Mapping of QTLs associated with biological nitrogen fixation traits in soybean

MARIA APARECIDA SANTOS1,2, ISAIAS OLÍVIO GERALDI2, ANTONIO AUGUSTO FRANCO GARCIA2, NÁGILA BORTOLATTO1,3, ALINE SCHIAVON1,3 and MARIANGELA HUNGRIA1

1Embrapa Soja, Londrina, Paraná, Brazil
2Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Dep. Genética, Piracicaba, São Paulo, Brazil
3Centro Universitário Filadélfia (UNIFIL), Center of Biology, Londrina, Paraná, Brazil


Biological nitrogen fixation (BNF) is a key process, but despite the economic and environmental importance, few studies about quantitative trait loci (QTL) controlling BNF traits are available, even in the economically important crop soybean Glycine max (L.) Merr. In this study, a population of 157 F2:7 RILs derived from crossing soybean cultivars Bossier (high BNF capacity) and Embrapa 20 (medium BNF capacity) was genotyped with 105 simple sequence repeat markers (SSRs). The genetic map obtained has 1231.2 cM and covers about 50% of the genome, with an average interval of 18.1 cM. Three traits, nodule number (NN), the ratio nodule dry weight (NDW)/NN and shoot dry weight (SDW) were used to evaluate BNF performance. A composite interval mapping for multiple traits method (mCIM) analysis mapped two QTLs for SDW (LGs E and L), three for NN (LGs B1, E and I), and one for NDW/NN (LG I); all QTLs were of small effect (R2-values ranging from 1.7% to 10.0%) and explained 15.4%, 13.8% and 6.5% of total variation for these three traits, respectively.

Mariangela Hungria, Embrapa Soja, Cx. Postal 231, 86001-970, Londrina, Paraná, Brazil. E-mail: mariangela.hungria@embrapa.br

Nitrogen (N), the element most limiting for crop growth, is usually supplied by application of fertilizer, but at substantial costs to farmers and with potentially adverse effects on the environment. Soybean Glycine max (L.) Merr., among other legumes, can obtain N through biological nitrogen fixation (BNF), and breeding for yield improvement can be performed considering BNF, inorganic-N supply, or both (COALE et al. 1985; KUMUDINI et al. 2008). The BNF results from the symbiosis between legumes with a diverse group of nitrogen fixing soil bacteria collectively known as rhizobia. This association is mutually beneficial: while bacteria provide a source of nitrogen to plant development, plants provide carbon sources for maintaining bacteria metabolism.

The establishment of the symbioses involves complex mechanisms starting with the mutual exchange of diffusible signal molecules. Plants secrete seed and root molecules – mainly flavonoids – that are sensed by specific rhizobia. In response, the rhizobia produce lipochitooligosaccharides namely Nod factors (NFs) that are recognized by LysM receptor-like kinase of the legumes. A new organ – the nodule – is then formed on the root and hosts the nitrogen fixing bacteria (SUBRAMANIAN et al. 2006; KOUCHI et al. 2010).

The elicitation and the development of an effective nodule are accompanied by the expression of specific genes in both rhizobia and their host-plants. The genetic mechanisms of the bacteria are best understood and several rhizobial genes called nod, nol and noe have been described (MASSON-BOVIN et al. 2009). On the other side, studies with the host-plants are more complex, due to biological characteristics and genome sizes. More recently, the integration of genetic and genomic approaches and the adoption of legume models with Medicago truncatula and Lotus japonicum opened a framework to the investigation of host genes essential for the BNF process. Twenty-six genes involved with Nod perception and different steps of nodule formation were characterized based on mutants, cloning and mapping strategies (KOUCHI et al. 2010). Furthermore, dozens of genes involved in nodule formation and in the BNF process called nodulin genes have been prospected by high throughput tools, such as expressed sequence tags and microarray analysis (BRECHENMACHER et al. 2008; LIBAULT et al. 2010).

In soybean, several loci controlling nodulation have been described since the 1950s. WILLIAMS and LYNCH (1954) reported the recessive gene rj, in a spontaneous non-nodulating mutant. Genes Rj1 (Caldwell 1966), Rj2 (Vest 1970) and Rj3 (Vest and Caldwell 1972) have natural occurrence and control strain specific restricted nodulation. In addition, genes rj and rj (Harper and Nickell 1995) – conferring non-nodulation ability – have been identified by chemical induced mutation and
correspond to the same loci of the \( r_j \). Supernodulating mutants (\( r_j \) and \( r_j^* \)) have also been generated by chemical mutagenesis (Vuong et al. 1996).

Some of these genes controlling nodulation in soybean have been cloned and their products have been described. The first gene related to nodulation control in soybean was cloned by Searle et al. (2003) using the \( r_j \) and \( r_j^* \) supernodulating mutants. By similar approaches the Nod factor receptor genes called NFR1\( \alpha \) and NFR1\( \beta \) (Indrasumunar et al. 2011) and NFR5\( \alpha \) and NFR5\( \beta \) (Indrasumunar et al. 2010) were cloned. Similar genes were first identified by mutant directed cloning in Medicago truncatula (Amor et al. 2003; Limpens et al. 2003), L. japonicus (Madsen et al. 2003; Radutoiu et al. 2003) and pea (Pisum sativum) (Zhiukon et al. 2008).

These discoveries provide important clues to understand the molecular mechanisms of host regulation of the symbiosis. However, more information with field orientation is necessary to access the potential of breeding for BNF. The occurrence of natural variability on nodulation and efficiency of BNF in soybean have been reported (Neuhäuser et al. 1988; Sinclair et al. 1991; Pazdernik et al. 1997; Bohrer and Hungria 1998; Hungria and Bohrer 2000). However, due to technical difficulties in evaluating phenotypic traits (nodulation and nitrogen fixation rates), the mechanisms and genetic control of these traits are poorly understood. Therefore, studies of genetic architecture of nodulation and nitrogen fixation are still required. An interesting and informative way to dissect the genetic architecture is to map quantitative trait loci (QTL). This approach has been broadly used in soybean to identify traits of interest, such as yield potential, disease resistance, contents of protein and oil in seeds (Soybase 2012).

Preliminary studies about genomic regions controlling BNF in soybean have been performed (Tanya et al. 2005; Nicolas et al. 2006; Santos et al. 2006). There is also information about other legumes, such as pea (Bourjon et al. 2010) and common bean (Phaseolus vulgaris L.) (Nodari et al. 1993; Tsai et al. 1998). These studies have demonstrated the potential of mapping QTL for BNF traits, but additional information is still required to improve the existent knowledge.

The objective of this study was to improve the genome cover reported by Santos et al. (2006) in this same population; these authors used 24 simple sequence repeat markers (SSRs) to identify QTLs related to BNF and covered a small percentage (5%) of the genome. Now we have analyzed 432 SSRs, from which we selected 105 SSRs; the method of composite interval mapping for multiple traits (mCIM) was employed, altogether resulting in 50% coverage of the genome.

### MATERIAL AND METHODS

#### Plant materials and biological nitrogen fixation traits

A population of 157 F\(_2\) inbred lines derived from a cross between Bossier (high BNF capacity) and Embrapa 20 (medium BNF capacity), developed by Nicolas et al. (2002), was used and the experimental design and phenotypic evaluations were as described before (Santos et al. 2006). In brief, plants were grown in pots containing 4 kg of non-sterile 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 (Feih et al. 1971), by 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). Six weeks after emergence plants were harvested for evaluation of nodulation (nodule number, NN; nodule dry weight, NDW, NDW/NN ratio) and plant growth (shoot dry weight, SDW). Phenotypic correlations for combinations of traits were estimated using the CORR procedure of SAS STAT 9.1, while genetic correlations were estimated according to Falconer and Mackay (1996).

#### Genotyping and construction of a genetic linkage map

Leaf tissue DNA extraction and genotyping methods have been reported by Santos et al. (2006). A total of 432 simple sequence repeat (SSR) markers covering the 20 chromosomes of soybean genome were screened against the parents. The population was genotyped with 105 SSRs. To analyze the linkages between markers and to create a genetic map, the program Mapmaker/EXP ver. 3.0 was used (Lander et al. 1987). Microsatellite markers were initially grouped using the default parameters that included the Kosambi’s function, the linkage LOD score greater than 3.0 and a maximum genetic distance of 37.2 cM. The commands “group”, “compare” and “map” were used for building the linkage groups. All non-anchored markers were added by using the “try” multipoint analysis command, and the results were checked by the “ripple” command with a window size of 5. Linkage groups were named using the designation of the consensus map (Cregan et al. 1999; Song et al. 2004).

#### QTL mapping of biological nitrogen fixation

To map QTL controlling BNF, we used the multiple-trait composite interval mapping (mCIM) proposed by Jiang and Zeng (1995), that considers all traits simultaneously. The hypothesis tested was that an interval flanked by two...
Consequently, for our data, there are 27.67 independent linkage group, excluding gaps greater than 37.5 cM. To a test (LR) is estimated by
\[ \chi^2 = 2 \ln \left( \frac{L_0}{L_i} \right) \]
where \( L_0 \) is the maximum likelihood under the null hypothesis that \( a_1 = a_2 = \ldots a_n = 0 \), where \( a_i \) is the additive effect of the QTL for trait \( i \), and \( L_i \) is the maximum likelihood under the alternative hypothesis that at least one \( a_i \neq 0 \). In other words, this is a test for the presence of a pleiotropic QTL controlling one or more traits in one determined interval. The mapping analysis was performed using the module JZmapqtl of the program QTL Cartographer ver. 1.17 (BASTEN et al. 2003). Model 6 was adopted and a limited number of background markers (BG) for the mCIM analysis was identified via the forward/backward stepwise regression option using conservative probability thresholds \( (P_{in} = 0.05; P_{out} = 0.05) \). Genome scan was performed using a window size of 5 cM and a walk speed of 1 cM.

The threshold for significance of QTL detection was adjusted for the number of independent tests in the genome scan, following what has been suggested by ZENG (1994), JIANG and ZENG (1995) and VIEIRA et al. (2000). Due to the properties of the multiple regression model used for mCIM, the effective number of independent tests depends on the size of the genome region at each side of the test interval in which marker cofactors are not fitted (window size). We used a window size of 37.5 cM (recombination fraction of 0.50 considering Kosambi’s map function) as independent. The number of independent tests was then estimated by \( \Sigma \left( \frac{T_i}{28.11} + 1 \right) \), where \( T_i \) is the total estimated map length in cM of the \( i \)th linkage group, excluding gaps greater than 37.5 cM. Consequently, for our data, there are 27.67 independent tests in total. Thus, a type I error rate of \( \alpha = 0.05 \) was used in the joint mapping analysis, corresponding to a \( \chi^2 \)-value of 17.52 (five degrees of freedom, one for each QTL position), equivalent to an LOD score of 3.80. Since for marginal analysis of each trait statistical tests are applied only in positions with putative QTL identified by the pleiotropic model, a type I error of \( \alpha = 0.05 \) (LOD score of 0.83) could be used (JIANG and ZENG 1995; VIEIRA et al. 2000).

The proportion of the variance (\( R^2 \)) explained by each particular QTL and the additive (a) effects were obtained from the mCIM analysis. The total phenotypic variance explained by two or more QTLs simultaneously for a given trait was determined by adjusting a model with all QTLs using a Windows version of QTLCartographer (WANG et al. 2005). Regarding the signals of additive effects, “+” indicates increasing allele from the Bossier cultivar; “−”, indicates decreasing allele from Embrapa 20. The use of mCIM for mapping QTLs for correlated traits is also important for evaluating the cause of correlation between traits (pleiotropy or linkage); in our study, pleiotropy was assumed if the map position for two marginal QTLs were equal or smaller than 3 cM.

RESULTS

Phenotypic analysis

SANTOS et al. (2006) reported significant variance among nodulation traits (NDW, NN, NDW/NN) and plant growth (SDW) means, indicating genetic variability of the RILs population. As expected (SOUZA et al. 2008), high values of coefficient of variation (>25.6) and medium to low values of broad-sense heritabilities (0.49 for SDW, 0.33 for NDW, 0.33 for NN and 0.27 for NDW/NN) were observed in the dataset. We estimated positive and high genetic and phenotypic correlations for the traits NDW and SDW \( (r_G = 0.83 \) and \( r_P = 0.66 \)), and for NDW and NN \( (r_G = 0.65 \) and \( r_P = 0.64 \)); medium for SDW and NN \( (r_G = 0.51 \) and \( r_P = 0.43 \)); and low for SDW and NDW/NN \( (r_G = 0.26 \) and \( r_P = 0.24 \)), NDW and NDW/NN \( (r_G = 0.23 \) and \( r_P = 0.33 \)). Negative correlations were verified between NN and NDW/NN \( (r_G = -0.60 \) and \( r_P = -0.40 \)).

Polymorphism and genetic mapping

A total of 432 SSR markers were used to detect polymorphisms between the two parental genotypes and 105 polymorphic markers were used to genotype the population of RILs. From these, 97 SSR markers were mapped in 20 LGs encompassing about 1231.2 cM with a mean distance of 18.1 cM between markers, representing a coverage of about 50% of the genome if we consider a consensus linkage of 2536 cM (SONG et al. 2004). Gaps of more than 30 cM between linked markers were detected in the LGs F, G, L, M and O (Fig. 1). The remaining eight markers (Satt421, Satt396, Satt180, Satt184, Satt395, Satt356, Sat_044 and Sat_020) were not linked to any group. In general, the order and relative distances of the SSR markers within LGs was similar to the information obtained in the public soybean genetic map (CREGAN et al. 1999; SONG et al. 2004). An exception was observed for the SSR marker GmEKN, located within a hypothetical nodulin gene and mapped in LG A2 by JEONG et al. (2006), while in our study it was mapped on LG D1b (LOD > 10). On average, five markers were mapped per LG, ranging from two on LGs C1, D1a, D2, G and K to eight on LG A1 and B1.

QTL mapping of BNF traits

As explained before, three traits were used in this analysis to avoid redundancy, NN, NDW/NN and SDW. Four genomic regions were significantly associated with BNF traits, mapped on LGs B1, E, L and I. Two QTLs (\( nm1 \cdot B1 \)
Fig. 1. Soybean linkage map based on a 157 F$_2$ RILs population derived from Bossier (high capacity of biological nitrogen fixation) × Embrapa 20 (medium capacity of biological nitrogen fixation). The linkage groups were named with designations of the consensus map of CREGAN et al. (1999) and SONG et al. (2004). The uninterrupted rectangles correspond to the covered regions of the genome and the disrupted rectangles represent not covered regions in the present study.
and \(sdw2-L\) were mapped with the LOD score threshold (LOD = 3.80) and two QTLs (\(bnf3-E\) and \(bnf4-I\)) were mapped with a LOD score slightly smaller than the threshold (LOD = 3.60), considered as probable QTLs. From these, two regions located on LGs E and I were significant only for one trait, while the regions located on LGs B1 and L were significant for more than one trait, with the regions located on LGs B1 and L being significant for more than one trait (Table 1, Fig. 2). The QTLs mapped jointly were named \(bnf3-E\) and \(bnf4-I\). The QTL \(bnf3-E\) located in the interval Satt573-Satt185 was significant for SDW and NN traits. The estimated additive effects of alleles were \(-0.29\) g plant\(^{-1}\) for SDW and \(-10.23\) nodules plant\(^{-1}\) for NN and explained 5.4% and 4.3% of the individual phenotypic variance (\(R^2\)), respectively. The QTL \(bnf4-I\) located in the interval Satt587-Satt354, was significant for NN and NDW/NN. The estimated additive effects of alleles were 8.92 nodules plant\(^{-1}\) and \(-0.21\) mg nodule\(^{-1}\) and explained 1.7% and 6.5% of the individual phenotypic variance. The correlation for BNF traits observed in QTLs \(bnf3-E\) and \(bnf4-I\) in this work was due to genetic linkage (Table 1).

Two specific QTLs controlling SDW were mapped on LGs E and L, and three QTLs controlling nodulation traits (NN, NDW/NN) were mapped on LG B1, E and I (Table 1). With LOD scores of 4.20 and 3.80, respectively, the QTLs \(sdw2-L\) and \(nn1-B1\) were the only ones that reached the empirical LOD score threshold. The additive effects of these QTLs were of 0.57 g plant\(^{-1}\) for SDW and of \(-13.8\) nodules plant\(^{-1}\) for NN, and explained 10.0% and 7.8% of individual phenotypic variance, respectively, corresponding to the highest values explained by the mapped QTLs in this study.

In summary, in this study we have mapped two QTLs for SDW (LGs E and L), three for NN (LGs B1, E and I), and one for NDW/NN (LG I). All QTLs were of small effect (\(R^2\)-values ranging from 1.7% to 10.0%) and explained 15.4%, 13.8% and 6.5% of total variation for these three traits, respectively, from the adjusted model with all QTLs.

**DISCUSSION**

The role of BNF in world crop production has been emphasized for several decades. In Brazil, estimates are of an economy of about US $ 7 billion per season only with the soybean crop. Furthermore, the environment is benefitted by the lower pollution of water reservoirs, and by the reduction of the emission of gases with greenhouse effects (Hungria et al. 2007). However, as pointed out by Graham and Vance (2000), there are several areas where scientists need to concentrate efforts to ensure BNF benefits, including environmental conditions (soil degradation,
Fig. 2. QTL mapped obtained using multiple-trait composite interval mapping (mCIM) involving four biological nitrogen fixation traits (nodule dry weight, NDW; nodule number, NN; nodule dry weight ratio, NDW/NN; and shoot dry weight, SDW) in a population of 157 recombinant inbred lines (RILs) of soybean. The blue triangles indicate the position of markers on the linkage groups; putative QTL mapped under the pleiotropic model (blue lines) are indicated by yellow triangles. Marginal analysis for each trait is represented by different colors on the figure.

soil acidification and salinization, phosphorus supply and crop rotation), strain and inoculant improvement, and plant breeding.

The importance of plant breeding in increasing global BNF has been highlighted a long time ago (LARUE 1978), and since then, our understanding of the molecular mechanisms involved in the interaction between legumes and rhizobia has greatly improved (KOUCHI et al. 2010). However, few progresses have been achieved towards establishing a procedure for assessment of BNF by breeding programs, that should include simple, rapid and non-destructive field-based techniques (LARUE 1978; GRAHAM and VANCE 2000). Nowadays, the molecular assisted selection is an outstanding alternative for this proposal.

We present here an effort to identify molecular markers associated with BNF traits. In agreement with the complex nature of BNF, the mapped QTLs explained a small proportion of the total phenotypic variation effect. In soybean, phenotypic evaluations showing considerable variation under different environmental conditions were reported by CREGAN and KEYSER (1986), BETTS and HERRIDGE (1987), BOHRER and HUNGRIA (1998) and HUNGRIA and BOHRER (2000). Furthermore, the proportion of the genetic/environment contribution to the phenotype within experimental populations is highly influenced by environmental conditions (RONIS et al. 1985; PAZDERNIK et al. 1996; NICOLAS et al. 2002; SANTOS et al. 2006). In addition, the nature of the BNF parameter is also distinct and quite variable, and coefficients of variation larger than 35% are often reported (BOHRER and HUNGRIA 1998; HUNGRIA and BOHRER 2000; SOUZA et al. 2008).

From the four regions controlling BNF identified in this study, two (LGs B1 and L) have been previously reported by NICOLAS et al. (2006), confirming the association of these regions with BNF traits. Furthermore, two new QTLs were mapped (LGs E and I) as a result of a greater coverage of the genome. On the other hand, some QTLs mapped on D1b, H and B2 (NICOLAS et al. 2006; SANTOS et al. 2006), and on A1, D1b, J, I and O (TANYA et al. 2005) were not identified in our study. We attributed this to the fact that these studies used a single marker approach (regression analysis), that could result in false positives, since cofactors were not included in the model. Additionally, it is likely that QTLs are highly dependent on specific environments, or that they do not always segregate in the mapping population.

The method of QTL mapping using a composite interval mapping for multiple traits method (mCIM) takes into account correlations among traits. This joint analysis improves the statistical power of the tests, the precision of parameter estimation, and allows inference of the genetic causes (genetic linkage or pleiotropy) of the occurrence of coincident QTLs (JIANG and ZENG 1995). We identified two coincident regions controlling more than one trait of BNF (LGs E and I). Genetic linkage was the cause of coincident QTLs, SDW and NN on LG E, and NN and NDW/NN on LG I (Table 1). In the case of linked QTLs, a selection based on molecular markers could
promote increases in NN and NDW/NN despite the negative correlation observed for these two traits.

The QTL *nn1-B1* was located in a region previously associated with nodulation traits by Nicolás et al. (2006). Coincidentally, this region has been associated with isoflavones (Toda et al. 2002; Kassem et al. 2006) and, as is well known, these metabolites are involved in molecular signaling between the host plant and the rhizobium, starting the process of nodule formation (Hungria and Stacey 1997; Subramanian et al. 2006). This QTL may result from a pleiotropic effect, or might be closely linked to genes involved in the metabolism of isoflavones.

Other interesting observation is that in the region of the QTLs *nn1-B1*, Indrasumunar et al. (2010) cloned a gene (Nod Factor Reception $5\alpha$ - NFR5$\alpha$) involved with the recognition of Nod factors secreted by Bradyrhizobium. This result corroborates with the evidence that this region host genes that control nodulation in soybean.

For decades we have performed extensive work showing that SDW can be considered an excellent parameter to estimate the contribution of BNF on soybean. These studies included experiments with 152 cultivars inoculated with Bradyrhizobium and non-inoculated controls with and without chemical N-fertilizers under greenhouse controlled conditions in sterile substrate (Bohrer and Hungria 1998) and pots containing non-sterile soils (Hungria and Bohrer 2000). The feasibility of using SDW to evaluated BNF in soybean was also confirmed under field conditions, in several trials performed in soils poor on N (Souza et al. 2008). In this study, for the SDW trait, we identified two QTLs (LGs E and L) that together explained 15.4% of the phenotypic variation (Table 1). The QTL *sdw2-L* showed the highest $R^2$, explaining 10% of the phenotypic variation for SDW and was also reported by Nicolás et al. (2006). Moreover, QTLs for plant height and lodging are described in this region (Chase et al. 1996), so we consider this QTL as highly significant for SDW.

The QTL *bnf4* coincides with a major QTL for seed protein content in LG I (Nichols et al. 2006). Soybean is one of the highest protein content crop species and thus shows a high demand on N. There are studies showing that the N derived from BNF is more easily moved to pods during the grain filling than is N-nitrate, both in soybean (Israël et al. 1985) and in common bean (Hungria et al. 1985; Hungria and Neves 1987).

The QTLs identified in this study may be very useful in future marker-assisted selections to improve BNF in soybean. Furthermore, the soybean genome recently published reports the identification of fifty-two genes related to BNF (28 nodulins and 24 regulatory genes) (Schmutz et al. 2010), and the crossing of information from the genome with QTL mapping can provide evidence about the function of these genes in the BNF process.

**Conclusions**

Nitrogen fixation is, undoubtedly, an important component of sustainable agricultural systems and is specifically important for soybean, which accumulates high amounts of N. However, BNF is dependent on genetics and environmental factors, and the evaluation of traits related to it is laborious; therefore, selection assisted by molecular markers may be an approach worthy of consideration for improvement of BNF in breeding programs. We reported here four genomic regions associated with BNF; two of them, B1 and L, have been described in other mapping populations and can be useful in marker-assisted selection. Subsequent efforts to saturate the genetic map may allow the identification of other QTLs controlling BNF.

**Acknowledgements** – This research was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CNPq-MAPA (577933/2008) and CNPq-Repensa (562008/2010-1). M. A. Santos had a PhD fellowship from CNPq. I. O. Geraldi, A. A. F. Garcia and M. Hungria have research fellowships from CNPq. Authors thank Rinaldo B. Conceição and Lígia M. O. Chueire for technical help and Dr. Allan R. J. Eaglesham for scientific and English suggestions.

**REFERENCES**


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


