University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Agronomy & Horticulture -- Faculty Publications | Agronomy and Horticulture Department

8-9-2020

Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research

Qijian Song
USDA-ARS, Beltsville, MD, Qijian.Song@usda.gov

Long Yan
Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, China

Charles Quigley
USDA-ARS, Beltsville, MD, chuck.quigley@usda.gov

Edward Fickus
USDA-ARS, Beltsville, MD, edward.fickus@usda.gov

He Wei
Henan Academy of Agricultural Sciences, Zhengzhou

See next page for additional authors

Follow this and additional works at: https://digitalcommons.unl.edu/agronomyfacpub

 Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Botany Commons, Horticulture Commons, Other Plant Sciences Commons, and the Plant Biology Commons

Song, Qijian; Yan, Long; Quigley, Charles; Fickus, Edward; Wei, He; Chen, Linfeng; Dong, Faming; Araya, Susan; Liu, Jinlong; Hyten, David; Pantalone, Vincent R.; and Nelson, Randall L., "Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research" (2020). *Agronomy & Horticulture -- Faculty Publications*. 1413.
https://digitalcommons.unl.edu/agronomyfacpub/1413

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Qijian Song, Long Yan, Charles Quigley, Edward Fickus, He Wei, Linfeng Chen, Faming Dong, Susan Araya, Jinlong Liu, David Hyten, Vincent R. Pantalone, and Randall L. Nelson

This article is available at DigitalCommons@University of Nebraska - Lincoln: https://digitalcommons.unl.edu/agronomyfacpub/1413

the plant journal

The Plant Journal (2020) **104,** 800–811 doi: 10.1111/tpj.14960

RESOURCE

Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research

Qijian Song[1,*], Long Yan[2], Charles Quigley[1], Edward Fickus[1], He Wei[3], Linfeng Chen[1], Faming Dong[1], Susan Araya[1], Jinlong Liu[1], David Hyten[4], Vincent Pantalone[5] and Randall L. Nelson[6]

[1]*Soybean Genomics and Improvement Lab., USDA-ARS, Beltsville, MD, USA,*

[2]*Shijiazhuang Branch Center of National Center for Soybean Improvement/the Key Laboratory of Crop Genetics and Breeding, Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, China,*

[3]*Institute of Industrial Crops, Henan Academy of Agricultural Sciences, Zhengzhou, Henan Province, China,*

[4]*Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA,*

[5]*Department of Plant Sciences, University of Tennessee, Knoxville, TN, USA, and*

[6]*Soybean/Maize Germplasm, Pathology and Genetics Research Unit and Department of Crop Sciences, USDA-ARS, University of Illinois, Urbana, IL, USA*

Received 10 April 2020; accepted 30 July 2020; published online 9 August 2020.
*For correspondence (e-mail Qijian.Song@usda.gov).

SUMMARY

The limited number of recombinant events in recombinant inbred lines suggests that for a biparental population with a limited number of recombinant inbred lines, it is unnecessary to genotype the lines with many markers. For genomic prediction and selection, previous studies have demonstrated that only 1000–2000 genome-wide common markers across all lines/accessions are needed to reach maximum efficiency of genomic prediction in populations. Evaluation of too many markers will not only increase the cost but also generate redundant information. We developed a soybean (*Glycine max*) assay, BARCSoySNP6K, containing 6000 markers, which were carefully chosen from the SoySNP50K assay based on their position in the soybean genome and haplotype block, polymorphism among accessions and genotyping quality. The assay includes 5000 single nucleotide polymorphisms (SNPs) from euchromatic and 1000 from heterochromatic regions. The percentage of SNPs with minor allele frequency >0.10 was 95% and 91% in the euchromatic and heterochromatic regions, respectively. Analysis of progeny from two large families genotyped with SoySNP50K versus BARCSoySNP6K showed that the position of the common markers and number of unique bins along linkage maps were consistent based on the SNPs genotyped with the two assays; however, the rate of redundant markers was dramatically reduced with the BARCSoySNP6K. The BARCSoySNP6K assay is proven as an excellent tool for detecting quantitative trait loci, genomic selection and assessing genetic relationships. The assay is commercialized by Illumina Inc. and being used by soybean breeders and geneticists and the list of SNPs in the assay is an ideal resource for SNP genotyping by targeted amplicon sequencing.

Keywords: soybean (*Glycine max*), single nucleotide polymorphisms, SNP assay, BARCSoySNP6K beadchips, haplotype block, QTL mapping, genomic selection, genomic prediction, breeding selection.

INTRODUCTION

Molecular markers are widely used for the purposes of quantitative trait locus (QTL) mapping (Wen *et al.*, 2014; Bandillo *et al.*, 2015; Zhang *et al.*, 2015; Diers *et al.*, 2018), map-based cloning (Watanabe *et al.*, 2011; Philippe *et al.*, 2013), estimation of genetic diversity (Li *et al.*, 2010; Van Inghelandt *et al.*, 2010), construction of genetic linkage maps (Harushima *et al.*, 1998; Song *et al.*, 2004; Song *et al.*, 2005; Choi *et al.*, 2007; Hyten *et al.*, 2010b; Song *et al.*, 2016) and genomic prediction (Chang *et al.*, 2016a; Jarquín *et al.*, 2019). Single nucleotide polymorphisms (SNPs) are the most abundant form of DNA polymorphism

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

in eukaryotic genomes (Kruglyak, 1997; Collins *et al.*, 1998) and are suitable for high-throughput genotyping (Yoon *et al.*, 2007; Barreiro *et al.*, 2009; Ding and Jin, 2009; Lin *et al.*, 2009). Thus, SNPs were embraced as an excellent source of genetic markers in soybean. Zhu *et al.* (2003) and Choi *et al.* (2007) successfully discovered over 5500 SNPs in more than 2000 genes or gene transcripts by polymerase chain reaction amplification of genic regions and sequencing of the resulting amplicons. In recent years, a large number of sequence variants in approximately 2000 soybean genomes were efficiently identified from DNA sequences generated with the next-generation sequencing technology (Kim *et al.*, 2010; Lam *et al.*, 2010; Hyten *et al.*, 2010a; Li *et al.*, 2013; Song *et al.*, 2013; Zhou *et al.*, 2015; Valliyodan *et al.*, 2016). The Infinium Beadchip assay and genotyping by sequencing (GBS) are the two approaches being commonly used for high-throughput SNP genotyping in soybean. GBS can quickly generate millions of reads for variant discovery. The approach was further developed by Elshire *et al.* (2011) who used methylation-sensitive restriction enzymes to digest genomic DNA of the parents individually and recombinant inbred lines (RILs) of a mapping population. Adapters were then ligated to the resulting restriction fragments. The resulting DNA sequence reads from high-throughput sequencer were then aligned to the reference sequence to detect the SNP allele present at thousands of loci in the population as well as in the parents. The GBS is being used for high-density genetic map construction and QTL mapping in soybean (Huang *et al.*, 2009). The Infinium Beadchip assay allows the assay of a large number of SNPs per DNA sample in parallel on a single silicon slide (http://www.illumina.com/). Song *et al.* (2013) developed a SoySNP50K Beadchip containing >52 000 SNPs that were selected to equalize the distance between selected SNPs in the euchromatic and heterochromatic regions, increase assay success rate and minimize the number of SNPs with low minor allele frequency (MAF). The SoySNP50K Beadchip was successfully used to analyze 18 489 annual *Glycine max* and 1160 *Glycine soja* accessions in the USDA Soybean Germplasm Collection and a number of RIL populations (Song *et al.*, 2015; Song *et al.*, 2016). Analysis of cultivated soybean including landrace and elite accessions with the SNPs showed extensive linkage disequilibrium (LD) and large haplotype blocks in the soybean genome (Song *et al.*, 2015). The high LD and haplotype blocks in the soybean genome greatly facilitated marker–trait association discovery in soybean and led to the identification of candidate genes/QTL controlling a range of traits based on the SoySNP50K dataset (Hwang *et al.*, 2014; Vaughn *et al.*, 2014; Wen *et al.*, 2014; Zeng *et al.*, 2014; Bandillo *et al.*, 2015; Dhanapal *et al.*, 2015; Ray *et al.*, 2015; Wen *et al.*, 2015; Zhang *et al.*, 2015; Chang *et al.*, 2016b; Hartman and Chang, 2017; Leamy *et al.*, 2017; Zhang *et al.*, 2016; Zeng *et al.*, 2017). Because soybean haplotype blocks contained 41%–48% of the genomic sequences in the euchromatic and >90% in heterochromatic regions of the soybean genome (Schmutz *et al.*, 2010; Song *et al.*, 2016), it is anticipated that many SNPs in the same haplotype blocks are most likely to generate identical segregation pattern particularly among RILs of biparental populations or among accessions sharing the same geographic origins.

It is well known that the rate of recombination in crop genomes is low. For example, in a soybean nested association mapping population consisting of 40 diverse families with 5600 RILs, the average number of recombination events (REs) was approximately 58 per RIL; however, 70% of the REs occurred in at least two RILs within a family and only 30% of the REs (approximately 18 REs per line) were unique to each RIL in a given family (Song *et al.*, 2017). In maize, the average number of REs in 25 nested association mapping families with 4699 RILs was 29 (Kump *et al.*, 2011). The limited number of REs suggests that for a biparental population with a limited number of RILs, it is unnecessary to genotype the RILs with too many markers, e.g., in a soybean population with 200 RILs and each RIL with 18 unique REs, a total of 3600 markers will saturate the linkage map for QTL mapping and enough to tag each recombinant in the population.

Genomic prediction and genomic selection are methods to predict plant phenotypes rapidly using genome-wide marker information and select lines based on predicted breeding values instead of phenotypes. Genomic selection has great potential to accelerate plant breeding. The application of genomic selection to breeding programs requires the lines or germplasm to be genotyped with the same set of markers as those in the prediction models. Previous studies have demonstrated that only 1000–2000 genome-wide markers assayed across all lines/accessions were needed to reach maximum efficiency of genomic prediction in the populations (Poland *et al.*, 2012; Bao *et al.*, 2014; Zhang *et al.*, 2016). Increasing markers will not improve prediction efficiency. Likewise, other applications including determination of variety difference for plant variety protection and analysis of pedigree among germplasm do not require many markers. A simple, cheap, quick and accurate genotyping tool will facilitate those applications.

Thus, our objective was to develop an efficient SNP assay with a set of 6000 core SNPs for soybean genetic and breeding research. These SNPs were selected from the SoySNP50K assay based on the SNP genotyping quality, SNP position in genomic regions and haplotype blocks, and SNP MAF among 18 489 *G. max* accessions. The assay covered genome-wide large haplotype blocks defined by SoySNP50K but reduced the number of SNPs that might provide redundant genotypic information or might be monomorphic in RIL populations and germplasms from a similar genetic background.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

RESULTS

SNP number and haplotype block size in the euchromatic and heterochromatic regions of *Glycine max* genome

Of the 42 509 SNPs in the SoySNP50K dataset, in total, 35 483 SNPs including 29 737 SNPs in the euchromatic and 5746 in the heterochromatic regions of the 20 chromosomes remained after SNPs with MAF <5% in the *G. max* population and SNPs with poor allele clustering were eliminated. Analyses identified 5139 haplotype blocks in euchromatic regions and 468 haplotype blocks in heterochromatic regions (Table S1). Among these, in total, 973 haplotype blocks in the euchromatic regions were >50 kb (Table 1) and 454 in the heterochromatic regions were >100 kb (Table 2). There were 10 489 and 4361 SNPs in these haplotype blocks in the euchromatic regions and the heterochromatic regions, respectively. In the euchromatic regions, the total sequence length of the haplotype blocks with size >50 kb was approximately 112 Mb and the total sequence length between adjacent blocks was 347 Mb. In the heterochromatic regions, the total sequence length of the haplotype blocks with size >100 kb was 281 Mb and the sequence length between adjacent blocks was 210 Mb.

Table 1 Distribution of haplotype block sizes in euchromatic regions of *G. max* genotypes

Haplotype block size (kb)	Number of haplotype blocks
<10	2088
≥10 and <20	1146
≥20 and <30	449
≥30 and <40	261
≥40 and <50	222
≥50 and <60	182
≥60 and <70	162
≥70 and <80	127
≥80 and <90	80
≥90 and <100	64
≥100 and <200	277
≥200 and <300	42
≥300 and <400	11
≥400 and <500	10
≥500	17
Total	973 (>50 kb)/5138

Table 2 Distribution of haplotype block sizes in heterochromatic regions of *G. max* genotypes

Block size (kb)	Number of haplotype blocks
<100	14
≥100 and <200	53
≥200 and <300	52
≥300 and <400	47
≥400 and <500	37
≥500 and <600	32
≥600 and <700	20
≥700 and <800	36
≥800 and <900	29
≥900	147
Total	454 (>100 kb)/468

Percentage of co-segregating SNPs from the large haplotype blocks in the two mapping populations

Of the 21 478 polymorphic SNPs mapped in the WP population, a total of 15 034 SNPs was in the haplotype blocks of >50 kb in the euchromatic region and >100 kb in the heterochromatic region, which included 13 837 SNPs in the 3193 haplotype blocks with ≥2 SNPs, and 1197 SNPs in the haplotype blocks with only one SNP. Of the 13 837 SNPs in the haplotype blocks with ≥2 SNPs, 8141 SNPs (58.9%) were co-segregating across RILs. In contrast, the percentage of co-segregating SNPs not in haplotype blocks was only 40.5% (2612 of 6444 SNPs).

In the EW population, in total, 9899 of the 10 753 SNPs were in the haplotype blocks, these include 9190 SNPs in the 1796 haplotype blocks with ≥2 SNPs and 709 SNPs in the haplotype block with only one SNP. Among the 9190 SNPs in the haplotype blocks with ≥2 SNPs, approximately 71.1% (6534 SNPs) were co-segregating. While the percentage of co-segregating SNPs not in the haplotype blocks was only 54.5% (852 of 1563 SNPs) in the EW population. The results suggested that elimination of SNPs in the same haplotype blocks was able to reduce the number of SNPs with identical segregation patterns in the RIL population.

Selection of SNPs in the euchromatic and heterochromatic regions

Because only one SNP was chosen from each large haplotype block, in total, 973 and 454 SNPs with the largest MAF among the SNPs in the same haplotype block were selected from euchromatic and heterochromatic regions, respectively. The remaining 4027 euchromatic and 546 heterochromatic SNPs were selected from the 24 994 SNPs that did not reside in large haplotype blocks using an in-house iteration algorithm (Table S2). The average MAF was 0.35 for selected SNPs versus 0.23 for preselected SNPs in the euchromatic regions, and 0.30 for selected SNPs versus 0.19 for preselected SNPs in the heterochromatic regions based on the 14 183 non-redundant *G. max* accessions. The percentage of SNPs with MAF >0.10 was 95% in the euchromatic and 91% in the heterochromatic regions (Figures 1 and 2), suggesting that a large proportion of the selected SNPs were anticipated to be highly polymorphic among randomly selected accessions or between biparent of *G. max*. As shown in Figure 3, the selected SNPs were well spread along each chromosome except for some regions with large haplotype blocks. The average density of SNPs in the regions not containing large haplotype blocks was 86 kb/SNP and 394 kb/SNP in

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

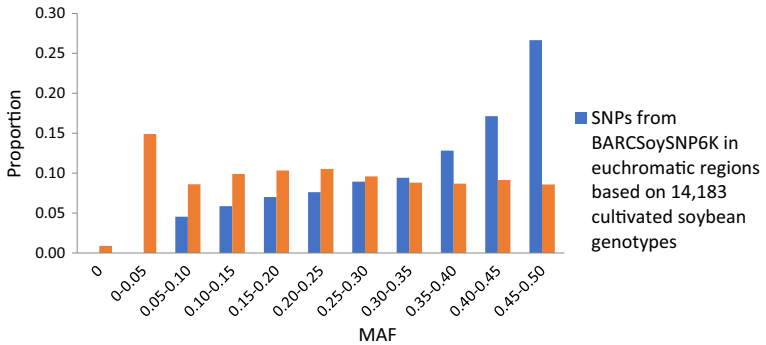

Figure 1. Distribution of minor allele frequency (MAF) between preselected and selected single nucleotide polymorphisms (SNPs) in euchromatic regions.

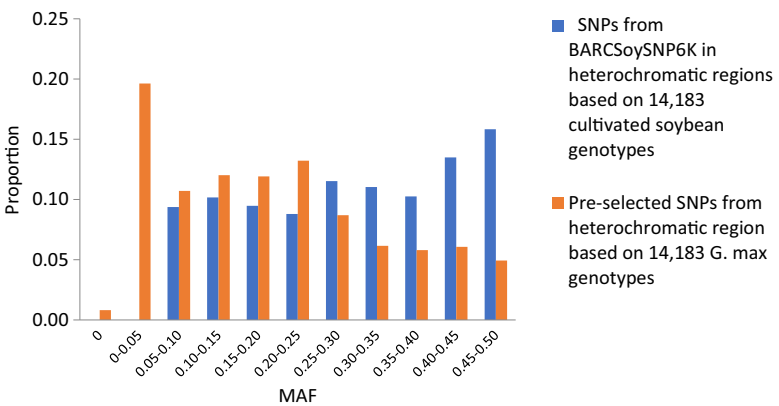

Figure 2. Distribution of minor allele frequency (MAF) between preselected and selected single nucleotide polymorphisms (SNPs) in heterochromatic regions.

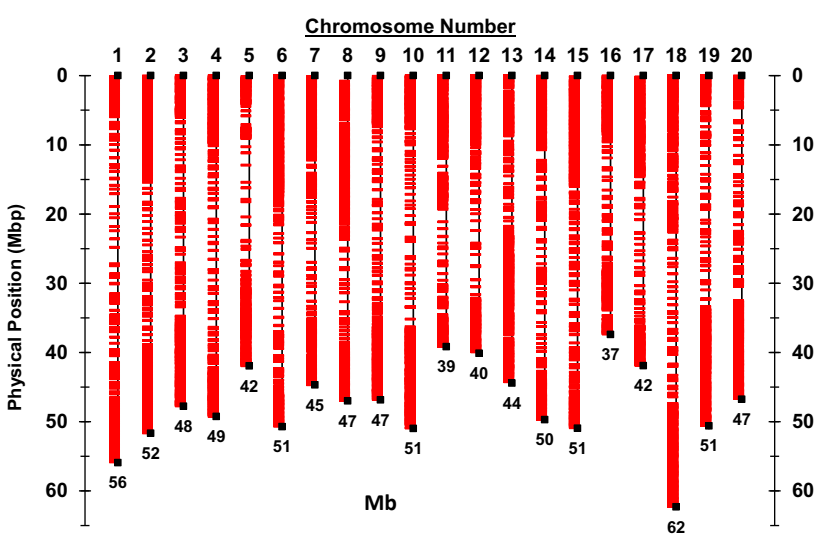

Figure 3. Physical position of 6000 single nucleotide polymorphisms in the BARCSoySNP6k assay.

the euchromatic and heterochromatic regions, respectively.

Comparing clusters of a diverse set of wild, landrace and elite soybeans based on the SNPs in BARCSoySNP6K and SoySNP50K assays

Clustering of the 96 wild, 96 landrace and 96 elite cultivars based on their SNPs in the BARCSoySNP6K assay showed that accessions from three different populations were grouped into different clusters (Figure 4). The mean pairwise genetic distance was 0.34, 0.40 and 0.28 among the 96 elites, 96 landraces and 96 wild soybeans, respectively, while the mean genetic distances were 0.29, 0.32 and 0.29, respectively when calculations were based on the 42 509 SNPs included in the SoySNP50K Beadchip. The selected set of SNPs had a high resolution to distinguish accessions

© 2020 The Authors.

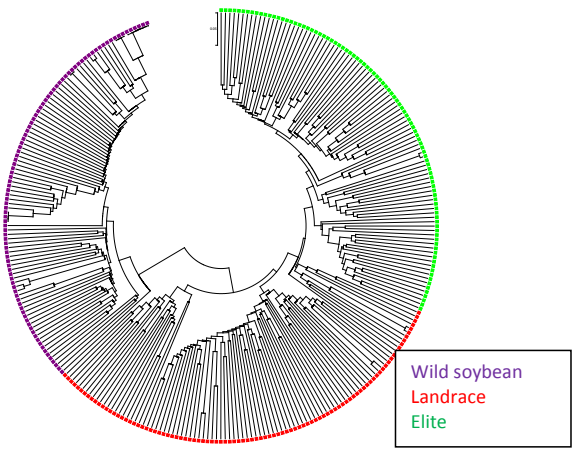

Figure 4. Cluster of 96 wild, 96 landrace and 96 elite cultivars based on alleles of single nucleotide polymorphisms included in the BARCSoySNP6K assay.

and a high proportion of the SNPs polymorphic in any pair of genotypes. Montel's correlation coefficient of the distance matrices among all accessions determined by BARCSoySNP6K versus SoySNP50K was significant (0.42 at $P < 0.0001$) (Figure 5).

Comparison of the linkage maps of SNPs genotyped with BARCSoySNP6K and SoySNP50K assays based on *G. max* x *G. max* and *G. max* x *G. soja* populations

In total, 3161 (53%) and 1621 (27%) of the 6000 SNPs in the BARCSoySNP6K assay were polymorphic in the WP and EW populations, respectively; only 9% of the SNPs in the WP and 14% in the EW showed identical segregation profiles among RILs. Only 20 linkage groups, which were corresponding to the 20 chromosomes, were created with a total map distance spanning 2344 cM in the WP and 2514 cM in the EW population, none of the SNPs were unlinked to one of the 20 linkage groups. The polymorphic SNPs were 21 478 (41%) in the WP and (10 753) 22% in EW population based on the SNPs from SoySNP50K

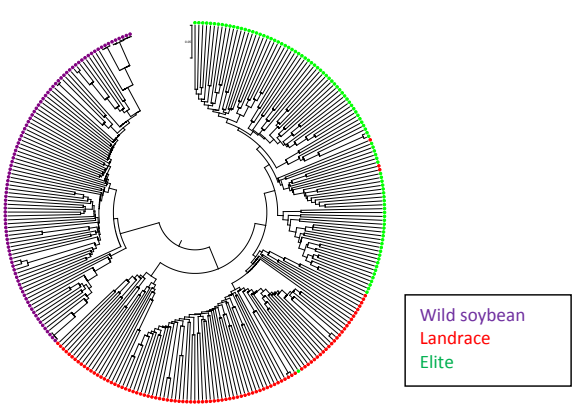

Figure 5. Cluster of 96 wild, 96 landrace and 96 elite cultivars based on alleles of single nucleotide polymorphisms included in the SoySNP50K assay.

containing 52 385 SNPs; however, 50% (10 753) and 62% (7386) of the polymorphic SNPs had the same segregation pattern among RILs in the WP and EW population, respectively. Analysis of the intervals between adjacent markers showed that 73% versus 98% of intervals in the WP population and 73% versus 96% of intervals in the EW population were <1 cM based on the SNPs in the BARCSoySNP6K versus SoySNP50K. The selection effectively excluded the SNPs with very small intervals (Figures 6 and 7). The total genetic distance of the linkage maps was 2445 cM in the WP and 2647 cM in the EW population based on the SNPs in SoySNP50K. These results showed that the percentage of SNPs with identical genetic profiles were dramatically reduced in the BARCSoySNP6K assay and the total length of genetic linkage maps based on the SNPs in the two assays was only slightly different. The order of markers on the linkage maps constructed with the SNPs in the BARCSoySNP6K versus SoySNP50K was consistent along all 20 chromosomes (Tables S3 and S4) with a Spearman's correlation coefficient of >0.999 ($P < 0.001$) in both the WP and EW populations.

Expected number of polymorphic SNPs in RIL population derived from randomly selected biparent in different populations

The average pairwise genetic distance was 0.34, 0.40 and 0.28 among the 96 elite, 96 landraces and wild soybean, respectively (Table 3). The average number of polymorphic markers between any random pair of accessions was projected to be 2000 in elite, 2402 in landrace and 3124 in cultivated soybean versus wild soybean. In addition, the number was expected to range from 1439 to 2559, 1812 to 2992 and 2746 to 3502 at the 95% probability level in elite, landrace and cultivated soybean versus wild soybean, respectively.

Inferring linkage map position of the BARCSoySNP6K SNPs based on the linkage map of WP population

Among the 6000 SNPs in the BARCSoySNP6K, in total, 3090 had genetic positions on the map derived from the WP population, we inferred genetic position for the 2862 SNPs not mapped in the WP linkage map (Table S5). The genetic location of the remaining 48 SNPs was not inferred because their physical positions in the Wm82a2v1 soybean assembly were not determined.

DISCUSSION

GBS is widely used for genotyping, however, BARCSoySNP6K Beadchip, as a different platform, has a number of advantages for genotyping RILs derived from a biparental cross: (i) low cost and high efficiency (no bioinformatics skill or high-performance computer is needed for data analysis); (ii) high quality with few or no missing SNP allele calls, thus no imputation of the SNP allele is

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

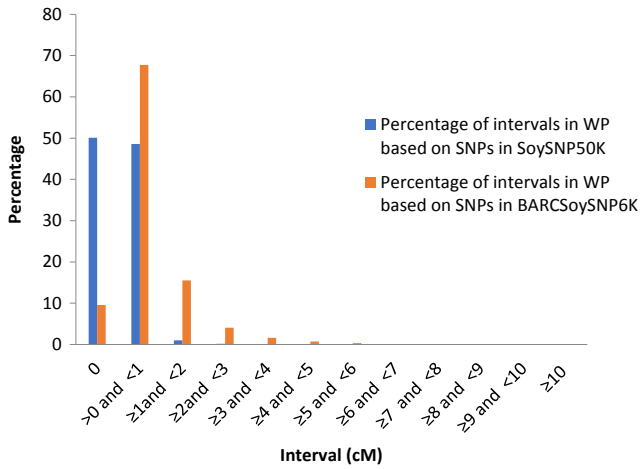

Figure 6. Percentage of intervals between adjacent single nucleotide polymorphisms (SNPs) in the linkage maps constructed based on the SNPs in SoySNP50K and BARCSoySNP6K in Williams 82 × PI 479752 population.

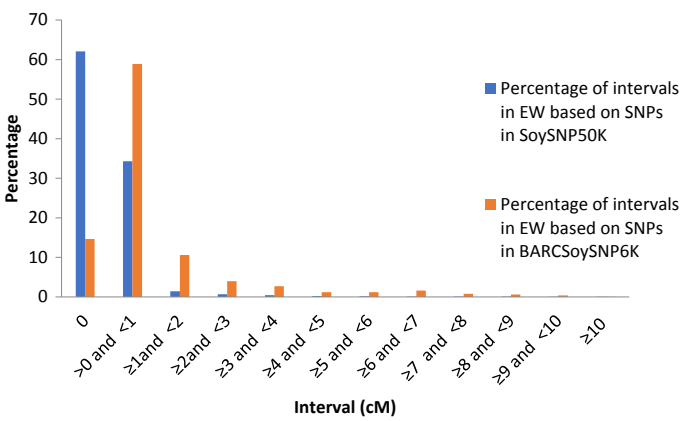

Figure 7. Percentage of intervals between adjacent single nucleotide polymorphisms (SNPs) in the linkage maps constructed based on the SNPs in SoySNP50K and BARCSoySNP6K in Essex × Williams 82 population.

Table 3 Average pairwise distance among 96 wild, 96 landrace and 96 elite soybean genotypes

	Average	Standard deviation
Among 96 landrance and 96 elites	0.40	0.058
Among 96 elite	0.34	0.048
Among 96 landrace	0.40	0.051
Among 96 wild	0.28	0.053

required; (iii) useful to genotype F_2 and F_3 RIL populations with high heterozygote rates, while the analysis of short reads from GBS usually generates inaccurate SNP alleles due to the difficulty in distinguishing variants caused by heterozygotic or homologous sequences; (iv) because the accessions of the USDA Soybean Germplasm Collection have been genotyped with the SoySNP50K Beadchip and the SNPs in the BARCSoySNP6K are a subset of the SNPs in the SoySNP50K, genotypic data from new germplasm can be integrated into the SoySNP50K dataset and facilitate germplasm comparison; (v) SNPs in the 6K were carefully selected to cover the major haplotype blocks and whole genome, and were identical on each chip, these common markers across studies will greatly facilitate genomic selection as well as a meta-analysis of genotypic datasets from different laboratories; and (vi) particularly useful to genotype RIL populations for QTL discovery due to low REs in plant genomes (Kump *et al.*, 2011; Song *et al.*, 2017), suggesting that genotyping RIL populations of limited sizes with a large number of markers is unnecessary because of high LD in these RIL populations. For example, Qi *et al.* (2014) generated 33.10 Gb of data by GBS in soybean, the average sequencing depth was greater than 42× for the two parents, and 3.92× for each of 149 lines, the resulting linkage maps included 5308 markers on 20 linkage groups, of these markers, in total, 3231 had unique genotypic profiles among the lines, the remaining 2077 markers co-segregated with one of the 3231 SNPs in the population. Li *et al.* (2015) analyzed three wheat RIL populations using reads from GBS and constructed genetic maps containing 28 644 SNPs, of these, approximately 13% (3757) had unique segregation patterns in the population, the remaining 87% (24 887) were redundant. Punnuri *et al.* (2016) analyzed an average of 2.2 million reads for each of 150 pearl millet RILs and mapped 16 650 SNPs, but 1189 had unique linkage map positions. Verma *et al.* (2015) obtained GBS sequences for 93 lines in chickpea and

© 2020 The Authors.

mapped 3363 SNPs with a unique SNP map position. Poland *et al*. (2012) analyzed GBS reads from 82 double haploid lines in barley and mapped 1019 SNPs with a unique map position.

SNP selection bias or ascertainment bias can result from the selection of highly discriminating SNPs utilizing one set of germplasm, but which might then be found to be less usefully discriminating among another unrelated set of germplasm (Achard *et al*., 2020). For the initial discovery of the SNPs to be included in the SoySNP50K assay, a set of eight genotypes, including six diverse cultivated and two wild soybean genotypes, were sequenced (Song *et al*., 2013). Subsequently, the profiles of 52 000 SNPs among >18 000 *G. max* and 1168 *G. soja* soybean accessions were analyzed and 42 508 SNPs that were all highly polymorphic among and between cultivated and wild soybeans were kept. The BACRSoySNP6K SNPs were selected from these 42 509 SNPs based on the criteria, including MAF, the quality of genotyping data, even genomic spacing, representative of both euchromatic and heterochromatic regions and haplotype block positions. Importantly, the selection of the 6K SNPs according to these criteria was based on all >18 000 *G. max* accessions in the USDA Soybean Germplasm Collection instead of specific subpopulations, thus eliminating possible SNP selection bias. This study showed that the BARCSoySNP6K assay was able to discriminate accessions among landrace, elites and wild soybean as we predicated. Similarly, Liu *et al*. (2017) found that associations among 577 Chinese and US soybean cultivars utilizing the 6K reflected the geographical origins and pedigrees of the cultivars, showing no indication of ascertainment bias within or among these sets of soybean germplasm, similar results were also reported by Achard *et al*. (2020).

The BARCSoySNP6K assay has been applied in various genetic research, e.g., it has been used to construct linkage maps (Lee *et al*., 2015), identify QTL/genes controlling a number of traits such as sudden death syndrome resistance (Wen *et al*., 2014; Lightfoot *et al*., 2018), aphid resistance (Bhusal *et al*., 2017; Zhang *et al*., 2017), charcoal rot resistance (Vinholes *et al*., 2019), *Phytophthora sojae*, *Pythium irregulare* and *Fusarium graminearum* resistance (Stasko *et al*., 2016; Million *et al*., 2019), salt tolerance (Do *et al*., 2018), waterlogging (Ye *et al*., 2018), iron deficiency chlorosis (Merry *et al*., 2019), nitrogen fixation (Huo *et al*., 2019), growth period (Liu *et al*., 2016), seed isoflavone content (Li *et al*., 2018), oil and fatty acids (Priolli *et al*., 2019), protein content (Nascimento *et al*., 2018) and yield (Ye *et al*., 2018). These studied not only confirmed previously identified QTL but also resulted in the discovery of new QTL, candidate genes or pathways controlling these traits. Wen *et al*. (2014) identified seven loci in previously mapped QTL intervals and 13 loci associated with sudden death syndrome, the identified loci explained an average of 54.5% of the phenotypic variance measured by different disease assessment criteria. Zhang *et al*. (2017) detected two major QTL, Rag6 and Rag3c, that were significantly associated with aphid-resistance alleles from E08934. Rag6 on chromosome 8 explained 19.5%–46.4% of the phenotypic variance and Rag3c on chromosome 16 explained 12.5%–22.9% of the phenotypic variance in different trials. Both Rag6 and Rag3c conferred antibiosis resistance to aphids and were subsequently confirmed in two validation populations with different genetic backgrounds, and Stasko *et al*. (2016) evaluated the resistance response to three isolates of *P. sojae*, one isolate of *Py. irregulare* and one isolate of *F. graminearum* in biparental families. They identified 10, 2 and 3 QTL conferring resistance to *P. sojae*, *Py. irregulare* and *F. graminearum*, respectively, and QTL for resistance toward both *Py. irregulare* and *F. graminearum* colocalized on chromosome 19. Million *et al*. (2019) identified a major quantitative disease resistance locus on chromosome 8 that contributed 38.5% of the phenotypic variance toward *F. graminearum* from a cross, together with other markers, they mapped this QTL to a 305 kb region harboring 36 genes. Ye *et al*. (2018) mapped a waterlogging tolerance QTL, qWT_Gm03, into a genomic region of <380 kbp in a RIL population. The tolerant allele of qWT_Gm03 promoted root growth under non-stress conditions and favorable root plasticity under waterlogging. They further found the involvement of auxin pathways regulating waterlogging tolerance. Huo *et al*. (2019) identified five SNP loci on chromosome 17, which were associated with shoot nitrogen concentrations under different environments, and led to the identification of two candidate genes (Glyma17g16400 and Glyma17g15600), which were further verified by gene expression analysis. The assay has also been used for other applications. Zhang *et al*. (2018) used the markers from 6K and other sources to develop advanced breeding lines carrying different aphid-resistance gene(s), and three breeding lines pyramided with multiple aphid-resistance genes, their resistance to diverse and dynamic aphid populations is expected to be strong and durable across geographic regions. Achard *et al*. (2020) analyzed 6K assay data for 322 soybean cultivars released during a three decadal period and granted plant variety protection. They concluded that the methodology of using molecular data from the 6K meets the criteria of maintaining existing levels of intellectual property protection. They also noted that the assay "makes the process more efficient, and potentially more harmonized globally, does not add costs and may reduce costs of conducting DUS testing for applicants and plant variety protection agencies, and does not require levels of uniformity that are unrealistic, overly expensive, unnecessary, or impractical to achieve." Eickholt *et al*. (2019) used the 6K assay to reveal the percentage of wild soybean alleles from individual interspecific breeding lines developed from the

© 2020 The Authors.

hybridization of lodging-resistant soybean cultivar and wild soybean, and to explore the extent of recombination occurred between the *G. max* and *G. soja* genomes. They thus released a group of 17 interspecific breeding lines containing 21%–40% alleles derived from wild soybean to expand the North American soybean breeding pool. Stewart-Brown *et al*. (2019) and Ma *et al*. (2016) studied the potential of genomic selection for soybean germplasm or elite breeding lines genotyped with 6K. Cross validation analysis showed that high predictive abilities for protein, oil could be achieved with these markers, while a larger training set size in combination with increased genetic relatedness between training and validation set could further improve predictive ability of seed yield. In addition, the 6K assay has been used to map agronomy traits in nested association mapping populations (Beche *et al*., 2020), to estimate population structure, pedigree and LD (Contreras-Soto *et al*., 2017), and characterize the mapping population for registration (Lee *et al*., 2017). The assay has been proven useful and efficient by the soybean breeders and geneticists working in the private and public sectors, and the assay was commercialized by Illumina (http://www.illumina.com/areas-of-interest/agrigenomics/consortia.html).

Because the BARCSoySNP6K assay was selected based upon extensive knowledge of haplotype block structure and the distance between selected SNPs in the euchromatic and heterochromatic regions, our analyses of genotypic data from numerous RIL populations from this laboratory and collaborators showed that all the polymorphic SNPs were clustered into 20 linkage groups, suggesting these SNPs well cover the soybean genome.

However, there are limitations for the assay applications, the assay may not be able to detect association between traits and markers via genome-wide association analysis particularly if the occurrence of targeted traits is rare in the population and the traits are associated with the rare allele because markers with low MAF among the *G. max* were excluded in the assay. In addition, the number of markers in the assay may be insufficient to fine map the gene/QTL controlling traits in a set of diverse germplasm via genome-wide association analysis or in a large RIL population-derived *G max* × *G. soja* cross via genetic linkage association analysis, because the size of LD blocks is smaller in the wild soybean than in cultivated soybean and in the diverse germplasm populations than in RIL populations. In case of genomic regions where the 6K marker density is insufficient for fine mapping, KASPer markers from the SNPs or short sequence repeat markers, e.g., the BARCSoySSR1.01 database containing >36 000 candidate markers in the soybean genome (Song *et al*., 2010), could be utilized.

The selected 6000 SNPs are also a valuable source for developing a targeted enrichment or targeted amplicon GBS assay in soybean (Turner *et al*., 2009; Mamanova *et al*., 2010; Niedzicka *et al*., 2016); therefore, the same set of SNPs can be genotyped each time. It will be an option to genotype materials if the array technology reaches its life span or costs more. The 6K assay is also a source for developing KASPer markers because the sequence flanking the SNPs have been screened for their specificity in the genome and their ability to distinguish among the homozygote alleles and heterozygote allele, examples of the success for designing the KASPer markers from 6K or SoySNP50K SNPs have been reported previously (Zhang *et al*., 2017; Merry *et al*., 2019).

EXPERIMENTAL PROCEDURES

SNP analysis of soybean germplasm

DNA from the seeds of 18 489 cultivated soybean accessions was extracted using the CTAB method (Keim *et al*., 1998). A high-throughput SNP assay, the SoySNP50K Illumina Infinium II Beadchip (Song *et al*., 2013; Song *et al*., 2015) was used for SNP genotyping and the genotyping was conducted on the Illumina platform following the Infinium® HD Assay Ultra Protocol (Illumina, Inc., San Diego, CA, USA). The SNP alleles were called using the GenomeStudio Genotyping Module v1.8.4 (Illumina, Inc.). The dataset containing 18 489 *G. max* × 42 509 SNPs described by Song *et al*. (2015) was used for the selection of a core set of SNPs.

Procedure for the selection of SNPs included in the BARCSoySNP6K assay

In total, 1000 and 5000 SNPs were selected from heterochromatic and euchromatic regions, respectively. The density of the selected SNPs was determined to be five times greater in the euchromatic than the heterochromatic regions because the ratio of recombination rate/Mb in the two regions of soybean genome was 5:1 (Song *et al*., 2013). Owing to redundancy of *G. max* accessions in the USDA Germplasm Collection, in total, 4306 *G. max* accessions had a similarity of >99.9% to one or more other accessions in the collection and hence were excluded in the subsequent analyses. SNPs with MAF <0.05 among the remaining 14 183 *G. max* accessions were eliminated. As the reliability of the SNP allele call is dependent on the intensity of fluorescent signals from both alleles in the SNP Graph displayed by the GenomeStudio software and because the fluorescent signal of SNPs may vary because of the complexity or the lack of specificity of SNP flanking sequence, the incidence of poor allele clustering and cluster compression may occur and lead to inaccurate allele calls; therefore, these SNPs were also excluded. Haplotype blocks in *G. max* were identified following the procedures described by Song *et al*. (2015). In general, the haplotype blocks were determined through estimates of D' for all pairwise combinations of SNPs within 1 Mb windows as per the definition of Gabriel *et al*. (2002). The software PLINK (Purcell *et al*., 2007) was used to calculate the pairwise LD (r^2) among SNPs and thus to identify haplotype blocks. In each haplotype block with a size of ≥50 kb in the euchromatic regions and ≥100 kb in the heterochromatic regions, SNPs with low MAF, and SNPs residing in the same haplotype block were eliminated, only one SNP with the highest MAF in each haplotype block was selected. In the segments of the genome not present in haplotype blocks, the objective of the selection was to keep the SNPs with the highest MAF and to equalize the distance between selected SNPs in the euchromatic and heterochromatic regions. The iteration algorithm of SNP selection previously described by Song *et al* (2013) was used.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

Comparison of clusters of diverse soybean accessions based on the SNPs genotyped with SoySNP50K Beadchip and the BARCSoySNP6K assay

Clusters for a group of 96 diverse landraces, 96 elite cultivars and 96 *G. soja* accessions were constructed based on the alleles genotyped with selected 6000 SNPs in BARCSoySNP6K and 42 509 SNPs in SoySNP50K Beadchip. The Mega software was used for clustering (Tamura *et al.*, 2007). The 96 diverse landraces were from China, Japan and Korea and the 96 elite cultivars were from North America and represent a diversity of publicly developed cultivars released from 1990 to 2000. The 96 *G. soja* accessions were from China, Korea, Japan and Russia and were selected based upon the wide ranges of latitude and longitude at which they were collected (Song *et al.*, 2013). Congruence of the distance matrices among accessions based on BARCSoySNP6K and SoySNP50K SNPs were measured by the *Z* statistics of the Mantel test (Mantel, 1967). An approximate randomization test was used to examine the significance of the calculated *Z* value (Dietz, 1983).

Construction of linkage maps based on the SNPs genotyped with SoySNP50K Beadchip and the BARCSoySNP6K assay

We previously constructed linkage maps based on 21 478 SNP loci mapped in the Williams 82 × *G. soja* (Sieb. & Zucc.) PI 479752 population (WP) with 1083 RILs and 11 922 loci mapped in the Essex × Williams 82 population (EW) with 922 RILs (Song *et al.*, 2016). To evaluate the linkage maps constructed based on the selected 6000 SNPs, genotypic data of the RILs in the WP and EW populations for the 6000 SNPs were used to create linkage maps of the two populations using JoinMap 4.0 (Van Ooijen, 2006), the same software used to construct the linkage maps of WP and EW based on SNPs in the BARCSoySNP6K assay. Procedures of SNP filtering and linkage analysis described previously by Song *et al* (2016) were followed.

Expected number of polymorphic SNPs in RIL populations derived from randomly selected biparent in landrace, elite and wild populations

Pairwise genetic distance among 96 diverse landraces, 96 elite cultivars and 96 *G. soja* accessions based on the selected 6000 SNPs were calculated using Mega 7.0.26 (Kumar *et al.*, 2008). The genetic distance was estimated as the ratio of the number of polymorphic SNPs versus total number of SNPs for the pair. The projected polymorphic SNP number between any pair of random accessions at a 95% confidence interval was calculated by the formula $C_1 = p - 1.96 \times SE$, $C_2 = p + 1.96 \times SE$. Where p is the average of all pairwise distance, SE is the standard error of the pairwise distance, and C_1 and C_2 are the lower and upper confidence limits of the projected average SNP number at 0.05 probability level, respectively.

Inference of linkage map position of the BARCSoySNP6K SNPs based on the linkage map derived from WP population

Because not all SNPs were polymorphic in the WP population, some SNPs had no genetic position on the linkage maps; however, these SNPs may be polymorphic in other populations, and we therefore inferred the location (cM) of these SNPs by linear interpolation between the WP mapped SNPs that flank them and by SNP physical positions in the Wm82a2v1 assembly using an in-house script. The linkage maps for all SNPs will facilitate QTL mapping in any RIL population characterized with the 6000 SNPs in the assay.

CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

ACKNOWLEDGEMENTS

This research was supported with funding from the United Soybean Board Project #8265. The support of the United Soybean Board is greatly appreciated. This research was also funded by the US Department of Agriculture-Agricultural Research Service, project number: 8042-21000-289-00D. We thank Rob Parry and Chris Pooley at USDA–ARS, Beltsville, MD, for their technical support in assembling the necessary hardware and software required for the data analysis.

AUTHOR CONTRIBUTIONS

RLN, VP and DH prepared plant materials. CVQ and EWF performed DNA extraction and molecular genotyping. LY performed linkage map analysis. HW, L C, FD, SA, JL and QS performed assay validation analyses and material preparation. QS provided project planning and coordination, designed the assay, performed data analysis and prepared the manuscript.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the manuscript and its supporting materials.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Positions of haplotype blocks based on 14 183 cultivated soybean.

Table S2. Detailed information of the SNPs in the BARCSoySNP6K assay.

Table S3. Linkage map position of SNPs in EW population based on BARCSoySNP6K an SoySNP50K assays.

Table S4. Linkage map position of SNPs in WP population based on BARCSoySNP6K an SoySNP50K assay.

Table S5. Inferred linkage map position of SNPs in BARCSoySNP6K.

REFERENCES

Achard, F., Butruille, M., Madjarac, S., Nelson, P.T., Duesing, J., Laffont, J.L., Nelson, B., Xiong, J., Mikel, M.A. and Smith, J.S.C. (2020) Single nucleotide polymorphisms facilitate dus testing of soybean cultivars for plant variety protection. *Crop Sci.* 1–24. https://doi.org/10.1002/csc2.20201

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J. and Lorenz, A. (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome*, **8**, 3.

Bao, Y., Vuong, T., Meinhardt, C., Tiffin, P., Denny, R., Chen, S., Nguyen, H.T., Orf, J.H. and Young, N.D. (2014) Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. *Plant Genome*, **7**, 3.

Barreiro, L.B., Henriques, R. and Mhlanga, M.M. (2009) High-throughput SNP genotyping: combining tag SNPs and molecular beacons. In *Single Nucleotide Polymorphisms*. Totowa, NJ: Humana Press, pp. 255–276.

© 2020 The Authors.

Beche, E., Gillman, J.D., Song, Q., Nelson, R., Beissinger, T., Decker, J., Shannon, G. and Scaboo, A.M. (2020) Nested association mapping of important agronomic traits in three interspecific soybean populations. *Theor. Appl. Genet.* **133**, 1039–1054.

Bhusal, S.J., Jiang, G.-L., Song, Q., Cregan, P.B., Wright, D. and Gonzalez-Hernandez, J.L. (2017) Genome-wide detection of genetic loci associated with soybean aphid resistance in soybean germplasm PI 603712. *Euphytica*, **213**, 144.

Chang, H.-X., Brown, P.J., Lipka, A.E., Domier, L.L. and Hartman, G.L. (2016a) Genome-wide association and genomic prediction identifies associated loci and predicts the sensitivity of Tobacco ringspot virus in soybean plant introductions. *BMC Genom.* **17**, 153.

Chang, H.-X., Lipka, A.E., Domier, L.L. and Hartman, G.L. (2016b) Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Phytopathology*, **106**, 1139–1151.

Choi, I.-Y., Hyten, D.L., Matukumalli, L.K., Song, Q., Chaky, J.M., Quigley, C.V., Chase, K., Lark, K.G., Reiter, R.S. and Yoon, M.-S. (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics*, **176**, 685–696.

Collins, F.S., Brooks, L.D. and Chakravarti, A. (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231.

Contreras-Soto, R.I., de Oliveira, M.B., Costenaro-da-Silva, D., Scapim, C.A. and Schuster, I. (2017) Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (Glycine max). *Euphytica*, **213**, 173.

Dhanapal, A.P., Ray, J.D., Singh, S.K., Hoyos-Villegas, V., Smith, J.R., Purcell, L.C., King, C.A., Cregan, P.B., Song, Q. and Fritschi, F.B. (2015) Genome-wide association study (GWAS) of carbon isotope ratio (δ13C) in diverse soybean [Glycine max (L.) Merr.] genotypes. *Theor. Appl. Genet.* **128**, 73–91.

Diers, B.W., Specht, J., Rainey, K.M., Cregan, P., Song, Q., Ramasubramanian, V., Graef, G., Nelson, R., Schapaugh, W. and Wang, D. (2018) Genetic architecture of soybean yield and agronomic traits. *G3*, **8**, 3367–3375.

Dietz, E.J. (1983) Permutation tests for association between two distance matrices. *Syst. Biol.* **32**, 21–26.

Ding, C. and Jin, S. (2009) High-throughput methods for SNP genotyping. In *Single Nucleotide Polymorphisms*. Totowa, NJ: Humana Press, pp. 245–254.

Do, T.D., Vuong, T.D., Dunn, D., Smothers, S., Patil, G., Yungbluth, D.C., Chen, P., Scaboo, A., Xu, D. and Carter, T.E. (2018) Mapping and confirmation of loci for salt tolerance in a novel soybean germplasm, Fiskeby III. *Theor. Appl. Genet.* **131**, 513–524.

Eickholt, D., Carter, T.E., Taliercio, E., Dickey, D., Dean, L.O., Delheimer, J. and Li, Z. (2019) Registration of USDA-max x soja Core Set-1: recovering 99% of wild soybean genome from PI 366122 in 17 agronomic interspecific germplasm lines. *J. Plant Regist.* **13**, 217–236.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

Gabriel, S.B., Schaffner, S.F., Nguyen, H. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.

Hartman, G. and Chang, H.-X. (2017) Characterization of insect resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Front. Plant Sci.* **8**, 670.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A. and Parco, A. (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics*, **148**, 479–494.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q. and Huang, T. (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068–1076.

Huo, X., Li, X., Du, H., Kong, Y., Tian, R., Li, W. and Zhang, C. (2019) Genetic loci and candidate genes of symbiotic nitrogen fixation–related characteristics revealed by a genome-wide association study in soybean. *Mol. Breeding*, **39**, 127.

Hwang, E.-Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J. and Cregan, P.B. (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genom.* **15**, 1.

Hyten, D.L., Cannon, S.B., Song, Q., Weeks, N., Fickus, E.W., Shoemaker, R.C., Specht, J.E., Farmer, A.D., May, G.D. and Cregan, P.B. (2010a) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genom.* **11**, 38.

Hyten, D.L., Choi, I.-Y., Song, Q., Specht, J.E., Carter, T.E., Shoemaker, R.C., Hwang, E.-Y., Matukumalli, L.K. and Cregan, P.B. (2010b) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. *Crop Sci.* **50**, 960–968.

Jarquín, D., Howard, R., Graef, G. and Lorenz, A. (2019) Response surface analysis of genomic prediction accuracy values using quality control covariates in soybean. *Evol. Bioinform.* **15**, 1176934319831307.

Keim, P., Olson, C. and Shoemaker, R.G.N. (1998) A rapid protocol for isolating soybean DNA. *Soybean Genet. Newsl.* **15**, 150–152.

Kim, M.Y., Lee, S., Van, K., Kim, T.-H., Jeong, S.-C., Choi, I.-Y., Kim, D.-S., Lee, Y.-S., Park, D. and Ma, J. (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. *Proc. Natl Acad. Sci. USA*, **107**, 22032–22037.

Kruglyak, L. (1997) The use of a genetic map of biallelic markers in linkage studies. *Nat. Genet.* **17**, 21–24.

Kumar, S., Nei, M., Dudley, J. and Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **9**, 299–306.

Kump, K.L., Bradbury, P.J., Wisser, R.J., Buckler, E.S., Belcher, A.R., Oropeza-Rosas, M.A., Zwonitzer, J.C., Kresovich, S., McMullen, M.D. and Ware, D. (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* **43**, 163–168.

Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.-L., Li, M.-W., He, W., Qin, N. and Wang, B. (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **42**, 1053–1059.

Leamy, L.J., Zhang, H., Li, C., Chen, C.Y. and Song, B.-H. (2017) A genome-wide association study of seed composition traits in wild soybean (Glycine soja). *BMC Genom.* **18**, 18.

Lee, S., Freewalt, K.R., McHale, L.K., Song, Q., Jun, T.-H., Michel, A.P., Dorrance, A.E. and Mian, M.R. (2015) A high-resolution genetic linkage map of soybean based on 357 recombinant inbred lines genotyped with BARCSoySNP6K. *Mol. Breeding*, **35**, 58.

Lee, S., Jun, T.-H., McHale, L.K., Michel, A.P., Dorrance, A.E., Song, Q. and Mian, M.A. (2017) Registration of Wyandor x PI 567301B soybean recombinant inbred line population. *J. Plant Regist.* **11**, 324–327.

Li, X., Kamala, S., Tian, R., Du, H., Li, W., Kong, Y. and Zhang, C. (2018) Identification and validation of quantitative trait loci controlling seed isoflavone content across multiple environments and backgrounds in soybean. *Mol. Breeding*, **38**, 8.

Li, Y., Zhao, S., Ma, J., Li, D., Yan, L., Li, J., Qi, X., Guo, X., Zhang, L. and He, W. (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genom.* **14**(1), 579.

Li, Y.H., Li, W., Zhang, C., Yang, L., Chang, R.Z., Gaut, B.S. and Qiu, L.J. (2010) Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (Glycine soja) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol.* **188**, 242–253.

Li, H., Vikram, P., Singh, R.P. *et al.* (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics*, **16**(1), 216.

Lightfoot, D.A., Lee, Y.C., Iqbal, M.J., Njiti, V., Kantartzi, S.K., Gibson, P. and Anderson, J. (2018) Confirmation of QTL that underlie resistance to soybean sudden death syndrome using NILs and SNPs. *Atlas J. Biol.* **2018**, 583–591.

Lin, C.H., Yeakley, J.M., McDaniel, T.K. and Shen, R. (2009) Medium-to high-throughput SNP genotyping using VeraCode microbeads. In *DNA and RNA Profiling in Human Blood*. Totowa, NJ: Humana Press, pp. 129–142.

Liu, Z., Li, H., Fan, X., Huang, W., Yang, J., Li, C., Wen, Z., Li, Y., Guan, R. and Guo, Y. (2016) Phenotypic characterization and genetic dissection of growth period traits in soybean (Glycine max) using association mapping. *PLoS One*, **11**, e0158602.

Liu, Z., Li, H., Wen, Z., Fan, X., Li, Y., Guan, R., Guo, Y., Wang, S., Wang, D. and Qiu, L. (2017) Comparison of genetic diversity between Chinese and

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,

American soybean (Glycine max (L.)) accessions revealed by high-density SNPs. *Frontiers. Plant Sci.* **8**, 2014.

Ma, Y., Reif, J.C., Jiang, Y., Wen, Z., Wang, D., Liu, Z., Guo, Y., Wei, S., Wang, S. and Yang, C. (2016) Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). *Mol. Breeding*, **36**, 113.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J. and Turner, D.J. (2010) Target-enrichment strategies for next-generation sequencing. *Nat. Methods*, **7**, 111–118.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Can. Res.* **27**, 209–220.

Merry, R., Butenhoff, K., Campbell, B.W., Michno, J.-M., Wang, D., Orf, J.H., Lorenz, A.J. and Stupar, R.M. (2019) Identification and fine-mapping of a soybean quantitative trait locus on chromosome 5 conferring tolerance to iron deficiency chlorosis. *Plant Genome*, **12**, 3.

Million, C.R., Wijeratne, S., Cassone, B.J., Lee, S., Mian, R., McHale, L.K. and Dorrance, A.E. (2019) Hybrid genome assembly of a major quantitative disease resistance locus in soybean toward fusarium graminearum. *Plant Genome*, **12**(2), 1–17.

Nascimento, D., Polo, L.R.T., Lazzari, F., Silva, G.J.d. and Schuster, I. (2018) Genomic Association between SNP Markers and QTLs for protein and oil content in grain weight in soybean (Glycine max). *J. Sci. Res. Rep.* **20**, 1–13.

Niedzicka, M., Fijarczyk, A., Dudek, K., Stuglik, M. and Babik, W. (2016) Molecular Inversion Probes for targeted resequencing in non-model organisms. *Sci. Rep.* **6**, 24051.

Philippe, R., Paux, E., Bertin, I. *et al.* (2013) A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Genome Biol.* **14**, R64.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H. and Sorrells, M. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, **5**, 103–113.

Priolli, R.H.G., Carvalho, C.R.L., Bajay, M.M., Pinheiro, J.B. and Vello, N.A. (2019) Genome analysis to identify SNPs associated with oil content and fatty acid components in soybean. *Euphytica*, **215**, 54.

Punnuri, S.M., Wallace, J.G., Knoll, J.E., Hyma, K.E., Mitchell, S.E., Buckler, E.S., Varshney, R.K. and Singh, B.P. (2016) Development of a high-density linkage map and tagging leaf spot resistance in pearl millet using genotyping-by-sequencing markers. *The Plant Genome*, **9**(2), 1–13.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W. and Daly, M.J. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575.

Qi, Z., Huang, L., Zhu, R. *et al.* (2014) A high-density genetic map for soybean based on specific length amplified fragment sequencing. *PLoS One*, **9**(8), e104871.

Ray, J.D., Dhanapal, A.P., Singh, S.K., Hoyos-Villegas, V., Smith, J.R., Purcell, L.C., King, C.A., Boykin, D., Cregan, P.B. and Song, Q. (2015) Genome-wide association study of ureide concentration in diverse maturity group IV soybean [Glycine max (L.) Merr.] accessions. *G3*, **5**, 2391–2403.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J. and Cheng, J. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One*, **8**, e54985.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2015) Fingerprinting soybean germplasm and its utility in genomic research. *G3*, **5**, 1999–2006.

Song, Q., Jenkins, J., Jia, G., Hyten, D.L., Pantalone, V., Jackson, S.A., Schmutz, J. and Cregan, P.B. (2016) Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1. 01. *BMC Genom.* **17**, 1.

Song, Q., Jia, G., Zhu, Y., Grant, D., Nelson, R.T., Hwang, E.-Y., Hyten, D.L. and Cregan, P.B. (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1. 0) in soybean. *Crop Sci.* **50**, 1950–1960.

Song, Q., Yan, L., Quigley, C., Jordan, B.D., Fickus, E., Schroeder, S., Song, B.-H., Charles An, Y.-Q., Hyten, D. and Nelson, R. (2017) Genetic characterization of the soybean nested association mapping population. *Plant Genome*, **10**(2), 1–14.

Song, Q.J., Marek, L.F., Shoemaker, R.C., Lark, K.G., Concibido, V.C., Delannay, X., Specht, J.E. and Cregan, P.B. (2004) A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* **109**, 122–128.

Song, Q.J., Shi, J.R., Singh, S., Fickus, E.W., Costa, J.M., Lewis, J., Gill, B.S., Ward, R. and Cregan, P.B. (2005) Development and mapping of microsatellite (SSR) markers in wheat. *Theor. Appl. Genet.* **110**, 550–560.

Stasko, A.K., Wickramasinghe, D., Nauth, B.J., Acharya, B., Ellis, M.L., Taylor, C.G., McHale, L.K. and Dorrance, A.E. (2016) High-density mapping of resistance QTL toward Phytophthora sojae, Pythium irregulare, and Fusarium graminearum in the same soybean population. *Crop Sci.* **56**(5), 2476–2492.

Stewart-Brown, B.B., Song, Q., Vaughn, J.N. and Li, Z. (2019) Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3*, **9**, 2253–2265.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

Turner, E.H., Lee, C., Ng, S.B., Nickerson, D.A. and Shendure, J. (2009) Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat. Methods*, **6**, 315–316.

Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., Chen, C., Li, Y., Joshi, T. and Song, L. (2016) Landscape of genomic diversity and trait discovery in soybean. *Sci. Rep.* **6**, 23598.

Van Inghelandt, D., Melchinger, A.E., Lebreton, C. and Stich, B. (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* **120**, 1289–1299.

Van Ooijen, J. (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. *Kyazma BV, Wageningen*, **33** (10), 1371.

Vaughn, J.N., Nelson, R.L., Song, Q., Cregan, P.B. and Li, Z. (2014) The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3*, **4**, 2283–2294.

Verma, S., Gupta, S., Bandhiwai, 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). *Scientific Reports*, **5**, 17512.

Vinholes, P., Rosado, R., Roberts, P., Borém, A. and Schuster, I. (2019) Single nucleotide polymorphism-based haplotypes associated with charcoal rot resistance in Brazilian soybean germplasm. *Agron. J.* **111**, 182–192.

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S. and Kitamura, K. (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics*, **188**, 395–407.

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 Genom.* **16**, 671.

Wen, Z., Tan, R., Yuan, J., Bales, C., Du, W., Zhang, S., Chilvers, M.I., Schmidt, C., Song, Q. and Cregan, P.B. (2014) Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genom.* **15**, 809.

Ye, H., Song, L., Chen, H., Valliyodan, B., Cheng, P., Ali, L., Vuong, T., Wu, C., Orlowski, J. and Buckley, B. (2018) A major natural genetic variation associated with root system architecture and plasticity improves waterlogging tolerance and yield in soybean. *Plant Cell Environ.* **41**, 2169–2182.

Yoon, M.S., Song, Q.J., Choi, I.Y., Specht, J.E., Hyten, D.L. and Cregan, P.B. (2007) BARCSoySNP23: a panel of 23 selected SNPs for soybean cultivar identification. *Theor. Appl. Genet.* **114**, 885–899.

Zeng, A., Chen, P., Korth, K., Hancock, F., Pereira, A., Brye, K., Wu, C. and Shi, A. (2017) Genome-wide association study (GWAS) of salt tolerance in worldwide soybean germplasm lines. *Mol. Breeding*, **37**, 30.

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., Florez-Palacios, L., Brye, K., Song, Q. and Cregan, P. (2014) Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Sci.* **54**, 554–564.

© 2020 The Authors.

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.

Zhang, J., Song, Q., Cregan, P.B., Nelson, R.L., Wang, X., Wu, J. and Jiang, G.-L. (2015) Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. *BMC Genom.* **16**, 217.

Zhang, S., Wen, Z., DiFonzo, C., Song, Q. and Wang, D. (2018) Pyramiding different aphid-resistance genes in elite soybean germplasm to combat dynamic aphid populations. *Mol. Breeding*, **38**, 29.

Zhang, S., Zhang, Z., Bales, C., Gu, C., DiFonzo, C., Li, M., Song, Q., Cregan, P., Yang, Z. and Wang, D. (2017) Mapping novel aphid resistance QTL from wild soybean, Glycine soja 85–32. *Theor. Appl. Genet.* **130**, 1941–1952.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y. and Ma, Y. (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408–414.

Zhu, Y.L., Song, Q.J., Hyten, D.L., Van Tassell, C.P., Matukumalli, L.K., Grimm, D.R., Hyatt, S.M., Fickus, E.W., Young, N.D. and Cregan, P.B. (2003) Single-nucleotide polymorphisms in soybean. *Genetics*, **163**, 1123–1134.

© 2020 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd,