2017

Genetic Characterization of the Soybean Nested Association Mapping Population

Qijian Song
USDA, Agricultural Research Service, qijian.song@ars.usda.gov

Long Yan
Hebei Academy of Agricultural and Forestry Sciences

Charles Quigley
USDA, Agricultural Research Service

Brandon D. Jordan
University of Illinois

Edward Fickus
USDA, Agricultural Research Service

See next page for additional authors

Published April 27, 2017

Genetic Characterization of the Soybean Nested Association Mapping Population

Qijian Song*, Long Yan, Charles Quigley, Brandon D. Jordan, Edward Fickus, Steve Schroeder, Bao-Hua Song, Yong-Qiang Charles An, David Hyten, Randall Nelson, Katy Rainey, William D Beavis, Jim Specht, Brian Diers, and Perry Cregan

Abstract

A set of nested association mapping (NAM) families was developed by crossing 40 diverse soybean [*Glycine max* (L.) Merr.] genotypes to the common cultivar. The 41 parents were deeply sequenced for SNP discovery. Based on the polymorphism of the single-nucleotide polymorphisms (SNPs) and other selection criteria, a set of SNPs was selected to be included in the SoyNAM6K BeadChip for genotyping the parents and 5600 RILs from the 40 families. Analysis of the SNP profiles of the RILs showed a low average recombination rate. We constructed genetic linkage maps for each family and a composite linkage map based on recombinant inbred lines (RILs) across the families and identified and annotated 525,772 high confidence SNPs that were used to impute the SNP alleles in the RILs. The segregation distortion in most families significantly favored the alleles from the female parent, and there was no significant difference of residual heterozygosity in the euchromatic vs. heterochromatic regions. The genotypic datasets for the RILs and parents are publicly available and are anticipated to be useful to map quantitative trait loci (QTL) controlling important traits in soybean.

Core Ideas

- 40 NAM families were developed and 5600 RILs in the families were characterized.
- The linkage maps for each family and a composite linkage map were constructed.
- More than a half million high-confidence SNPs were identified and annotated.
- Segregation distortion in most families favored alleles from the female parent.
- The REs in the soybean genome is low.

Q.J. Song, C. Quigley, E. Fickus, and P. Cregan, Soybean Genomics and Improvement Lab., USDA–ARS, 10300 Baltimore Ave, Building 006, Beltsville, MD 20705; L. Yan, Institute of Cereal and Oil Crops, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, 050035, China; B.D. Jordan, B. Diers, Dep. of Crop Sciences, Univ. of Illinois, 268 National Soybean Research Center, 1101 West Peabody Drive, Urban, IL 61801; S. Schroeder, USDA–ARS, Animal Genomics and Improvement Laboratory, Beltsville, MD, 20705; B.H. Song, Dep. of Biological Sciences, Univ. of North Carolina at Charlotte, Charlotte, NC 28223; Y.Q.C. An, USDA–ARS, Plant Genetics Research Unit at Donald Danforth Plant Sciences Center, Saint Louis, Missouri; D. Hyten, Dep. of Agronomy and Horticulture, Univ. of Nebraska, Lincoln, NE 68583; J. Specht, Dep. of Agronomy and Horticulture, Univ. of Nebraska, Lincoln, NE 68583, USA; R. Nelson, USDA–ARS, Soybean/Maize Germplasm, Pathology and Genetics Research Unit and Dep. of Crop Sciences, Univ. of Illinois, Urbana, IL 61801; K. Rainey, Dep. of Agronomy, Purdue Univ., Lilly 2-351, West Lafayette, IN 47906; W.D. Beavis, Dep. of Agronomy, Iowa State Univ., Ames, Iowa 50011. Received 26 Oct. 2016. Accepted 23 Jan. 2017. *Corresponding author (Qijian.song@ars.usda.gov). Assigned to Associate Editor Robert Stupar.

Abbreviations: GWAS, genome-wide association studies; MG, maturity group; NAM, nested association mapping; QTL, quantitative trait loci; RE, recombination event; RH, residual heterozygousity; RIL, recombinant inbred line; SNP, single-nucleotide polymorphism.

Published in Plant Genome
Volume 10. doi: 10.3835/plantgenome2016.10.0109

Many important traits related to yield, yield components, seed quality, and stress resistance in crops are controlled by multiple QTL. The performance of these complex traits is affected by the environment, interaction between the environment and QTL, and interactions among the QTL. Single-family-based analysis of genotypes together with phenotypes is the most commonly used method to map the QTL associated with these traits. In soybean, >180 phenotypic traits, such as seed yield (Chung et al., 2003; Palomeque et al., 2009; Wang et al., 2004, 2014), yield components (He et al., 2014; Jeong et al., 2012; Kato et al., 2014), morphological traits (Lee et al., 2014, Yamanaka et al., 2001), seed composition (Brummer et al., 1997; Warrington et al., 2015), resistance to diseases (Pham et al., 2013; Wu et al., 2009) and pests (Rector et al., 1998, Terry et al., 2000, Zhang et al., 2009), and abiotic stresses (Abdel-Haleem et al., 2012; Lee et al., 2004), have been analyzed with this approach as documented in SoyBase (http://www.soybase.org/search/qtllist_by_symbol.php). Linkage mapping does not require high marker density because of the limited number of recombination events (REs) that occur during selfing and the limited number of lines typically used in such studies. As a result, the resolution of linkage mapping is usually poor. Population-based genome-wide association studies (GWASs) can provide better resolution by exploiting historical REs in a population and should also include more causative loci than would be expected to segregate in a biparental population (Nordborg and Tavaré, 2002). With the advent of high-throughput genotyping and sequencing technology, accessions can be efficiently assayed with high-density markers, thus, a GWAS can be performed on well-developed phenotypic datasets such that exist for germplasm collections. For example, the USDA's Soybean Germplasm Collection includes ~20,000 domesticated and wild soybean accessions introduced from other countries or developed in the United States. This collection was genotyped with the SoySNP50K BeadChip containing >52,000 SNPs (Song et al., 2013, 2015). With this genotypic dataset, GWAS have detected numerous loci associated with a number of traits (Dhanapal et al., 2015a,b; Hwang et al., 2014; Rincker et al., 2016; Vaughn et al., 2014; Wen et al., 2014; Zhang et al., 2015).

Nested association mapping was proposed to increase the resolution of QTL mapping by capturing the history of REs from GWAS and increase power of genome-wide association analysis through introduction of linkage mapping (Yu and Buckler, 2006; Yu et al., 2008). Unlike traditional QTL mapping, which only uses limited genetic variation represented by only two parental individuals, NAM can increase genetic variation across contributing parental lines, increase genetic resolution, reduce linkage disequilibrium, and control population structure through design (Rafalski, 2010). Nested association mapping populations are developed by crossing multiple diverse founders to a common parent followed by the development of RILs or progenies in each family, which are then genotyped with low-density markers, while the parents are genotyped with high-density markers. The high-density genotypic information can then be projected onto the progenies. Thus, the association of the imputed genotypic data with the phenotypic data can be analyzed using the methods described previously (Buckler et al., 2009, Kump et al., 2011, Yu and Buckler, 2006, Yu et al., 2008). A NAM population was first created in maize (*Zea mays* L.) (Buckler et al., 2009), which consisted of 25 segregating families derived from crosses involving B73 and a wide diversity of unrelated maize lines to produce 5000 RILs. The RILs were genotyped with a total of 1106 SNPs (McMullen et al., 2009), and the parents were sequenced with next-generation sequencing to produce a maize hapmap with 1.6 million SNPs (Gore et al., 2009). Nested association mapping populations have also been developed in other crop species including barley (*Hordeum vulgare* L.) (Schnaithmann et al., 2014) and sorghum [*Sorghum bicolor* (L.) Moench] (Jordan et al., 2011). The NAM design has been used successfully in fine-mapping QTL controlling a number of traits such as leaf rust resistance in barley (Schnaithmann et al., 2014) and flowering time (Buckler et al., 2009), southern and northern leaf blight (Kump et al., 2011; Poland et al., 2011), leaf architecture (Tian et al., 2011), kernel composition (Cook et al., 2012), and stalk strength (Peiffer et al., 2013) in maize.

The Ilumina Beadchip assay is ideal for analyzing thousands of SNPs in a large number of genotypes quickly and cost-effectively. It provides reproducible and high-quality data with limited missing data. The technology is especially useful for genotyping RILs with limited recombination events or in early generations with a high rate of heterozygotes. In addition, genotyping different families with the same set of SNPs in the Beadchip will facilitate identification and comparison of the QTL over different crosses of the RIL populations.

The objective of this study was to create a community resource for dissecting complex traits through the development and genetic characterization of a soybean NAM population.

Materials and Methods

Selection of a Diverse Set of Parents for Nested Association Mapping Population Development

To select a set of diverse parents from soybean maturity groups (MGs) I to V to cross with the hub parent 'IA3023', a high-yielding MG III cultivar developed by Dr. Walter Fehr at Iowa State University, a total of 120 soybean genotypes from the United States, China, Korea, Japan, and other countries (Supplemental Table S1) were identified based on their high yield, diverse ancestry, or drought tolerance. At the time of parent selection for population development, the SoySNP50K had not yet been developed, so the genotypes were analyzed with the Illumina GoldenGate assay containing 1536 SNPs (Hyten et al., 2010). The genotypic dataset was then

used to calculate the pair-wise distance among the 120 genotypes. The distance between a pair of genotypes was defined as the proportion of SNP differences among all SNPs. A dendrogram was then constructed using software MEGA5.1 (Tamura et al., 2011) and a subset of NAM parents was obtained by selecting genotypes based on clusters of the genotypes as well as their agronomic performance, seed yield, and drought tolerance.

DNA Sequence Analysis using the Illumina HiSeq 2000

DNA was isolated from leaf tissue of each high-yielding or exotic parent. Genomic DNA was randomly fragmentased for 20 min using NEBNext dsDNA fragmentase (New England Biolabs). Procedures for DNA end repair and DNA size selection were completed according to the description of the kit (kit number M0348L, NEBNext dsDNA Fragmentase). DNA with fragments of ~500 bp was selected. Paired-end sequencing was used to obtain 150 bp of sequence from each end of the genomic fragments using the Illumina HiSeq 2000. The Illumina CASAVA V1.80 software was used to map the short reads to the whole-genome sequence of the cultivar Williams 82, Glyma1.01 (Schmutz et al., 2010) and to identify SNPs. At least three reads covered each locus. When the second assembly of the soybean whole-genome sequence (Wm82.a2.v1) was available (http://www.phytozome.org/), the positions of these SNPs in the new assembly were also determined.

Design of the Illumina Infinium SoyNAM6K BeadChip

The design of the SoyNAM6K BeadChip, which contained a different set of 6000 SNPs than the BARCSoySNP6K Beadchip, followed the procedures previously described by Song et al. (2013). The SNPs with A/T or G/C alleles, SNPs with Ns in the 60 nt of flanking sequence, SNPs residing within 25 nt of another SNP, or SNPs with 25 nt of a flanking sequence that was not unique in the genome were eliminated. The SNP design scores were evaluated using Illumina's Assay Design Tool (http://www.illumina.com/downloads/Illumina_Assay_Design_Tool.pdf) and SNPs with a design score <0.6 were excluded. For the purpose of identifying a set of SNP markers that would segregate in >70% of the 40 NAM families, SNPs carrying the IA3023 parental allele in >30% of the 40 non-IA3023 parents were eliminated. The algorithm and script to maximize the distance between adjacent SNPs were then performed as described by Song et al. (2013), except that the SNP selection index was defined as the product of the SNP design score and the proportion of the 40 non-IA3023 parents that did not carry the IA3023 allele.

Genotyping Recombinant Inbred Lines of the Nested Association Mapping Families with the SoyNAM6K BeadChip

The 5600 F_5–derived RILs made up of 140 RILs from each of the 40 families in the NAM population and the parents were genotyped with the SoyNAM6K BeadChip using the Illumina Infinium HD Assay platform (Illumina, Inc.). The procedures described in the Infinium II assay protocol were followed for the preparation of DNA and the bead assay, hybridization, staining of samples and image scanning. The SNP alleles were called using the GenomeStudio Genotyping Module v1.8.4 (Illumina, Inc.). Only those SNPs with two or three discrete clusters and both alleles with high signal intensity on the SNP Graph Alt were included for the analysis.

In addition, a pair-wise distance matrix among the 140 RILs within each NAM family was calculated and used to identify lines that were not RILs but were the result of self-hybridization.

Genotyping the Parents of the Nested Association Mapping Population with the SoySNP50K BeadChip

The 41 parents were also assayed with the SoySNP50K BeadChip (Song et al., 2013) using the procedure described above. These SNPs, together with the SNPs obtained via genome sequencing were projected onto the RILs using the SoyNAM6K BeadChip marker results.

Construction of Genetic Linkage Maps

Linkage maps were constructed for each of the RIL families based on the dataset of RILs genotyped with the SoyNAM6K BeadChip. The SNP loci with segregation distortion significant at the $p = 0.001$ probability level were eliminated before linkage analysis. A chi-square goodness of fit test was used to identify the loci with segregation distortion based on the expected ratio of 0.46875:0.0625:0.46875 for homozygote A/heterozygotes/homozygote B in the F_5–derived RIL family, respectively. For the purpose of reducing the time to calculate linkage distance and the linkage map position of SNPs, loci with identical segregation patterns were identified, and only one SNP from each identical segregation group was included in the genetic linkage analysis. The position of the excluded loci was then assigned to the same position as the representative locus of the group. Maps were constructed using the software JoinMap 4.0 (Van Ooijen, 2006).

A composite linkage map across all NAM families was also created using the JMP Genomics Genetics software in SAS JMP Genomics 7 (SAS Institute, 2014). Prior to the analysis, markers that did not segregate within a family were coded as missing data for the family. The map for each chromosome was created separately using the subset of markers annotated as corresponding to each chromosome based on the Wm82.a2.v1 assembly. The physical position of the SNPs on the Wm82.a2.v1 assembly was determined by alignment of the 60-bp sequence flanking the SNPs on the Glyma1.01 assembly. Recombination rates were determined using the Recombination and Linkage Groups process under the category Linkage Maps and QTL. The major parameters used for this process were as follows: *p*-value cutoff for segregation

test plots = 0.05, linkage grouping method = automated hierarchical clustering, and automated linkage group clustering method = average. The order of the markers was determined using the Linkage Map Order process. The parameters used for this process are as follows: map function = Kosambi, order algorithm = map order optimization, nearby marker recombination constraint = 0.4, and apply stopping rules for map order optimization.

Counting Recombination Events

Because of the limited number of DNA crossovers that occurred on each chromosome of the RILs, SNP alleles from one parent are usually consistent along a chromosome until a RE is reached, and then the allele changes to the variant from the other parent. Thus, the number of REs was counted based on the number of allele pattern changes along the 20 chromosomes of each RIL whether it is in a single SNP or multiple SNPs. The number of unique REs within a family is the number of the REs that have unique crossover breakpoints among all of the RILs in the family.

Single-Nucleotide Polymorphism Annotation

High-confidence SNPs used to impute the SNP alleles in the RILs of the NAM families were annotated according to the gene models of the soybean whole-genome sequence assembly (Wm82.a2.v1) provided by the Joint Genome Institute, US Department of Energy (https://phytozome.jgi.doe.gov/pz/portal.html). A perl script was developed based on the methods of Ding et al. (2008) and McLendon et al. (2008) and was used to annotate the putative functions of SNPs (Goettel et al., 2014). The SNPs within gene boundaries were annotated according to the genic region in which they were found. For SNPs identified in coding sequence, the alleles were substituted into the relevant codon and tested for amino acid changes.

Results

Selection of a Diverse Set of Forty Nested Association Mapping Parents to Cross with IA3023

Of the 1536 SNPs in the GoldenGate assay, a total of 1364 were polymorphic among the 120 genotypes. A neighbor-joining tree derived from the analysis of the 120 genotypes (Supplemental Table S1) with the 1364 SNPs was constructed and a total of 39 diverse genotypes from major clusters were selected to cross with IA3023 for the creation of RIL families (Fig. 1). In addition, LG00-3372, which was not analyzed with the GoldenGate SNPs, was included as the 40th parent. Genomic DNA of these 40 parents was sequenced for SNP discovery. Based on the GoldenGate analysis, the average distance among the 39 selected parents was 0.29 ranging from 0.10 to 0.42. The average distance of the 39 parents to the IA3023 parent was 0.18 and ranged from 0.10 to 0.31. Among the 40 parents, one was from MG I, 10 from MG II, 20 from MG III, eight from MG IV, and one from MG V. These included 17 high-yielding experimental lines or cultivars and 15 experimental lines with exotic ancestry (LG prefix). These lines were developed in breeding programs in Illinois, Indiana, Ohio, Missouri, Nebraska, Michigan, and Tennessee. There were also eight accessions from the USDA Soybean Germplasm Collection that originated from five countries that were selected for high yield under drought conditions (Supplemental Table S1). The 40 genotypes were crossed to the hub parent IA3023 and 140 F_5–derived RILs per family were developed through single-seed descent. IA3023 served as the female parent in 38 of the 40 matings, but in two matings (NAM02 and NAM06), IA3023 was served as the male parent.

Sequencing of Nested Association Mapping Parents for Single-Nucleotide Polymorphism Discovery

Approximately 497,089 Mb of DNA sequence data were obtained from the 41 NAM parents. This included 123,132 Mb from the common parent IA3023 with 127× sequence coverage and 373,957 Mb from the other 40 parental lines with sequence coverage from 4× to 15× (Table 1). The paired-end sequence reads from each genotype were aligned to the soybean whole-genome sequence Glyma1.01. Following alignment, a total of 5,232,558 variants were identified. The percentage of SNPs with minor allele frequency <0.05 and 0.10 were 13.7 and 33.5%, respectively (Fig. 2). There were 1,033,848 SNP singletons among the 41 parents. The average proportion of singletons was 2.4% per parent but ranged from 1.4% in U03-100612 to 5.0% in PI518751. Of the 5,232,558 SNPs, 2,581,741 SNPs had ambiguous calls of <30% and the presence of alleles in at least two of the 41 genotypes.

Development of the SoyNAM6K BeadChip

After elimination of the SNPs based on the criteria for the BeadChip design, a total of 30,174 SNPs remained and were candidates for the selection of 6000 SNPs to be included in the SoyNAM6K BeadChip (Supplemental Table S2). The number of the selected 6000 SNPs from each chromosome varied from 212 to 472 with a density of 0.12 to 0.22 Mb SNP^{-1}. Approximately 90% of these SNPs were chosen from euchromatic regions of the soybean genome (Table 2). Of the 6000 SNPs submitted to Illumina, 5303 SNPs were present as bead types on the SoyNAM6K BeadChip. The SoyNAM6K BeadChip was first evaluated with a set of 96 DNA samples that included the RILs from the NAM02 family and their parents IA3023 and TN05-3027. Genotypic data were obtained from all the bead types included on the SoyNAM6K BeadChip.

Genotyping 5600 RILs of the Nested Association Mapping Population with the SoyNAM6K BeadChip

A total of 140 RILs from each of the 40 NAM families were genotyped with the SoyNAM6K BeadChip. Successful genotyping data were obtained for 4312 SNPs of

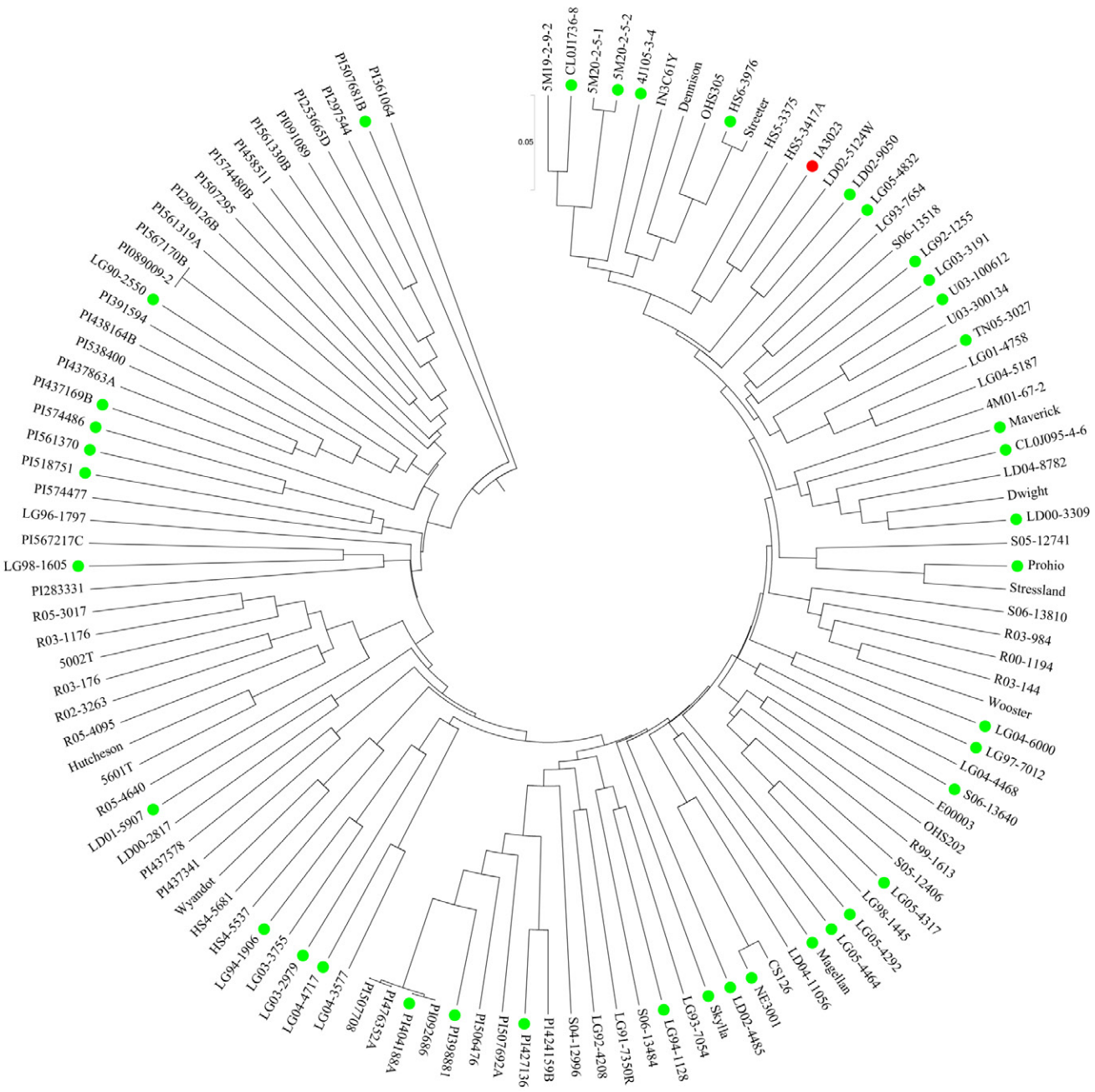

Fig. 1. Phylogenic tree with 120 soybean genotypes based on 1364 SNPs. The genotypes with a green dot were used as parents to cross with IA3023 and to develop families for the nested association mapping population.

the 5303 bead types included on the SoyNAM6K BeadChip. Analysis of the 140 RILs of each family showed that 424 RILs had either identical SNP genotypes as their female parent or had a high rate of segregating loci that were nonparental. This included all 140 RILs from the NAM46 family, as 20.5% of the 3845 polymorphic loci in the NAM46 family carried nonparental alleles, suggesting that the lines in the family were not from the designated parents. Thus, there were a total of 5176 RILs after elimination of the 424 RILs. The remaining number of RILs varied from 98 to 140 among the 39 NAM families (Table 3). Within each family, the ratio of RILs with two different homozygous alleles was evaluated for each locus. After elimination of nonpolymorphic SNPs and SNPs with severe segregation distortion, that is, MAF < 10%, the number of polymorphic loci per NAM family ranged from 2470 (NAM25) to 3791 (NAM48) (Table 3).

Recombination Events and Genetic Linkage Maps of the Nested Association Mapping Population

The polymorphic SNPs were used to determine the REs in each family. A total of 302,329 REs were observed among the 5176 RILs in the NAM population and the average number of REs per line varied from 43.5 in NAM09 to

Table 1. Sequence data (Mb) and sequence coverage obtained for each of the 41 nested association mapping (NAM) parents.

Parents	NAM family ID	Yield	Sequence coverage	Parents	NAM family ID	Yield	Sequence coverage
		Mb	×			Mb	×
4J105-3-4	NAM03	12,112	12.5	LG94-1128	NAM33	6025	6.2
5M20-2-5-2	NAM04	8210	8.5	LG94-1906	NAM34	8126	8.4
CL0J095-4-6	NAM05	9641	9.9	LG97-7012	NAM36	6003	6.2
CL0J173-6-8	NAM06	6375	6.6	LG98-1605	NAM37	5344	5.5
HS6-3976	NAM08	9026	9.3	Magellan	NAM14	3890	4.0
IA3023		123,132	127.0	Maverick	NAM15	5750	5.9
LD00-3309	NAM10	12,443	12.8	NE3001	NAM18	8535	8.8
LD01-5907	NAM11	5915	6.1	PI398881	NAM40	10,717	11.0
LD02-4485	NAM12	7565	7.8	PI404188A	NAM54	11,533	11.9
LD02-9050	NAM13	6191	6.4	PI427136	NAM41	13,702	14.1
LG00-3372	NAM38	6819	7.0	PI437169B	NAM42	15,076	15.5
LG03-2979	NAM24	13,641	14.1	PI507681B	NAM46	11,330	11.7
LG03-3191	NAM25	11,802	12.2	PI518751	NAM48	14,566	15.0
LG04-4717	NAM26	9997	10.3	PI561370	NAM50	13,670	14.1
LG04-6000	NAM39	6605	6.8	PI574486	NAM64	9345	9.6
LG05-4292	NAM27	11,725	12.1	Prohio	NAM09	10,322	10.6
LG05-4317	NAM28	9464	9.8	S06-13640	NAM17	6206	6.4
LG05-4464	NAM29	9571	9.9	Skylla	NAM22	7004	7.2
LG05-4832	NAM30	9412	9.7	TN05-3027	NAM02	11,035	11.4
LG90-2550	NAM31	5441	5.6	U03-100612	NAM23	14,596	15.0
LG92-1255	NAM32	9227	9.5	Total		497,089	

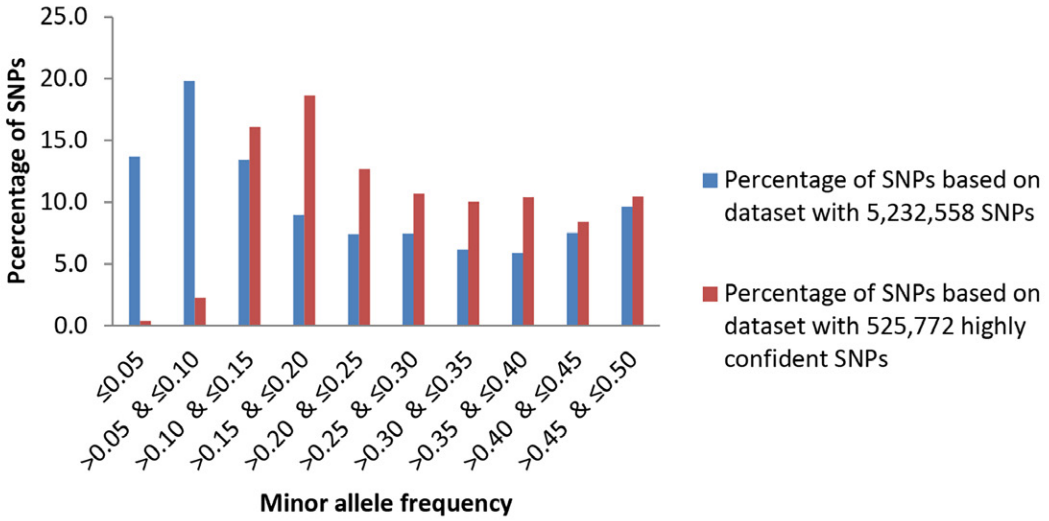

Fig. 2. Distribution of minor allele frequency of single-nucleotide polymorphisms (SNPs) based on the datasets with 5,232,558 and 525,772 SNPs.

77.0 in NAM41. The average number of REs/RILs in the 5176 RILs was 58.5 with a standard deviation of 32. The number of REs for the majority of the RILs was quite consistent, for example, the number of REs was <100 for 4911 RILs (Fig. 3). There was only a small proportion of the RILs (1.6%) with REs >164, a threshold at the 0.001 probability level (Supplemental Table S3). These RILs were observed in 28 of the 39 NAM families, and the number varied from 1 to 9 among the families. These RILs were excluded from the linkage map analysis. Of the 302,329 REs, a total of 89,742 (30%) were unique, meaning that these REs only occurred in one RIL within a family, and 212,587 (70%) overlapped in at least two RILs within a family (Table 3). The average number of unique REs per RIL was 17.5 among the 39 NAM families.

The length of the genetic linkage maps of the NAM families varied from 942.1 to 1763.8 cM (Table 3; Supplemental Table S4) and the total length of the composite linkage maps based on the RILs from all 39 families was 1735 cM (Supplemental Table S5). Some of the difference

Table 2. Density of single-nucleotide polymorphisms (SNPs) selected for inclusion in the SoyNAM6K BeadChip in euchromatic and heterochromatic regions of the 20 soybean chromosomes and the mean density of SNPs in each chromosome of the Wm82.a2.v1 assembly.

Chromosome	No. of SNPs	No. of SNPs in euchromatic regions	No. of SNPs in heterochromatic regions	Density of SNP markers in euchromatic region	Density of SNP markers in heterochromatic region	Density of SNP markers in each chromosome	Density of SNPs in the long arm of euchromatic region	Density of SNPs in the short arm of euchromatic region
				Mb SNP^{-1}				
Gm01	260	178	82	0.083	0.501	0.22	2.025	0.039
Gm02	318	316	2	0.083	12.670	0.16	0.073	0.107
Gm03	286	227	59	0.083	0.490	0.17	0.144	0.048
Gm04	237	226	11	0.083	2.763	0.21	0.208	0.048
Gm05	275	274	1	0.083	19.139	0.15	0.095	0.063
Gm06	323	265	58	0.083	0.494	0.16	0.072	0.324
Gm07	351	331	20	0.083	0.854	0.13	0.059	0.330
Gm08	376	375	1	0.083	15.787	0.12	0.062	2.770
Gm09	270	211	59	0.083	0.496	0.17	0.088	0.077
Gm10	345	291	54	0.083	0.495	0.15	0.140	0.041
Gm11	294	293	1	0.083	14.805	0.13	0.105	0.067
Gm12	212	206	6	0.083	3.829	0.19	0.149	0.056
Gm13	360	355	5	0.083	2.970	0.12	0.083	.
Gm14	279	244	35	0.083	0.839	0.18	0.304	0.046
Gm15	290	281	9	0.083	3.062	0.18	0.074	0.149
Gm16	240	212	28	0.084	0.703	0.16	0.184	0.052
Gm17	258	243	15	0.083	1.444	0.16	0.170	0.037
Gm18	472	442	30	0.083	0.856	0.13	0.061	0.154
Gm19	337	329	8	0.083	2.902	0.15	0.097	0.064
Gm20	217	214	3	0.083	9.663	0.22	2.431	0.015
Total	6000	5513	487					

in map length can be explained by variation in the number of markers segregation in families. For example, the family NAM25 had the shortest map length and the lowest number of segregating markers. The order of the markers was largely consistent among families (Supplemental Fig. S1). The order of the loci in composite linkage map was generally consistent with their physical position in the Williams 82 whole-genome sequence (Wm82.a2.v1 assembly) based on either each family (Supplemental Table S4) or all families (Supplemental Fig. S2).

Segregation Distortion of Single-Nucleotide Polymorphisms among Families and Genomic Regions

Of the 126,859 polymorphic loci observed in 39 families, a total of 4722 (or 3.75%) SNPs exhibited significant segregation distortion at $P < 0.01$. The correlation of the segregation distortion percentage between the euchromatic and heterochromatic regions was significant among families ($r = 0.9405$, $P < 0.0001$). The average percentage of the segregation distortion in the euchomatic and heterochromatic regions across families was 3.7 and 4.2%, respectively ($t = 0.026$, $P = 0.979$) (Table 4), thus significant differences in segregation distortion in the two regions were not detected. However, segregation bias varied among families and occurred nonrandomly within the euchromatic and heterochromatic regions (Table 4; Fig. 4), for example, percentage of SNPs with segregation distortion was 13.2 and 53.5% in the NAM04 and NAM25, respectively, but ranged from 0.39 to 8.97% among other families. Clusters of segregation distortion loci were observed in the euchromatic regions of chromosome 1 and 2, and heterochromatic regions of chromosome 3 and 7 (Fig. 4). Among the 39 families, 26 had more SNPs with an over representation of the IA3023 allele for loci with significant segregation distortion than the non-IA3023 parent allele, while only 13 families had more SNPs with non-IA3023 parent alleles than IA3023 alleles. The percentage of SNPs with segregation distortion favoring the non-IA3023 parent were from families for which the non-IA3023 parent was from MG I, II, III, IV, and V and had segregation of 1.2, 2.1, 1.7, 4.2, and 0.7%, respectively.

Residual Heterozygousity in Euchromatic and Heterochromatic Regions

The residual heterozygousity (RH) in the F_5 plants used to derive the RILs averaged 6.5% in the euchromatic regions and 6.6% in the heterochromatic regions of the NAM families, which was not significantly different. This closely matches the expected rate of 6.25% for the F_5 plants. In addition, the correlation of the RH between the two regions across the NAM families (Table 5) was 0.91, which was highly significant.

Table 3. Number of recombinant inbred lines (RILs), polymorphic loci, recombination events (REs), and total length of genetic distance in each nested association mapping (NAM) family.

Population ID	No. of RILs remaining after quality control	No. of polymorphic loci in each family	Total number of RE	Average number of REs per RIL	No. of unique REs among RILs	Proportion of total unique REs among RILs in family	No. of REs occurring in at least two RILs in family	No. of unique REs per RIL in family	Total length of genetic distance
									cM
NAM02	98	3191	6461	65.9	2427	0.38	4034	25	1224.2
NAM03	137	3253	7712	56.2	1820	0.24	5892	13	1289.0
NAM04	136	3204	7249	53.3	1768	0.24	5481	13	1252.2
NAM05	139	3389	6928	49.8	1709	0.25	5219	12	1343.4
NAM06	140	3494	6730	48.1	1516	0.23	5214	11	1159.8
NAM08	138	3267	6463	46.8	1376	0.21	5087	10	1046.3
NAM09	137	2864	5970	43.5	989	0.17	4981	7	1040.9
NAM10	139	3410	7558	54.3	1786	0.24	5772	13	1392.2
NAM11	124	3113	6863	55.3	1680	0.24	5183	14	1298.1
NAM12	138	3528	7562	54.7	1989	0.26	5573	14	1467.1
NAM13	137	3113	6926	50.5	1599	0.23	5327	12	981.2
NAM14	137	3088	7655	55.8	2167	0.28	5488	16	1171.6
NAM15	139	3435	9933	71.4	2836	0.29	7097	20	1340.3
NAM17	136	3084	6607	48.5	1716	0.26	4891	13	1218.7
NAM18	136	3203	7311	52.9	2819	0.39	4492	21	1376.2
NAM22	138	3212	7268	51.9	1527	0.21	5741	11	1448.3
NAM23	140	3393	9219	65.8	1629	0.18	7590	12	1332.0
NAM24	140	3754	7983	65.9	2758	0.35	5225	20	1486.0
NAM25	121	2470	6738	62.3	2065	0.31	4673	17	942.1
NAM26	108	3130	6799	51.9	2541	0.37	4258	24	1345.2
NAM27	131	2631	10,062	72.3	2165	0.22	7897	17	1199.7
NAM28	139	3369	7415	53.7	3028	0.41	4387	22	1425.3
NAM29	138	3322	8457	61.2	2444	0.29	6013	18	1276.5
NAM30	138	3454	7287	57.3	2853	0.39	4434	21	1363.9
NAM31	127	3317	6218	45.1	2596	0.42	3622	20	1521.4
NAM32	138	2849	7875	58.7	2118	0.27	5757	15	1055.1
NAM33	134	3376	8492	63.3	2660	0.31	5832	20	1583.6
NAM34	134	3157	7622	55.6	2605	0.34	5017	19	1504.7
NAM36	137	3111	9621	73.4	2490	0.26	7131	18	1300.9
NAM37	131	3298	7368	55.8	2932	0.40	4436	22	1637.7
NAM38	132	3146	7105	53	2523	0.36	4582	19	1106.3
NAM39	134	2887	8250	60.2	2354	0.29	5896	18	1060.3
NAM40	137	3751	5554	55.5	2977	0.54	2577	22	1401.0
NAM41	100	3007	10245	77	2166	0.21	8079	22	1058.8
NAM42	133	3405	10047	72.2	3065	0.31	6982	23	1633.8
NAM48	139	3791	9009	68.2	3303	0.37	5706	24	1692.6
NAM50	132	3163	8929	63.7	2765	0.31	6164	21	1763.8
NAM54	140	3678	7863	63.4	3126	0.40	4737	22	1420.1
NAM64	124	3552	8975	65.9	2855	0.32	6120	23	1729.9
Total	5176	126,859	302,329	2280.3	89742	–	212,587	682	51,890.0
Average	129.4	3171.5	7558.2	58.5	2301.1	0.3	5450.9	17.5	1330.5

Identification of Single-Nucleotide Polymorphisms among the Parents to Project the Alleles in Lines of the Families

Among the 5,232,558 SNPs identified by sequence analysis, a total of 491,612 were high confidence SNPs that had a high MAF and with unique 60-bp flanking sequence in the soybean genome. The rate of missing alleles or heterozygous alleles of these SNPs was <10% among the 40 NAM parents. As the sequence flanking these SNPs was specific in the soybean genome, the SNPs were unlikely to have resulted from the alignment of sequence reads from homeologs in the soybean genome.

Genotyping of the NAM parents with the SoySNP50K BeadChip showed that 42,509 SNPs were polymorphic among the NAM parents. Of the 491,612 SNPs obtained via genome sequencing or genotyped with

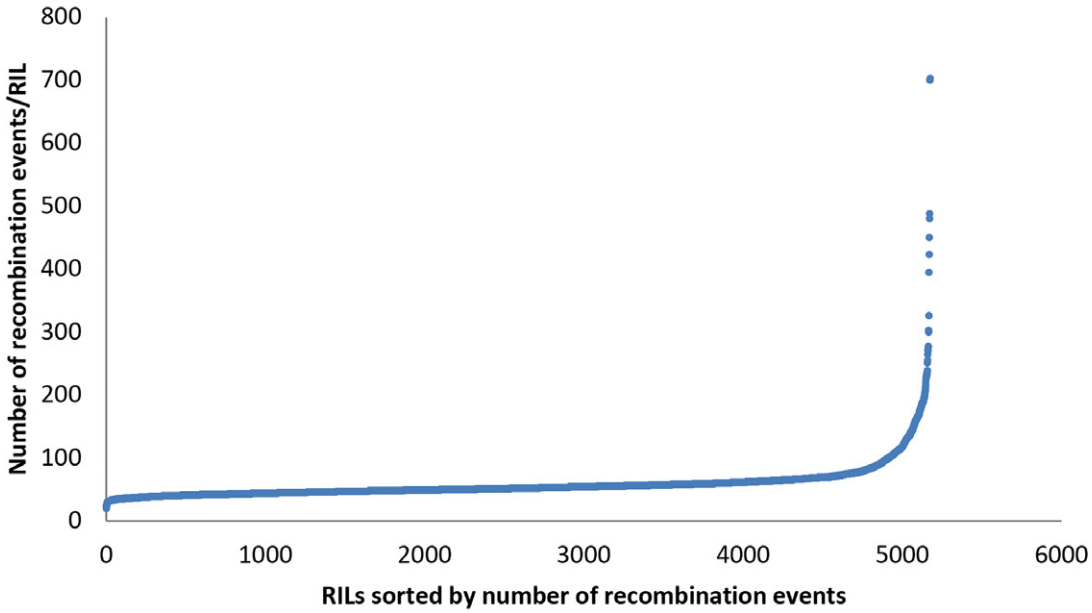

Fig. 3. Number of recombination events vs. each recombinant inbred line (RIL) sorted by the number of recombination events.

SoySNP50K BeadChip, only 8349 were common to the two analyses, thus, a total of 525,772 unique SNPs from the two datasets could be used to project the SNP alleles in the RILs of the NAM families. The low number of the common SNPs in the two analyses was due to the elimination of the SNPs with low MAF among the parents in the genome sequencing analysis. The SNPs with MAF of 0.05 and 0.10 were 0.3 and 2.5%, respectively (Fig. 2), thus, 97.5% of the SNPs were highly polymorphic among the parents. In addition, of the 42,509 SNPs in the SoySNP50K, 170 were common with the SoyNAM6K SNPs.

Annotation of the 525,772 SNPs showed that 73% of the SNPs were in intergenic regions and 27% in genic regions. Of the SNPs in the genic regions, 61 and 24% were in introns and coding DNA sequence, respectively, and 46 and 54% of the coding DNA sequence SNPs were synonymous and nonsynonymous SNPs, respectively (Supplemental Fig. S3).

Data Availability

The short read sequences for the 41 NAM parents are available at the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) with accession number SRP042221 and the genotyping datasets for the NAM RILs and parents, as well as for the NAM parents assayed with SoySNP50K BeadChip are available at SoyBase, the USDA–ARS Soybean Genetics and Genomics Database (http://www.soybase.org/SoyNAM/soynamdetails.php). The SNPs for genotyping the parents and RILs were deposited at the NCBI dbSNP (http://www.ncbi.nlm.nih.gov/snp).

Discussion

The NAM design takes advantage of genetic variation, historical REs among multiple parental lines, the power of linkage within families, a large population size, and a high density of markers for the progeny. Thus, it is powerful for complex trait dissection (Li et al., 2011). For the purpose of fine mapping in NAM families, the NAM parents are genotyped using high-throughput sequencing or bead array analysis, while segregating progenies are genotyped with low-density markers (Yu et al., 2008). Parental genotypes with large numbers of SNPs are then projected to their progeny. In maize, 26 parents with half a million SNPs were successfully projected to their 5000 progenies using 1106 SNPs genotyped for the parents and progeny (Guo and Beavis, 2011). Approximately 90% of the genotyped SNPs were assigned linkage map positions using linear interpolation between the maize Accessioned Gold Path and the NAM linkage maps. Approximately 70% of the SNPs provided high probability estimates of genotypes in almost 5000 RILs (Guo and Beavis, 2011). The most recent report on the maize NAM showed that 7000 SNPs genotyped through genotyping-by-sequencing in the NAM RILs were used to project the 28.9 million SNPs from the NAM parents onto the 5000 RIL progeny (Wallace et al., 2014).

In this study, a total of 140 RILs from each of the 40 NAM families were genotyped with 4312 SNP markers. In addition, the 41 parents of the NAM families were sequenced and genotyped with the SoySNP50K BeadChip containing 42,509 SNPs. The resulting 525,772 high confidence SNPs will be used to project the SNP alleles present in the RILs. The RILs in the soybean NAM population and their parents have been evaluated for yield, morphological traits, agronomic traits, seed composition, physiological traits, and disease resistance in multiple environments in MG II to IV by public soybean breeders from Ohio to Nebraska, and thus, the genotyping datasets of the RILs and the NAM parents are being used to map the genes or QTL controlling these traits. It is anticipated that the large

Table 4. Number of polymorphic loci and loci with segregation distortion in euchromatic and heterochromatic regions of the nested association mapping (NAM) families.

Family	No. of polymorphic loci in the euchromatic regions	No. of polymorphic loci in the heterochromatic regions	No. of SNPs with segregation distortion in the euchromatic regions	No. of SNPs with segregation distortion in the heterochromatic regions	Percentage of SNPs with segregation distortion in euchromatic regions	Percentage of SNPs with segregation distortion in heterochromatic regions	Percentage of SNPs with segregation distortion in euchromatic and heterochromatic regions	No. of SNPs with segregation distortion SNPs and with more of the non-IA3023 parent genotype	No. of SNPs with segregation distortion SNPs and with more of the IA3023 parent genotype
NAM02	2975	206	21	0	0.71	0.00	0.66	12	9
NAM03	3014	201	143	12	4.74	5.97	4.82	59	96†
NAM04	3020	181	393	29	13.01	16.02	13.18	97	307†
NAM05	3149	234	14	9	0.44	3.85	0.68	17	6
NAM06	3239	249	75	3	2.32	1.20	2.24	53	25
NAM08	3055	191	89	22	2.91	11.52	3.42	78	33
NAM09	2698	160	114	1	4.23	0.63	4.02	102	13
NAM10	3171	229	65	10	2.05	4.37	2.21	32	43†
NAM11	2875	214	44	17	1.53	7.94	1.97	18	43†
NAM12	3334	179	41	13	1.23	7.26	1.54	28	26
NAM13	2867	213	140	22	4.88	10.33	5.26	115	47
NAM14	2899	164	103	11	3.55	6.71	3.72	71	43
NAM15	3174	214	144	3	4.54	1.40	4.34	26	120†
NAM17	2886	179	92	1	3.19	0.56	3.03	1	91†
NAM18	3061	140	283	4	9.25	2.86	8.97	186	95
NAM22	2962	163	61	1	2.06	0.61	1.98	13	47†
NAM23	3134	244	40	2	1.28	0.82	1.24	7	19†
NAM24	3546	201	198	3	5.58	1.49	5.36	127	69
NAM25	2188	138	1158	87	52.93	63.04	53.53	611	616†
NAM26	2858	210	12	0	0.42	0.00	0.39	2	10†
NAM27	2477	130	44	1	1.78	0.77	1.73	11	34†
NAM28	3183	179	90	2	2.83	1.12	2.74	7	85†
NAM29	3082	172	20	3	0.65	1.74	0.71	8	15†
NAM30	3293	157	61	0	1.85	0.00	1.77	51	10
NAM31	3072	239	14	1	0.46	0.42	0.45	3	12†
NAM32	2676	165	59	3	2.20	1.82	2.18	6	56†
NAM33	3174	194	107	4	3.37	2.06	3.30	9	102†
NAM34	2875	239	32	3	1.11	1.26	1.12	6	29†
NAM36	2935	161	17	0	0.58	0.00	0.55	2	15†
NAM37	3052	236	178	9	5.83	3.81	5.69	25	162†
NAM38	2826	221	32	0	1.13	0.00	1.05	3	29†
NAM39	2754	124	65	0	2.36	0.00	2.26	15	50†
NAM40	3539	207	79	3	2.23	1.45	2.19	64	18
NAM41	2828	177	53	2	1.87	1.13	1.83	47	8
NAM42	3225	173	31	15	0.96	8.67	1.35	4	42†
NAM48	3569	217	80	3	2.24	1.38	2.19	13	65†
NAM50	2926	229	31	5	1.06	2.18	1.14	3	33†
NAM54	3414	260	136	2	3.98	0.77	3.76	19	119†
NAM64	3331	210	42	15	1.26	7.14	1.61	3	54†
Total	118336	7600	4401	321	158.61	182.31	160.18	1954	2696
Mean	3034.3	194.9	112.8	8.2	4.07	4.67	4.11	50.1	69.1

† Families containing more segregation distortion SNPs with IA3023 alleles.

sample size and dense markers will result in the definition of numerous QTL controlling these important traits. Furthermore, the NAM collaborators will continue to grow a portion of the NAM lines each year to be sure viable seeds will be available. This work will be supported by charging a small fee to cover the cost of shipping and regenerating the seed. Because the seeds of the RILs and NAM parents are publicly available for distribution (http://www.soybase.org/SoyNAM/SoyNAM_seed_0315.htm), new traits can be evaluated by interested researchers.

We observed that 91 RILs from a number of different families had identical marker scores. All of these RILs were

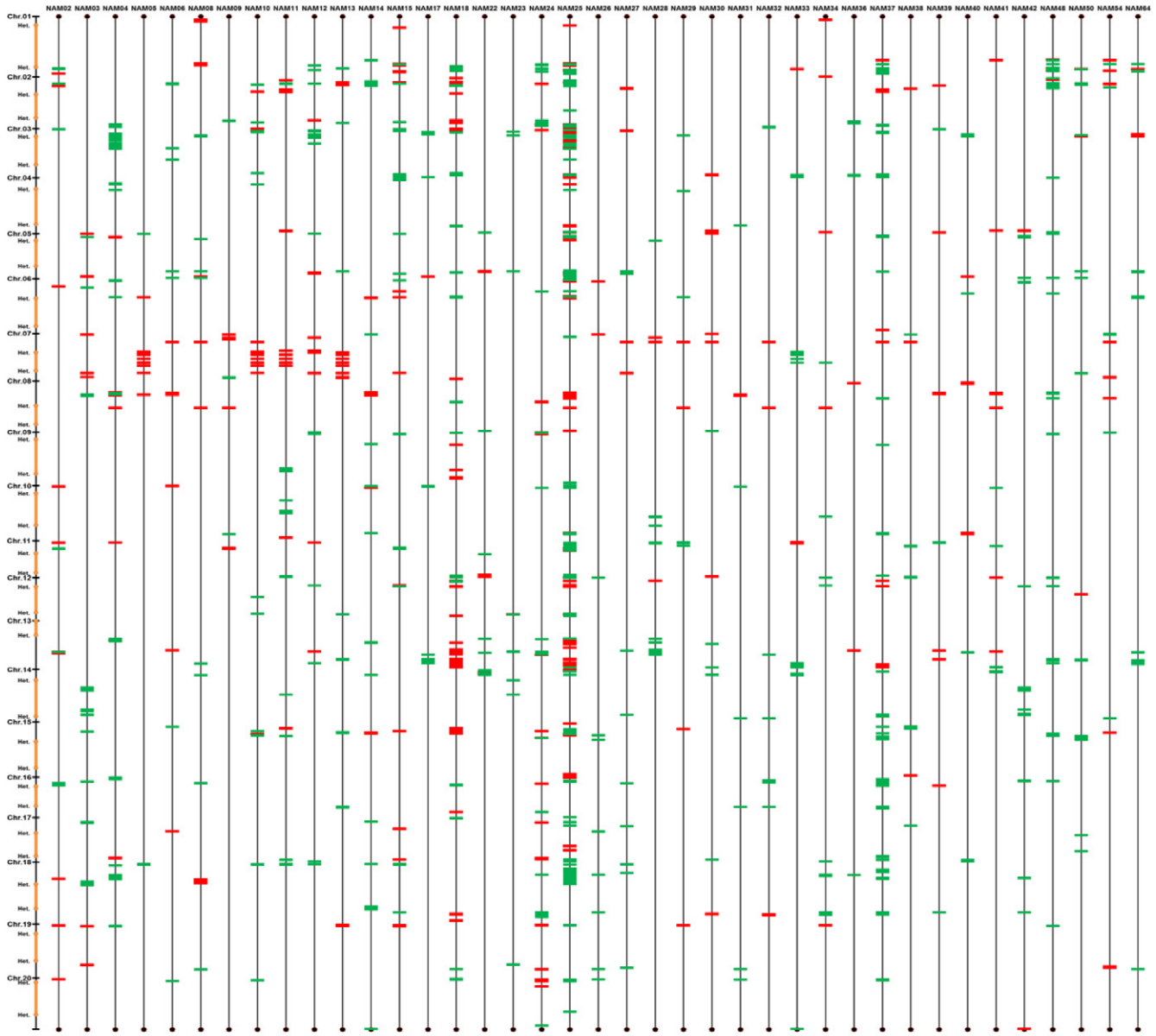

Fig. 4. Distribution of single-nucleotide polymorphisms (SNPs) with segregation distortion in euchromatic and heterochromatic regions of chromosomes across families. Red bar, SNP with bias favoring the IA3023 allele; green bar, SNP with bias favoring the non-IA3023 parent allele.

genetically identical to their female parent in each population. Examination of morphological traits further verified that these RILs were likely selfed progeny of the female parent, thus, these RILs were eliminated from the dataset. In the NAM46 family, ~20% of the 3845 polymorphic loci contained nonparental alleles, and it thus appeared that one of the parents was incorrect. To identify the correct male parent, we inferred the SNP alleles of the non-IA3023 parent based on SNP alleles in the RILs and searched the accessions in the USDA Soybean Germplasm Collection genotyped with SoySNP50K BeadChip for an accession that matched the inferred genotype. However, we failed to find an accession matching the inferred genotype. In the accession that most closely matched the inferred genotype, 8% of the loci differed from the inferred parent. Thus, we eliminated the genotypic data for the entire NAM46 family.

Although a significant difference for segregation distortion between euchromatic and heterochromatic regions was not found in the soybean genome, segregation distortion varied among families and regions of the genome. We did not find candidate QTL or genes for the distortion in the NAM04 and NAM25 families or in chromosomal regions with high distortion rates. The maturity of IA3023 and non-IA3023 parents may not be a factor either, as IA3023 is a MG III cultivar and the percentage of segregation bias in the families with the second parents from MGs I and V was even lower than for families with MG II, III, and IV parents. In two-thirds of the families, we observed a bias of the allele from the IA3023 parent being favored in distorted loci and in only approximately one-third of the families were the non-IA3023 parent alleles favored. This suggests the exclusion of pollen during selfing or the effect of unknown

Table 5. Percentage of heterozygotes in euchromatic and heterochromatic regions by family.

Family	No. of homozygotes with non-IA3023 alleles in euchromatic regions	No. of heterozygotes in euchromatic regions	No. of homozygotes with IA3023 alleles in euchromatic regions	No. of homozygotes with non-IA3023 alleles in heterochromatic regions	No. of heterozygotes in heterochromatic regions	No. of homozygotes with IA3023 alleles in heterochromatic regions	Percentage of heterozygotes in euchromatic regions	Percentage of heterozygotes in heterochromatic regions
NAM02	137,376	18,164	130,571	9,478	1046	8,904	6.3	5.4
NAM03	188,469	29,946	191,339	11,839	1578	13,159	7.3	5.9
NAM04	182,766	31,773	192,983	10,400	1774	11,753	7.8	7.4
NAM05	205,415	29,670	199,743	15,073	2316	14,178	6.8	7.3
NAM06	212,523	33,573	204,892	15,658	2654	15,528	7.4	7.8
NAM08	191,085	29,414	198,974	13,107	1612	10,875	7.0	6.3
NAM09	177,937	23,231	166,656	10,077	1176	10,052	6.3	5.5
NAM10	198,273	33,514	204,878	13,786	2251	14,749	7.7	7.3
NAM11	161,501	27,643	164,739	11,481	2165	12,022	7.8	8.4
NAM12	214,352	31,303	210,570	11,076	1658	11,259	6.9	6.9
NAM13	183,664	28,315	177,599	12,647	2464	13,030	7.3	8.8
NAM14	181,339	30,663	178,459	10,158	2074	9,930	7.9	9.4
NAM15	199,386	29,286	202,641	13,125	2098	13,030	6.8	7.4
NAM17	171,979	30,382	186,886	10,605	2033	11,367	7.8	8.5
NAM18	190,891	34,747	177,333	7,981	1603	8,380	8.6	8.9
NAM22	189,155	32,733	184,624	9,771	1505	10,876	8.1	6.8
NAM23	206,480	31,527	198,523	15,663	2084	15,575	7.2	6.3
NAM24	218,330	38,385	228,951	12,165	2030	12,586	7.9	7.6
NAM25	110,672	34,437	113,319	6,971	2100	6,322	13.3	13.6
NAM26	138,123	15,769	146,163	10,072	1055	10,773	5.3	4.8
NAM27	147,170	16,572	151,658	7,593	844	7,981	5.3	5.1
NAM28	157,236	19,393	168,471	8,699	895	9,327	5.6	4.7
NAM29	199,634	23,375	195,512	10,400	1303	10,962	5.6	5.7
NAM30	212,150	26,723	204,617	9,949	1569	9,511	6.0	7.5
NAM31	178,736	20,038	182,849	13,479	1717	13,985	5.3	5.9
NAM32	170,639	18,572	172,970	10,088	1191	10,422	5.1	5.5
NAM33	196,959	20,190	198,420	11,541	1372	12,345	4.9	5.4
NAM34	179,375	19,454	177,061	14,827	1556	14,700	5.2	5.0
NAM36	185,667	16,839	188,461	10,716	869	9,736	4.3	4.1
NAM37	172,152	26,610	197,506	13,509	1984	14,495	6.7	6.6
NAM38	167,609	19,300	177,628	12,418	1687	13,557	5.3	6.1
NAM39	172,478	18,277	168,238	7,450	805	7,890	5.1	5.0
NAM40	220,893	20,788	232,400	11,883	1305	13,888	4.4	4.8
NAM41	133,353	16,567	133,503	8,262	1016	8,303	5.8	5.8
NAM42	190,180	25,534	202,135	9,491	1335	10,924	6.1	6.1
NAM48	216,265	32,361	230,948	13,012	2149	13,666	6.7	7.5
NAM50	175,428	17,434	182,168	13,513	1681	13,398	4.6	5.9
NAM54	226,887	24,489	212,636	16,055	2043	16,041	5.3	6.0
NAM64	190,964	18,690	195,562	10,878	1408	12,896	4.6	5.6
Total	7,153,491	995,681	7,232,586	444,896	64,005	458,375	253.4	258.7
Mean	183,422.8	25,530.3	185,450.9	11,407.6	1641.2	11,753.2	6.5	6.6

genetic factors may play a role. In maize, genetic factors such as Gametophyte factor 1-strong allele (Ga1-S), gametophyte factor (ga2), and sugary 1 (su1) in specific families were related to pollen fertilization or seed germination and were reported to distort segregation (McMullen et al., 2009).

We also observed that the percentage of heterozygous SNP loci in the F_5 plants used to derive RILs in the NAM population was 6.5%, which was close to the expected rate of 6.25% for F_5 plants. In maize, heterozygosity in $F_{5:6}$ NAM progenies was >8% (Hung et al., 2012). Unlike the previous report in the outcrossing crop maize (McMullen et al., 2009), no significant difference for RH was observed between the euchromatic and heterochromatic regions in the soybean genome among all NAM families. This difference between crops is possibly because of differences in reproduction as soybean is a self-pollinated crop and maize is open pollinated. Given the strong heterosis observed in maize, heterochromatic regions with low

recombination rate have higher levels of heterozygosity than euchromatic regions (McMullen et al., 2009).

In the NAM population, the total number of REs was 302,329 and the average number of REs was ~58 per RIL. However, 70% of the REs occurred in at least two RILs within a family and only 30% of the REs (~18 REs per line) were unique to one RIL in a given family. In maize, the total number of REs in 25 NAM families with 4699 RILs was 136,000 and the average number of REs in a line was 29. While the average REs per RIL was higher in soybean than in the maize NAM population (58.5 vs. 29) (Kump et al., 2011), the number is still relatively small. The limited number of REs suggests that a design such as NAM with a large number of progeny is critical for fine mapping of genes and QTL. This also suggests that for a biparental population with a limited number of RILs, it is unnecessary to genotype the RILs with a large number of markers. Although the order of most markers along the linkage maps was consistent among families, some inconsistencies were observed. The inconsistency usually occurred in the regions with sparse markers or regions associated with genotyping errors. The total length of the composite genetic linkage maps for the soybean NAM population was 1736 cM based on all of the RILs in the families. This is larger than the composite map of 1402 cM based on 4699 RILs genotyped with 1106 SNP loci in maize (McMullen et al., 2009). The plot of the composite linkage position vs. physical position of the SNPs not only showed the consistency of the order of SNPs on the linkage map and along the chromosomes of the Wm82.a2.v1 assembly, but also showed the gaps in the soybean genome where no SNPs were included in the SoyNAM6K. Although most of these gaps were in the heterochromatic regions with extremely low recombination rates, some were in the genomic regions with nearly fixed alleles among the 41 NAM parents.

Supplemental Information Available

Supplemental information is available with the online version of this manuscript.

Supplemental Table S1. Origins and maturity group of 121 soybean lines considered as NAM parents accessions.

Supplemental Table S2. ssID, position on the Glyma1.01 and Wm82.a2.v1 assemblies and 60bp sequence flanking SNPs selected for the SoyNAM6K BeadChip.

Supplemental Table S3. Number of recombination events in each recombinant inbred line.

Supplemental Table S4. Linkage group and position of SNPs in each NAM family.

Supplemental Table S5. Composite genetic linkage group and position of the SNPs polymorphic in 39 families of the soybean NAM population.

Supplemental Fig. S1. Consistency of linkage maps of each chromosome across families.

Supplemental Fig. S2. Plot of composite linkage position vs. physical position (Wm82.a2.v1) of the SNPs based on 39 families of the NAM population.

Supplemental Fig. S3. Distribution of SNPs in the genic regions.

Conflict of Interest

The authors declare that there is no conflict of interest.

Acknowledgments

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; David Grant at USDA–ARS, Ames, IA, for making the dataset available at the Soybase. This research was supported with funding from the United Soybean Board Project 2241. The support of the United Soybean Board is greatly appreciated. B.D. and J.S. built consensus among public soybean breeders to conduct a coordinated NAM project. B.D., W.D.B., K.R. and J.S. designed the SoyNAM to minimize the confounding impact of maturity on yield and other agronomic traits in the primary soybean growing region of North America. R.N. selected and provided lines for parents, J.S., and B.D. created NAM populations and provided RILs, Q.S., D.H., J.S., B.D. and P.C. provided planning and coordination for genotyping, C.Q. and E.F. prepared DNA sequence libraries and performed molecular genotyping, Q.S. and S.S. performed DNA sequence analysis, Q. S. and Y.A. performed annotation analysis, Q.S., L.Y., B.D.J., B.S., W.D.B., K.R., J.S., B.D. and P.C. performed data analyses. Q.S. and P.C. designed the BeadChip and prepared the manuscript.

References

Abdel-Haleem, H., T.E. Carter, Jr., L.C. Purcell, C.A. King, L.L. Ries, P. Chen, et al. 2012. Mapping of quantitative trait loci for canopy-wilting trait in soybean (*Glycine max* L. Merr). Theor. Appl. Genet. 125:837–846. doi:10.1007/s00122-012-1876-9

Brummer, E., G. Graef, J. Orf, J. Wilcox, and R. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370–378. doi:10.2135/cropsci1997.0011183X003700020011x

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, et al. 2009. The genetic architecture of maize flowering time. Science 325:714–718. doi:10.1126/science.1174276

Chung, J., H. Babka, G. Graef, P. Staswick, D. Lee, P. Cregan, et al. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053–1067. doi:10.2135/cropsci2003.1053

Cook, J.P., M.D. McMullen, J.B. Holland, F. Tian, P. Bradbury, J. Ross-Ibarra, et al. 2012. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. 158:824–834. doi:10.1104/pp.111.185033

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, et al. 2015a. Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] genotypes. Theor. Appl. Genet. 128:73–91. doi:10.1007/s00122-014-2413-9

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, et al. 2015b. Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. Plant Genome 8:3. doi:10.3835/plantgenome2014.11.0086

Ding, L., G. Getz, D.A. Wheeler, E.R. Mardis, M.D. McLellan, K. Cibulskis, et al. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455:1069–1075. doi:10.1038/nature07423

Goettel, W., E. Xia, R. Upchurch, M.L. Wang, P. Chen, and Y.Q.C. An. 2014. Identification and characterization of transcript polymorphisms in soybean lines varying in oil composition and content. BMC Genomics 15:1. doi:10.1186/1471-2164-15-299

Gore, M.A., J.M. Chia, R.J. Elshire, Q. Sun, E.S. Ersoz, B.L. Hurwitz, et al. 2009. A first-generation haplotype map of maize. Science 326:1115–1117. doi:10.1126/science.1177837

Guo, B., and W.D. Beavis. 2011. In silico genotyping of the maize nested association mapping population. Mol. Breed. 27:107–113. doi:10.1007/s11032-010-9503-4

He, Q., H. Yang, S. Xiang, W. Wang, G. Xing, T. Zhao, et al. 2014. QTL mapping for the number of branches and pods using wild chromosome segment substitution lines in soybean [*Glycine max* (L.) Merr.]. Plant Genet. Resour. 12:S172–S177. doi:10.1017/S1479262114000495

Hung, H., C. Browne, K. Guill, N. Coles, M. Eller, A. Garcia, et al. 2012. The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity 108:490–499. doi:10.1038/hdy.2011.103

Hwang, E.Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, et al. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1. doi:10.1186/1471-2164-15-1

Hyten, D.L., I.K. Choi, Q.J. Song, J.E. Specht, T.E. Carter, and R.C. Shoemaker. 2010. A high-density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for QTL mapping. Crop Sci. 50. doi:10.2135/cropsci2009.06.0360

Jeong, N., S.J. Suh, M.H. Kim, S. Lee, J.K. Moon, H.S. Kim, et al. 2012. *Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. Plant Cell 24:4807–4818. doi:10.1105/tpc.112.104968

Jordan, D., E. Mace, A. Cruickshank, C. Hunt, and R. Henzell. 2011. Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. Crop Sci. 51:1444–1457. doi:10.2135/cropsci2010.06.0326

Kato, S., T. Sayama, K. Fujii, S. Yumoto, Y. Kono, T.Y. Hwang, et al. 2014. A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. Theor. Appl. Genet. 127:1365–1374. doi:10.1007/s00122-014-2304-0

Kump, K.L., P.J. Bradbury, R.J. Wisser, E.S. Buckler, A.R. Belcher, M.A. Oropeza-Rosas, et al. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43:163–168. doi:10.1038/ng.747

Lee, G., H. Boerma, M. Villagarcia, X. Zhou, T. Carter, Jr., Z. Li, et al. 2004. A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor. Appl. Genet. 109:1610–1619. doi:10.1007/s00122-004-1783-9

Lee, S., T. Jun, A.P. Michel, and M.R. Mian. 2014. SNP markers linked to QTL conditioning plant height, lodging, and maturity in soybean. Euphytica 203:521.

Li, H., P. Bradbury, E. Ersoz, E.S. Buckler, and J. Wang. 2011. Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. PLoS ONE 6:E17573. doi:10.1371/journal.pone.0017573

McLendon, R., A. Friedman, D. Bigner, E.G. Van Meir, D.J. Brat, G.M. Mastrogianakis, et al. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068. doi:10.1038/nature07385

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, et al. 2009. Genetic properties of the maize nested association mapping population. Science 325:737–740. doi:10.1126/science.1174320

Nordborg, M., and S. Tavaré. 2002. Linkage disequilibrium: What history has to tell us. Trends Genet. 18:83–90. doi:10.1016/S0168-9525(02)02557-X

Palomeque, L., L. Li-Jun, W. Li, B. Hedges, E.R. Cober, and I. Rajcan. 2009. QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted × high-yielding exotic soybean lines. Theor. Appl. Genet. 119:417–427. doi:10.1007/s00122-009-1049-7

Peiffer, J.A., S.A. Flint-Garcia, N. De Leon, M.D. McMullen, S.M. Kaeppler, and E.S. Buckler. 2013. The genetic architecture of maize stalk strength. PLoS ONE 8:E67066. doi:10.1371/journal.pone.0067066

Pham, A.-T., K. McNally, H. Abdel-Haleem, H.R. Boerma, and Z. Li. 2013. Fine mapping and identification of candidate genes controlling the resistance to southern root-knot nematode in PI 96354. Theor. Appl. Genet. 126:1825–1838. doi:10.1007/s00122-013-2095-8

Poland, J.A., P.J. Bradbury, E.S. Buckler, and R.J. Nelson. 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc. Natl. Acad. Sci. USA 108:6893–6898. doi:10.1073/pnas.1010894108

Rafalski, J.A. 2010. Association genetics in crop improvement. Curr. Opin. Plant Biol. 13:174–180. doi:10.1016/j.pbi.2009.12.004

Rector, B., J. All, W. Parrott, and H. Boerma. 1998. Identification of molecular markers linked to quantitative trait loci for soybean resistance to corn earworm. Theor. Appl. Genet. 96:786–790. doi:10.1007/s001220050803

Rincker, K.L., A.E. Lipka, and B.W. Diers. 2016. Genome-wide association study of brown stem rot resistance in soybean across multiple populations. Plant Genome 9:2. doi:10.3835/plantgenome2015.08.0064

SAS Institute. 2014. JMP Genomics, version 7. SAS Inst. Inc., Cary, NC.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, et al. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. doi:10.1038/nature08670

Schnaithmann, F., D. Kopahnke, and K. Pillen. 2014. A first step toward the development of a barley NAM population and its utilization to detect QTLs conferring leaf rust seedling resistance. Theor. Appl. Genet. 127:1513–1525. doi:10.1007/s00122-014-2315-x

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, et al. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8:E54985. doi:10.1371/journal.pone.0054985

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson et al. 2015. Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes, Genomes, Genet. 50:1999–2006.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739. doi:10.1093/molbev/msr121

Terry, L., K. Chase, T. Jarvik, J. Orf, L. Mansur, and K. Lark. 2000. Soybean quantitative trait loci for resistance to insects. Crop Sci. 40:375–382. doi:10.2135/cropsci2000.402375x

Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, et al. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. 43:159–162. doi:10.1038/ng.746

Van Ooijen, J. 2006. JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, the Netherlands.

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

Wallace, J.G., P.J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, and E.S. Buckler. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. 10:E1004845. doi:10.1371/journal.pgen.1004845

Wang, D., G. Graef, A. Procopiuk, and B. Diers. 2004. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor. Appl. Genet. 108:458–467. doi:10.1007/s00122-003-1449-z

Wang, X., G.L. Jiang, M. Green, R.A. Scott, Q. Song, D.L. Hyten, et al. 2014. Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Mol. Genet. Genomics 289:935–949. doi:10.1007/s00438-014-0865-x

Warrington, C., H. Abdel-Haleem, D. Hyten, P. Cregan, J. Orf, A. Killam, et al. 2015. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. Theor. Appl. Genet. 128:839–850. doi:10.1007/s00122-015-2474-4

Wen, Z., R. Tan, J. Yuan, C. Bales, W. Du, S. Zhang, et al. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics 15:809. doi:10.1186/1471-2164-15-809

Wu, X., S. Blake, D.A. Sleper, J.G. Shannon, P. Cregan, and H.T. Nguyen. 2009. QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor. Appl. Genet. 118:1093–1105. doi:10.1007/s00122-009-0965-x

Yamanaka, N., S. Ninomiya, M. Hoshi, Y. Tsubokura, M. Yano, Y. Nagamura, et al. 2001. An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. DNA Res. 8:61–72. doi:10.1093/dnares/8.2.61

Yu, J., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17:155–160. doi:10.1016/j.copbio.2006.02.003

Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551. doi:10.1534/genetics.107.074245

Zhang, G., C. Gu, and D. Wang. 2009. Molecular mapping of soybean aphid resistance genes in PI 567541B. Theor. Appl. Genet. 118:473–482. doi:10.1007/s00122-008-0914-0

Zhang, J., Q. Song, P.B. Cregan, R.L. Nelson, X. Wang, J. Wu, et al. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics 16:217. doi:10.1186/s12864-015-1441-4