesian LASSO, BayesB, Kinship GAUSS, and Bayes C π for four traits, i.e., seed yield, 100 seed weight, days to maturity, and days to 50% flowering. They reported prediction accuracies ranging between 0.138 (seed yield) to 0.192 (100 seed weight).

Li et al. (2018) showed low prediction accuracies using rrBLUP, Bayesian LASSO, and BRR for grain yield/ha, seed number per plant, 100 seed weight, and early vigor score in chickpea, which can be attributed to the small size of TP.

In common bean, cooking time (CKT), seed weight, and water absorption capacity were evaluated using 922 lines consisting of four populations (a Mesoamerican 8-parental MAGIC population, a biparental RIL, an Andean, and a Mesoamerican breeding line

(MIP) panel (Diaz et al., 2020). Six models based on additive effects (BRR, BayesA, BayesB, BayesC, Bayesian Lasso, and gBLUP) and a Bayesian reproducing kernel Hilbert spaces regression (RKHS) models based on both additive and non-additive effects were used. They reported prediction accuracies for CKT ranging from MIP (0.22) to MAGIC population (0.55). A recent study showed prediction abilities ranging between 0.6 and 0.8 were shown in common bean for four agronomic traits under several environmental stresses (Keller et al., 2020)

In peanut or groundnut, 281 Kersting's groundnut lines were used for GWAS-assisted GSs for several traits, including seed traits, 100 seed weight, leaf length, days to 50% flowering, and days to maturity using 493 SNPs and rrBLUP model (Akohoue et al., 2020). They recorded prediction accuracies ranging from 0.42 to 0.79 for 100 seed weight, seed length and width, days to maturity, and days to 50% flowering. A low prediction accuracy of 0.11–0.20 was reported for traits including plant architecture traits such as height and diameter, petiole length, leaf width, number of seeds, grain yield, number of pods per plant, and number of seeds per pod.

Recently, genomic resources have been made available in some minor legume crops (Varshney et al., 2019; Bohra et al., 2020). In peas, the predictive abilities based on Bayesian LASSO model were 0.28, 0.30, 0.64, and 0.65 for lodging susceptibility, yield, seed weight, and onset of flowering, respectively (Annicchiarico et al., 2019).

Several GS methods were used for predicting GEBVs in legume crops (Figure 2), but the progress still needs to catch up compared to grain crops, including wheat, rice, and maize. However, the GS approach proved helpful and could be applied in the early stages of legume breeding programs to identify promising progenies and parents based on the predicted breeding values.

Applications in plant breeding

Since the last decade, GWAS has been successfully used in major legume crops to dissect or identify the genetic bases for various agronomic traits (Bajaj et al., 2015b; Kujur et al., 2015a; Wen et al., 2015; Zhang et al., 2015), quality traits (Hwang et al., 2014; Upadhyaya et al., 2015; Shaibu et al., 2019a), biotic (Butenhoff, 2015; Zhang et al., 2019; Banoo et al., 2020; Zhang et al., 2020; Faridi et al., 2021), and abiotic stress (Thudi et al., 2002; Thudi et al., 2014; Dhanapal et al., 2015; Assefa et al., 2020; Kohli et al., 2020). A plethora of GWASs have been conducted using different types of markers (SSR, SNPs, etc.) in some of the major legumes like soybean (Butenhoff, 2015; Zhang et al., 2015; Zhang et al., 2019; Assefa et al., 2020), chickpea (Kujur et al., 2015b; Upadhyaya et al., 2017; Kohli et al., 2020), groundnut (Zhang et al., 2017c; Shaibu et al., 2019a; Wang et al., 2019), and minor legume crops (Ali et al., 2016; Faridi et al., 2021). Brief details of GWASs conducted in legume crops are given in Table 1.

Revolutionization and rapid development in genomic techniques in the recent past have accelerated molecular studies not only in model crops but also in other crops like legumes (Varshney et al., 2005). Sequencing and availability of reference genomes have also made it feasible for the researchers to identify the alleles/QTL association with the desired trait in any germplasm. Although the GWAS approach can be used in any crop with extensive phenotyping and genotyping, it has

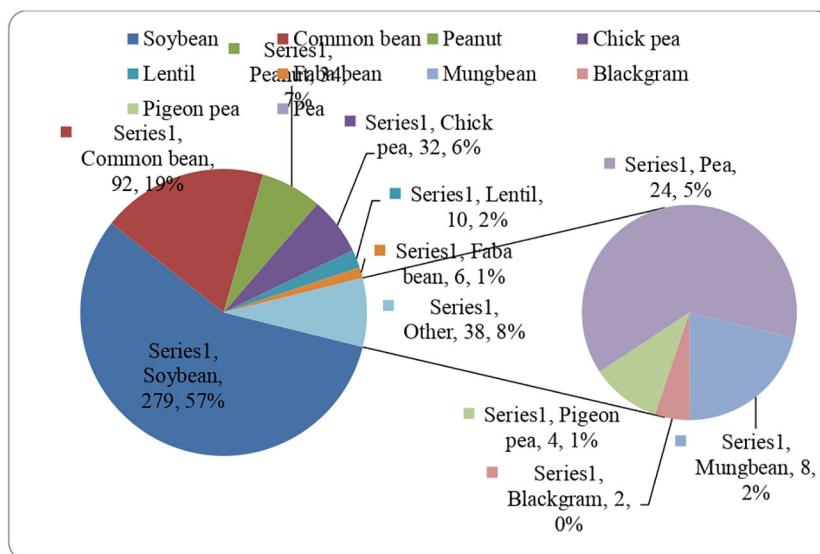


FIGURE 2
Number of GWAS studies conducted in different leguminous crops from timeline 2012 to 2023.

been used in major legume crops like soybean, chickpea, and groundnut (Mousavi-Derazmahalleh et al., 2019). These major legume crops' research community has sufficient funding for high-throughput genotyping and phenotyping. As these crops cover a significantly larger area across the globe, significantly diverse and classified germplasms are available for these crops (Mousavi-Derazmahalleh et al., 2019). However, GWAS has its limitations like false-positive association and exclusion of a significant association. All the limitations can be overcome by accurate phenotyping, large enough diverse germplasm, multilocation trials for phenotyping, and accuracy in genotyping. The use of the best suitable model, method, and bioinformatic tools also determines the accuracy of GWAS. The development of model tools for legume crops can trigger the GWAS in major and minor legume crops.

Conclusion and future perspectives

Legumes are an essential component of human nutrition and play a vital role in sustainable agriculture due to their protein-rich content, soil quality improvement, and reduced environmental impact. With the increasing global population and changing climatic conditions, there is a pressing need to develop high-yielding, disease-resistant legume cultivars that can meet the nutritional needs of the growing population. Improvement in the nutritional and production quality of legume crops with the use of conventional breeding methods is not at the required rate. Whole-genome sequencing is available only for a few major crops. As per the availability, low (RFLP) to high-throughput (SNP) markers have been used in various crops for AM, QTL mapping, or GWAS. NGS has become feasible in the model and even in non-model crops with improved efficiency and affordable sequencing methods. GWAS has been used in major legume crops to identify

the genomic region linked with desired characteristics of the plant. It is yet to be exploited in minor legumes with sufficient germplasm/population. The availability of reference genomes and rapid development in genomic techniques has made it feasible for researchers to identify the alleles/QTL association with the desired trait in any germplasm. The use of suitable models, methods, and bioinformatic tools determines the accuracy of GWAS. The development of model tools for legume crops can trigger GWAS in major and minor legume crops. Authentication or precision of identified marker-trait association is required for their utilization in plant breeding programs or MAS/BAC programs. Using NGS and other high-throughput techniques for sequencing will make it possible to develop a genomic-assisted crop improvement program in legumes. Rapid development can be gained concerning agronomic traits, biotic/abiotic stress tolerance, and after-use quality improvement. Legume yield potential is meager compared to other major crops; this yield plateau can be broken in legumes for climate change problems using GWAS with multi-location phenotyping. The integration of these new and improved technologies with traditional breeding methods will help to accelerate the development of new legume cultivars with improved yield and nutritional qualities.

Author contributions

PS, PK conceptualized the review study and edited the manuscript. PY contributed to advanced methods and tools for GWAS. SS put forth experimental populations for association mapping studies in various crops. GK was associated with GWAS-assisted genomic selection. SSG and MP contributed to new high throughput-genotyping technologies in plants. VS and TT assisted in collecting data for applications of GWAS in plant

breeding and drafting the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

References

- Abady, S., Shimelis, H., Janila, P., Yaduru, S., Shayanowako, A. I., Deshmukh, D., et al. (2021). Assessment of the genetic diversity and population structure of groundnut germplasm collections using phenotypic traits and SNP markers: implications for drought tolerance breeding. *PLoS One* 16 (11), e0259883. doi: 10.1371/journal.pone.0259883
- Abou-Khater, L., Maalouf, F., Jighly, A., Alsamman, A. M., Rubiales, D., Rispaill, N., et al. (2022). Genomic regions associated with herbicide tolerance in a worldwide faba bean (*Vicia faba* L.) collection. *Sci. Rep.* 12, 1–13. doi: 10.1038/s41598-021-03861-0
- Afzal, M., Alghamdi, S. S., Nawaz, H., Migdadi, H. H., Altaf, M., El-Harty, E., et al. (2022). Genome-wide identification, and expression analysis of CC-NB-ARC-LRR (NB-ARC) disease-resistant family members from soybean (*Glycine max* L.) reveal their response to biotic stress. *J. King Saud Univ* 34, 1758.
- Agarwal, C., Chen, W., Varshney, R. K., and Vandemark, G. (2022). Linkage QTL mapping and genome-wide association study on resistance in chickpea to *Pythium ultimum*. *Front. Genet.* 13, Art-945787.
- Agarwal, C., Vandemark, G. J., Chen, W., and Varshney, R. K. (2019). *Genome-wide association studies and QTL mapping in chickpeas for resistance to ascochyta blight and pythium ultimum*. San Antonio, TX: International Annual Meeting: Embracing the Digital Environment.
- Ahmed, S. M., Alsamman, A. M., Jighly, A., Mubarak, M. H., Al-Shamaa, K., Istanbuli, T., et al. (2021). Genome-wide association analysis of chickpea germplasm differing for salinity tolerance based on DArTseq markers. *PLoS One* 16, e0260709. doi: 10.1371/journal.pone.0260709
- Aitken, K. S., McNeil, M. D., Hermann, S., Bundock, P. C., Kilian, A., Heller-Uszynska, K., et al. (2014). A comprehensive genetic map of sugarcane that provides enhanced map coverage and integrates high-throughput diversity array technology (DArT) markers. *BMC Genomics* 15 (1), 1–12. doi: 10.1186/1471-2164-15-152
- Akhoue, F., Achigan-Dako, E. G., Sneller, C., Van Deynze, A., and Sibiy, J. (2020). Genetic diversity, SNP-trait associations, and genomic selection accuracy in a West African collection of kersting's groundnut [*Macrotyloma geocarpum* (Harms) maréchal & baudet]. *PLoS One* 15, e0234769. doi: 10.1371/journal.pone.0234769
- Aldemir, S., Ateş, D., TEMEL, H. Y., Yağmur, B., Alsaleh, A., Kahrman, A., et al. (2017). QTLs for iron concentration in seeds of the cultivated lentil (*Lens culinaris* medic.) via genotyping by sequencing. *Turkish J. Agric. Forestry* 41 (4), 243–255. doi: 10.3906/tar-1610-33
- Ali, M. B., Welna, G. C., Sallam, A., Martsch, R., Balko, C., Gebser, B., et al. (2016). Association analyses to genetically improve drought and freezing tolerance of faba bean (*Vicia faba* L.). *Crop Sci.* 56, 1036–1048. doi: 10.2135/cropsci2015.08.0503
- An, B., Gao, X., Chang, T., Xia, J., Wang, X., Miao, J., et al. (2020). Genome-wide association studies using binned genotypes. *Heredity* 124 (2), 288–298. doi: 10.1038/s41437-019-0279-y
- Annicchiarico, P., Nazzicari, N., Pecetti, L., Romani, M., and Russi, L. (2019). Pea genomic selection for Italian environments. *BMC Genomics* 20, 603. doi: 10.1186/s12864-019-5920-x
- Arora, S., Mahato, A. K., Singh, S., Mandal, P., Bhutani, S., Dutta, S., et al. (2017). A high-density intraspecific SNP linkage map of pigeon pea (*Cajanus cajan* L. millsp.). *PLoS One* 12 (6), e0179747.
- Assefa, T., Zhang, J., Chowda-Reddy, R. V., Moran Lauter, A. N., Singh, A., O'Rourke, J. A., et al. (2020). Deconstructing the genetic architecture of iron deficiency chlorosis in soybean using genome-wide approaches. *BMC Plant Biol.* 20, 1–13. doi: 10.1186/s12870-020-2237-5
- Awlia, M., Nigro, A., Fajkus, J., Schmoekel, S. M., Negrao, S., Santelia, D., et al. (2016). High-throughput non-destructive phenotyping of traits that contribute to salinity tolerance in arabidopsis thaliana. *Front. Plant Sci.* 7, 1414. doi: 10.3389/fpls.2016.01414
- Bajaj, D., Das, S., Badoni, S., Kumar, V., Singh, M., Bansal, K. C., et al. (2015a). Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci. Rep.* 5 (1), 1–7. doi: 10.1038/srep12468
- Bajaj, D., Das, S., Upadhyaya, H. D., Ranjan, R., Badoni, S., Kumar, V., et al. (2015b). A genome-wide combinatorial strategy dissects complex genetic architecture of seed coat color in chickpea. *Front. Plant Sci.* 6, 979. doi: 10.3389/fpls.2015.00979
- Bajaj, D., Upadhyaya, H. D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Identification of candidate genes for dissecting complex branch number traits in chickpea. *Plant Sci.* 245, 61–70. doi: 10.1016/j.plantsci.2016.01.004
- Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., and Jin, Y. (2016). Nested association mapping of stem rust resistance in wheat using genotyping by sequencing. *PLoS One* 11, e0155760. doi: 10.1371/journal.pone.0155760
- Bandeira E Sousa, M., Cuevas, J., de Oliveira Couto, E. G., Pérez-Rodríguez, P., Jarquin, D., Fritsche- Neto, R., et al. (2017). Genomic-enabled prediction in maize using kernel models with genotype × environment interaction. *G3 (Bethesda)* 7, 1995–2014. doi: 10.1534/g3.117.042341
- Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6, 11. doi: 10.1186/1939-8433-6-11
- Banoo, A., Nabi, A., Rasool, R. S., Shah, M. D., Ahmad, M., Sofi, P. A., et al. (2020). North-western Himalayan common beans: population structure and mapping of quantitative anthracnose resistance through genome-wide association study. *Front. Plant Sci.* 11, 571618. doi: 10.3389/fpls.2020.571618
- Bastarache, L., and Denny, J. C. (2022). And roden, dPhenome-wide association studies. *JAMA* 327 (1), 75–76. doi: 10.1001/jama.2021.20356
- Basu, U., Srivastava, R., Bajaj, D., Thakro, V., Daware, A., Malik, N., et al. (2018). Genome-wide generation and genotyping of informative SNPs to scan molecular signatures for seed yield in chickpea. *Sci. Rep.* 8 (1), 13240. doi: 10.1038/s41598-018-29926-1
- Basu, U., Bajaj, D., Sharma, A., Malik, N., Daware, A., Narnoliya, L., et al. (2019). Genetic dissection of photosynthetic efficiency traits for enhancing seed yield in chickpea. *Plant Cell Environ.* 42 (1), 158–173. doi: 10.1111/pce.13319
- Behravan, H., Hartikainen, J. M., Tengström, M., et al. (2018). Machine learning identifies interacting genetic variants contributing to breast cancer risk: A case study in finnish cases and controls. *Sci. Rep.* 8 (1), 13149. doi: 10.1038/s41598-018-31573-5
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. doi: 10.1038/nature07517
- Bhagyawant, S. S., and Srivastava, N. (2008). Genetic fingerprinting of chickpea (*Cicer arietinum* L.) germplasm using ISSR markers and their relationships. *Afr. J. Biotech.* 7 (24), 4428–4431. doi: 10.5897/AJB08.973
- Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Theron, A., et al. (2016). Development and validation of the axiom® Apple480K SNP genotyping array. *Plant J.* 86 (1), 62–74. doi: 10.1111/tbj.13145
- Bohra, A., Jha, U. C., Satheesh Naik, S. J., Mehta, S., Tiwari, A., Maurya, A. K., et al. (2021). “Genomics: Shaping legume improvement,” in *Genetic enhancement in major food legumes* (Cham: Springer), 49–89. doi: 10.1007/978-3-030-64500-7_3
- Bohra, A., Saxena, K. B., Varshney, R. K., and Saxena, R. K. (2020). Genomics-assisted breeding for pigeon pea improvement. *Theor. Appl. Genet.* 133, 1721–1737. doi: 10.1007/s00122-020-03563-7
- Breria, C. M., Hsieh, C. H., Yen, T. B., Yen, J. Y., Noble, T. J., and Schafleitner, R. (2020). A SNPbased genome-wide association study to mine genetic loci associated to salinity tolerance in mungbean (*Vigna radiata* L.). *Genes* 11 (7), 759.
- Butenhoff, K. J. (2015). *QTL mapping and gwas identify sources of iron deficiency chlorosis and canopy wilt tolerance in the fiskeby iii x mandarin (ottawa) soybean population (Doctoral dissertation, university of Minnesota)*. Available at: <https://hdl.handle.net/11299/170730>.
- Carlson, C. H., Gouker, F. E., Crowell, C. R., Evans, L., DiFazio, S. P., Smart, C. D., et al. (2019). Joint linkage and association mapping of complex traits in shrub willow (*Salix purpurea* L.). *Ann. Bot.* 124 (4), 701–716. doi: 10.1093/aob/mcz047
- Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., et al. (2008). Community resources and strategies for association mapping in sorghum. *Crop Sci.* 48 (1), 30–40. doi: 10.2135/cropsci2007.02.0080

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Chang, H. X., Lipka, A. E., Domier, L. L., and Hartman, G. L. (2016). Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Phytopathology* 106 (10), 1139–1151. doi: 10.1094/PHYTO-01-16-0042-FI
- Che, Z., Yan, H., Liu, H., Yang, H., Du, H., Yang, Y., et al. (2020). Genome-wide association study for soybean mosaic virus SC3 resistance in soybean. *Mol. Breed.* 40, 1–14. doi: 10.1007/s11032-020-01149-1
- Chen, H., Liu, X., Zhang, H., Yuan, X., Gu, H., and Cui, X. (2018). Advances in salinity tolerance of soybean: genetic diversity, heredity, and gene identification contribute to improving salinity tolerance. *J. Integr. Agric.* 17, 2215–2221. doi: 10.1016/S2095-3119(17)61864-1
- Choudhary, P., Khanna, S. M., Jain, P. K., Bharadwaj, C., Kumar, J., Lakhera, P. C., et al. (2012). Genetic structure and diversity analysis of the primary gene pool of chickpea using SSR markers. *Genet. Mol. Res.* 11 (2), 891–905. doi: 10.4238/2012.April.10.5
- Contreras-Soto, R. I., Mora, F., de Oliveira, M. A. R., Higashi, W., Scapim, C. A., and Schuster, I. (2017). A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS One* 12, e0171105. doi: 10.1371/journal.pone.0171105
- Courtois, B., Audebert, A., Dardou, A., Roques, S., Ghneim-Herrera, T., Droc, G., et al. (2013). Genome-wide association mapping of root traits in a japonica rice panel. *PLoS One* 8 (11), e78037. doi: 10.1371/journal.pone.0078037
- Cui, C., Mei, H., Liu, Y., Zhang, H., and Zheng, Y. (2017). Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01189
- Das, S., Porter, L. D., Ma, Y., Coyne, C. J., Chaves-Cordoba, B., and Naidu, R. A. (2022). Resistance in lentil (*Lens culinaris*) genetic resources to the pea aphid (*Acyrtosiphon pisum*). *Entomologia Experimentalis Applicata* 170, 755–769. doi: 10.1111/eea.13202
- Davey, J., and Blaxter, M. L. (2011). RADSeq: next-generation population genetics. *Briefings Funct. Genomics* 10, 108.
- de Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313
- Deokar, A. A., Ramsay, L., Sharpe, A. G., Diapari, M., Sindhu, A., Bett, K., et al. (2014). Genome-wide SNP identification in chickpea for use in the development of a high-density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics* 15 (1), 1–19. doi: 10.1186/1471-2164-15-708
- Dhanapal, A. P., Ray, J. D., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2015). Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] genotypes. *Theor. Appl. Genet.* 128, 73–91. doi: 10.1007/s00122-014-2413-9
- Diaz, S., Ariza-Suarez, D., Ramdeen, R., Aparicio, J., Arunachalam, N., Hernandez, C., et al. (2020). Genetic architecture and genomic prediction of cooking time in common bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.622213
- Do, T. D., Vuong, T. D., Dunn, D., Clubb, M., Valliyodan, B., and Patil, G. (2019). Identification of new loci for salt tolerance in soybean by high-resolution genome-wide association mapping. *BMC Genomics* 20, 1–16. doi: 10.1186/s12864-019-5662-9
- Doldi, M. L., Vollmann, J., and Lelley, T. (1997). Genetic diversity in soybean as determined by RAPD and microsatellite analysis. *Plant Breed.* 116 (4), pp.331–pp.335. doi: 10.1111/j.1439-0523.1997.tb01007.x
- Dorani, F., Hu, T., Woods, M. O., and Zhai, G. (2018). Ensemble learning for detecting gene-gene interactions in colorectal cancer. *PeerJ* 6, e5854. doi: 10.7717/peerj.5854
- Eathington, S. R., Crosbie, T. M., Edwards, M. D., Reiter, R. S., and Bull, J. K. (2007). Molecular markers in a commercial breeding program. *Crop Sci.* 47, S-154-S-163. doi: 10.2135/cropsci2007.04.00151PBS
- 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 (5), e19379. doi: 10.1371/journal.pone.0019379
- Farahani, S., Maleki, M., Ford, R., Mehrabi, R., Kanouni, H., Kema, G. H. J., et al. (2022). Genome-wide association mapping for isolate-specific resistance to ascochyta blight in chickpea (*Cicer arietinum* L.). *Physiol. Mol. Plant Pathol.* 121, 101883. doi: 10.1016/j.pmpp.2022.101883
- Faridi, R., Koopman, B., Schierholt, A., Ali, M. B., Apel, S., and Link, W. (2021). Genetic study of the resistance of faba bean (*Vicia faba*) against the fungus ascochyta blight through a genome-wide association analysis. *Plant Breed.* 140 (3), 442–452. doi: 10.1111/pbr.12918
- Fayaz, H., Tyagi, S., Wani, A. A., Pandey, R., Akhtar, S., and Bhat, M. A. (2022). Genome-wide association analysis to delineate high-quality SNPs for seed micronutrient density in chickpea (*Cicer arietinum* L.). *Sci. Rep.* 12 (1), 11357. doi: 10.1038/s41598-022-14487-1
- Friedli, M., Kirchgessner, N., Grieder, C., Liebisch, F., Mannale, M., and Walter, A. (2016). Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. *Plant Methods* 12 (1), 1–15. doi: 10.1186/s13007-016-0109-7
- Furbank, R. T., and Tester, M. (2011). Phenomics—technologies to relieve the phenotyping. *Trends Plant Sci.* 16 (12), 635–644. doi: 10.1016/j.tplants.2011.09.005
- Gali, K. K., Sackville, A., Tafesse, E. G., Lachagari, V. B. R., McPhee, K., Hybl, M., et al. (2019). Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01538
- Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., et al. (2020). Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). *Plant Biotechnol. J.* 18 (6), 1457–1471. doi: 10.1111/pbi.13311
- Gaur, R., Azam, S., Jeena, G., Khan, A. W., Choudhary, S., Jain, M., et al. (2012). High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). *DNA Res.* 19 (5), 357–373. doi: 10.1093/dnares/dss018
- Gela, T., Ramsay, L., Haile, T. A., Vandenberg, A., and Bett, K. (2021). Identification of anthracnose race 1 resistance loci in lentil by integrating linkage mapping and genome-wide association study. *Plant Genome* 14, e20131. doi: 10.1002/tpg2.20131
- Gianola, D., and van Kaam, J. B. C. H. M. (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178, 2289–2303. doi: 10.1534/genetics.107.084285
- Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., et al. (2006). PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* 169, 623–635. doi: 10.1111/j.1469-8137.2005.01609.x
- Hajiloo, M., Damavandi, B., HooshSadat, M., Sangi, F., Mackey, J. R., Cass, C. E., et al. (2013). “Breast cancer prediction using genome wide single nucleotide polymorphism data” *BMC Bioinf.* 14 (13), S35. doi: 10.1186/1471-2105-14-S13-S3
- Hall, M. A., Wallace, J., Lucas, A., Kim, D., Basile, A. O., Verma, S. S., et al. (2017). PLATO software provides analytic framework for investigating complexity beyond genome-wide association studies. *Nat. Commun.* 8, 1167. doi: 10.1038/s41467-017-0802-2
- Hasanuzzaman, M., Araujo, S., and Gill, S. S. (2020). *The plant family fabaceae: biology and physiological responses to environmental stresses* (Singapore: Springer). doi: 10.1007/978-981-15-4752-2
- Heslot, N., Yang, H.-P., Sorrells, M. E., and Jannink, J.-L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52, 146. doi: 10.2135/cropsci2011.06.0297
- Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). Large-Scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10 (6), 716–732. doi: 10.1111/j.1467-7652.2012.00710.x
- Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67. doi: 10.1080/00401706.1970.10488634
- L. Horst and G. Wenzel (Eds.) (2007). *Molecular marker systems in plant breeding and crop improvement*.
- Hang, J. (2022). *Genome-wide association study of seed protein and amino acid contents in cultivated lentils as determined by near-infrared reflectance spectroscopy (Master's thesis)*. Available at: <http://hdl.handle.net/1993/36250>
- Hong, Y., Pandey, M. K., Lu, Q., Liu, H., Gangurde, S. S., Li, S., et al. (2021). Genetic diversity and distinctness based on morphological and SSR markers in peanut. *Agron. J.* 113 (6), 4648–4660. doi: 10.1002/aj2.20671
- Hu, D., Kan, G., Hu, W., Li, Y., Hao, D., Li, X., et al. (2019a). Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. *Front. Plant Sci.* 10, 811. doi: 10.3389/fpls.2019.00811
- Hu, J., Maalouf, F., Zhang, Z., and Yu, L.-X. (2019b). “Genome-wide association study of the resilience to high temperature of faba bean (*Vicia faba* L.) germplasm,” in *2019 ASHS Annual Conference (ASHS)*.
- Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8 (2), giy154. doi: 10.1093/gigascience/giy154
- Humphrik, J. F., Lazar, D., Fürst, T., Husíková, A., Hýbl, M., and Spíchal, L. (2015). Automated integrative high-throughput phenotyping of plant shoots: a case study of the cold-tolerance of pea (*Pisum sativum* L.). *Plant Methods* 11 (1), 1–11. doi: 10.1186/s13007-015-0063-9
- Huynh, B.-L., Ehlers, J. D., Huang, B. E., Munoz, A. M., Lonardi, S., Santos, J. R. P., et al. (2018). A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. walp.). *Plant J.* 93, 1129–1142. doi: 10.1111/tpj.13827
- Hwang, E.-Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15, 1–12. doi: 10.1186/1471-2164-15-1
- Iqura, E., Humira, S., and François, B. (2015). Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biol.* 15 (1), 1–12. doi: 10.1186/s12870-014-0408-y
- Iruela, M., Rubio, J., Cubero, J. I., Gil, J., and Millan, T. (2002). Phylogenetic analysis in the genus cicer and cultivated chickpea using RAPD and ISSR markers. *Theor. Appl. Genet.* 104 (4), 643–651. doi: 10.1007/s001220100751

- Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29 (4), 1–7. doi: 10.1093/nar/29.4.e25
- Jadhav, M. P., Gangurde, S. S., Hake, A. A., Yadawad, A., Mahadevaiah, S. S., Pattanashetti, S. K., et al. (2021). Genotyping-by-sequencing based genetic mapping identified major and consistent genomic regions for productivity and quality traits in peanut. *Front. Plant Sci.* 12, 668020. doi: 10.3389/fpls.2021.668020
- 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. Genom.* 290, 559–571. doi: 10.1007/s00438-014-0932-3
- Jha, U. C., Jha, R., Thakro, V., Anurag, K., Sanjeev, G., Harsh, N., et al. (2021). Discerning molecular diversity and association mapping for phenological, physiological and yield traits under high temperature stress in chickpea (*Cicer arietinum* L.). *J. Gen.* 100, 4. doi: 10.1007/s12041-020-01254-2
- Jia, C., Zhao, F., Wang, X., Han, J., Zhao, H., Liu, G., et al. (2018). Genomic prediction for 25 agronomic and quality traits in alfalfa (*Medicago sativa*). *Front. Plant Sci.* 9, 1220. doi: 10.3389/fpls.2018.01220
- Jiang, L., Zheng, Z., Qi, T., Kemper, K. E., Wray, N. R., Visscher, P. M., et al. (2019). A resource-efficient tool for mixed model association analysis of Large-scale data. *Nat. Genet.* 51 (12), 1749–1555. doi: 10.1038/s41588-019-0530-8
- Johnson, N., Boatwright, J. L., Bridges, W., Thavarajah, P., Kumar, S., Shipe, E., et al. (2021). Genome-wide association mapping of lentil (*Lens culinaris* Medikus) prebiotic carbohydrates toward improved human health and crop stress tolerance. *Sci. Rep.* 11, 1–12. doi: 10.1038/s41598-021-93475-3
- Kalve, S., Gali, K. K., and Tar’an, B. (2022). Genome-wide association analysis of stress tolerance indices in an interspecific population of chickpea. *Front. Plant Sci.* 13, 933277. doi: 10.3389/fpls.2022.933277
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178 (3), 1709–1235. doi: 10.1534/genetics.107.080101
- Karthika, R., Jean, C. C., Ping, Z., Gopesh, S., Dorrie, M., Nurul, A., et al. (2021). Genetic diversity and GWAS of agronomic traits using an ICARDA lentil (*Lens culinaris* Medik.). *Reference Plus collection. Plant Gen. Res.* 19 (4), 279–288. doi: 10.1017/S147926212100006X
- Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H. F., et al. (2020). Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01001
- Khan, M. A., Tong, F., Wang, W., He, J., Zhao, T., and Gai, J. (2018). Analysis of QTL-allele system conferring drought tolerance at seedling stage in a nested association mapping population of soybean [*Glycine max* (L.) Merr.] using a novel GWAS procedure. *Planta* 248, 947–962. doi: 10.1007/s00425-018-2952-4
- Khazaei, H., Fedoruk, M., Caron, C. T., Vandenberg, A., and Bett, K. E. (2018). Single nucleotide polymorphism markers associated with seed quality characteristics of cultivated lentil. *Plant Genome* 11, 170051. doi: 10.3835/plantgenome2017.06.0051
- Khazaei, H., Podder, R., Caron, C. T., Kundu, S. S., Diapari, M., Vandenberg, A., et al. (2017). Marker-trait association analysis of iron and zinc concentration in lentil (*Lens culinaris* Medik.) seeds. *Plant Genome* 10, plantgenome2017-02. doi: 10.3835/plantgenome2017.02.0007
- Khera, P., Upadhyaya, H. D., Pandey, M. K., Roorkiwal, M., Sriswathi, M., Janila, P., et al. (2013). Single nucleotide polymorphism-based genetic diversity in the reference set of peanut (*Arachis* spp.) by developing and applying cost-effective kompetitive allele specific polymerase chain reaction genotyping assays. *Plant Genome* 6 (3), plantgenome2013-06. doi: 10.3835/plantgenome2013.06.0019
- Kim, K. H., Kim, J.-Y., Lim, W.-J., Jeong, S., Lee, H.-Y., Cho, Y., et al. (2020). Genome-wide association and epistatic interactions of flowering time in soybean cultivar. *PLoS One* 15, e0228114. doi: 10.1371/journal.pone.0228114
- Kohli, P. S., Kumar Verma, P., Verma, R., Parida, S. K., Thakur, J. K., and Giri, J. (2020). Genome-wide association study for phosphate deficiency responsive root hair elongation in chickpea. *Funct. Integr. Genomics* 20, 775–786. doi: 10.1007/s10142-020-00749-6
- Kujur, A., Bajaj, D., Upadhyaya, H., Das, S., Ranjan, R., Shree, T., et al. (2015a). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5 (1), 1–20. doi: 10.1038/srep11166
- Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M. S., et al. (2015b). Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci. Rep.* 5 (1), 1–13. doi: 10.1038/srep09468
- Kulwal, P. L., and Singh, R. (2021). “Association mapping in plants,” in *Crop breeding* (New York, NY: Humana), 105–117.
- Kumar, K., Anjoy, P., Sahu, S., Durgesh, K., Das, A., Tribhuvan, K. U., et al. (2022). Single trait versus principal component based association analysis for flowering related traits in pigeonpea. *Sci. Rep.* 12 (1), 10453. doi: 10.1038/s41598-022-14568-1
- Kumar, J., Choudhary, A. K., Solanki, R. K., and Pratap, A. (2011). Towards marker-assisted selection in pulses: a review. *Plant Breed.* 130 (3), 297–313. doi: 10.1111/j.1439-0523.2011.01851.x
- Kumar, J., Gupta, S., Biradar, R. S., Gupta, P., Dubey, S., and Singh, N. P. (2018a). Association of functional markers with flowering time in lentil. *J. Appl. Genet.* 59, 9–21. doi: 10.1007/s13353-017-0419-0
- Kumar, J., Gupta, S., Gupta, D. S., and Singh, N. P. (2018b). Identification of QTLs for agronomic traits using association mapping in lentil. *Euphytica* 214, 1–15. doi: 10.1007/s10681-018-2155-x
- Kumar, H., Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., Meena, M. C., et al. (2019). Genetic dissection of grain iron and zinc concentrations in lentil (*Lens culinaris* Medik.). *J. Genet.* 98, 1–14. doi: 10.1007/s12041-019-1112-3
- Kump, K., Bradbury, P., Wissler, R., Edward, S., Araby, R., Marco, A., et al. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* 43, 163–168. doi: 10.1038/ng.747
- LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res.* 37 (13), 4181–4193. doi: 10.1093/nar/37.13.4181
- Lee, S., Kerns, S., Ostrer, H., Rosenstein, B., Deasy, J. O., and Oh, J. H. (2018). Machine learning on a genome-wide association study to predict late genitourinary toxicity after prostate radiation therapy. *Int. J. Radiat. OncologyBiologyPhysics* 101 (1), 128–355. doi: 10.1016/j.ijrobp.2018.01.054
- Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., and Fritz, S. (2011). Improvedlasso for genomic selection. *Genet. Res. (Camb)* 93, 77–87. doi: 10.1017/S0016672310000534
- Li, Y., 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, 359. doi: 10.3389/fpls.2017.00359
- Li, N., Shi, J., Wang, X., Liu, G., and Wang, H. (2014). A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (*Brassica napus* L.). *BMC Plant Biol.* 14, 114. doi: 10.1186/1471-2229-14-114
- Li, D., Zhao, X., Han, Y., Li, W., and Xie, F. (2019). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* 111 (1), 90–95. doi: 10.1016/j.ygeno.2018.01.004
- Li, Y., Ruperao, P., Batley, J., Edwards, D., Khan, T., Colmer, T. D., et al. (2018). Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00190
- Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y.-M., Todhunter, R. J., et al. (2014). Enrichment of statistical power for genome-wide association studies. *BMC Biol.* 12, 73. doi: 10.1186/s12915-014-0073-5
- Li, X., Wei, Y., Acharya, A., Hansen, J. L., Crawford, J. L., Viands, D. R., et al. (2015). Genomic prediction of biomass yield in two selection cycles of a tetraploid alfalfa breeding population. *Plant Genome* 8, 0. doi: 10.3835/plantgenome2014.12.0090
- Li, Y., Ruperao, P., Batley, J., Edwards, D., Khan, T., Colmer, T. D., et al. (2018). Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00190
- Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., et al. (2004). Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin. Chem.* 50 (7), 1156–1164. doi: 10.1373/clinchem.2004.032136
- Lin, J., Lan, Z., Hou, W., Yang, C., Wang, D., Zhang, M., et al. (2020). Identification and fine mapping of a genetic locus underlying soybean tolerance to SMV infections. *Plant Sci.* 292, 110367. doi: 10.1016/j.plantsci.2019.110367
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8 (10), 833–835. doi: 10.1038/nmeth.1681
- Liu, X., Qin, D., Piersanti, A., Zhang, Q., Miceli, C., and Wang, P. (2020). Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. *BMC Plant Biol.* 20 (1), 399. doi: 10.1186/s12870-020-02607-w
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12 (2), e1005767. doi: 10.1371/journal.pgen.1005767
- Liu, Y., Wang, D., He, F., Wang, J., Joshi, T., and Xu, D. (2019). Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. *Front. Genet.* 10, 109110. doi: 10.3389/fgene.2019.01091
- Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., et al. (2010). Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19585–19590. doi: 10.1073/pnas.1006105107
- Ma, Y., Marzougui, A., Coyne, C. J., Sankaran, S., Main, D., Porter, L. D., et al. (2020). Dissecting the genetic architecture of aphanomyces root rot resistance in lentil by QTL mapping and genome-wide association study. *Int. J. Mol. Sci.* 21, 2129. doi: 10.3390/ijms21062129
- Maalouf, F., Abou-Khater, L., Babiker, Z., Jighly, A., Alsamman, A. M., Hu, J., et al. (2022). Genetic dissection of heat stress tolerance in faba bean (*Vicia faba* L.) using GWAS. *Plants* 11, 1108. doi: 10.3390/plants11091108

- Maciukiewicz, M., Marshe, V. S., Hauschild, A.-C., Foster, J. A., Rotzinger, S., Kennedy, J. L., et al. (2018). GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. *J. Psychiatr. Res.* 99 (April), 62–68. doi: 10.1016/j.jpsychires.2017.12.009
- Maenhout, S., De Baets, B., Haesaert, G., and Van Bockstaele, E. (2007). Support vector machine regression for the prediction of maize hybrid performance. *Theor. Appl. Genet.* 115, 1003–1013. doi: 10.1007/s00122-007-0627-9
- Marinangeli, C., Curran, J., Barr, S., Slavin, J., Puri, S., Swaminathan, S., et al. (2017). Enhancing nutrition with pulses: defining a recommended serving size for adults. *Nutr. Rev.* 75, 990–1006. doi: 10.1093/nutrit/nux058
- Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., et al. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics* 16 (1), 290. doi: 10.1186/s12864-015-1459-7
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- McCouch, S. R., Wright, M. H., Tung, C. W., Maron, L. G., McNally, K. L., Fitzgerald, M., et al. (2016). Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7 (1), 10532. doi: 10.1038/ncomms10532
- Mieth, B., Kloft, M., Rodríguez, J. A., Sonnenburg, Sören, Vobruba, R., Morcillo-Suárez, C., et al. (2016). Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies. *Sci. Rep.* 6 (1), 36671. doi: 10.1038/srep36671
- Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007
- Mittag, F., Büchel, F., Saad, M., Jahn, A., Schulte, C., Bochdanovits, Z., et al. (2012). Use of support vector machines for disease risk prediction in genome-wide association studies: concerns and opportunities. *Hum. Mutat.* 33 (12), 1708–1718. doi: 10.1002/humu.22161
- Mousavi-Derazmahalleh, M., Bayer, P. E., Hane, J. K., Valliyodan, B., Nguyen, H. T., Nelson, M. N., et al. (2019). Adapting legume crops to climate change using genomic approaches. *Plant Cell & Environ.* 42, 6–19. doi: 10.1111/pce.13203
- Muraya, M. M., Chu, J., Zhao, Y., Junker, A., Klukas, C., Reif, J. C., et al. (2017). Genetic variation of growth dynamics in maize (*Zea mays* L.) revealed through automated non-invasive phenotyping. *Plant J.* 89 (2), 366–380.
- Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., et al. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21, 2194–2202. doi: 10.1105/tpc.109.068437
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yildiz, M., et al. (2018). DNA Molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* 32 (2), 261–285. doi: 10.1080/13102818.2017.1400401
- Nair, R. J., and Pandey, M. K. (2021). Role of molecular markers in crop breeding: a review. *Agric. Rev.* 1, 1–8. doi: 10.18805/ag.R-2322
- Ndjonjop, M. N., Semagn, K., Zhang, J., Arnaud, C. G., Kpeki, S. B., Alphonse, G., et al. (2018). Development of species diagnostic SNP markers for quality control genotyping in four rice (*Oryza* L.) species. *Mol. Breed.* 38, 131. doi: 10.1007/s11032-018-0885-z
- Neupane, S., Wright, D. M., Martinez, R. O., Butler, J., Weller, J., and Bett, K. (2022). Focusing the GWAS lens on days to flower using latent variable phenotypes derived from global multi-environment trials. *The Plant Genome* 16, e20269. doi: 10.1101/2022.03.10.483676
- Nguyen, T.-T., Huang, J. Z., Wu, Q., Nguyen, T. T., and Li, M. J. (2015). Genome-wide association data classification and SNPs selection using two-stage quality-based random forests. *BMC Genomics* 16 (2), S55. doi: 10.1186/1471-2164-16-S2-S5
- Nguyen, T. T., Taylor, P. W. J., Redden, R. J., and Ford, R. (2004). Genetic diversity estimates in cicer using AFLP analysis. *Plant Breed.* 123 (2), 173–179. doi: 10.1046/j.1439-0523.2003.00942.x
- Nice, L. M., Steffenson, B. J., Brown-Guedira, G. L., Akhunov, E. D., Liu, C., Kono, T. J. Y., et al. (2016). Development and genetic characterization of an advanced backcross-nested association mapping (AB-NAM) population of wild cultivated barley. *Genetics* 203, 1453. doi: 10.1534/genetics.116.190736
- Nkhata, W., Shimelis, H., Melis, R., Chirwa, R., Mzengeza, T., Mathew, I., et al. (2021). Genomewide association analysis of bean fly resistance and agro-morphological traits in common bean. *PLoS One* 16 (4), e0250729. doi: 10.1371/journal.pone.0250729
- Oetjens, M. T., Brown-Gentry, K., Goodloe, R., Dilks, H. H., and Crawford, D. C. (2016). Population stratification in the context of diverse epidemiologic surveys sans genome-wide data. *Front. Genet.* 7. doi: 10.3389/fgene.2016.00076
- Ogut, J. O., Piepho, H.-P., and Schulz-Streck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* 5 Suppl 3, S11. doi: 10.1186/1753-6561-5-S3-S11
- Oh, J., Kerns, S., Ostrer, H., Simon, N. P., Barry, R., and Joseph, O. D. (2017). Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes. *Sci. Rep.* 7, 43381. doi: 10.1038/srep43381
- Olukolu, B. A., Mayes, S., Stadler, F., Quat Ng, N., Fawole, I., Dominique, D., et al. (2012). Genetic diversity in bambara groundnut (*Vigna subterranea* (L.) verdc.) as revealed by phenotypic descriptors and DaRT marker analysis. *Genet. Resour. Crop Evol.* 59 (3), 347–358. doi: 10.1007/s10722-011-9686-5
- Orsak, A., Deokar, A., and Tar'an, B. (2019). 27. genome-wide association study for seed quality traits in chickpea. *FOR THE PEOPLE AND THE PLANET* 40.
- Packer, R. J., Williams, A. T., Hennah, W., Eisenberg, M. T., Shrine, N., Fawcett, K. A., et al. (2023). DeepPheWAS: an R package for phenotype generation and association analysis for phenome-wide association studies. *Bioinformatics* 39 (4), btad073. doi: 10.1093/bioinformatics/btad073
- Pan, L., He, J., Zhao, T., Xing, G., Wang, Y., Yu, D., et al. (2018). Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure. *Theor. Appl. Genet.* 131, 2581–2599. doi: 10.1007/s00122-018-3174-7
- Pandey, M. K., Gangurde, S. S., Sharma, V., Pattanashetti, S. K., Naidu, G. K., Faye, I., et al. (2020). Improved genetic map identified major QTLs for drought tolerance and iron deficiency tolerance-related traits in groundnut. *Genes* 12 (1), 37.
- Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., et al. (2017). Development and evaluation of a high-density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7, 40577. doi: 10.1038/srep40577
- Pandey, V., Nutter, R. C., and Prediger, E. (2008). "Applied biosystems SOLID system: ligation-based sequencing," in *Next generation genome sequencing: towards personalized medicine*. Ed. M. Janitz (Germany: Wiley-VCH, Weinheim), 431–444.
- Parida, S. K., Kujur, A., Bajaj, D., Das, S., Srivastava, R., Badoni, S., et al. (2017). Integrative genome-wide association studies (GWAS) to understand complex genetic architecture of quantitative traits in chickpea.
- Patil, G., Do, T., Vuong, T., Babu, V., Dong Lee, J., Juhi, C., et al. (2016). Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Sci. Rep.* 6, 19199. doi: 10.1038/srep19199
- Pawar, S., Pandit, E., Mohanty, I. C., Saha, D., and Pradhan, S. K. (2021). Population genetic structure and association mapping for iron toxicity tolerance in rice. *PLoS One* 16 (3), e0246232. doi: 10.1371/journal.pone.0246232
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *De novo* SNP discovery and genotyping in model and non-model species. *PLoS One* 7 (5), e37135. doi: 10.1371/journal.pone.0037135
- Pividori, M., Rajagopal, P. S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., et al. (2020). PhenomeXcan: mapping the genome to the phenome through the transcriptome. *Sci. Adv.* 6 (37), eaba2083. doi: 10.1126/sciadv.aba2083
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5 (3). doi: 10.3835/plantgenome2012.06.0006
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38 (8), 904–995. doi: 10.1038/ng1847
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959. doi: 10.1093/genetics/155.2.945
- Puspitasari, W., Allemann, B., Angra, D., Appleyard, H., Ecke, W., Möllers, C., et al. (2022). NIRS for vicine and convicine content of faba bean seed allowed GWAS to prepare for marker-assisted adjustment of seed quality of German winter faba beans. *J. Cultivated Plants* 74 (01-02). doi: 10.5073/JFK.2022.01-02.01
- Raju, N. L., Gnanesh, B. N., Lekha, P., Jayashree, B., Pandey, S., Hiremath, P. J., et al. (2010). The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). *BMC Plant Biol.* 10 (1), 1–22. doi: 10.1186/1471-2229-10-45
- Raman, R., Warren, A., Krysinska-Kaczmarek, M., Rohan, M., Sharma, N., Dron, N., et al. (2022). Genome-wide association analyses track genomic regions for resistance to ascochyta blight in Australian chickpea breeding germplasm. *Front. Plant Sci.* 1314. doi: 10.3389/fpls.2022.877266
- Rasool, S., Mahajan, R., Nazir, M., Bhat, K. A., Shikari, A. B., Ali, G., et al. (2022). SSR and GBS based GWAS study for identification of QTLs associated with nutritional elemental in common bean (*Phaseolus vulgaris* L.). *Scientia Hort.* 306, 111470. doi: 10.1016/j.scienta.2022.111470
- Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2019). Genomewide association study and genomic selection for soybean chlorophyll content associated with soybean cyst nematode tolerance. *BMC Genomics* 20, 1–18. doi: 10.1186/s12864-019-6275-z
- Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2020). Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. romagnoni, Alberto, Simon jégou, kristel van Steen, gilles wainrib, and Jean-Pierre hugot. 2019. "Comparative performances of machine learning methods for classifying crohn disease patients using genome-wide genotyping data" *Sci. Rep.* 9 (1), 10351. doi: 10.1038/s41598-019-46649-z
- Reif, J. C., Liu, W., Gowda, M., et al. (2010). Genetic basis of agronomically important traits in sugar beet (*Beta vulgaris* L.) investigated with joint linkage

- association mapping. *Theor. Appl. Genet.* 121, 1489–1499. doi: 10.1007/s00122-010-1405-7
- Revilla, P., Rodríguez, V. M., Ordás, A., Rincet, R., Charcosset, A., Giauffret, C., et al. (2016). Association mapping for cold tolerance in two large maize inbred panels. *BMC Plant Biol.* 16 (1), 127. doi: 10.1186/s12870-016-0816-2
- Rincker, K., Lipka, A. E., and Diers, B. W. (2016). Genome-wide association study of brown stem rot soybean across multiple populations. *Plant Genome* 9 (2). doi: 10.3835/plantgenome2015.08.0064
- Roorikwal, 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
- Romagnoni, A., Jégou, S., Van Steen, K., Gilles, W., Hugot, J., et al. (2019). Comparative performances of machine learning methods for classifying crohn disease patients using genome-wide genotyping data. *Sci. Rep.* 9, 10351. doi: 10.1038/s41598-019-46649-z
- Roorikwal, M., Jain, A., Kale, S. M., Doddamani, D., Chitkani, A., Thudi, M., et al. (2018). Development and evaluation of high density SNP array (Axiom® CicerSNP array) for high resolution genetic mapping and breeding applications in chickpea. *Plant Biotechnol. J.* 16, 890–901. doi: 10.1111/pbi.12836
- Roorikwal, 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. doi: 10.3389/fpls.2016.01666
- Roorikwal, M., Von Wettberg, E. J., Upadhyaya, H. D., Warschewsky, E., Rathore, A., and Varshney, R. K. (2014). Exploring germplasm diversity to understand the domestication process in cicer spp. using SNP and DArT markers. *PLoS One* 9 (7), e102016. doi: 10.1371/journal.pone.0102016
- Rothberg, J., and Leamon, J. (2008). The development and impact of 454 sequencing. *Nat. Biotechnol.* 26, 1117–1124. doi: 10.1038/nbt1485
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Higuchi, R. H., Horn, G. T., Mullis, K. B., et al. (1985). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–492. doi: 10.1126/science.2448875
- Salgotra, R. K., and Stewart, C. N. (2022). Genetic augmentation of legume crops using genomic resources and genotyping platforms for nutritional food security. *Plants* 11 (14), 1866. doi: 10.3390/plants11141866
- Sallam, A., Moursi, Y. S., Martsch, R., and Eltahir, S. (2022). Genome-wide association mapping for root traits associated with frost tolerance in faba beans using KASP-SNP markers. *Front. Genet.* 13, 907267. doi: 10.3389/fgene.2022.907267
- Samineni, S., Mahendrakar, M. D., Hottii, A., Chand, U., Rathore, A., and Gaur, P. M. (2022). Impact of heat and drought stresses on grain nutrient content in chickpea: genome-wide marker-trait associations for protein, Fe and Zn. *Environ. Exp. Bot.* 194, 104688. doi: 10.1016/j.envexpbot.2021.104688
- Sandhu, K. S., Shiv, A., Kaur, G., Meena, M. R., Raja, A. K., Vengavasi, K., et al. (2022). Integrated approach in genomic selection to accelerate genetic gain in sugarcane. *Plants* 11, 2139. doi: 10.3390/plants11162139
- Saxena, R. K., Kale, S. M., Kumar, V., Parupalli, S., Joshi, S., Singh, V. K., et al. (2017a). Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea. *Sci. Rep.* 7, 1813. doi: 10.1038/s41598-017-01535-4
- Saxena, R. K., Patel, K., Kumar, C. V. S., Tyagi, K., Saxena, K. B., and Varshney, R. K. (2018). Molecular mapping and inheritance of restoration of fertility (Rf) in A4 hybrid system in pigeonpea (*Cajanus cajan* (L.) millsp.). *Theor. Appl. Genet.* 131, 1605–1614. doi: 10.1007/s00122-018-3101-y
- Saxena, R. K., Prathima, C., Saxena, K. B., Hoisington, D. A., Singh, N. K., and Varshney, R. K. (2010). Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). *Plant Breed.* 129 (2), 142–148. doi: 10.1111/j.1439-0523.2009.01680.x
- Saxena, R. K., Rathore, A., Bohra, A., Yadav, P., Das, R. R., Khan, A. W., et al. (2018b). Development and application of high-density axiom® cajanus SNP array with 56 K SNPs to understand the genome architecture of released cultivars and founder genotypes for redefining future pigeonpea breeding programs. *Plant Genome*. doi: 10.3835/plantgenome2018.01.0005
- Saxena, R. K., Singh, V. K., Kale, S. M., Tathineni, R., Parupalli, S., Kumar, V., et al. (2017b). Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea. *Sci. Rep.* 7, 1911. doi: 10.1038/s41598-017-01537-2
- Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü, Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44 (7), 825–830. doi: 10.1038/ng.2314
- Shafiekhani, A., Kadam, S., Fritsch, F. B., and DeSouza, G. N. (2017). Vinobot and vinoculer: two robotic platforms for high-throughput field phenotyping. *Sensors* 17 (1), 214. doi: 10.3390/s17010214
- Shaibu, A. S., Motagi, B., and Ayittah, P. (2019a). Genome wide association studies for fatty acids, mineral and proximate compositions in groundnut (*Arachis hypogaea* L.) seeds. *Pertanika J. Trop. Agric. Sci.* 42 (3), 939–955.
- Shaibu, A., Sneller, C., Motagi, B., Chepkoech, J., Chepnetich, M., Miko, Z., et al. (2019b). Genome wide association studies for four physiological traits in groundnut (*Arachis hypogaea* L.) minicore collection. *Agronomy* 10, 192. doi: 10.3390/agronomy10020192
- Shaibu, A. S., Sneller, C., Motagi, B. N., Chepkoech, J., Chepnetich, M., Miko, Z. L., et al. (2020). Genome-wide detection of SNP markers associated with four physiological traits in groundnut (*Arachis hypogaea* L.) mini core collection. *Agronomy* 10, 192. doi: 10.3390/agronomy10020192
- Shao, Z., Shao, J., Huo, X., Li, W., Kong, Y., Du, H., et al. (2022). Identification of closely associated SNPs and candidate genes with seed size and shape via deep re-sequencing GWAS in soybean. *Theor. Appl. Genet.* 135 (7), 2341–2351. doi: 10.1007/s00122-022-04116-w
- Shi, L., Shi, T., Broadley, M. R., White, P. J., Long, Y., Meng, J., et al. (2013). High-throughput root phenotyping screens identify genetic loci associated with root architectural traits in brassica napus under contrasting phosphate availabilities. *Ann. Bot.* 112 (2), 381–389. doi: 10.1093/aob/mcs245
- Shilpa, H. B., and Lohithaswa, H. C. (2021). Discovery of SNPs in important legumes through comparative genome analysis and conversion of SNPs into PCR-based markers. *J. Genet.* 100 (2), 1–12. doi: 10.1007/s12041-021-01320-3
- Shingote, P. R., Gotarkar, D. N., Kale, R. R., Limbalkar, O. M., and Wasule, D. L. (2022). Recent advances and applicability of GBS, GWAS, and GS in soybean. *Genotyping by Sequencing Crop Improvement*, 218–249. doi: 10.1002/9781119745686.ch10
- Shrestha, A., Cosenza, F., van Inghelandt, D., Wu, P. Y., Li, J., Casale, F. A., et al. (2022). The double round-robin population unravels the genetic architecture of grain size in barley. *J. Exp. Bot.* 73, 7344–61. doi: 10.1093/jxb/erac369
- Singh, C. M., Pratap, A., Gupta, S., Biradar, R. S., and Singh, N. P. (2020). Association mapping for mungbean yellow mosaic India virus resistance in mungbean (*Vigna radiata* L. Wilczek). *3 Biotech* 10 (2), 33. doi: 10.1007/s13205-019-2035-7
- Singh, A., Dikshit, H. K., Mishra, G. P., Aski, M., and Kumar, S. (2019). Association mapping for grain diameter and weight in lentil using SSR markers. *Plant Gene* 20, 100204. doi: 10.1016/j.plgene.2019.100204
- Singh, V. K., Khan, A. W., Saxena, R. K., Kumar, V., Kale, S. M., Sinha, P., et al. (2016). Next-generation sequencing for identification of candidate genes for fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). *Plant Biotechnol. J.* 14 (5), 1183–1194. doi: 10.1111/pbi.12470
- Singh, A., Sharma, V., Dikshit, H. K., Aski, M., Kumar, H., Thirunavukkarasu, N., et al. (2017). Association mapping unveils favorable alleles for grain iron and zinc concentrations in lentil (*Lens culinaris* subsp. *culinaris*). *PLoS One* 12, e0188296. doi: 10.1371/journal.pone.0188296
- Singh, L., Dhillon, G. S., Sarabjit, K., Sandeep, D. K., Amandeep, K., Malik, P., et al. (2022). Genome-wide association study for yield and yield-related traits in diverse blackgram panel (*Vigna mungo* L. hepper) reveals novel putative alleles for future breeding programs. *Front. Genet.* 13. doi: 10.3389/fgene.2022.849016
- Singh, R., Singhal, V., and Randhawa, G. J. (2008). Molecular analysis of chickpea (*Cicer arietinum* L.) cultivars using AFLP and STMS markers. *J. Plant Biochem. Biotechnol.* 17 (2), 167–171. doi: 10.1007/BF03263279
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8 (1), e54985. doi: 10.1371/journal.pone.0054985
- Souframanien, J., and Gopalakrishna, T. (2004). A comparative analysis of genetic diversity in blackgram genotypes using RAPD and ISSR markers. *Theor. Appl. Genet.* 109 (8), 1687–1693. doi: 10.1007/s00122-004-1797-3
- Srungrarapu, R., Mahendrakar, M. D., Mohammad, L. A., Chand, U., Jagarlamudi, V. R., Kondamudi, K. P., et al. (2022). Genome-wide association analysis reveals trait-linked markers for grain nutrient and agronomic traits in diverse set of chickpea germplasm. *Cells* 11, 2457. doi: 10.3390/cells11152457
- Stagnari, F., Albino, M., Angelica, G., and Michele, P. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. *Chem. Biol. Technol. Agric.* 4, 2. doi: 10.1186/s40538-016-0085-1
- Stojakowski, S. A., Milczarski, P., Hanek, M., Bolibok-Bragoszewska, H., Myśków, B., Kilian, A., et al. (2011). DArT markers tightly linked with the *Rf1* gene controlling restoration of Male fertility in the CMS-c system in cultivated rye (*Secale cereale* L.). *J. Appl. Genet.* 52 (3), 313–318. doi: 10.1007/s13353-011-0049-x
- Sudheesh, S., Kahrood, H. V., Braich, S., Dron, N., Hobson, K., Cogan, N. O., et al. (2021). Application of genomics approaches for the improvement in ascochyta blight resistance in chickpea. *Agronomy* 11 (10), 1937. doi: 10.3390/agronomy11101937
- Sui, M., Wang, Y., Bao, Y., Wang, X., Li, R., Lv, Y., et al. (2020). Genome-wide association analysis of sucrose concentration in soybean (*Glycine max* L.) seed based on high-throughput sequencing. *Plant Genome* 13, e20059. doi: 10.1002/tpg2.20059
- Sul, J. H., Bilow, M., Yang, W.-Y., Kostem, E., Furlotte, N., He, D., et al. (2016). Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. *PLoS Genet.* 12, e1005849. doi: 10.1371/journal.pgen.1005849
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8 (3), e58700. doi: 10.1371/journal.pone.0058700
- Tafesse, E. G., Gali, K. K., Lachagari, V. B. R., Bueckert, R., and Warkentin, T. D. (2021). Genome-wide association mapping for heat and drought adaptive traits in pea. *Genes* 12 (12), 1897. doi: 10.3390/genes12121897

- Tafesse, E. G., Gali, K. K., Lachagari, V. B. R., Bueckert, R., and Warkentin, T. D. (2021). Genome-wide association mapping for heat and drought adaptive traits in pea. *Genes* 12 (12), 1897. doi: 10.3390/genes12121897
- Talebi, R., Fayaz, F., Mardi, M., Pirsyedi, S. M., and Najj, A. M. (2008). Genetic relationships among chickpea (*Cicer arietinum*) elite lines based on RAPD and agronomic markers. *Int. J. Agric. Biol.* 8, 301–305.
- Tanabata, T., Shibaya, T., Hori, K., Ebana, K., and Yano, M. (2012). SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol.* 160 (4), 1871–1880. doi: 10.1104/pp.112.205120
- Thompson, J. A., Nelson, R. L., and Vodkin, L. O. (1998). Identification of diverse soybean germplasm using RAPD markers. *Crop Sci.* 38 (5), pp.1348–1355. doi: 10.2135/cropsci1998.0011183X003800050033x
- Thudi, M., Khan, A. W., Kumar, V., Gaur, P. M., Katta, K., Garg, V., et al. (2016). Whole genome re-sequencing reveals genome-wide variations among parental lines of 16 mapping populations in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 16 Suppl 1, 10. doi: 10.1186/s12870-015-0690-3
- Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, U. G., Pillay, M., et al. (2002). Genetic diversity in musa acuminata colla and musa balbisiana colla and some of their natural hybrids using AFLP markers. *Theor. Appl. Genet.* 104 (8), 1246–1252. doi: 10.1007/s00122-002-0914-4
- 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 (5), e96758. doi: 10.1371/journal.pone.0096758
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Ser. B (Methodological)* 58 (1), 267–288. <http://www.jstor.org/stable/2346178>.
- Topp, C. N., Iyer-Pascuzzi, A. S., Anderson, J. T., Lee, C. R., Zurek, P. R., Symonova, O., et al. (2013). 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. *Proc. Nat. Acad. Sci.* 110 (18), E1695–E1704. doi: 10.1073/pnas.1304354110
- Torkamaneh, D., Laroche, J., Boyle, B., Hyten, D. L., and Belzile, F. (2021). A bumper crop of SNPs in soybean through high-density genotyping-by-sequencing (HD-GBS). *Plant Biotechnol. J.* 19 (5), 860. doi: 10.1111/2fppi.13551
- Tran, D. T., Stekete, C. J., Boehm, J. D., Noe, J., and Li, Z. (2019). Genome-wide association analysis pinpoints additional major genomic regions conferring resistance to soybean cyst nematode (*Heterodera glycines* ichinohe). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00401
- Ude, G., Pillay, M., Nwakanma, D., and Tenkouano, A. (2002). Genetic diversity in musa acuminata colla and musa balbisiana colla and some of their natural hybrids using AFLP markers. *Theor. Appl. Genet.* 104 (8), 1246–1252. doi: 10.1007/s00122-002-0914-4
- Upadhyaya, H. D., Bajaj, D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016a). Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci. Rep.* 6, 1–12. doi: 10.1038/srep24050
- Upadhyaya, H. D., Bajaj, D., Das, S., Saxena, M. S., Badoni, S., Kumar, V., et al. (2015). A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea. *Plant Mol. Biol.* 89, 403–420. doi: 10.1007/s11103-015-0377-z
- Upadhyaya, H. D., Bajaj, D., Narnoliya, L., Das, S., Kumar, V., Gowda, C. L. L., et al. (2016b). Genome-wide scans for delineation of candidate genes regulating seed-protein content in chickpea. *Front. Plant Sci.* 7, 302. doi: 10.3389/fpls.2016.00302
- Upadhyaya, H. D., Bajaj, D., Srivastava, R., Daware, A., Basu, U., Tripathi, S., et al. (2017). Genetic dissection of plant growth habit in chickpea. *Funct. Integr. Genomics* 17, 711–723. doi: 10.1007/s10142-017-0566-8
- Unterseer, S., Bauer, E., Haberger, G., Seidel, M., Knaak, C., Ouzunova, M., et al. (2014). A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15 (1), 1–15. doi: 10.1186/1471-2164-15-823
- 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 (3), 240–246. doi: 10.1038/nbt.2491
- Varshney, R. K. (2016). Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* 242, 98–107. doi: 10.1016/j.plantsci.2015.09.009
- Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630. doi: 10.1016/j.tplants.2005.10.004
- Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* 132 (3), 797–816. doi: 10.1007/s00122-018-3252-x
- Varshney, R. K., Roorkiwal, M., and Sorrells, M. E. (2017). Genomic selection for crop improvement: new molecular breeding strategies for crop improvement 1–258. doi: 10.1007/978-3-319-63170-7
- Varshney, R. K., Roorkiwal, M., Sun, S., Prasad, B., Annapurna, C., Mahendar, T., et al. (2021). A chickpea genetic variation map based on the sequencing of 3,366 genomes. *Nature* 599, 622–627. doi: 10.1038/s41586-021-04066-1
- Varshney, R. K., Saxena, R. K., Upadhyaya, H. D., Khan, A. W., Yu, Y., Kim, C., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082–1088. doi: 10.1038/ng.3872
- Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., and Bhatia, S. (2015). High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using genotyping-by-Sequencing (GBS). *Sci. Rep.* 5 (1), 17512. doi: 10.1038/srep17512
- Vuong, T. D., Sonah, H., Meinhardt, C. G., Deshmukh, R., Kadam, S., Nelson, R. L., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* 16, 593. doi: 10.1186/s12864-015-1811-y
- Wu, X., Islam, A. S. M. F., Limpot, N., Mackasmel, L., Mierzwa, J., Cortés, A. J., et al. (2020). Genome-wide SNP identification and association mapping for seed mineral concentration in mung bean (*Vigna radiata* L.). *Fron. Gen.* 11. doi: 10.3389/fgene.2020.00656
- 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
- Wang, J., Yan, C., Li, Y., Li, C., Zhao, X., Yuan, C., et al. (2019). GWAS discovery of candidate genes for yield-related traits in peanut and support from earlier QTL mapping studies. *Genes* 10, 803. doi: 10.3390/genes10100803
- Wang, L., Yang, Y., Zhang, S., Che, Z., Yuan, W., and Yu, D. (2020). GWAS reveals two novel loci for photosynthesis-related traits in soybean. *Mol. Genet. Genomics* 295, 705–716. doi: 10.1007/s00438-020-01661-1
- Wang, H., Yue, T., Yang, J., Wu, W., and Xing, E. P. (2019). Deep mixed model for marginal epistasis detection and population stratification correction in genome-wide association studies. *BMC Bioinf.* 20 (23), 6565. doi: 10.1186/s12859-019-3300-9
- Warsame, A. O., Angra, D., and O'Sullivan, D. (2019). Identification of a candidate gene controlling hilum colour in faba bean. *For People Planet* 31, 1–44.
- Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44 (3), 388–396.
- Wen, Z., Boyse, J. F., Song, Q., Cregan, P. B., and Wang, D. (2015). Genomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics* 16, 1–14. doi: 10.1186/s12864-015-1872-y
- Wu, X., Ren, C., Joshi, T., Vuong, T., Xu, D., and Nguyen, H. T. (2010). SNP discovery by high-throughput sequencing in soybean. *BMC Genomics* 11 (1), 1–10. doi: 10.1186/1471-2164-11-469
- Wurschum, T., Liu, W., Gowda, M., Maurer, H. P., Fischer, S., Schechert, A., et al. (2012). Comparison of biometrical models for joint linkage association mapping. *Heredity* 108, 332–340. doi: 10.1038/hdy.2011.78
- Yadav, S., Jackson, P., Wei, X., Ross, E. M., Aitken, K., Deomano, E., et al. (2020). Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy* 10, 585. doi: 10.3390/agronomy10040585
- Yadav, K., Yadav, S. K., Yadav, A., Pandey, V. P., and Dwivedi, U. N. (2014). Comparative analysis of genetic diversity among cultivated pigeonpea (*Cajanus cajan* (L.) millsp.) and its wild relatives (*C. albicans* and *C. lineatus*) using randomly amplified polymorphic DNA (RAPD) and inter simple sequence repeat (ISSR) fingerprinting. *Am. J. Plant Sci.* 5 (11), 1665. doi: 10.3390/agronomy10040585
- Yan, L., Hofmann, N., Li, S., Ferreira, M. E., Song, B., Jiang, G., et al. (2017). Identification of QTL with large effect on seed weight in a selective population of soybean with genome-wide association and fixation index analyses. *BMC Genomics* 18, 1–11. doi: 10.1186/s12864-017-3922-0
- Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5 (1), 1–9. doi: 10.1038/ncomms6087
- Yang, S. Y., Saxena, R. K., Kulwal, P. L., Ash, G. J., Dubey, A., Harper, J. D., et al. (2011). The first genetic map of pigeon pea based on diversity arrays technology (DART) markers. *J. Genet.* 90 (1), 103–109. doi: 10.1007/s12041-011-0050-5
- Yang, Y., Wang, L., Zhang, D., Cheng, H., Wang, Q., Yang, H., et al. (2020). GWAS identifies two novel loci for photosynthetic traits related to phosphorus efficiency in soybean. *Mol. Breed.* 40, 1–14. doi: 10.1007/s11032-020-01112-0
- You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. (2018). Development and applications of a high throughput genotyping tool for polyploid crops: single nucleotide polymorphism (SNP) array. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00104
- Yu, Z., Chang, F., Lv, W., Sharmin, R. A., Wang, Z., Kong, J., et al. (2019). Identification of QTN and candidate gene for seed-flooding tolerance in soybean [*Glycine max* (L.) merr.] using genome-wide association study (GWAS). *Genes* 10, 957.
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38 (2), 203–208. doi: 10.1038/ng1702
- Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc.: Ser. B (Stat. Methodology)* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x

- Zatybekov, A., Abugalieva, S., Didorenko, S., Gerasimova, Y., Sidorik, I., Anuarbek, S., et al. (2017). GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. *BMC Plant Biol.* 17 (1), 1–8. doi: 10.1186/s12870-017-1125-0
- Zatybekov, A., Abugalieva, S., Didorenko, S., Rsaliyev, A., and Turuspekov, Y. (2018). GWAS of a soybean breeding collection from South East and South Kazakhstan for resistance to fungal diseases. *Vavilov Journal of Genetics and Breeding* 22 (5), 536–543. doi: 10.18699/vj18.392
- Zavinon, F., Adoukonou-Sagbadja, H., Keilwagen, J., Lehnert, H., Ordon, F., and Perovic, D. (2020). Genetic diversity and population structure in beninese pigeon pea [*Cajanus cajan* (L.) huth] landraces collection revealed by SSR and genome wide SNP markers. *Genet. Resour. Crop Evol.* 67 (1), 191–208. doi: 10.1007/s10722-019-00864-9
- Zeng, A., Chen, P., Korth, K., Hancock, F., Pereira, A., Brye, K., et al. (2017). Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. *Mol. Breed.* 37, 1–14. doi: 10.1007/s11032-017-0634-8
- Zhang, H., Chu, Y., Dang, P., Tang, Y., Jiang, T., Clevenger, J. P., et al. (2020). Identification of QTLs for resistance to leaf spots in cultivated peanut (*Arachis hypogaea* L.) through GWAS analysis. *Theor. Appl. Genet.* 133, 2051–2061. doi: 10.1007/s00122-020-03576-2
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546
- Zhang, X., Huang, C., Wu, D., Qiao, F., Li, W., Duan, L., et al. (2017a). High-throughput phenotyping and QTL mapping reveal the genetic architecture of maize plant growth. *Plant Physiol.* 173 (3), 1554–1564. doi: 10.1104/pp.16.01516
- Zhang, C., McGee, R. J., Vandemark, G. J., and Sankaran, S. (2021). Crop performance evaluation of chickpea and dry pea breeding lines across seasons and locations using phenomics data. *Front. Plant Sci.* 12, 640259. doi: 10.3389/fpls.2021.640259
- Zhang, J., Song, Q., Cregan, P. B., and Jiang, G.-L. (2016). Genome-wide association study, genomic prediction, and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129, 117–130. doi: 10.1007/s00122-015-2614-x
- Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16, 1–11. doi: 10.1186/s12864-015-1441-4
- Zhang, Z., Todhunter, R. J., Buckler, E. S., and Van Vleck, L. D. (2007). Technical note: use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J. Anim. Sci.* 85, 881–885. doi: 10.2527/jas.2006-656
- Zhang, J., Wen, Z., Li, W., Zhang, Y., Zhang, L., Dai, H., et al. (2017b). Genome-wide association study for soybean cyst nematode resistance in Chinese elite soybean cultivars. *Mol. Breed.* 37, 1–10. doi: 10.1007/s11032-017-0665-1
- Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., et al. (2019). A combined linkage and GWAS analysis identify QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* 20, 5915. doi: 10.3390/ijms20235915
- Zhang, Z., Xu, Y., Liu, J., and Kwok, C. K. (2012). “Identify predictive SNP groups in genome-wide association study: a sparse learning approach. *Procedia computer science*,” in *Proceedings of the 3rd International Conference on Computational Systems-Biology and Bioinformatics*, 11 (January), 107–114. doi: 10.1016/j.procs.2012.09.012
- Zhang, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021). Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134, 1329–1341. doi: 10.1007/s00122-021-03774-6
- Zhang, X., Zhang, J., He, X., Wang, Y., Ma, X., and Yin, D. (2017c). Genome-wide association study of major agronomic traits related to domestication in peanut. *Front. Plant Sci.* 8, 611. doi: 10.3389/fpls.2017.01611
- Zhao, J., Bayer, P. E., Ruperao, P., Saxena, R. K., Khan, A. W., Goliz, A. A., et al. (2020). Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnol. J.* 18 (9), 1946–1954. doi: 10.1111/pbi.13354
- Zhao, X., Jiang, H., Feng, L., Qu, Y., Teng, W., Qiu, L., et al. (2019). Genome-wide association and transcriptional studies reveal novel genes for unsaturated fatty acid synthesis in a panel of soybean accessions. *BMC Genomics* 20, 1–16. doi: 10.1186/s12864-019-5449-z
- Zhou, X., Xia, Y., Liao, J., Liu, K., Li, Q., Dong, Y., et al. (2016). Quantitative trait locus analysis of late leaf spot resistance and plant-type-related traits in cultivated peanut (*Arachis hypogaea* L.) under multi-environments. *PLoS One* 11 (11), e0166873.
- Zhou, X., Xia, Y., Ren, X., Chen, Y., Huang, L., Huang, S., et al. (2014). Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). *BMC Genomics* 15, 1–14.
- Zou, K., Kang, D., Kim, K.-S., and Jun, T.-H. (2020). Screening of salinity tolerance and genome-wide association study in 249 peanut accessions (*Arachis hypogaea* L.). *Plant Breed. Biotechnol.* 8, 434–438. doi: 10.9787/PBB.2020.8.4.434