

RESEARCH ARTICLE

Open Access



# Deconstructing the genetic architecture of iron deficiency chlorosis in soybean using genome-wide approaches

Teshale Assefa<sup>1</sup>, Jiaoping Zhang<sup>1</sup>, R. V. Chowda-Reddy<sup>1</sup>, Adrienne N. Moran Lauter<sup>2</sup>, Arti Singh<sup>1</sup> , Jamie A. O'Rourke<sup>2</sup>, Michelle A. Graham<sup>2</sup> and Asheesh K. Singh<sup>1\*</sup>

## Abstract

**Background:** Iron (Fe) is an essential micronutrient for plant growth and development. Iron deficiency chlorosis (IDC), caused by calcareous soils or high soil pH, can limit iron availability, negatively affecting soybean (*Glycine max*) yield. This study leverages genome-wide association study (GWAS) and a genome-wide epistatic study (GWES) with previous gene expression studies to identify regions of the soybean genome important in iron deficiency tolerance.

**Results:** A GWAS and a GWES were performed using 460 diverse soybean PI lines from 27 countries, in field and hydroponic iron stress conditions, using more than 36,000 single nucleotide polymorphism (SNP) markers. Combining this approach with available RNA-sequencing data identified significant markers, genomic regions, and novel genes associated with or responding to iron deficiency. Sixty-nine genomic regions associated with IDC tolerance were identified across 19 chromosomes via the GWAS, including the major-effect quantitative trait locus (QTL) on chromosome Gm03. Cluster analysis of significant SNPs in this region deconstructed this historically prominent QTL into four distinct linkage blocks, enabling the identification of multiple candidate genes for iron chlorosis tolerance. The complementary GWES identified SNPs in this region interacting with nine other genomic regions, providing the first evidence of epistatic interactions impacting iron deficiency tolerance.

**Conclusions:** This study demonstrates that integrating cutting edge genome wide association (GWA), genome wide epistasis (GWE), and gene expression studies is a powerful strategy to identify novel iron tolerance QTL and candidate loci from diverse germplasm. Crops, unlike model species, have undergone selection for thousands of years, constraining and/or enhancing stress responses. Leveraging genomics-enabled approaches to study these adaptations is essential for future crop improvement.

**Keywords:** Gene expression, Germplasm, GWAS, GWES, Iron deficiency chlorosis (IDC), Quantitative trait locus (QTL), Soybean

## Background

Iron (Fe) is an essential micronutrient required for multiple metabolic processes in plants including photosynthesis, respiration, and electron transport [1]. While Fe is one of the most abundant elements in the earth's crust, aerobic conditions, high pH, and/or calcareous soils, make it insoluble and unavailable for plant use. Of the world's cultivated soils, roughly 30% are classified as calcareous [2], including

farmlands in the upper Midwestern United States where soybean is a major crop. Fe deficiency negatively affects plant growth and yield [3]. Studies in the model organism *Arabidopsis thaliana* have demonstrated that the expression of a number of underlying genes is regulated by Fe availability [4–6]. Additional genes involved in Fe acquisition and homeostasis have been the subject of numerous studies in *Arabidopsis* [4–9]; however, work in model species has not translated to improved Fe efficiency in crop species. Unlike *Arabidopsis*, which has not been domesticated, soybean was domesticated more than 5000 years ago [10]. Following

\* Correspondence: [singhak@iastate.edu](mailto:singhak@iastate.edu)

<sup>1</sup>Department of Agronomy, Iowa State University, Ames, IA, USA  
Full list of author information is available at the end of the article



domestication, soybean and other crops have been under continued selection for yield, biotic and abiotic stress tolerance, making it likely that they have developed novel strategies for dealing with Fe deficiency stress through long-term selection and mutation strategies. Therefore, a critical need exists to study Fe deficiency responses within a crop species and across a broad range of diverse genotypes.

Genome-wide association studies (GWAS) have been used to detect quantitative trait loci (QTLs) associated with important agronomic traits in rice, maize, wheat, and soybean [11–15]. Incorporating diverse plant introduction (PI) germplasm accessions in GWAS increases the likelihood of identifying novel genes and rare alleles [16]. A previous iron deficiency chlorosis (IDC) GWAS in soybean identified QTL on seven chromosomes [13]. This study was performed in two populations of advanced breeding lines developed by public and private breeding programs for the upper Midwest. However, the narrow genetic base for U.S. commercial soybean cultivars [17] limited the likelihood of identifying novel mechanisms and natural genetic variants for IDC tolerance that could be used for future soybean improvement.

GWAS usually focus on additive genetic effects. However, epistatic interactions also contribute to genetic variation [18, 19]. Detection of genome-wide epistatic (GWE) interactions represents a complementary approach to traditional genetic studies and is essential to understanding the genetic architecture of quantitative traits [20]. Studies in human and animal systems have revealed that epistatic interactions have major effects on the genetic architecture of complex disease traits [21, 22]. While the utility of GWE studies (GWES) is acknowledged in the crop research community, they have not been widely integrated.

The objective of this study was to use GWAS and GWES to examine the genetic architecture of soybean responses to Fe deficiency using a diverse panel of germplasm accessions genotyped with a dense coverage of single-nucleotide polymorphism (SNP) markers. Unlike traditional QTL or expression analyses, this combinatorial approach allowed us to identify novel genes and mechanisms conferring tolerance to IDC within the soybean germplasm collection. By coupling these studies with over 30 years of IDC research, we have deconstructed the major IDC QTL on soybean chromosome 3 (Gm03) into multiple discrete regions contributing to IDC tolerance. The identification of these candidate genes from GWAS and GWES will expand our understanding of IDC tolerance mechanisms and identify important genetic markers for soybean breeding and improvement.

## Results

### Phenotype

The phenotypic variation in IDC score among the PI accessions, averaged across replicates, ranged from highly

resistant (1.0) to highly susceptible (4.7) in the 2014 field study (Additional file 1: Figure S1a) and from 1.0 to 4.6 in the 2015 field study, indicating significant differences among the PI lines for the expression of IDC symptoms (Additional file 4). The population had an approximately normal distribution for visual score and SPAD measurements at each time point. The correlation between the IDC ratings at the different time points varied from 0.71 to 0.86 in 2014 and from 0.95 to 0.98 in 2015 with the greatest correlation observed between T2 and T3 for both field and hydroponic studies (Additional file 1: Figure S1b). The estimated broad-sense heritability [23] for visual scores was 82, 64, and 52% for the first, second, and third time points, respectively. Comparing any two samples results in a positive correlation, supporting the inclusion of all collected data in our analyses.

### GWAS

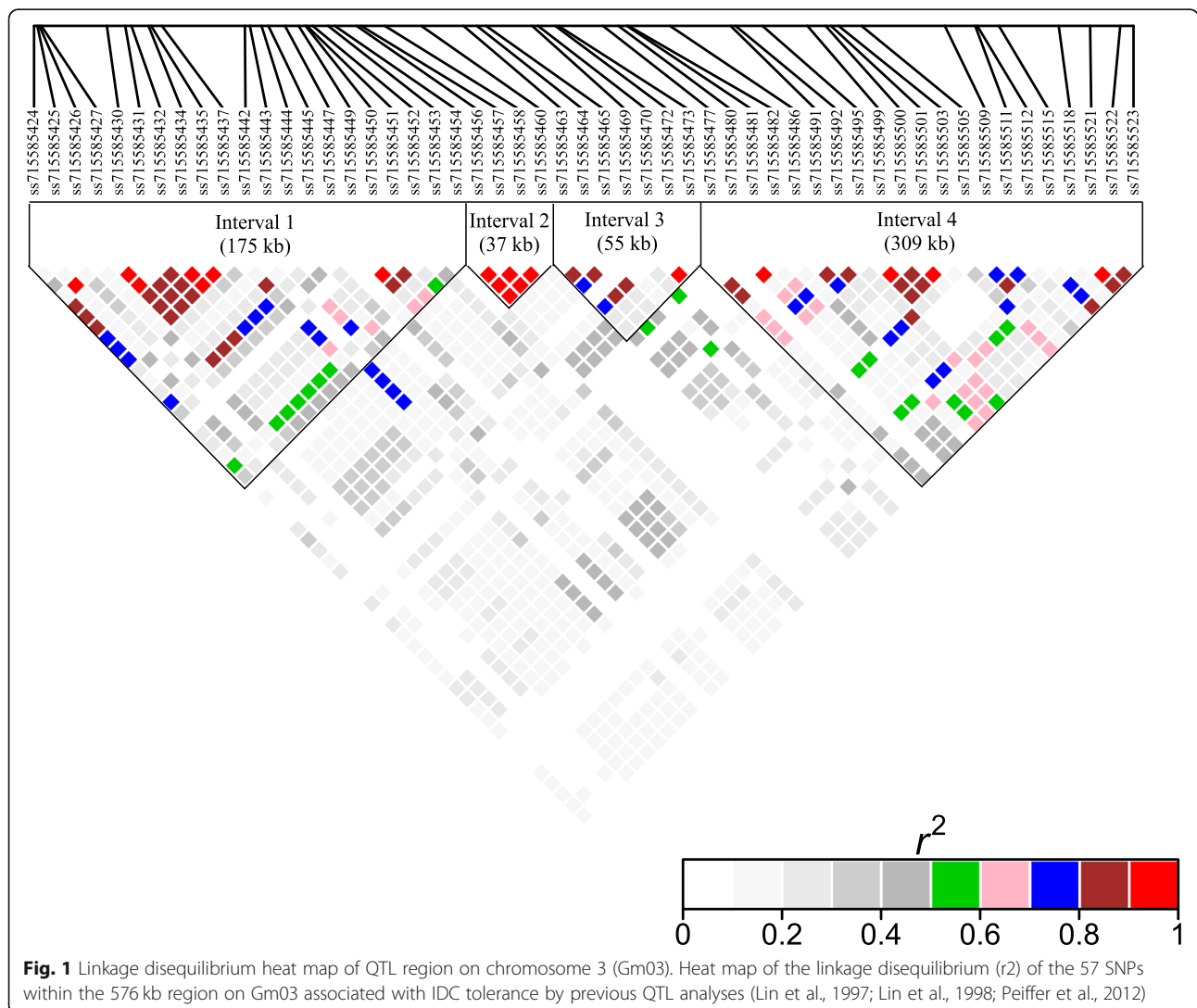
A total of 97 unique SNPs were identified across all experiments, with 43, 32 and 48 unique SNPs identified from field conditions in 2014 and 2015, and hydroponics, respectively (Additional file 5, Additional file 2: Figure S2 and Additional file 3: Figure S3). In the 2014 and 2015 field data, 10 (23%) and 23 SNPs (72%) were identified in two or more time points or phenotyping methods, respectively. In the hydroponic conditions, 6 SNPs (13%) were identified in two or more time points or phenotyping methods. Twelve of the 97 total SNPs (12%) were identified across multiple years in the field data and of these, four (4%) were identified across multiple years of field data and in hydroponics. SNPs detected in more than one environment are listed in Additional file 6.

The 97 unique SNPs identified by GWAS were used to define 69 genomic regions of interest for IDC, each containing at least one significant SNP for at least one experimental measurement. On average, these regions were 35 kb in length with a standard deviation of 25.5 kb. A total of 278 genes in the Williams 82 reference genome were identified from the 69 regions of interest [24] (Additional file 7). To identify high priority candidate genes, we mined available RNA-seq data from leaves and roots of the iron efficient line Clark at 30, 60 and 120 min after iron stress [25] and at 1 and 6 h after iron stress [26]. In addition, we mined data from leaves of the iron inefficient line IsoClark 21 days post gene silencing of GmRPA3c, a previously identified and characterized iron response gene, in iron sufficient and deficient conditions [27]. Of the 278 candidate GWAS genes, 65 were significantly differentially expressed in at least one of the previous RNA-seq studies (Additional file 7). The majority of the differentially expressed genes (48), were identified within 2 h of iron stress application, in hydroponic conditions. Only 7 genes were identified across multiple studies, reflecting the dynamic nature of

the soybean iron stress response. Of the 65 differentially expressed genes, 26 were differently expressed in leaves, 36 were differentially expressed in roots, and 5 were differentially expressed in leaves and roots.

Interestingly, the twelve SNPs identified across multiple years in the field data were located within a 392 kb window on Gm03 (from 34,364,354 to 35,757,151). Given that this region corresponds to a historical IDC QTL [28–31], we were interested in examining linkage across this region. Clustering ( $r^2 > 0.5$ ) of significant SNPs on Gm03 (34,227,914–34,955,422) identified four distinct linkage blocks (genomic intervals) within this 730 kb region (Fig. 1). Interval 1 on Gm03 spans the SNPs *ss715585424* and *ss715585454* (corresponding to genes *Glyma.03 g127900* - *Glyma.03 g129500*). Of the 17 genes in this interval, ten overlapped with the 120 kb IDC introgression identified by Peiffer et al. [31]. Five genes (*Glyma.03 g127900*, *Glyma.03 g128200*, *Glyma.03 g128300*, *Glyma.03 g129300* and *Glyma.03 g129400*) were significantly differentially expression in

response to iron stress, four in roots and one in leaves (Additional file 8). Three of these have also been associated with abiotic stress and defense responses in other species (*Glyma.03 g128200*, *Glyma.03 g128300* and *Glyma.03 g129400*). Interval 2 spans SNPs *ss715585456* and *ss715585460* (corresponding to five genes: *Glyma.03 g129600* - *Glyma.03 g130000*). Within this interval both *Glyma.03 g129600* and *Glyma.03 g130000* were significantly differentially expressed in response to iron stress in roots. Interval 3 spans SNPs *ss715585463* - *ss715585473* (corresponding to six genes: *Glyma.03 g130100* - *Glyma.03 g130600*) including the final segment of the 120 kb IDC introgression reported by Peiffer et al. [31]. Of the six genes, three were differentially expressed in response to iron stress including *Glyma.03 g130200*, *Glyma.03 g130400* and *Glyma.03 g130600*. Interval 4 spans SNPs *ss715585477* - *ss715585531* (corresponding to 30 genes: *Glyma.03 g130700* to *Glyma.03 g133400*). Of these, six genes were differentially expressed in iron stress RNA-seq data

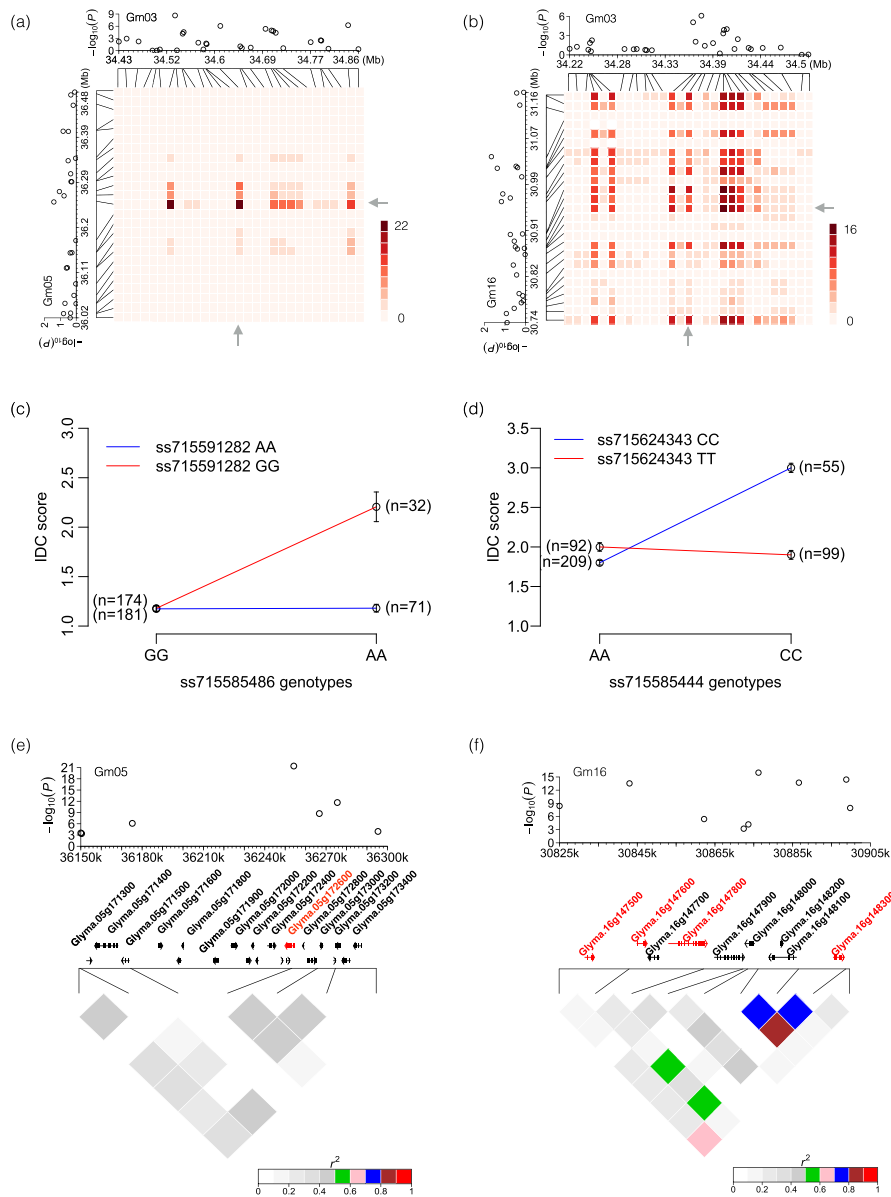


including *Glyma.03 g130900-Glyma.03 g131200*, *Glyma.03 g132700* and *Glyma.03 g132900*.

**GWES**

Epistatic tests identified 20 SNP: SNP interactions between five SNPs (*ss715585442*, *ss715585444*, *ss715585450*,

*ss715585469*, and *ss715585486*) on Gm03 which interacted with 12 SNPs on other chromosomes (Fig. 2, Additional file 9). The same approach used to identify candidate genes in GWAS QTL was used to identify candidate genes within GWES regions. We identified 50 candidate genes (Additional file 10 and Additional file 11), of



**Fig. 2** Epistatic interactions between soybean chromosome 3 (Gm03) and chromosomes 5 (Gm05) and 16 (Gm16). **a** Heat map of SNP: SNP interactions between five SNPs on Gm03 and *ss715585486* (Gm05). **b** Heat map of SNP: SNP interactions between three SNPs on Gm03 and *ss715585444* (Gm16). Colors represent the  $-\log_{10}$ -transformed  $P$  value of each interaction. Statistically significant interactions are highlighted with arrows. **c** and **d** Impact of the AA allele of *ss715585486* (Gm05) and the TT allele of *ss715585444* (Gm16) on IDC score. Allelic combinations are provided on the X-axis with IDC visual scores on the Y-axis. These alleles confer a one-point improvement on the IDC scale. Lines that are heterozygous at the related loci were ruled out. **e** and **f** Candidate gene prediction for epistatic loci. The top panels show  $-\log_{10}$  transformed  $P$ -values of the SNP: SNP interactions plotted in chromosomal position. Middle panels show all putative genes within the regions of interest; candidate genes of interest are highlighted in red. Linkage disequilibrium ( $r^2$ ) of chromosomal regions of interest for epistatic interactions are plotted in bottom panels

which 13 were differentially expressed in the different iron stress responsive RNA-seq data sets. All five SNPs on Gm03 exhibited significant interactions with *ss715591282* on Gm05 (Fig. 2a, c and e). The SNP is not located within any previously identified IDC QTL. The AA allele of this SNP confers a one point improvement on the IDC scale compared to the GG allele (Fig. 2c). Three of the five SNPs from Gm03 interact with *ss715624343* (Fig. 2b, d and f); this SNP lies on soybean Gm16, which has not previously been reported as an IDC QTL. Analysis shows that a TT allele of *ss715624343* accounts for a one-point increase on the five-point chlorosis rating scale compared to the CC allele.

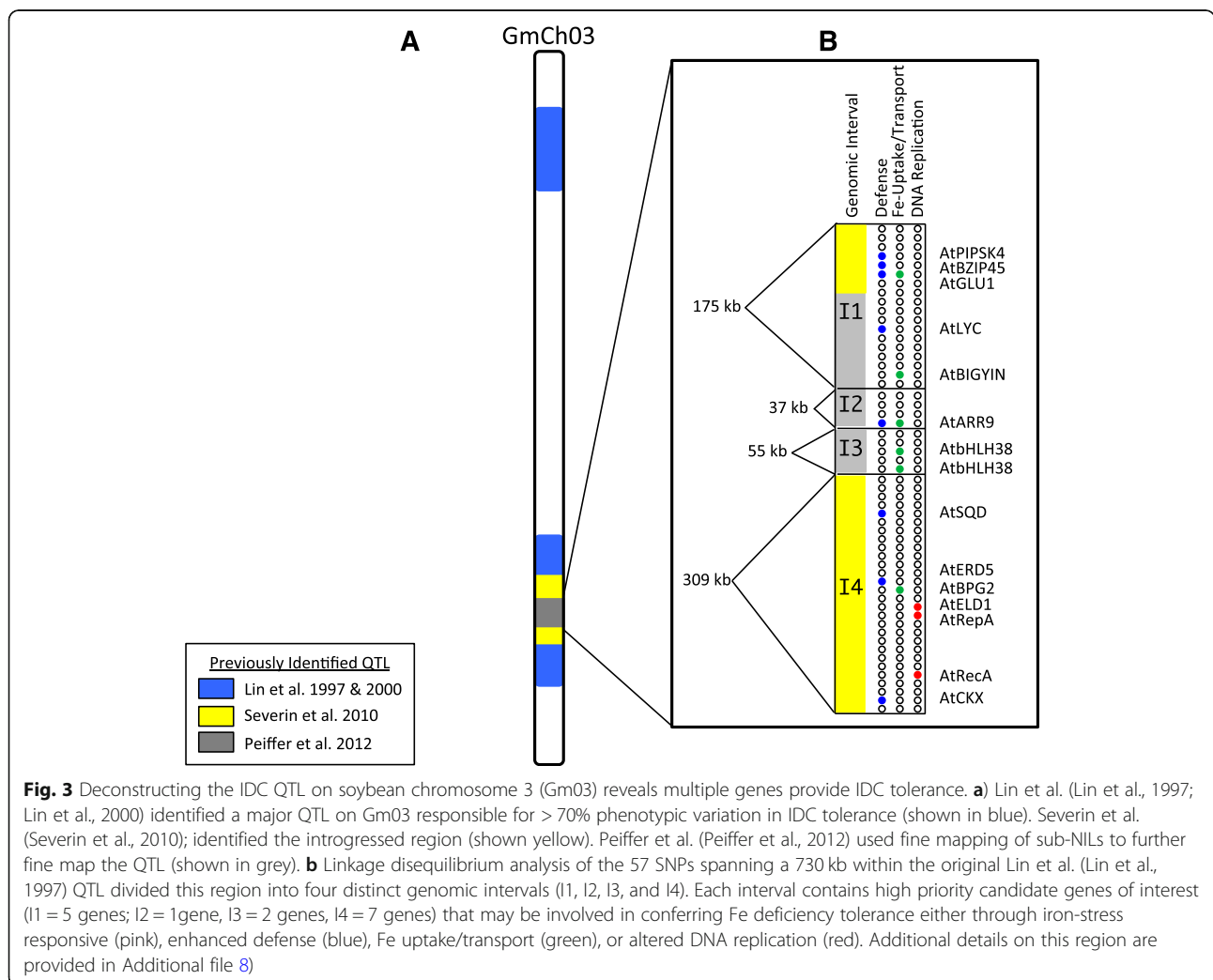
In order to validate GWES results, we took advantage of StringDB, which stores known interactions between proteins [32]. This confirmed an interaction between candidate gene *Glyma.03 g129400* (AtBIGYIN), which responds to iron stress (Additional file 7), and *Glyma.16 g148100* (At5g55610, Additional file 11). In Arabidopsis, both of these proteins are located within the outer

mitochondrial membrane and are thought to play a role in mitochondrial signaling during stress [33].

### Discussion

#### Historical discovery of the IDC QTL on soybean Gm03

Over the last 35 years, several different approaches have been used to identify genes conferring tolerance to IDC in soybean. Cianzio et al. [34] and Cianzio and Fehr [35] were the first to demonstrate the genetic inheritance of IDC tolerance in soybean. In 1997 and 1998, Lin et al. [28, 36] used field and hydroponic studies to identify an IDC QTL on soybean Gm03 that explained 70% of the phenotypic variation (Fig. 3). In 2010, Severin et al. [30] used next-generation sequencing data to identify the introgression from Fe inefficient T203 into iron-efficient Clark, used to develop inefficient IsoClark. Peiffer et al. [31] fine mapped the introgressed region, using Clark and IsoClark sub-NILs to identify a 120 kb region conferring IDC tolerance. Of the eighteen genes in this region, two candidate genes in soybean were identified



as having the greatest homology to the AtbHLH38 transcription factor in *Arabidopsis* that regulates Fe uptake in the root [37].

#### Expression studies of IDC tolerance in soybean

While mapping studies suggested that a single gene likely controlled IDC response in Clark, expression studies suggested the involvement of multiple regulatory cascades. In 2009, O'Rourke et al. [38] compared gene expression between Clark and Isoclark leaves following Fe stress treatment. This study suggested that Isoclark contained a mutation responsible for regulating the expression of downstream Fe uptake and transport genes. It also suggested that regulation of DNA replication and defense was an important component of Clark's Fe stress response. Atwood et al. [27] used virus-induced gene silencing (VIGS) to repress the expression of GmRPA3c (Replication Protein A, subunit 3c) in Isoclark to mirror the expression observed in Clark. GmRPA3c silencing improved Isoclark performance during Fe stress and resulted in massive transcriptional reprogramming of genes involved in iron uptake and homeostasis, defense/cell death, DNA replication, abiotic stress, regulation of the circadian clock, and autophagy. Differential expression of DNA replication genes has now been observed as early as one hour after Fe stress treatment in soybean [26]. The involvement of DNA replication in abiotic and biotic stress responses has now been reported in multiple crop species including maize, barley, and onion [39–43].

The differential expression of Fe, defense, and DNA replication cascades in soybean suggest the involvement of at least two regulatory genes in soybean's Fe stress response. Peiffer et al. [31] identified two candidate BHLH38 transcription factors in the IDC QTL on Gm03 induced by Fe stress in roots. Multiple constructs have been developed and used to silence these genes, but silenced Clark plants had no significant phenotypic changes when grown under Fe stress conditions. While negative results do not preclude the involvement of these soybean BHLH38 genes in Clark's Fe stress response, it again suggests the involvement of additional genes. Further, there is no evidence that AtBHLH38 regulates the expression of DNA replication or defense genes, nor that regulation of these genes is a component of the *Arabidopsis* Fe stress response [6, 44, 45]. Therefore, it is likely that regulation of the DNA replication and defense machinery, in response to abiotic stress, is a unique adaptation in crop species. Since only one major IDC QTL has been identified in soybean, it would suggest that multiple candidate genes reside within this region.

#### Splitting the IDC QTL on soybeanGm03

Of the 69 genomic regions identified in this study, eight overlapped with the previously identified IDC QTL [28,

29] and introgressed regions [30] on Gm03. Within this region, 13 of 16 significant and unique SNPs were identified across three or more experimental conditions. Given the physical distribution of the SNPs across this 730 kb region, the linkage disequilibrium (LD) between these SNPs was examined. This approach broke the previously described IDC QTL into four distinct intervals of 175, 37, 55 and 309 kb (intervals 1–4, respectively, Figs. 1 and 3), demonstrating that multiple genes within this window on Gm03 contribute to IDC tolerance. Intervals 2 and 3 completely overlapped with the 120 kb introgression identified in a previous study [31]. In contrast, interval 4 did not overlap with the Peiffer IDC introgression at all. Within each of the intervals, high priority candidate genes involved in signal transduction were identified that could explain the hallmarks of soybean's Fe stress response: defense, DNA replication, and Fe uptake/transport.

Interval 1 contained four high priority candidate genes of interest: *Glyma.03 g128200* (AtTGA6), *Glyma.03 g128300* (AtGLU1), *Glyma.03 g128900* (AtLYC), and *Glyma.03 g19400* (AtBIGYIN). AtTGA6 regulates crosstalk between salicylic acid (SA) and jasmonic acid (JA)/ethylene defense responses [46]. All three hormones are also involved in the regulation of Fe deficiency responses [47, 48]. AtGLU1 encodes a ferredoxin-dependent glutamate synthase. AtGLU1 knock-downs are slightly chlorotic [49]. Gene expression analyses of AtGLU1 reveal extensive transcriptional reprogramming including repression of photosynthesis-related genes and induction of abiotic stress-associated genes. Similarly, AtLYC is involved in non-photochemical quenching under high light conditions [50]. Transformation of AtLYC in tobacco resulted in increased tolerance to salt stress. AtBIGYIN regulates mitochondrial size and number [51]. Mitochondria are essential for the synthesis of Fe-S clusters, which are linked to intracellular Fe homeostasis [52]. All of these genes, except *Glyma.03 g128900* (AtLYC), were differentially expressed in roots 30 min after iron stress (Additional file 8).

Interval 2 contained a single high priority candidate gene. *Glyma.03 g130000* was differentially expressed in response to Fe stress in the roots at 30 min (Additional file 8). *Glyma.03 g130000* encodes a homolog of AtRR4 (Response Regulator 4), a component of the cytokinin-signaling pathway. Cytokinin signaling regulates a broad range of nutrient and environmental stress responses [53] and is directly or indirectly involved in regulating Fe uptake genes [54]. AtARR4 is also directly involved in the regulation of the circadian clock and the cell cycle. The circadian clock in turn regulates the Fe homeostasis genes AtIRT1, AtBHLH38, and AtFERRITIN1 [55]. Similarly, Atwood et al. [27] and Moran Lauter et al. [26] observed

differential expression of circadian clock genes in response to Fe stress in soybean.

Interval 3 contained the two AtBHLH38 (*Glyma.03 g130400* and *Glyma.03 g130600*) transcription factors characterized by Peiffer et al. [31]. AtBHLH38 interacts with FIT and directly enhances the expression of downstream Fe genes FRO2 and IRT1 [37]. Both genes are differentially expressed in response to Fe in the roots at 30 min (Additional file 8). *Glyma.03 g130600* was also differentially expressed in the roots at 1 h (Additional file 8). Soybean lines containing a 12 bp deletion in *Glyma.03 g130400* were unable to induce the expression of the Fe uptake genes GmFRO2 and GmFIT1 under Fe stress conditions.

Interval 4 is likely one of the most novel regions because our analyses placed it outside the 120 kb introgression identified by Peiffer et al. [31]. This region contains four high priority candidate genes: *Glyma.03 g130900* (AtSDP1), *Glyma.03 g131100* (AtSQD1), *Glyma.03 g132400* (At1G52950) and *Glyma.03 g133000* (AtRECA). Homologs of SDP1 play an essential role in lipid metabolism and cell survival during stress conditions in yeasts, mammals and plants [56]. While AtSQD1 mutants have no obvious phenotypes, AtSQD2 mutants exhibit severe chlorosis under phosphate starvation conditions [57]. Both *Glyma.03 g130900* and *Glyma.03 g131100* were repressed by iron stress in leaves at 120 min. *Glyma.03 g131100* was induced by iron stress in roots at 30 min (Additional file 8). *Glyma.03 g132400* is homologous to RPA subunit 1 and contains the RPA1 domain (PTHR23273). RPA is a heterotrimeric protein made of three subunits (RPA1, RPA2 and RPA3) which bind single-stranded DNA during DNA repair and replication [58]. Atwood et al. [27] demonstrated that the majority of RPA subunits respond to Fe stress in soybean. Further, silencing of RPA subunit 3c (GmRPA3c) restored IDC tolerance in Isoclark. AtRECA regulates DNA repair through homologous recombination and is also an essential component of DNA replication [59]. Since RECA is targeted to the chloroplast, loss of RECA results in leaf abnormalities associated with the lack of DNA repair. RPA and AtRECA are both important components of the DNA replication and repair machinery (KEGG: ath03030, ath03420, ath03430, ath03440). Biotic and abiotic stresses, like IDC, result in the release of reactive oxygen species that can damage DNA [43]. Recognition of DNA damage results in inhibition of cell proliferation, allowing repair to occur. This also inhibits growth, reducing the need for iron.

#### GWAS identifies novel IDC genes throughout the soybean genome

Population-based association studies involving unrelated individuals are promising approaches for identifying the

genetic basis of complex traits such as IDC tolerance. Previous soybean IDC studies were limited by their reliance on homology to IDC-associated genes in a model species or on differential gene expression in response to Fe stress in a limited number of genotypes, conditions, or tissues. Using multiple phenotyping methods, developmental stages, growth conditions, and a diverse germplasm panel, would allow us to discover multiple mechanisms governing IDC tolerance in the soybean germplasm collection.

SNPs significantly associated with Fe deficiency were identified on every chromosome except Gm04. The genomic regions on Gm03 and Gm19 reside within previously identified Fe efficiency QTL [28, 60, 61]. The remaining significant SNPs and the associated 228 genes identified throughout the genome were unique to this study. Fe specific stress response genes included genes involved in heavy metal sensing, uptake, and homeostasis. AtURH2 (*Glyma.20 g000200*) is involved in Fe sensing in yeast and mammals [62], and expression of AtOSX3 (*Glyma.06 g056600*) confers tolerance to high metal levels and oxidizing chemicals [63]. *Glyma.08 g347000* is homologous to AtMRP3, a multi-drug resistance-associated protein induced by multiple heavy metals, but not Fe [64].

Disease and stress-related genes associated with significant SNPs had functions conferring tolerance to various stresses including heat stress (*Glyma.05 g001200* and *Glyma.06 G056400* [48, 64, 65]), cold stress (*Glyma.05 G000200* and *Glyma.12 G235700* [48, 65, 66]), cadmium tolerance (*Glyma.09 G110400*, [64]), shade avoidance (*Glyma.14 G032200*, [67]), salt stress (*Glyma.09 G051200*, [68]), and phosphate deficiency (*Glyma.09 G110200*, [69]). Genes contributing to disease and pathogen responses included four genes on Gm19 encoding canonical disease resistance genes (*Glyma.19 G139400*, *Glyma.19 G139500*, *Glyma.19 G139600*, and *Glyma.19 G139700*), though none have been associated with specific diseases. Additional disease response genes include the homologs to AtCDR1 (*Glyma.12 G235400*), a constitutive disease response gene involved in *Pseudomonas syringae* responses, AtPCC1 (*Glyma.15 G256200* and *Glyma.15 G256300*) which is involved in RPP7-dependent resistance to downy mildew [70–72], and AtLECRK (*Glyma.02 g043000*), involved in resistance to *Phytophthora spp.*

DNA replication and cell growth associated genes with significant SNPs included *Glyma.01 G193400* (AtCYCD5), *Glyma.02 G074800* (AtGAI), and *Glyma.05 G168600* (AtEBS1), all important components regulating plant growth and the cell cycle [73–75]. In addition, significant SNPs were also identified in two sucrose transporter (AtSUC2) homologs (*Glyma.02 g075000* and *Glyma.02 g075100*). Increased sugar accumulation in roots induces

reductase activity and the expression of Fe acquisition genes [76]. A family of SWEET sugar transporters and two SUC2 genes were differentially expressed after 1 h of Fe deficiency stress in soybean, confirming the importance of sugar signaling in soybean's Fe stress response [26]. The identification of genes involved in Fe specific, general biotic/abiotic stress responses and DNA replication in this experiment confirms that Fe deficiency responses in soybean occur through multiple novel mechanisms.

### Iron deficiency and epistasis

Epistatic interaction analyses identified five SNPs on Gm03 that fell within three of the four genomic intervals depicted in Fig. 3 (three SNPs in interval one, one SNP each in interval three and four). Cumulatively, these 5 SNPs had epistatic interactions with 13 different genomic locations, representing 50 candidate interacting genes. Each of these SNPs interacts with SNP *ss715591282*, which is located on Gm05 (Fig. 2a, Additional file 9). This SNP is not located within any previously identified IDC QTL ([SoyBase.org](http://SoyBase.org)). Further, it is not associated with any QTL identified in this GWAS study. The AA allele of this SNP confers a more than one point improvement on the IDC scale compared to the GG allele (Fig. 2c). A one-point increase on a five-point IDC rating scale corresponds to a 20% reduction in yield [3]. Genomic analyses of the SNP identified nine candidate genes, three of which were differentially expressed in response to iron stress. *Glyma.05 g172300* (AtPSBY) encodes a component of photosystem II, which produces reactive oxygen species in response to stress [77]. The function of the other two iron responsive genes (*Glyma.05 g172400* and *Glyma.05 g172800*) is unknown. It is worth noting the SNP itself is located within *Glyma.05 g172600* (Fig. 2e). The Arabidopsis homolog of this gene is Actin Regulated Protein 3 (AtARP3), involved in regulating Ca<sup>2+</sup> signaling in response to salt stress [78]. Though calcium signaling has not yet been characterized in IDC responses, it is a conserved signaling response known to be involved in multiple nutrient deficiencies including phosphate, potassium, boron, and salt. It is possible that any of these nine genes in this region are involved in the epistatic interaction associated with SNP *ss715591282*.

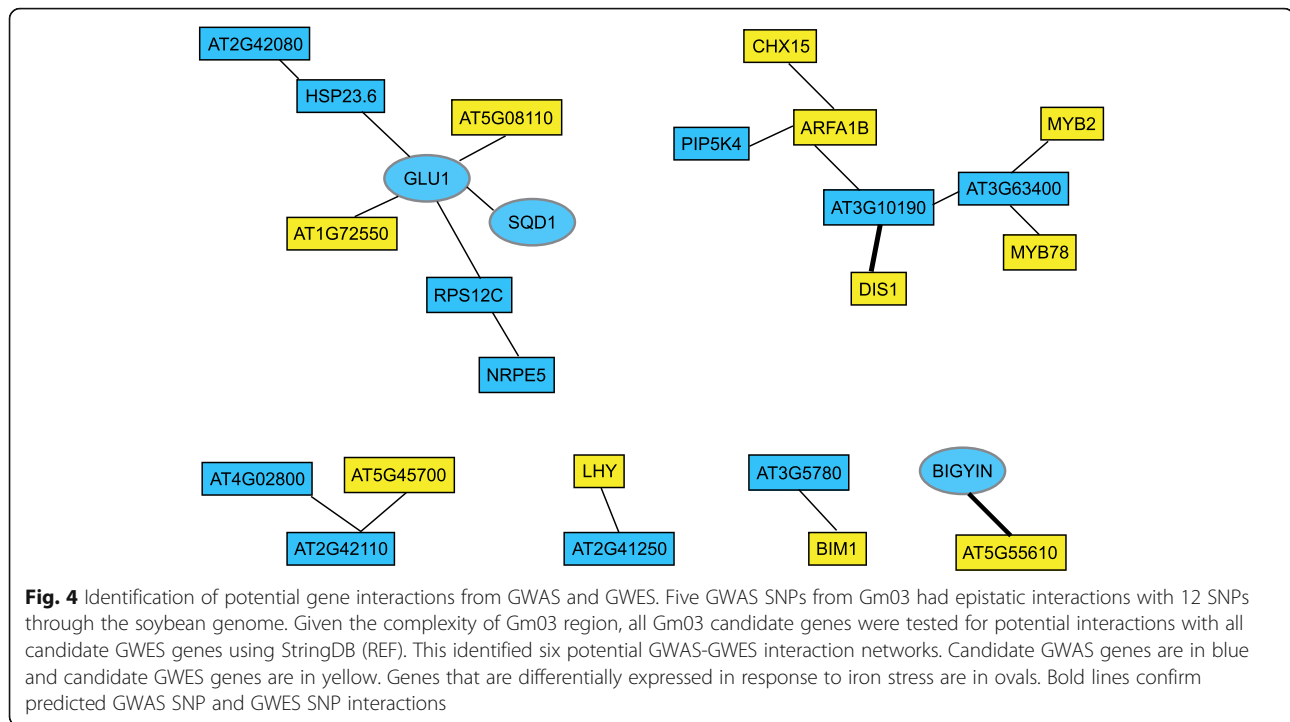
An additional epistatic interaction involves three SNPs on interval one that interact with SNP *ss715624343* on Gm16 (Fig. 2b). No IDC QTL have been previously discovered on Gm16 ([SoyBase.org](http://SoyBase.org)). While three candidate genes were identified in this region, none were differentially expressed in response to iron stress or had obvious function related to biotic or abiotic stress responses. Remarkably, the TT allele of this SNP confers a > one point improvement on the IDC scale compared to the CC allele (Fig. 2d).

Within the GWAS candidate genes for SNP *ss715585450*, we identified *Glyma.03 g129400* (AtBIGYIN), which responded to iron stress. Using StringDB [32], we were able to identify a candidate epistatic gene corresponding to SNP *ss715624343*, *Glyma.16 g148100*. Given the complexity of the Gm03 regions and that three distinct non-consecutive GWAS SNPs were able to interact with multiple genomic locations, we decided to use StringDB to test the interactions of all Gm03 candidate GWAS genes with all the candidate GWES genes. This identified six potential networks, each containing at least one GWAS and GWES candidate gene (Fig. 4). Of the six networks, two contained iron stress responsive genes. Of particular interest was the network containing the GWAS iron stress responsive gene *Glyma.03 g128300* (AtGLU1). AtGLU1 encodes a ferredoxin-dependent glutamate synthase. Knock-down mutants display leaf chlorosis and activation of multiple stress responses [49]. In the network, AtGLU1 interacts with the GWES candidate genes *Glyma.05 g127900* (At1G72550), a tRNA synthase, and *Glyma.06 g206600* (At5G08110/AtHRQ1), which is required for genome stability and repair [79]. These results provide further support for our epistatic analyses. Exploiting these novel findings could result in improved crop performance under stress conditions.

### Conclusions and perspectives

In this report, we identified a significant number of molecular markers, genomic regions and candidate genes responding to iron deficiency. Leveraging genomics-enabled approaches to study iron deficiency chlorosis is essential for future soybean improvement programs under multiple objectives [80]. Genome wide studies will benefit from digital and automated phenotyping [81, 82, 83]. Also, markers reported in this study might help in future soybean genomic studies. Our genome-wide studies leveraged thousands of SNPs, a diverse soybean germplasm panel, multiple phenotyping methods, developmental stages and growth conditions. This approach allowed us to identify novel IDC QTL that could be used for future soybean improvement. By integrating gene expression data from RNA-seq studies of soybean iron stress responses, we were able to identify high priority candidate genes. The novel GWES study allowed us to identify novel gene networks contributing to iron stress responses in soybean. In addition, we were able to dissect the historical IDC QTL on Gm03 into four distinct genomic intervals. For more than 30 years, it was thought that a single candidate gene controlled this QTL accounting for as much as 70% of the observed phenotypic variation in Fe deficiency tolerance. Only through an interdisciplinary approach that combined 30 years of breeding, gene expression, and new genome-wide





association studies could we demonstrate that this region contains multiple candidate genes. Linkages to the same biochemical and molecular pathways suggest there are multiple avenues for generating tolerance to IDC in soybean that can be leveraged for future crop improvement.

## Methods

### Plant materials

This study included 460 soybean PI accessions from 27 countries obtained from the USDA National Plant Germplasm System ([www.ars-grin.gov](http://www.ars-grin.gov), Additional file 4). Accessions were classified by maturity group with 31, 36, and 33% classified as maturity groups I, II, and III, respectively. A randomized complete block design was used with two replicates in 2014 and four replicates in 2015. In 2014, three to four seeds of each genotype were hand planted in 0.3 m long plots with a 0.91 m plot-to-plot distance (alleyway) and 0.76 row to row spacing. In 2015, five seeds were hand planted in 0.3 m long hill plots with 0.61 m plot to plot distance (alleyway) and 0.76 m row to row spacing. Weeds that emerged after sowing were controlled by hand-weeding.

The PI accessions were also evaluated under hydroponic conditions at the Iowa State University Agronomy Department greenhouse in 2015 under 16-h photoperiods. Five seeds of each PI accession were germinated on germination paper for 7 days. Uniform seedlings of each accession were then transplanted into a hydroponic system containing 240 L of an Fe deficient medium supplemented with a daily nutrient solution as described by Chaney et al. [84]. The

experiment was set up as a randomized complete block design with two replicates. Soybean IDC checks Clark (IDC tolerant) and IsoClark (IDC susceptible) were included.

### Experimental field and soil testing

An experimental plot previously used for IDC studies was selected for the field experiments at the Bruner farm, Iowa State University. Soil sampling was conducted using a soil probe (JMC Soil Samplers, Newton, IA) and each sample was analyzed at the Soil and Plant Analysis Laboratory at Iowa State University. The soil parameters measured were pH (1:1, H<sub>2</sub>O: soil) [85], calcium carbonate content [86], and Fe content [87].

### Iron deficiency chlorosis evaluation

In the field, IDC symptoms were rated on a visual scale of 1 (no chlorosis) to 5 (severe chlorosis, stunting, and necrosis) at T1 (representing V2 to V3), T2 (representing V5 to V6), and T3 (representing R1, approximately 2 weeks after T2) (Additional file 1: Figure S1a) [34, 88]. Chlorophyll concentration was measured using a Soil Plant Analysis Development (SPAD) meter in the field at T1 and T2. In the hydroponic system, visual phenotypes were recorded at T1, T2 and T3 representing the V1, V2 and V3 trifoliolate stages, respectively. SPAD measurements were taken at V1 and V2.

### Genotyping and quality control

All 460 PI accessions were genotyped using the Illumina Infinium SoySNP50K BeadChip as described in previous

studies [89], and data was obtained from SoyBase (<https://soybase.org/snps/>) which contains 42,506 high confidence SNPs. BEAGLE genetic analysis software (version 3.3.1, [90]) was used with the default settings to impute missing data. Markers missing at a frequency greater than 10% were removed from further analyses. After imputation, SNPs with a minor allele frequency less than 5% were removed from the data set. A total of 36,139 SNPs was used for GWA and GWE analyses.

### Linkage disequilibrium

The LD between markers was calculated as the squared allelic frequency correlation coefficient ( $r^2$ ) using the R package *synbreed* [91]. The  $r^2$  value was calculated independently for euchromatic and heterochromatic regions because of significant differences in the recombination rates between the two regions [92]. The physical lengths of euchromatin and heterochromatin for each chromosome were determined from SoyBase (<https://soybase.org/SequenceIntro.php>, version Williams82.a2. v1, [24]). The  $r^2$  values for SNPs with a pairwise distance less than 10 Mbp in either euchromatic or heterochromatic regions were plotted on a LD decay graph [93] using R [94]. The rate for the LD decay was determined as the chromosomal distance at the point where the average  $r^2$  dropped by half [95].

### GWAS and GWES

The IDC phenotypic data for each PI accession at each time point were analyzed using the best linear unbiased prediction (BLUP) and the R package *lme4* [96] to reduce the effects of environmental variation. The mixed linear model accounting for familial relationship was fitted for each time point by using the genome association and prediction integrated tool (GAPIT) R package [97, 98]. Using the Bayesian information criterion test of model fitness, no population structure was detected likely due to the use of genetically diverse core collection genotypes in this study (Additional file 12, [99, 100]). The empirical significance level  $P < 0.001$ , determined by 1000 permutations, was used as the threshold for SNPs-trait associations [92]. Considering the low decay rate of the genome-wide LD in soybean [92], we used LD  $r^2 > 0.5$  with the most significant SNP to cluster nearby SNPs to form the QTL, and the most significant SNP was picked to represent each locus, which resulted in multiple loci at the previously reported major-effect locus on Gm03. The genome-wide epistatic interactions between SNP pairs were analyzed using the software PLINK version 1.07 [101, 102]. To correct the multiple comparisons of SNPs, a Bonferroni threshold of  $\alpha = 0.05$  was used [103].

### Prediction of candidate genes

To identify the GWAS and GWES genomic intervals containing candidate genes, the two closest non-significant SNPs on either side of a significant SNP were identified (Additional file 5). The non-significant SNPs were then queried against the SoyBase genome browser ([www.soybase.org/gb2/gbrowse/gmax2.0/](http://www.soybase.org/gb2/gbrowse/gmax2.0/)) to determine their position in the genome and to identify all genes located between SNPs. Overlapping SNP intervals were combined into a single genomic region. The identified genes were then annotated using the SoyBase annotation tool ([www.soybase.org/genomeannotation/](http://www.soybase.org/genomeannotation/)), which provided the best Arabidopsis homolog (The Arabidopsis Information Resource version 10, [www.arabidopsis.org](http://www.arabidopsis.org)). Literature searches for Arabidopsis homologs were used to identify genes with functions related to Fe deficiency and abiotic stress tolerance. The soybean genes located within a SNP interval were also queried against other soybean Fe deficiency genes reported in previous publications, to link candidate genes with genes differentially expressed in response to IDC and previously reported IDC QTL [15, 26–28, 30, 31, 36, 38, 61]. When necessary, the SoyBase Gene Model Correspondence Lookup (<https://www.soybase.org/correspondence/>) was used to compare gene expression data from different genome assemblies. To examine potential candidate gene interactions, the Arabidopsis homologs of candidate genes from all Gm03 GWAS and all GWES were examined using StringDB version 10.5 [28].

### Statistical analysis

The model for IDC visual scores and SPAD measurement data collected at each time point was  $y_{ijk} = \mu + g_i + b_j + e_{ijk}$ , where  $\mu$  is the total mean,  $g_i$  is the genetic effect of the  $i$ th genotype,  $b_j$  is the block effect, and  $e_{ijk}$  is the residual effect including random error and possible interaction between genotype and block. Broad-sense heritability estimates for IDC were calculated on an entry mean basis using the equation,  $H^2 = \sigma_g^2 / [\sigma_g^2 + (\sigma_{g \times y}^2 / y) + (\sigma_e^2 / ry)]$ , where  $\sigma_g^2$  = genotypic variance,  $\sigma_{g \times y}^2$  = genotype by year interaction,  $y$  = number of years, and  $r$  = number of replicates [21]. Estimation of variance components was computed using SAS version 9.3 (SAS Institute Inc., Cary, NC).

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-020-2237-5>.

**Additional file 1: Figure S1.** (a and b). Field chlorosis ratings and correlations across treatments and time.

**Additional file 2: Figure S2.** Manhattan plots of IDC GWAS for each time point.

**Additional file 3: Figure S3.** Venn diagrams illustrating the distribution of significant SNPs associated with IDC detected at different time points.

**Additional file 4.** Soybean Plant Introduction (PI) names, maturity group, and IDC scores for lines utilized in GWAS and GWES analyses.

**Additional file 5.** Identification of genomic regions containing SNPs significantly associated with IDC.

**Additional file 6.** Statistics of the peak SNPs of the QTL associated with soybean IDC tolerance across environments.

**Additional file 7.** Characterization of candidate genes in GWAS QTL.

**Additional file 8.** Characterization of candidate genes in the historical IDC QTL on soybean Gm03.

**Additional file 9.** Significantly associated epistatic (SNP-SNP) interactions for iron deficiency chlorosis.

**Additional file 10.** Identification of genomic regions containing epistatic SNPs (GWES) associated with IDC.

**Additional file 11.** Characterization of candidate genes in GWES regions.

**Additional file 12.** Bayesian Information Criterion values of mixed linear model with principal components (PCs) applied for association analysis of iron deficiency chlorosis in soybean population.

#### Abbreviations

BLUP: Best linear unbiased prediction; GWAS: Genome-wide association study; GWE: Genome wide epistasis; GWES: Genome wide epistatic study; IDC: Iron deficiency chlorosis; LD: Linkage disequilibrium; QTL: Quantitative trait loci; SNP: Single nucleotide polymorphism

#### Acknowledgements

We sincerely acknowledge research assistance provided by Jae Brungardt, Brian Scott, and members of Singh soybean group at Iowa State University. We thank Jennifer Hicks for reviewing the manuscript.

#### Authors' contributions

A.S. and A.K.S. designed research; T.A., A.S. and A.K.S. performed research; T. A., J.Z., A.S., J.A.O., M.A.G. and A.K.S. analyzed data; T.A., J.Z., A.N.M.L., R.V.C.R., A.S., J.A.O., M.A.G. and A.K.S. wrote manuscript. All authors read and approved the final version of the manuscript.

#### Funding

This research was supported in part by the Iowa Soybean Association (AKS and AS), Monsanto Chair in Soybean Breeding (AKS), R F Baker Center for Plant Breeding (AKS), the North Central Soybean Research Program (MAG and JAO), United States Department of Agriculture CRIS IOW04403 and grant 2017-67021-25965 (AS and AKS), and the United States Department of Agriculture, Agricultural Research Service Project 3625-21220-005-00D and 5030-21220-006-00D (MAG and JAO). The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. The funding agencies had no role in the study design, data collection, analysis, interpretation of data, writing the manuscript, and the decision to publish.

#### Availability of data and materials

The datasets supporting the conclusions of the present study are included within this article (and its additional files). Data not included in the supplemental files can be obtained by request to the corresponding author. Seed used in experiments can be obtained from the USDA National Plant Germplasm System.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests

#### Author details

<sup>1</sup>Department of Agronomy, Iowa State University, Ames, IA, USA. <sup>2</sup>United States Department of Agriculture, Agricultural Research Service, Corn Insects and Crop Genetics Research Unit and Department of Agronomy, Iowa State University, Ames, IA, USA.

Received: 18 August 2019 Accepted: 3 January 2020

Published online: 28 January 2020

#### References

- Zheng SJ. Iron homeostasis and iron acquisition in plants: maintenance, functions and consequences. *Ann Bot.* 2010;105(5):799–800.
- George E, Horst W, Neumann E. Adaptation of plants to adverse chemical soil conditions. In: Marschner H, editor. Mineral nutrition of higher plants. San Diego: Academic Press Limited; 1995. p. 409–55.
- Froehlich D, Fehr W. Agronomic performance of soybeans with differing levels of iron deficiency chlorosis on calcareous soil. *Crop Sci.* 1981;21(3):438–41.
- Bauer P, Ling HQ, Guerinot ML. FIT, the FER-like iron deficiency induced transcription factor in Arabidopsis. *Plant Physiol Biochem.* 2007;45(5):260–1.
- Long TA, Tsukagoshi H, Busch W, Lahner B, Salt DE, Benfey PN. The bHLH transcription factor POPEYE regulates response to iron deficiency in Arabidopsis roots. *Plant Cell.* 2010;22(7):2219–36.
- Yan JY, Li CX, Sun L, Ren JY, Li GX, Ding ZJ, et al. A WRKY transcription factor regulates Fe translocation under Fe deficiency in Arabidopsis. *Plant Physiol.* 2016;171(3):2017–20.
- Henriques R, Jásik J, Klein M, Martinoia E, Feller U, Schell J, et al. Knock-out of Arabidopsis metal transporter gene IRT1 results in iron deficiency accompanied by cell differentiation defects. *Plant Mol Biol.* 2002;50(4–5):587–97.
- Rogers EE, Guerinot ML. FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in Arabidopsis. *Plant Cell.* 2002;14(8):1787–99.
- Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat JF, et al. IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell.* 2002;14(6):1223–33.
- Hymowitz T. Speciation and cytogenetics. Soybeans: Improvement, Production and Uses. In: Boerma HR, Specht JE, editors. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. 3rd ed. Madison; 2004. p. 97–129.
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet.* 2010;42(11):961–7.
- Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, et al. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet.* 2011;43(2):163–8.
- Mamidi S, Lee RK, Goos JR, McClean PE. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PLoS One.* 2014;9(9):e107469.
- Maccaferri M, Zhang J, Bulli P, Abate Z, Chao S, Cantu D, et al. A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. tritici) in a worldwide collection of hexaploid spring wheat (*Triticum aestivum* L.). *G3 Genes|Genomes|Genetics.* 2015; 5:449–65.
- Zhang J, Naik HS, Assefa T, Sarkar S, Reddy RVC, Singh A, et al. Computer vision and machine learning for robust phenotyping in genome-wide studies. *Sci Rep.* 2017;7:44048.
- Frankel OH. Genetic perspectives of germplasm conservation. Genetic manipulation: impact on man and society. Edited by: Arber, W, Illmensee K, PeacockWJ, Starlinger P. Cambridge: Cambridge University Press; 1984.161–170.
- Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci U S A.* 2006;103(45):16666–71.
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010;11(6):446–50.
- Ordovas JM, Robertson R, Cléirigh EN. Gene–gene and gene–environment interactions defining lipid-related traits. *Curr Opin Lipidol.* 2011;22(2):129–36.

20. Moellers TC, Singh A, Zhang J, Brungardt J, Kabbage M, Mueller DS, et al. Main and epistatic loci studies in soybean for *Sclerotinia sclerotiorum* resistance reveal multiple modes of resistance in multi-environments. *Sci Rep.* 2017;7(1):3554.
21. Cheverud JM, Routman EJ. Epistasis and its contribution to genetic variance components. *Genetics.* 1995;139(3):1455–61.
22. Hu X, Liu Q, Zhang Z, Li Z, Wang S, He L, et al. SHEsisEpi, a GPU-enhanced genome-wide SNP-SNP interaction scanning algorithm, efficiently reveals the risk genetic epistasis in bipolar disorder. *Cell Res.* 2010;20(7):854–7.
23. Hallauer AR, Carena MJ, Miranda Filho JB. *Quantitative Genetics in Maize Breeding.* 3rd ed. New York: Springer; 2010.
24. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. *Nature.* 2010;463(7278):178–83.
25. Moran Lauter AN, Rutter L, Cook D, O'Rourke JA, Graham MA. Examining short-term responses to a long-term problem: RNA-seq analyses of soybean responses to iron deficiency chlorosis. *Front Plant Sci Submitt.* 2019.
26. Moran Lauter AN, Peiffer GA, Yin T, Whitham SA, Cook D, Shoemaker RC, et al. Identification of candidate genes involved in early iron deficiency chlorosis signaling in soybean (*Glycine max*) roots and leaves. *BMC Genomics.* 2014;15:702.
27. Atwood SE, O'Rourke JA, Peiffer GA, Yin T, Majumder M, Zhang C, et al. Replication protein A subunit 3 and the iron efficiency response in soybean. *Plant Cell Environ.* 2014;37(1):213–34.
28. Lin S, Cianzio S, Shoemaker R. Mapping genetic loci for iron deficiency chlorosis in soybean. *Mol Breed.* 1997;3(3):219–29.
29. Lin SF, Grant D, Cianzio S, Shoemaker R. Molecular characterization of iron deficiency chlorosis in soybean. *J Plant Nutr.* 2000;23(11–12):1929–39.
30. Severin AJ, Peiffer GA, Xu WW, Hyten DL, Bucciarelli B, O'Rourke JA, et al. An integrative approach to genomic introgression mapping. *Plant Physiol.* 2010;154(1):3–12.
31. Peiffer GA, King KE, Severin AJ, May GD, Cianzio SR, Lin SF, et al. Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. *Plant Physiol.* 2012;158(4):1745–54.
32. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447–52.
33. Duncan O, van der Merwe MJ, Daley DO, Whelan J. The outer mitochondrial membrane in higher plants. *Trends Plant Sci.* 2013;18:207–17.
34. Cianzio SR, Fehr WR, Anderson I. Genotypic evaluation for iron deficiency chlorosis in soybeans by visual scores and chlorophyll concentration. *Crop Sci.* 1979;19(5):644–6.
35. Cianzio S, Fehr W. Variation in the inheritance of resistance to iron deficiency chlorosis in soybeans. *Crop Sci.* 1982;22(2):433–4.
36. Lin SF, Baumer JS, Ivers D, Cianzio SR, Shoemaker RC. Field and nutrient solution tests measure similar mechanisms controlling iron deficiency chlorosis in soybean. *Crop Sci.* 1998;38(1):254–9.
37. Yuan Y, Wu H, Wang N, Li J, Zhao W, Du J, et al. FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in *Arabidopsis*. *Cell Res.* 2008;18(3):385–97.
38. O'Rourke JA, Nelson RT, Grant D, Schmutz J, Grimwood J, Cannon S, et al. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics.* 2009;10:376.
39. Yu LX, Setter TL. Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. *Plant Physiol.* 2003;131(2):568–82.
40. Oshino T, Abiko M, Saito R, Ichiishi E, Endo M, Kawagishi-Kobayashi M, et al. Premature progression of anther early developmental programs accompanied by comprehensive alterations in transcription during high-temperature injury in barley. *Mol Gen Genomics.* 2007;278(1):31–42.
41. Panda BB, Achary VMM. Mitogen-activated protein kinase signal transduction and DNA repair network are involved in aluminum-induced DNA damage and adaptive response in root cells of *Allium cepa* L. *Front Plant Sci.* 2014;5:256.
42. Frey FP, Urbany C, Hüttel B, Reinhardt R, Stich B. Genome-wide expression profiling and phenotypic evaluation of European maize inbreds at seedling stage in response to heat stress. *BMC Genomics.* 2015;16:123.
43. Hu Z, Cools T, De Veylder L. Mechanisms used by plants to cope with DNA damage. *Annu Rev Plant Biol.* 2016;67:439–62.
44. Stein RJ, Waters BM. Use of natural variation reveals core genes in the transcriptome of iron-deficient *Arabidopsis thaliana* roots. *J Exp Bot.* 2012;63:1039–55.
45. Waters BM, McInturf SA, Stein RJ. Rosette iron deficiency transcript and microRNA profiling reveals links between copper and iron homeostasis in *Arabidopsis thaliana*. *J Exp Bot.* 2012;63:5903–18.
46. Zander M, La Camera S, Lamotte O, Métraux JP, Gatz C. *Arabidopsis thaliana* class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. *Plant J.* 2010;61(2):200–10.
47. Brumbarova T, Bauer P, Ivanov R. Molecular mechanisms governing *Arabidopsis* iron uptake. *Trends Plant Sci.* 2015;20(2):124–33.
48. Shen C, Yang Y, Liu K, Zhang L, Guo H, Sun T, et al. Involvement of endogenous salicylic acid in iron-deficiency responses in *Arabidopsis*. *J Exp Bot.* 2016;67(14):4179–93.
49. Kissen R, Winge P, Tran DHT, Jørstad TS, Størseth TR, Christensen T, et al. Transcriptional profiling of an Fd-GOGAT1/GLU1 mutant in *Arabidopsis thaliana* reveals a multiple stress response and extensive reprogramming of the transcriptome. *BMC Genomics.* 2010;11:190.
50. Chen X, Han H, Jiang P, Nie L, Bao H, Fan P, et al. Transformation of  $\beta$ -lycopene cyclase genes from *Salicornia europaea* and *Arabidopsis* conferred salt tolerance in *Arabidopsis* and tobacco. *Plant Cell Physiol.* 2011;52(5):909–21.
51. Scott I, Tobin AK, Logan DC. BIGYIN, an orthologue of human and yeast FIS1 genes functions in the control of mitochondrial size and number in *Arabidopsis thaliana*. *J Exp Bot.* 2006;57(6):1275–80.
52. Chen OS, Hemenway S, Kaplan J. Inhibition of Fe-S cluster biosynthesis decreases mitochondrial iron export: evidence that Yfh1p affects Fe-S cluster synthesis. *Proc Natl Acad Sci U S A.* 2002;99(19):12321–6.
53. Werner T, Schmölling T. Cytokinin action in plant development. *Curr Opin Plant Biol.* 2009;12(5):527–38.
54. Séguéla M, Briat JF, Vert G, Curie C. Cytokinins negatively regulate the root iron uptake machinery in *Arabidopsis* through a growth-dependent pathway. *Plant J.* 2008;55(2):289–300.
55. Hong S, Kim SA, Guerinot ML, McClung CR. Reciprocal interaction of the circadian clock with the iron homeostasis network in *Arabidopsis*. *Plant Physiol.* 2013;161(2):893–903.
56. Fan J, Yu L, Xu C. A central role for triacylglycerol in membrane lipid breakdown, fatty acid  $\beta$ -oxidation, and plant survival under extended darkness. *Plant Physiol.* 2017;174:1517–30.
57. Okazaki Y, Otsuki H, Narisawa T, Kobayashi M, Sawai S, Kamide Y, et al. A new class of plant lipid is essential for protection against phosphorus depletion. *Nat Commun.* 2013;4:1510.
58. Wold MS. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Plant Biol.* 1997;66(1):61–92.
59. Rowan BA, Oldenburg DJ, Bendich AJ. RecA maintains the integrity of chloroplast DNA molecules in *Arabidopsis*. *J Exp Bot.* 2010;61(10):2575–88.
60. Diers BW, Cianzio SR, Shoemaker RC. Possible identification of quantitative trait loci affecting iron efficiency in soybean. *J Plant Nutr.* 1992;15(10):2127–36.
61. Charlson DV, Bailey TB, Cianzio SR, Shoemaker RC. Molecular marker Satt481 is associated with iron-deficiency chlorosis resistance in a soybean breeding population. *Crop Sci.* 2005;45(6):2394–9.
62. Mühlhoff U, Molik S, Godoy JR, Uzarska MA, Richter N, Seubert A, et al. Cytosolic monothiol glutaredoxins function in intracellular iron sensing and trafficking via their bound iron-sulfur cluster. *Cell Metab.* 2010;12(4):373–85.
63. Blanvillain R, Kim JH, Wu S, Lima A, Ow DW. OXIDATIVE STRESS 3 is a chromatin-associated factor involved in tolerance to heavy metals and oxidative stress. *Plant J.* 2009;57(4):654–65.
64. Zientara K, Wawrzyńska A, Łukomska J, López-Moya JR, Liszewska F, Assunção AG, et al. Activity of the AtMRP3 promoter in transgenic *Arabidopsis thaliana* and *Nicotiana tabacum* plants is increased by cadmium, nickel, arsenic, cobalt and lead but not by zinc and iron. *J Biotechnol.* 2009;139(3):258–63.
65. Larkindale J, Vierling E. Core genome responses involved in acclimation to high temperature. *Plant Physiol.* 2008;146(2):748–61.
66. Mastrangelo AM, Belloni S, Barilli S, Ruperti B, Di Fonzo N, Stanca AM, et al. Low temperature promotes intron retention in two e-cor genes of durum wheat. *Planta.* 2005;221(5):705–15.
67. Pacín M, Semmoloni M, Legris M, Finlayson SA, Casal JJ. Convergence of CONSTITUTIVE PHOTOMORPHOGENESIS 1 and PHYTOCHROME INTERACTING FACTOR signalling during shade avoidance. *New Phytol.* 2016;211(3):967–79.
68. Renault H, El Amrani A, Berger A, Mouille G, SOUBIGOU-TACONNAT L, Bouchereau A, et al.  $\gamma$ -Aminobutyric acid transaminase deficiency impairs central carbon metabolism and leads to cell wall defects during salt stress in *Arabidopsis* roots. *Plant Cell Environ.* 2013;36(5):1009–18.

69. Hewitt MM, Carr JM, Williamson CL, Slocum RD. Effects of phosphate limitation on expression of genes involved in pyrimidine synthesis and salvaging in Arabidopsis. *Plant Physiol Biochem* 4. 2005;3(2):91–9.
70. Sauerbrunn N, Schlaich NL. PCC1: a merging point for pathogen defence and circadian signalling in Arabidopsis. *Planta*. 2004;218(4):552–61.
71. Xia Y, Suzuki H, Borevitz J, Blount J, Guo Z, Patel K, et al. An extracellular aspartic protease functions in Arabidopsis disease resistance signaling. *EMBO J*. 2004;23(4):980–8.
72. Eulgem T, Tsuchiya T, Wang XJ, Beasley B, Cuzick A, Tör M, et al. EDM2 is required for RPP7-dependent disease resistance in Arabidopsis and affects RPP7 transcript levels. *Plant J*. 2007;49(5):829–39.
73. Boruc J, Mylle E, Duda M, De Clercq R, Rombauts S, Geelen D, et al. Systematic localization of the Arabidopsis core cell cycle proteins reveals novel cell division complexes. *Plant Physiol*. 2010;152(2):553–65.
74. Sterken R, Kiekens R, Boruc J, Zhang F, Vercauteren A, Vercauteren I, et al. Combined linkage and association mapping reveals CYCD5; 1 as a quantitative trait gene for endoreduplication in Arabidopsis. *Proc Natl Acad Sci U S A*. 2012;109(12):4678–83.
75. Blanco-Herrera F, Moreno AA, Tapia R, Reyes F, Araya M, D'Alessio C, et al. The UDP-glucose: glycoprotein glucosyltransferase (UGGT), a key enzyme in ER quality control, plays a significant role in plant growth as well as biotic and abiotic stress in *Arabidopsis thaliana*. *BMC Plant Biol*. 2015;15.
76. Lin XY, Ye YQ, Fan SK, Jin CW, Zheng SJ. Increased sucrose accumulation regulates iron-deficiency responses by promoting auxin signaling in Arabidopsis plants. *Plant Physiol*. 2016;170(2):907–20.
77. Pospíšil P. Production of reactive oxygen species by photosystem II as a response to light and temperature stress. *Front Plant Sci*. 2016;7:1950.
78. Zhao Y, Pan Z, Zhang Y, Qu X, Zhang Y, Yang Y, et al. The actin-related protein2/3 complex regulates mitochondrial-associated calcium signaling during salt stress in Arabidopsis. *Plant Cell*. 2013;25(11):4544–59.
79. Röhrig S, Dorn A, Enderle J, Schindele A, Herrmann NJ, Knoll A, et al. The RecQ-like helicase HRQ1 is involved in DNA crosslink repair in Arabidopsis in a common pathway with the Fanconi anemia-associated nuclease FAN1 and the postreplicative repair ATPase RAD5A. *New Phytol*. 2018;218:1478–90.
80. Deniz Akdemir, William Beavis, Roberto Fritsche-Neto, Asheesh K. Singh, Julio Isidro-Sánchez, (2019) Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity* 122 (5):672–683.
81. Sambuddha Ghosal, David Blystone, Asheesh K. Singh, Baskar Ganapathysubramanian, Arti Singh, Soumik Sarkar, (2018) An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences* 115 (18):4613–4618.
82. Tianshuang Gao, Hamid Emadi, Homagni Saha, Jiaoping Zhang, Alec Lofquist, Arti Singh, Baskar Ganapathysubramanian, Soumik Sarkar, Asheesh Singh, Sourabh Bhattacharya, (2018) A Novel Multirobot System for Plant Phenotyping. *Robotics* 7 (4):61.
83. Asheesh Kumar Singh, Baskar Ganapathysubramanian, Soumik Sarkar, Arti Singh, (2018) Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science* 23 (10):883–898.
84. Chaney RL, Coulombe BA, Bell PF, Angle JS. Detailed method to screen dicot cultivars for resistance to Fe-chlorosis using FeDTPA and bicarbonate in nutrient solutions. *J Plant Nutr*. 1992;15(10):2063–83.
85. Watson ME, Brown JR. pH and lime requirement. In Brown JR ed. *Recommended Chemical Soil Test Procedures for the North Central Region*. North Central Regional Research Publication. 1998:13–6.
86. Sherrod L, Dunn G, Peterson G, Kolberg R. Inorganic carbon analysis by modified pressure-calimeter method. *Soil Sci Soc Am J*. 2002;66(1):299–305.
87. Whitney DA. Micronutrients: zinc, iron, manganese and copper. In Brown JR ed. *Recommended Chemical Soil Test Procedures for the North Central Region*. North Central Regional Research Publication. 1998:1–4.
88. Fehr WR, Caviness CE. Stages of soybean development. Stages of soybean development. 1977. Ames: Iowa State University of Science and Technology, Iowa Agriculture and Home Economics Experiment Station. Special Report. 87.
89. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, et al. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One*. 2013;8(1).
90. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81(5):1084–97.
91. Wimmer V, Albrecht T, Aunger HJ, Schön CC. synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics*. 2012;28(15):2086–7.
92. Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, et al. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics*. 16:217.
93. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A*. 2001;98(20):11479–84.
94. R Core Team. R: a language and environment for statistical computing. Vienna; 2013.
95. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*. 2013;45(1):43–50.
96. Bates D, Maechler M, Bolker B. Linear mixed-effects models using S4 classes. R package version. 2005;0:98–1.
97. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*. 2010;42(4):355–60.
98. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012; 28(18):2397–9.
99. Coser SM, Chowda Reddy RV, Zhang J, Mueller DS, Mengistu A, Wise KA, et al. Genetic architecture of charcoal rot (*Macrophomia phaseolina*) resistance in soybean revealed using a diverse panel. *Front Plant Sci*. 2017;8: 1626.
100. Oliveira MF, Nelson RL, Geraldi IO, Cruz CD, de Toledo JFF. Establishing a soybean germplasm core collection. *Field Crop Res*. 2010;119(2–3):277–89.
101. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
102. Zhang J, Singh A, Mueller DS, Singh AK. Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *Plant J*. 84(6):1124–36.
103. Bonferroni C. Teoria statistica delle classi e calcolo delle probabilita. *Pubbl del R Ist Super di Sci Econ e Commerciali di Firenze*. 1936;8:3–62.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

