Identification of QTLs Associated with Biological Nitrogen Fixation Traits in Soybean Using a Genotyping-by-Sequencing Approach

Anna Karolina Grunvald,★ Adalgisa Ribeiro Torres, André Luiz de Lima Passianotto, Maria Aparecida Santos, Martine Jean, François Belzile, and Mariangela Hungria

ABSTRACT

Although biological N fixation (BNF) is a key process for economic cultivation of soybean [*Glycine max* (L.) Merr.], the trait is often neglected in breeding programs, mainly due to difficulties in evaluating nodulation and N fixation parameters. We used a genotyping-by-sequencing approach to provide a dense genome-wide marker coverage with 1448 single nucleotide polymorphisms (SNPs) distributed broadly across the chromosomes of a soybean population. The mapping population was composed of 113 $F_{7:8}$ recombinant inbred lines, obtained by single-seed descent method, derived from crossing of soybean cultivars 'Bossier' (high BNF capacity) and 'Embrapa 20' (medium BNF capacity). The traits evaluated were nodule number (NN), nodule dry weight (NDW), average NDW (ANDW; i.e., NDW/NN), and shoot dry weight (SDW). A genetic map was constructed with 1448 SNPs that generated 35 linkage groups totaling 1793 cM, thus covering ∼72% of the genome based on the consensus linkage map. The estimated heritability for NN, NDW, ANDW, and SDW was 0.41, 0.30, 0.30, and 49.11, respectively. Inclusive composite interval mapping identified two significant quantitative trait loci (QTLs) for ANDW at 67 cM on chromosome 13, with a logarithm of odds (LOD) score of 4.66. The additive effect of this locus was −0.31 mg nodule⁻¹, and it explained 18.13% of the phenotypic variation. Another QTL was identified for SDW at 24 cM on chromosome 19 and exhibited an LOD score of 3.93. The additive effect was of 0.57 g plant⁻¹, and it accounted for 14.93% of the phenotypic variation. Additive genetic effects contribute to the final phenotype. In the present study, Embrapa 20 was the major contributor to ANDW, and Bossier was the major contributor to SDW. The latter was previously reported using simple sequence repeat markers.

A.K. Grunvald, A.R. Torres, A.L.d.L. Passianotto, M.A. Santos, and M. Hungria, Biotecnologia de Solos, Embrapa Soja, Londrina, Paraná, Brazil; M. Jean, F. Belzile, Dép. de Phytologie and Institut de Biologie Intégrative et dês Systèmes (IBIS), Univ. Laval, Quebec City, G1V 0A6, Canada. Received 2 Feb. 2018. Accepted 14 Aug. 2018. ★Corresponding author (annakgrunvald@gmail.com). Assigned to Associate Editor Zenglu Li.

Abbreviations: ANDW, average nodule dry weight; BNF, biological nitrogen fixation; CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico; IBIS, Institut de Biologie Intégrative et des Systèmes; ICIM, inclusive composite interval mapping; IGST-GBS, the Institut de Biologie Intégrative et des Systèmes Genotype-by-Sequencing Tool; GBS, genotyping-by-sequencing; LD, linkage disequilibrium; LG, linkage group; LOD, logarithm of odds; NDW, nodule dry weight; NN, nodule number; QTL, quantitative trait locus; RIL, recombinant inbred line; SDW, shoot dry weight; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

Several legumes, when in symbiosis with bacteria collectively known as rhizobia, have the ability to assimilate atmospheric N_2 in a process called biological N fixation (BNF). Legume–rhizobia symbioses potentially contribute both to environmental and agricultural sustainability by reducing greenhouse gas emissions, nitrogenous water pollution, and fossil energy consumption (Serraj et al., 1999; Kaewsuralikhit et al., 2005; Hungria and Mendes, 2015).

Soybean [*Glycine max* (L.) Merr.] is the most important legume cultivated worldwide. Symbiotic associations between soybean and *Bradyrhizobium japonicum*, *Bradyrhizobium diazoefficiens*, and *Bradyrhizobium elkanii* strains are economically and environmentally important; for example, savings on N fertilizer in Brazil with the soybean crop are estimated at about US$15 billion per cropping season (Hungria and Mendes, 2015).

Evaluation of phenotypic traits related to BNF is labor and time consuming, and thus it has been often neglected in

breeding programs (Dwivedi et al., 2015). To overcome this problem, marker-assisted selection might represent a promising alternative approach. However, the identification of molecular markers requires knowledge about the genomic regions and genes related to the traits of interest (Sonah et al., 2015).

Molecular markers have been developed for most major crop species based on several methods. Considering BNF traits in soybean, simple sequence repeat (SSR) markers have been used in some studies (Tanya et al., 2005; Nicolás et al., 2006; Santos et al., 2006, 2013), and the results obtained allowed mapping of quantitative trait loci (QTLs) but have also indicated the need for additional information to improve the genome coverage (Hwang et al., 2014). Tanya et al. (2005) evaluated nodule fresh and dry weight per plant, plant dry weight, acetylene reduction activity in 136 F_2–derived recombinant inbred lines (RILs). A genetic linkage map was constructed using 85 SSR markers. Five QTLs were associated with the number of nodules per plant. Nicolás et al. (2006) used 45 SSR markers for nodulation (nodule number [NN] and nodule dry weight [NDW]) and plant growth (shoot dry weight [SDW]) phenotypes in $F_{2:3}$ lines. A total of 21 SSR loci were mapped. Santos et al. (2006) and Santos et al. (2013) genotyped a population of 157 $F_{2:7}$ RILs by using 24 and 105 SSRs, respectively. Three traits (NN, the ratio NDW, and SDW) were used in both the studies to evaluate BNF performance. A small percentage of the genome was covered (5 and 50%, respectively). Hwang et al. (2014) provided more information by combining SSR and single nucleotide polymorphism (SNP) markers when genotyping 97 RILs of soybean, evaluated under field experiments, for the parameters of NN, nodule weight, and nodule size.

Among the molecular markers, SNPs are considered as highly suitable because they occur at high density within genomes (Gaur et al., 2012). The development of new sequencing technologies has greatly increased the discovery of SNPs in many species (Poland et al., 2012); however, whole-genome sequencing is still an expensive endeavor (Metzker, 2010). Nodule number, individual nodule weight, and nodule size are major components that likely contribute to increased total nodule weight per plant (Hwang et al., 2014). The population studied by Tanya et al. (2005) showed positive correlations between N_2 fixation activity from an acetylene reduction assay and NN or nodule fresh or dry weight. Pazdernik et al. (1996) reported that the correlation coefficients between acetylene reduction activity and NN per plant or fresh weight per plant were 0.45 and 0.86, respectively. Additionally, Greder et al. (1986) and Burias and Planchon (1990) reported positive association between nodule weight per plant and yield. These results indirectly imply that QTLs for nodule traits can contribute to increased N_2 fixation and yield.

To reduce the cost while still retaining a sufficient number of SNPs, several methods have been developed that involve sequencing a small fraction of the entire genome (van Orsouw et al., 2007; Trebbi et al., 2011; Sonah et al., 2015). A technically simple, highly multiplexed genotyping-by-sequencing (GBS) approach was developed (Elshire et al., 2011), with potential to be applied to all species at a low cost per sample. It is based on high-throughput next-generation sequencing of genomic subsets targeted by restriction enzymes.

In this study, we used the GBS approach to achieve denser and more extensive genome coverage, among a population of RILs previously studied by our group with SSR markers (Santos et al., 2006, 2013), with the aim of finding QTLs controlling BNF traits. Besides finding QTLs, our goal was to verify the feasibility of using GBS as a tool for soybean breeding programs aimed at improving BNF traits.

MATERIALS AND METHODS
Plant Material and Population Development

A soybean population was obtained by crossing cultivars 'Bossier' and 'Embrapa 20'. Bossier resulted from a natural mutation of the North American cultivar 'Lee' (PI 548656). Lee resulted from a cross of 'S-100' (maternal parent) × 'CNS' (paternal parent). Another name for this cultivar is D49-2524. Embrapa 20 (syn. Doko RC) was released in 1999 by Embrapa Soybean and resulted from a backcrossing of cultivar 'Doko' (Doko RC4, maternal parent) × IAC-7 (IAC-7R, paternal parent), with the main objective of incorporation resistance to *Cercospora sojina*. Both Doko and IAC-7 resulted from selections within populations of RB 72-1. Bossier and Embrapa 20 has been previously classified as having high and medium BNF capacity, respectively, by Nicolás et al. (2002) and in other studies by our group.

The population was developed in summer 1998, as described previously by Nicolás et al. (2002). A cross between Bossier and Embrapa 20 was made in at Embrapa Soybean (Paraná State, Brazil). The F_2 to F_7 generations were advanced by single-seed descent method. The F_7 plants were harvested individually, and the seeds were used for phenotypic data collection, as described by Santos et al. (2006, 2013). In the present study, we sowed one seed from $F_{7:8}$ RILs to get leaf tissue for DNA extraction. However, some lines were lost due to some problems with germination. For this reason, we performed the study with a population of 131 $F_{7:8}$ RILs.

Phenotype of Traits Associated with Biological Nitrogen Fixation

The lines of soybean were grown in pots containing 4 kg of nonsterile soil and sand, with one plant per pot under greenhouse conditions, and the experiment was performed in a completely randomized design with eight replications (one plant per replication). The plants were inoculated at the V2 stage of development, adding 1 mL of inoculum containing two strains, *Bradyrhizobium japonicum* SEMIA 5079 (= CPAC 15 and belonging to the same serogroup as SEMIA 566 and USDA

123) and *Bradyrhizobium elkanii* SEMIA 587 (1:1, v/v) (Fehr et al., 1971). Six weeks after emergence, plants were harvested for evaluation of BNF parameters: NN, NDW, average NDW (ANDW), and SDW.

DNA Extraction
DNA was extracted from two to three trifoliate leaves from each of the RILs using the DNeasy 96 plant kit (QIAGEN), as per the manufacturer's instructions. DNA was quantified using a Qubit spectrophotometer (Invitrogen) and normalized to 10 ng µL^{-1} for GBS library preparation.

Library Preparation and Sequencing
Two multiplex (66- and 67-plex) *Ape* KI GBS libraries were prepared by the Plateforme d'Analyses Génomiques (Institut de Biologie Intégrative et des Systèmes [IBIS], Laval University, Quebec City, QC, Canada) by using 100 ng of each DNA sample (131 RILs and the two parental genotypes); the libraries were prepared as described by Elshire et al. (2011) and Sonah et al. (2013). Each GBS library was sequenced on a single chip of an Ion Torrent ion proton sequencer at the Plateforme d'Analyses Génomiques at Laval University.

SNP Calling and Filtration
Read processing and SNP calling were performed using the IBIS Genotype-by-Sequencing Tool (IGST-GBS pipeline; Sonah et al., 2013) with the soybean Glyma1.01 assembly as the reference genome. Loci displaying extreme heterozygosity (outliers were defined as heterozygosity values that exceeded the third quartile by more than three times the interquartile range; i.e., Quartile 3 + the interquartile range × 3) were excluded from the analysis. The same filter for extreme heterozygosity was also applied to lines, as well as an additional filter to eliminate lines with large amounts of missing data (>90%). Low-quality heterozygous genotypes (genotype quality < 30) were converted to missing data by an in-house Python script. The raw SNPs were then filtered using VCFtools (Danecek et al., 2011) to keep only the SNPs located on chromosomes (and not scaffolds) having a minimum read depth of two, no more than 80% missing data, and a minimum minor allele frequency of 12.5%. Finally, missing genotypes were imputed using Beagle version 4.0 (Browning and Browning, 2007).

Map Construction and QTL Analysis
The genetic map was obtained using QTL IciMapping (version 3.2) (Wang et al., 2012). The Kosambi mapping function was used to convert the recombination frequency into centimorgans. Markers in complete linkage disequilibrium (LD; i.e., exhibiting an identical segregation pattern) were grouped into bin loci using the BIN tool. Linkage groups (LGs) were assembled based on a minimum logarithm of odds (LOD) threshold of 3.0. Loci assigned to a LG that differed from their known chromosomal position on the physical map were excluded from the analysis. Each LG was first ordered with the *nnTwoOpt* algorithm, followed by a fine-tuning of marker order by rippling using the SARF (sum of adjacent recombination frequencies) criterion and a window size of eight. Afterward, the genetic map was inspected visually, and markers with two or more

putative genotyping errors (putative double crossovers or isolated heterozygous calls) were excluded from the analysis. We performed QTL analysis using the inclusive composite interval mapping (ICIM) method, implemented in the same program used to construct the linkage map, QTL IciMapping.

RESULTS
Phenotypic Analysis and QTL Mapping of Biological Nitrogen Fixation Traits
In the present study, BNF performance was assessed by measuring nodulation traits NN, NDW, ANDW, and SDW on a collection of 131 $F_{7:8}$ RILs. The observed values for the two parents were as follows for Bossier and Embrapa 20, respectively: NN = 206.4 and 183.6 nodules plant^{-1}, NDW = 699.9 and 530.0 mg plant^{-1}, ANDW = 4.04 and 2.75 mg nodule^{-1}, and SDW = 10.19 and 9.11 g plant^{-1}. These results confirmed the higher BNF capacity of Bossier. For the collection of RILs, as shown in Fig. 1, NN ranged between 110.6 and 322.8 nodules plant^{-1}, with a mean of 216.5 ± 4.18 nodules plant^{-1} (mean error); NDW ranged between 337.8 and 1278.0 mg nodule^{-1}, with a mean of 757.6 ± 15.95 mg nodule^{-1}; ANDW ranged between 1.95 to 5.8 mg nodule^{-1}, with a mean of 3.81 ± 0.060 mg nodule^{-1}; and SDW between 6.49 and 15.43 g plant^{-1}, with a mean of 11.5 ± 0.13 g plant^{-1}. In all cases, the range of values observed in the RILs extended those of the two parents. Such transgressive segregation suggests that each parent contributed both positive and negative alleles at the loci controlling these traits. As expected (Bohrer and Hungria, 1998), high values of coefficient of variation were obtained for NN, NDW, ANDW (i.e., NDW/NN), and SDW: 47.51, 51.85, 46.67, and 27.51%, respectively. Medium to low values of estimated heritability for NN, NDW, ANDW, and SDW were obtained: 0.41, 0.30, 0.30, and 49.11, respectively. Finally, the distribution of values for all four traits resembles a normal distribution (Fig. 1), suggesting that the genetic control of these traits is complex and potentially results from the contribution of multiple genes. As pointed out by Hwang et al. (2014), a composite interval mapping model for QTL analysis generally assumes that traits follow a normal distribution.

Construction of a Genetic Map Using GBS-Derived SNPs
A total of 90.4 and 85.3 million Ion Torrent single-end reads were obtained for each of the two GBS libraries (66- and 67-plex), with means of 1.1 million (range = 189,942–3,521,829) and 1.2 million (range = 305,900–3,313,911) reads per sample, respectively. Using the IGST-GBS pipeline, these reads were used to identify a list of polymorphic SNP loci and to call genotypes at each SNP locus in each RIL and the two parents. The initial catalog of SNPs was subjected to the filtration steps described above (see Material and Methods), resulting
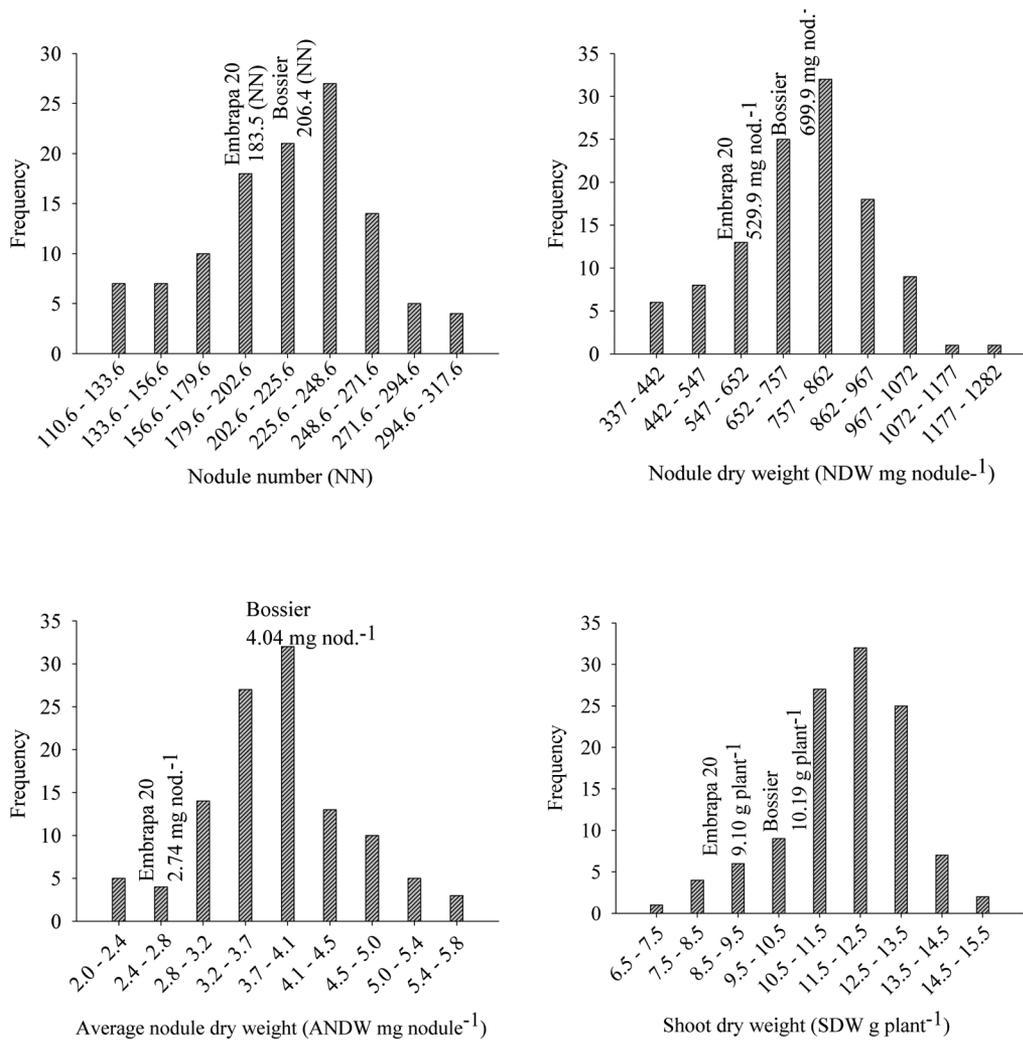
Fig. 1. Phenotypic evaluation of biological N fixation (BNF) performance for shoot dry weight (SDW), nodule number (NN), nodule dry weight (NDW), and average nodule dry weight (ANDW; i.e., NDW/NN) using a population of 131 $F_{7:8}$ recombinant inbred lines derived from a cross between Bossier (high BNF capacity) and Embrapa 20 (medium BNF capacity).

in a final set of 4885 high-quality SNP markers for map construction. Four RILs with >90% missing data (due to an insufficient number of reads) were removed from the dataset. Additionally, 14 RILs displaying anomalously high levels of heterozygosity (mean of 5.3% relative to a mean of 1.7% for the rest of the lines) were also excluded from the analysis. Therefore, the final mapping population was composed of 113 RILs.

After assessing the redundancy between the SNP markers with the BIN tool, 80.5% (3932) of the SNP markers were found to be in complete LD with one or more other markers. Such SNP markers sharing an identical segregation pattern among the RILs were grouped into 673 bin loci (containing 2 to 134 markers, average = 5.8). However, three bin loci were composed of SNP loci whose genomic positions were located on two chromosomes (in all cases, Gm04 and Gm18) and were excluded from the analysis. After an initial mapping analysis, an additional 13 loci with inconsistent chromosomal assignment, as well as 160 loci with low-quality genotypes, were excluded from the analysis.

After filtering of the markers, a soybean linkage map based on 113 soybean RILs $F_{7:8}$ was constructed with a total of 1450 SNPs (one SNP from each bin locus [Supplemental Table S1], and all SNPs with a unique segregation pattern). The resulting map is composed of 35 LGs, containing 2 to 103 markers, with an average of 39.1 SNPs (Fig. 2, Supplemental Table S2). The LGs themselves contained no gaps of >20 cM; however, many chromosomes were split into different LGs due to a lack of polymorphism between the parents in the intervening chromosomal segments (data not shown). Among the LGs formed, six chromosomes (Gm03, Gm07, Gm11, Gm13, Gm16, and Gm20) were composed of a single LG, 11 (Gm1, Gm2, Gm4, Gm5, Gm8, Gm10, Gm12, Gm14, Gm15, Gm17, and Gm19) were split into two LGs, and three (Gm6, Gm9, and Gm18) comprised three LGs. These LGs can be assembled to form the gapped chromosomes shown in Fig. 2 based on the known physical location of the SNPs, and the total genetic distance for each chromosome can be estimated (Table 1). The greatest map length was obtained for chromosome 4 (101.15 cM) and the
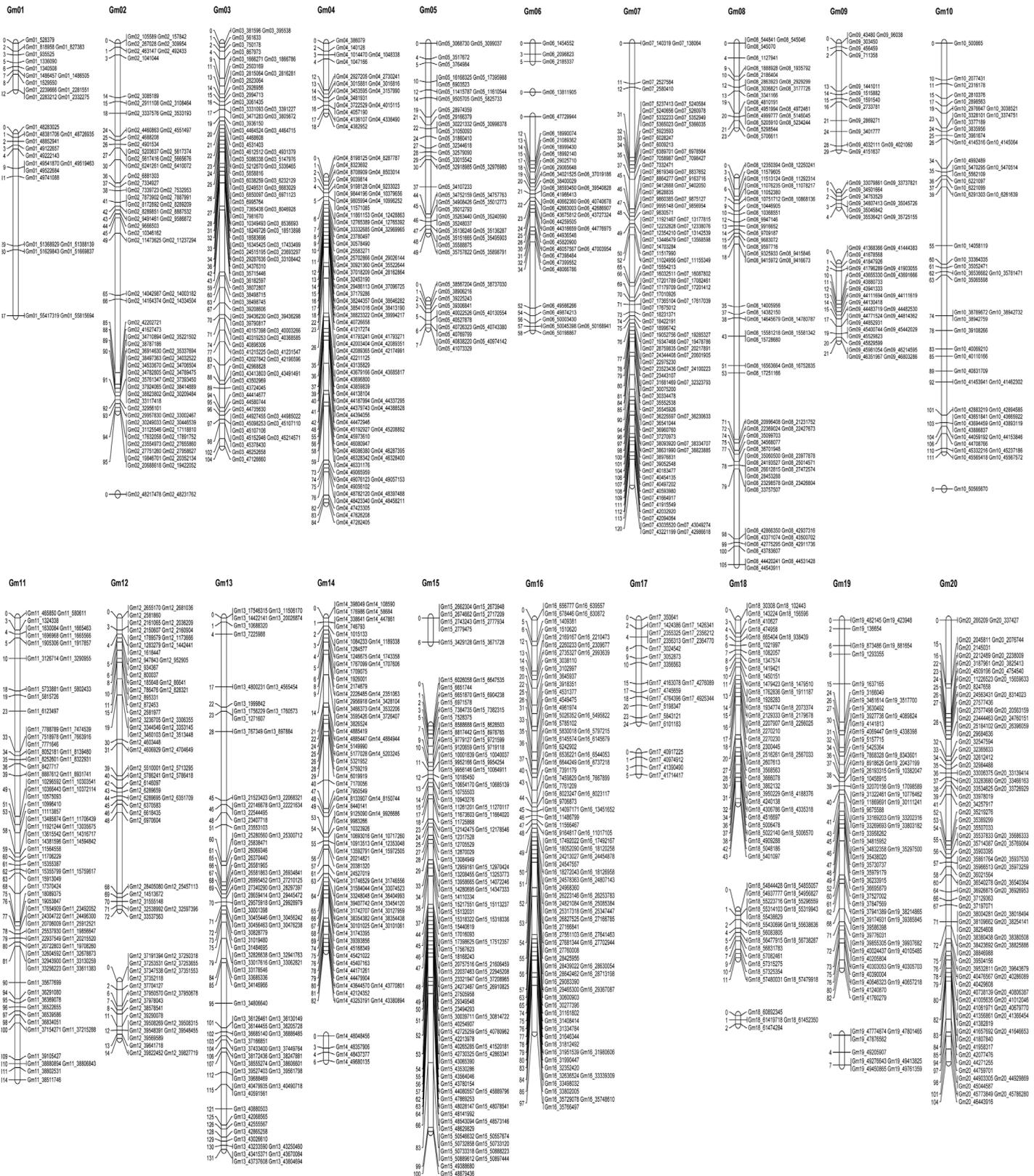
Fig. 2. Genetic linkage map of a soybean population composed of 113 F$_{7:8}$ inbred lines derived from a cross between Bossier (high biological N fixation [BNF] capacity) and Embrapa 20 (medium BNF capacity).

smallest for chromosome 17 (28.46 cM). Across all chromosomes, this resulted in a total genetic distance of 1792.6 cM, providing coverage of ~72% of the consensus linkage map (Cregan et al., 1999; Schmutz et al., 2010; Table 1).

The physical distance separating different LGs from the same chromosome was generally <20 Mb. In four cases, the gap between two LGs from the same chromosome was larger (Gm18 = 49 Mb, Gm01 = 46 Mb, Gm17

Chromosome	No. of SNPs†	Length
		cM
Gm01	29	59.10
Gm02	81	95.50
Gm03	88	103.56
Gm04	97	101.15
Gm05	52	56.41
Gm06	38	62.83
Gm07	99	120.19
Gm08	79	128.13
Gm09	49	54.00
Gm10	51	110.91
Gm11	82	114.19
Gm12	71	85.69
Gm13	83	131.35
Gm14	84	88.16
Gm15	112	106.35
Gm16	96	97.30
Gm17	22	28.46
Gm18	69	56.55
Gm19	73	88.67
Gm20	93	104.09
Total	1448	1792.59

† SNP, single nucleotide polymorphism.

= 34 Mb, Gm09 = 29 Mb). Overall, 93.1% of the loci were exactly or almost exactly in the same order in the physical and genetic map.

QTL Analysis

After filtering of SNP markers, a linkage map was constructed based on the four traits of BNF evaluated (NN, NDW, ANDW, and SDW) for 113 out of the 131 soybean RILs $F_{7:8}$ evaluated. We performed QTL mapping for each of the four traits using ICIM. As shown in Fig. 3 and summarized in Table 2, we identified one significant QTL for ANDW and another for SDW. With LOD scores of 4.66 and 3.93, respectively, these QTLs for ANDW and SDW were the only ones that reached the empirical LOD score threshold. For the ANDW trait, the QTL was identified at 67 cM on chromosome 13. The additive effect of this locus was −0.31 mg nodule^{-1}, and it explained 18.13% of the phenotypic variation. The additive genetic effects indicated major contribution of parental Embrapa 20. The QTL for SDW was identified at 24 cM on chromosome 19. The additive effect was of 0.57 g plant^{-1}, and it accounted for 14.93% of the phenotypic variation. For this trait, the parental Bossier had greater contribution.

DISCUSSION

The detection and the exploitation of genetic variation are critical for plant breeding. Molecular markers have been used extensively for the development of saturated genetic and physical maps and for the identification of genes or QTLs controlling traits of economic importance (Varshney et al., 2009). If a molecular marker is linked to a trait of interest, it can be used in marker-assisted selection programs, one of the goals of QTL mapping in plants (Varshney et al., 2009; Semagn et al., 2010).

Conventionally, QTL mapping has been performed using a segregating biparental population. In our study, a soybean population composed of 113 $F_{7:8}$ RILs, derived from a cross between Bossier and Embrapa 20 characterized by high and medium BNF capacity, respectively, was investigated by using a GBS approach to search for QTLs controlling BNF traits. We were able to obtain a more extensive genome coverage of a RIL population than previously studied by Tanya et al. (2005), Nicolás et al. (2006) and Santos et al. (2006, 2013). The present study resulted in a genetic map totaling 1793 cM, corresponding to ~72% of the length of soybean consensus linkage map derived from multiple segregating populations (2536 cM) (Song et al., 2004). In the population with the same parental lines, but using 157 $F_{2:7}$ recombinant RILs, Santos et al. (2006) covered ~5% of the genome using 24 SSR markers, and later 50% of the genome when expanded to 105 SSR markers (Santos et al., 2013). Therefore, the genetic map coverage achieved in the present study, with a set of 1448 SNP markers, represents a marked improvement over previous studies on this population of RILs.

The comparison of the genome coverage obtained in our study (1792.6 cM) with other studies points to both larger and smaller genetic maps. For example, Akond et al. (2015), using soybean population derived from the cross between 'Hamilton' and 'Spencer' covered 1423.72 cM with 1502 SNPs. On the other hand, greater length for a genetic map derived from a single segregating population was reported by Hwang et al. (2014), with 2241.3 cM obtained with 664 markers (171 SSRs + 493 SNPs) and 97 RILs. In Hwang et al. (2014), many maps were integrated and composed from multiple marker types and several populations. According to Akond et al. (2013, 2015), these maps are not comparable to the map construction with only one population and one type of marker.

Soybean is an important oilseed crop, but very few high-density genetic maps have been published for this species (Qi et al., 2014). Hyten et al. (2010, 2008) built two high-density integrated genetic linkage maps of soybean-based genome sequencing and high-throughput SNP genotyping. In the map obtained in our study, chromosomes were often composed of two different types of segments. In genomic segments where the two parents were polymorphic, dense marker coverage was achieved and no gaps of >20 cM were obtained. Therefore, in these genomic regions, sufficient marker coverage was achieved for the detection of QTLs. Other segments of chromosomes were completely devoid of any polymorphic markers, presumably because the two parental lines share these segments from a common
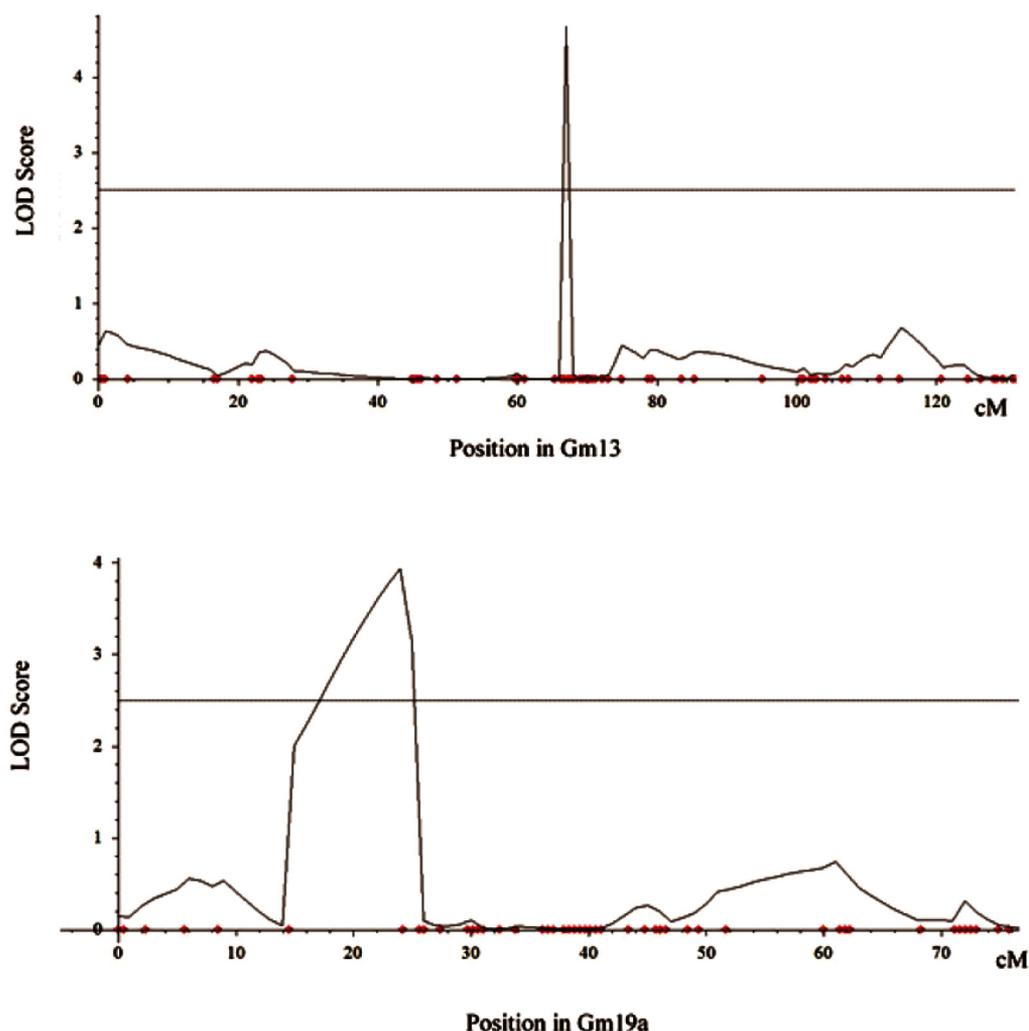
Fig. 3. Quantitative trait loci identified by inclusive composite interval mapping for average nodule dry weight on chromosome 13 and shoot dry weight on chromosome 19 using a population of 113 F$_{7:8}$ inbred lines derived from a cross between Bossier (high biological N fixation [BNF] capacity) and Embrapa 20 (medium BNF capacity).

ancestor. In such genomic regions, the parents presumably carry the same alleles at all loci, and they cannot contribute to the phenotypic contrast observed between the parental lines. Such low polymorphism between parental genotypes has been pointed out in several studies comparing modern soybean genotypes (Hiromoto and Vello, 1986; Gizlice et al., 1994; Abdelnoor et al., 1995).

Inclusive composite interval mapping allowed us to identify one QTL for ANDW on chromosome 13 and another for SDW on chromosome 19. Previously, Santos et al. (2013), using another method (composite interval method, CIM), also identified one QTL for SDW on chromosome 19, but in a different position. Simulations showed that ICIM is computationally less intensive, has increased

Table 2. Quantitative trait loci identified by inclusive composite interval mapping for biological N fixation (BNF) traits in a population of 113 F$_{7:8}$ recombinant inbred lines derived from a cross between Bossier (high BNF capacity) and Embrapa 20 (medium BNF capacity).

| Trait† | Chr. | Position | Significant position | | Total SNPs‡ | LOD§ | PVE¶ | Add# |
			Left	Right				
		cM					%	
ANDW	13	67	26,581,863	26,934,841	4	4.66	18.13	−0.31
SDW	19	24	1,637,165	3,166,049	10	3.93	14.93	0.57

†ANDW, average nodule dry weight; SDW, shoot dry weight.

‡ SNP, single nucleotide polymorphism.

§ LOD, logarithm of odds.

¶ PVE, phenotypic variation explained.

Add, additive effect.

detection power, reduces false discovery rate, and allows less biased estimates of QTL effects (Li et al., 2007).

The same mapping population was previously used for mapping of BNF traits using SSR markers (Santos et al., 2013). Except SDW, the QTLs identified for other BNF traits could not be validated in present study, despite using a high-density marker map. Although all SSR markers were indeed assigned to the expected LG and chromosome, based on their known physical position, we observed important distortions occurring in the integrated map. These discrepancies do not seem to be concentrated in a subset of individuals, leading us to hypothesize that these are not due to discrepancies in the identification or naming of individual RILs, but rather due to genotyping errors in the earlier SSR data. When a limited number of SSR markers or loci are used, these errors do not prevent the production of a linkage map, but when added to a much denser set of SNP markers, these genotyping errors cause conflicts that are difficult to resolve for the mapping software and lead to significant map distortion.

It is worth mentioning that in the region where the QTLs for SDW were identified in our study (1,637,165 [start] and 3,166,049 [end]), there are some genes that might be related to plant growth. For example, Glyma19g022500 (located at 2,519,866 to 2,522,550) encodes a gibberellin-regulated protein involved in several developmental and physiological processes in plants, including leaf growth. In a previous study, Hao et al. (2011) reported that the expression of a gene encoding a gibberellin-regulated protein (Glyma10g40580) was upregulated 33.32-fold under low-N conditions. The authors demonstrated differential transcript abundance and regulation in response to low-N stress between two soybean cultivars, one tolerant and one sensitive to low-N conditions. According to Bottini et al. (2004), gibberellin can be produced by endophytic bacteria and plant-growth-promoting rhizobacteria and had a beneficial effect on growth and yield. It is noteworthy that SDW is considered an excellent parameter to estimate the contribution of BNF in soybean, as shown under both greenhouse and field conditions (Bohrer and Hungria, 1998; Hungria and Bohrer, 2000; Souza et al., 2008).

In relation to the ANDW trait, the QTL identified on chromosome 13 has not been reported before. This QTL was located between SNP markers at the positions 26,581,863 and 26,934,841. The left marker was located 675 kb from the leftmost markers of a chitinase gene (Glyma13g22350) located at 25,904,742 to 25,907,117. Usually, chitinase activity in plants is low, but it can be induced in soybean when in symbiosis with *Bradyrhizobium* (Xie et al., 1999). In the nodulation process, chitinases have been shown to differentially cleave nodulation factors, and it has been proposed that they modulate the activity of these key morphogenetic signals in the development of the symbiosis (Salzer et al., 2000).

Both QTLs identified in our study explained a sizable amount of the phenotypic variation for ANDW and SDW, 18.13 and 14.93%, respectively. The Embrapa20 allele contributed to increase ANDW, whereas the Bossier allele contributed to higher SDW (Fig. 4). The fact that the individual QTLs identified here explained <20% of the phenotypic variation is typical of what has been observed in other QTL studies on BNF and is in agreement with the complex nature of this trait (Santos et al., 2013; Hwang et al., 2014).

The general means for the parameters related to BNF evaluated in our study were higher than the mean of the parental Bossier (high BNF capacity). Studies of quantitative traits in segregating populations sometimes report the presence of phenotypes that are very distant from those of the parental lines (de Vicente and Tanksley, 1993; Rieseberg et al., 1999). This phenomenon, known as transgressive segregation, is a major mechanism by which extreme or novel adaptations observed in new hybrid ecotypes or species are thought to arise. It is a phenomenon specific of segregating generations and refers to the fraction of individuals that exceed parental phenotypic values in either a negative or positive direction (Rieseberg et al., 1999). As pointed out by Hwang et al. (2014), a composite interval mapping
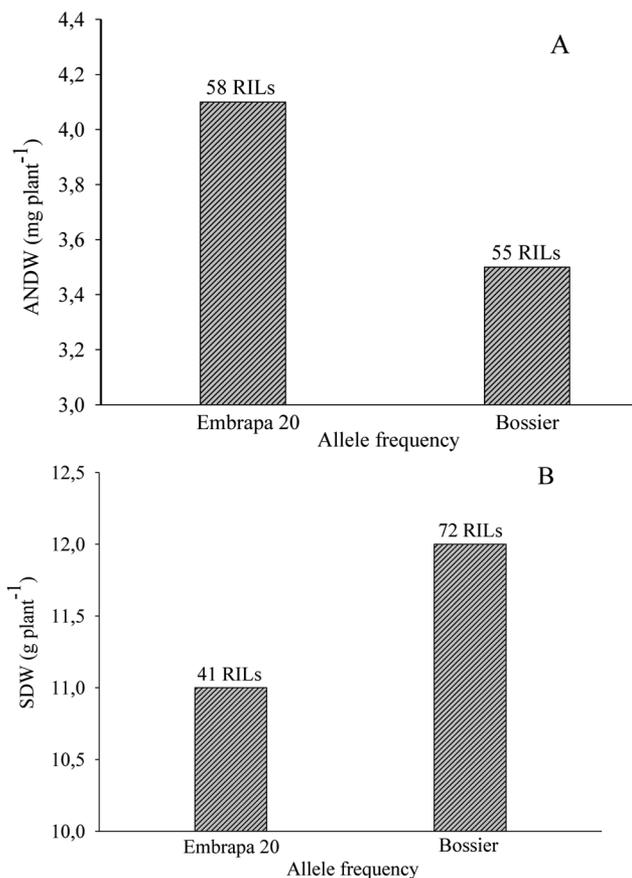

Fig. 4. (A) Average nodule dry weight (ANDW) of 113 $F_{7:8}$ recombinant inbred lines (RILs) contrasted for the single nucleotide polymorphism (SNP) on chromosome 13, and (B) shoot dry weight (SDW) of 113 $F_{7:8}$ RILs contrasted for the SNP on chromosome 19.

model for QTL analysis generally assumes that traits follow a normal distribution. Furthermore, due to the distinct and relatively variable nature of the BNF parameter, coefficients of variation>35% are often reported for conditions (Bohrer and Hungria, 1998; Hungria and Bohrer, 2000; Souza et al., 2008).

Soybean is an important crop worldwide, and one important feature is its high capacity for BNF, an attractive choice of economic and environmental importance (Hungria and Mendes, 2015). Nevertheless, phenotypic evaluation of BNF traits is time and labor consuming and can show considerable variability depending on the environmental conditions (Cregan and Keyser, 1986; Betts and Herridge, 1987; Bohrer and Hungria, 1998; Hungria and Bohrer, 2000).

The use of molecular markers is currently mandatory in plant breeding programs, and we have demonstrated that GBS can represent a highly cost-effective and high-throughput genotyping approach to search for molecular markers related to BNF in the soybean genome.

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Material Available

Supplemental material for this article is available online.

Acknowledgments

References

Abdelnoor, R.V., E.G. Barros, and M.A. Moreira. 1995. Determination of genetic diversity within Brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. Rev. Bras. Genet. 18:265–276.

Akond, M., S. Liu, L. Schoener, J.A. Anderson, S.K. Kantartzi, K. Meksem, et al. 2013. SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium Bead Chip genotyping array. J. Plant Genome Sci. 1:80–89.

Akond, M.A., S. Liu, S.K. Kantartzi, K. Meksem, N. Bellaloi, D.A. Lightfoot, et al. 2015. A SNP genetic linkage map based on the 'Hamilton' by 'Spencer' recombinant inbred line population identified QTL for seed isoflavone contents in soybean. Plant Breed. 134:580–588. doi:10.1111/pbr.12298

Betts, J.H., and D.F. Herridge. 1987. Isolation of soybean lines capable of nodulation and nitrogen fixation under high levels of nitrate supply. Crop Sci. 27:1156–1161. doi:10.2135/cropsci1987.0011183X002700060014x

Bohrer, T.R.J., and M. Hungria. 1998. Avaliação de cultivares de soja quanto à fixação biológica do nitrogênio. Pesqi. Agropecu. Bras. 33:937–952.

Bottini, R., F. Cassán, and P. Piccoli. 2004. Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. Appl. Microbiol. Biotechnol. 65:497–503. doi:10.1007/s00253-004-1696-1

Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987

Burias, N., and C. Planchon. 1990. Increasing soybean productivity through selection for nitrogen fixation. Agron. J.82:1031–1034. doi:10.2134/agronj1990.00021962008200060001x

Cregan, P.B., and H.H. Keyser. 1986. Host restriction of nodulation by *Bradyrhizobium japonicum* strain USDA 123 in soybean. Crop Sci. 26:911–916. doi:10.2135/cropsci1986.0011183X002600050016x

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, et al. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464–1490. doi:10.2135/cropsci1999.3951464x

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al. 2011. The variant cell format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

de Vicente, M.C., and S.D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596.

Dwivedi, S.L., K.L. Sahraeat, H.D. Upadhyaya, A. Mengoni, M. Galardini, M.M. Bazzicalupo, et al. 2015. Chapter one: Advances in host plant and *Rhizobium* genomics to enhance symbiotic nitrogen fixation in grain legumes. Adv. Agron. 129:1–116. doi:10.1016/bs.agron.2014.09.001

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

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Crop Sci. 11:929–931. doi:10.2135/cropsci1971.0011183X001100060051x

Gaur, R., S. Azam, G. Jeena, A.W. Khan, S. Choudhary, M. Jain, et al. 2012. High through put SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). DNA Res. 19:357–373. doi:10.1093/dnares/dss018

Gizlice, Z., T.E. Carter, and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143–1151. doi:10.2135/cropsci1994.0011183X003400050001x

Greder, R.R., J.H. Orf, and J.W. Lambert. 1986. Heritabilities and associations of nodule mass and recovery of *Bradyrhizobium japonicum* serogroup USDA 110 in soybean. Crop Sci. 26:33–37. doi:10.2135/cropsci1986.0011183X002600010007x

Hao, Q.N., X.A. Zhou, A.H. Sha, C. Wang, R. Zhou, and S.L. Chen. 2011. Identification of genes associated with nitrogen use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics 12:525. doi:10.1186/1471-2164-12-525

Hiromoto, D.M., and N.A. Vello. 1986. The genetic base of Brazilian soybean (*Glycine max* (L.) Merrill) cultivars. Rev. Bras. Genet. 2:295–306.

Hungria, M., and T.R.J. Bohrer. 2000. Variability of nodulation and dinitrogen fixation capacity among soybean cultivars. Biol. Fertil. Soils 31:45–52. doi:10.1007/s003740050622

Hungria, M., and I.C. Mendes. 2015. Nitrogen fixation with soybean: The perfect symbiosis? In: F.J. De Bruijn, editor, Biological nitrogen fixation. John Wiley & Sons, Hoboken, NJ. p. 1009–1024. doi:10.1002/9781119053095.ch99

Hwang, S., J.D. Ray, P.B. Cregan, C.A. King, M.K. Davies, and L.C. Purcell. 2014. Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (*Glycine max* L.[Merr.]). Euphytica 195:419–434. doi:10.1007/s10681-013-1005-0

Hyten, D.L., I. Choi, Q. Song, J.E. Specht, T.E. Carter, R.C. Shoemaker, et al. 2010. 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. doi:10.2135/cropsci2009.06.0360

Hyten, D.L., Q.J. Song, I.Y. Choi, M.S. Yoon, J.E. Specht, L.K. Matukumalli, et al. 2008. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. 116:945–952. doi:10.1007/s00122-008-0726-2

Kaewsuralikhit, S., T. Yokoyama, H. Kouchi, and Y. Arima. 2005. Comprehensive analysis of plant gene expression in soybean root nodules at different growth stages. Soil Sci. Plant Nutr. 51:535–547. doi:10.1111/j.1747-0765.2005.tb00062.x

Li, H., G. Ye, and J. Wang. 2007. A modified algorithm for the improvement of composite interval mapping. Genetics 175:361–374. doi:10.1534/genetics.106.066811

Metzker, M.L. 2010. Sequencing technologies: The next generation. Nat. Rev. Genet. 11:31–46. doi:10.1038/nrg2626

Nicolás, M.F., C.A.A. Arias, and M. Hungria. 2002. Genetics of nodulation and nitrogen fixation in Brazilian soybean cultivars. Biol. Fertil. Soils 36:109–117. doi:10.1007/s00374-002-0511-3

Nicolás, M.F., M. Hungria, and C.A.A. Arias. 2006. Identification of quantitative trait loci controlling nodulation and shoot mass in progenies from two Brazilian soybean cultivars. Field Crops Res. 95:355–366. doi:10.1016/j.fcr.2005.04.012

Pazdernik, D.L., P.H. Graham, C.P. Vance, and J.H. Orf. 1996. Host genetic variation in the early nodulation and dinitrogen fixation of soybean. Crop Sci. 36:1102–1107. doi:10.2135/cropsci1996.0011183X003600050005x

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. doi:10.1371/journal.pone.0032253

Qi, Z., L. Huang, R. Zhu, D. Xin, C. Liu, X. Han, et al. 2014. A high-density genetic map for soybean based on specific length amplified fragment sequencing. PLoS One 9:e104871. doi:10.1371/journal.pone.0104871 [erratum: 9(11):e114349].

Rieseberg, L.H., M.A. Archer, and R.K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity 83:363–372. doi:10.1038/sj.hdy.6886170

Salzer, P., A. Bonanomi, K. Beyer, R. Vögeli-Lange, R.A. Aeschbacher, J. Lange, et al. 2000. Differential expression of eight chitinase genes in *Medicago truncatula* roots during mycorrhiza formation, nodulation, and pathogen infection. Mol. Plant Microbe Interact. 13:763–777. doi:10.1094/MPMI.2000.13.7.763

Santos, M.A., I.O. Geraldi, A.A.F. Garcia, N. Bortolatto, A. Schiavon, and M. Hungria. 2013. Mapping of QTLs associated with biological nitrogen fixation traits in soybean. Hereditas 150:17–25. doi:10.1111/j.1601-5223.2013.02275.x

Santos, M.A., M.F. Nicolás, and M. Hungria. 2006. Identificação de QTL associados à simbiose entre *Bradyrhizobium japonicum, B. elkanii* e soja. Pesqi. Agropecu. Bras. 41:67–75. doi:10.1590/S0100-204X2006000100010

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 [erratum: 465:120].

Semagn, K., Å. Bjørnstad, and Y. Xu. 2010. The genetic dissection of quantitative traits in crops. Electron. J. Biotechnol. 13:16–17. doi:10.2225/vol13-issue5-fulltext-14

Serraj, R., T.R. Sinclair, and L.C. Purcell. 1999. Symbiotic N_2 fixation response to drought. J. Exp. Bot. 50:143–155.

Sonah, H., M. Bastien, E. Iquira, A. Tardivel, G. Legare, B. Boyle, et al. 2013. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One 8:e54603. doi:10.1371/journal.pone.0054603

Sonah, H., L. O'Donoughue, E. Cober, I. Rajcan, and F. Belzile. 2015. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol. J. 13:211–221. doi:10.1111/pbi.12249

Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, et al. 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109:122–128. doi:10.1007/s00122-004-1602-3

Souza, R.A., M. Hungria, J.C. Franchini, C.D. Maciel, R.J. Campo, and D.A.M. Zaia. 2008. Conjunto mínimo de parâmetros para avaliação da microbiota do solo e da fixação biológica do nitrogênio pela soja. Pesqi. Agropecu. Bras. 43:83–91. doi:10.1590/S0100-204X2008000100011

Tanya, P., P. Srinives, T. Toojinda, P. Nakhon, A. Vanavichit, and S.H. Lee. 2005. Identification of SSR markers associated with N_2 fixation components in soybean [*Glycine max* (L) Merr.]. Korean J. Genet. 27:351–359.

Trebbi, D., M. Maccaferri, P. de Heer, A. Sørensen, S. Giuliani, S. Salvi, et al. 2011. High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). Theor. Appl. Genet. 123:555–569. doi:10.1007/s00122-011-1607-7

van Orsouw, N.J., R.C.J. Hogers, A. Janssen, F. Yalcin, S. Snoeijers, E. Verstege, et al. 2007. Complexity reduction of polymorphic sequences (CRoPS™): A novel approach for large-scale polymorphism discovery in complex genomes. PLoS One 2:e1172. doi:10.1371/journal.pone.0001172

Varshney, R.K., S.N. Nayak, G.D. May, and S.A. Jackson. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. 27:522–530. doi:10.1016/j.tibtech.2009.05.006

Wang, J., H. Li, L. Zhang, and L. Meng. 2012. Users' manual of QTL IciMapping version 3.2. Quant. Genet. Group, Inst. Crop Sci., Chinese Acad. Agric. Sci., Beijing.

Xie, Z.P., C. Staehelin, A. Wiemken, W.J. Broughton, J. Müller, and T. Boller. 1999. Symbiosis-stimulated chitinase isoenzymes of soybean (*Glycine max* (L.) Merr.). J. Exp. Bot. 50:327–333. doi:10.1093/jxb/50.332.327