Theor Appl Genet (2015) 128:1999–2017
DOI 10.1007/s00122-015-2562-5

ORIGINAL ARTICLE

Genome-wide association analysis of symbiotic nitrogen fixation in common bean

Kelvin Kamfwa[1] · Karen A. Cichy[2] · James D. Kelly[1]

Received: 17 April 2015 / Accepted: 13 June 2015 / Published online: 2 July 2015

Abstract

Key message **Significant SNPs and candidate genes for symbiotic nitrogen fixation (SNF) and related traits were identified on Pv03, Pv07 and Pv09 chromosomes of common bean.**

Abstract A genome-wide association study (GWAS) was conducted to explore the genetic basis of variation for symbiotic nitrogen fixation (SNF) and related traits in the Andean Diversity Panel (ADP) comprising 259 common bean (*Phaseolus vulgaris*) genotypes. The ADP was evaluated for SNF and related traits in both greenhouse and field experiments. After accounting for population structure and cryptic relatedness, significant SNPs were identified on chromosomes Pv03, Pv07 and Pv09 for nitrogen derived from atmosphere (Ndfa) in the shoot at flowering, and for Ndfa in seed. The SNPs for Ndfa in shoot and Ndfa in seed co-localized on Pv03 and Pv09. Two genes *Phvul.007G050500* and *Phvul.009G136200* that code for leucine-rich repeat receptor-like protein kinases (LRR-RLK) were identified as candidate genes for Ndfa. LRR-RLK genes play a key role in signal transduction required for nodule formation. Significant SNPs identified in this study could potentially be used in marker-assisted breeding to accelerate genetic improvement of common bean for SNF.

Abbreviations

ADP	Andean Diversity Panel
BLASTn	Basic local alignment search tool for nucleotide
GWAS	Genome-wide association study
Ndfa	Nitrogen derived from the atmosphere
GH	Greenhouse
MLM	Mixed linear model
LD	Linkage disequilibrium
LRR-RLK	Leucine-rich repeat receptor-like protein kinase
N	Nitrogen
Pv	*Phaseolus vulgaris* chromosome
SNF	Symbiotic nitrogen fixation
SNP	Single nucleotide polymorphism

Introduction

Nitrogen (N) is frequently the most limiting nutrient for crop productivity (Boddey et al. 2009; Vance 2001). The two major sources of N for crop production are synthetic fertilizers and symbiotic nitrogen fixation (SNF) by legumes (Peoples et al. 2009a). SNF is the result of a symbiotic relationship between legumes and a diverse group of bacteria called Rhizobium (Graham 2009). This process begins with exchange of molecular signals between the legume root system and Rhizobium in the soil. The legume releases metabolites, usually flavonoids from its roots into the soil. This triggers the release of Nod factors (lipochitooligosaccharides) from the Rhizobium which when perceived by the plant induces the formation of an infection

Communicated by I. Rajcan.

✉ James D. Kelly
kellyj@msu.edu

[1] Department of Plant, Soil and Microbial Sciences, Michigan State University, 1066 Bogue St, East Lansing, MI 48824, USA

[2] USDA-ARS, Sugarbeet and Bean Research Unit, Michigan State University, 1066 Bogue St, East Lansing, MI 48824, USA

thread and subsequently specialized organs on the roots called nodules, which contain the Rhizobium (Gage 2009). When nodules are fully developed and functional, Rhizobium reduces atmospheric N_2 to ammonia, which is assimilated into forms of N that the plant can use (Strodtman and Emerich 2009). In return for the fixed N, the plant supplies the bacteria with photo-assimilates. SNF plays a significant role in crop productivity by providing N that is needed for plant growth and seed yield. In addition, SNF plays an important role in maintaining and enhancing soil fertility in a sustainable manner (Jensen and Hauggaard-Nielsen 2003). In low-input agricultural systems such as those in Africa and Latin America, SNF makes it possible to successfully grow grain legume crops with minimal N fertilizers (Mafongoya et al. 2009). This is important because most farmers in these regions do not have access or cannot afford fertilizers (Mafongoya et al. 2009). In countries where farmers can afford fertilizer and it is easily accessible, SNF still plays an important role because it reduces the amount of fertilizer applied thereby reducing the cost of producing the crop and potential ground water pollution (Vance 2001). In addition, SNF supports crop productivity in organic farming systems where artificial fertilizers cannot be applied.

Common bean (*Phaseolus vulgaris*) is the most important legume for direct consumption and a staple for millions of people in East Africa and South America (Akibode and Maredia 2012; Broughton et al. 2003). Common bean is considered weak in SNF in comparison with other major grain legumes (Bliss 1993). Reasons attributed to this shortcoming include the short growing season for most common bean genotypes that limits the supply of photo-assimilates to nodules (Graham et al. 2003). Depending on the environment and genotype, estimates of N fixed by common bean range from 0 to 165 kg ha^{-1} with an average of 55 kg ha^{-1}, which is considered low when compared with other major grain and pasture legumes (Giller 2001; Graham et al. 2003; Unkovich and Pate 2000). Whereas the amounts of N fixed by soybean (*Glycine max*) are adequate for successful production without synthetic N fertilizer, common bean requires that fixed N be supplemented with N fertilizer (Giller 2001). Indeed, common bean seed yield response to N fertilizer application tends to be significant (Giller 2001; Herridge and Redden 1999). Enhancing the SNF process in common bean has potential to improve its overall productivity. Genetic variability for SNF and associated traits within common bean has been widely reported (Buttery et al. 1997; Elizondo Barron et al. 1999; Graham and Rosas 1977; Graham 1981; Herridge and Redden 1999; Pereira et al. 1993), suggesting that genetic improvement would be feasible. Efforts to improve SNF in common bean from the Middle American gene pool have been successful and resulted in release of cultivars with enhanced SNF (Bliss et al. 1989). However, sustained success in developing Andean cultivars with enhanced SNF has been elusive. Most bean breeding programs do not routinely select for enhanced SNF because phenotyping for SNF is laborious and expensive especially when ^{15}N-isotope method (Shearer and Kohl 1986) is used. Genetic improvement for SNF has been hampered by its genetic complexity. Several plant traits including nodulation, photosynthesis, biomass accumulation, photo-assimilate partitioning to the nodules are involved in SNF. Many genes control these traits, and the environment significantly affects their expression, which limits the genetic enhancement of SNF. Understanding the genetic architecture of SNF in terms of the genomic regions and/or genes involved and their effects is critical to enhancing our knowledge of its genetic control.

Developing molecular markers that can be used by breeders to indirectly select for SNF would circumvent the challenges of direct selection for SNF, and accelerate the development of common bean cultivars with enhanced SNF. Given the importance of SNF, few studies to understand the genetic architecture of SNF in common bean and other economically important legumes exist. Only four previous QTL mapping studies on SNF and related traits in common bean have been published (Nodari et al. 1993; Ramaekers et al. 2013; Souza et al. 2000; Tsai et al. 1998). Three of these studies used a population of recombinant inbred lines (RILs) from a cross of BAT 93 × Jalo EEP558 grown in the greenhouse (GH) to identify QTL for nodule number on Pv01, Pv02, Pv03, Pv05, Pv06, Pv07, Pv09, Pv10 and Pv11 (Nodari et al. 1993; Souza et al. 2000; Tsai et al. 1998). A common theme among these studies is the use of nodule number to indirectly assay for SNF. SNF was directly assayed using ^{15}N natural abundance method (Shearer and Kohl 1986) to map QTL for SNF in only one study (Ramaekers et al. 2013). QTL for percentage of N derived from atmosphere (%Ndfa) were identified on Pv01, Pv04, and Pv10, and nodule dry weight on Pv03 in a population of 83 $F_{5:8}$ RILs from a G2333 × G19839 cross grown in the field (Ramaekers et al. 2013). When the same population was evaluated in the GH, QTL for total N were identified on Pv04 and Pv10 (Ramaekers et al. 2013).

Enhancing SNF in Andean beans could be one way of improving their productivity as the yields of Andean beans continue to lag behind those of Middle American beans (Vandemark et al. 2014). Andean beans are the most widely grown bean types in Africa (Beebe 2012), mostly by small-scale farmers who cannot afford fertilizer and are reliant on SNF as a source of N. Though, Middle American genotypes with enhanced SNF have been identified, introgressing SNF genes from Middle American germplasm is normally constrained by genetic incompatibilities and difficulties in recovering the large Andean seed size in progenies (Singh and Gutiérrez 1984). Identifying superior genotypes for

SNF from within the Andean gene pool could circumvent these challenges. In this study, we used a genome-wide association study (GWAS) to explore the genetic architecture of SNF in a panel comprising 259 Andean bean genotypes with the goal to identify traits and genomic regions associated with improved SNF in Andean beans.

Materials and methods

Plant materials

A subset of the Andean Diversity Panel (ADP) comprising 259 Andean bean genotypes from Africa, South America, North America, Central America, Caribbean, Asia and Europe was used in this study (Cichy et al. 2015). The genotypes in the ADP included varieties, elite lines and landraces. The ADP was evaluated in replicated greenhouse (GH) and field experiments in Michigan, USA. More details about the makeup of the diversity panel can be found in Cichy et al. (2015). This panel was used recently to identify QTL for agronomic traits and a candidate gene for days to flowering and maturity in common bean (Kamfwa et al. 2015). Two Andean non-nodulating mutants (nonods), G51396A and G51493A were included as checks in both greenhouse and field experiments. Both G51396A and G51493A have determinate growth habit. In addition, the no-nods were used to calculate nitrogen derived from atmosphere (Ndfa) in field experiments.

Greenhouse experiments

The ADP was evaluated for SNF and related traits in the greenhouse at Michigan State University, East Lansing, Michigan (MI), USA in 2012 and 2014 hereafter referred to as GH_2012 and GH_2014, respectively. In both GH_2012 and GH_2014, 259 Andean genotypes including two nonods genotypes were planted in a randomized complete block design with two replications. Before planting, seeds were sterilized in sodium hypochlorite and then rinsed in distilled water. Then eight seeds for two replications of each genotype were inoculated with *Rhizobium tropici* strain CIAT 899 by submerging seeds for 2 min in a broth culture of Rhizobium made from yeast extract mannitol media (Vincent 1970). Four seeds of each genotype were planted in a 4-l plastic pot filled with perlite and vermiculite in a 2:1 (v/v) ratio. Ten days after planting, thinning was done to leave two plants in each pot. A second inoculation was done by applying 1 ml of CIAT 899 broth. The N-free nutrient solution (Broughton and Dilworth 1970) was applied once per day through drip irrigation until flowering, when plants were harvested. Throughout the experiment, 13 h of supplemental light per day was provided, and temperature was maintained at 24 °C in the GH. At 32 days after planting, chlorophyll content was measured using a soil and plant analysis development (SPAD) chlorophyll meter (SPAD-502Plus) on one fully developed leaf of each of the two plants in the pot (Uddling et al. 2007). An average of these two values was computed. SPAD meter measures the absorbance of the leaf in the red and near-infra red regions. Based on these two absorbance values, the meter calculates a numerical value proportional to chlorophyll content in the leaf (Uddling et al. 2007). At flowering, plants were harvested by carefully shaking off the perlite/vermiculite media. Roots were carefully washed in water to avoid losing the nodules. The plant was then separated into roots, nodules and shoot. After this nodules and the shoot were dried in the oven at 60 °C for 72 h. The nodule dry weight and shoot biomass of the two plants for each genotype were recorded. The shoot was then ground with a Christy-Turner Lab Mill to pass through a 1 mm screen. About 5 mg of ground tissue were prepared and shipped to University of California, Davis Stable Isotope Facility for ^{15}N natural abundance and total N analyses. Both GH_2012 and GH_2014 experiments were handled similarly.

Field experiments

The ADP was evaluated for SNF and related traits in the field at the Montcalm Research Farm near Entrican, MI, USA, in 2012 and 2013 growing seasons, hereafter referred to as Field_2012 and Field_2013, respectively. In Field_2012, 259 ADP genotypes were evaluated, whereas the number was reduced to 237 in Field_2013. The reduction resulted from the elimination of genotypes that showed lack of adaptation to temperate field conditions in Michigan in Field_2012 and also due to limited seed quantities for some genotypes. The farm is located in central Michigan where Andean beans are produced on course textured sandy soils. The soil type is a combination of Eutric Glossoboralfs (coarse-loamy, mixed) and Alfic Fragiorthods (coarse-loamy, mixed, frigid) and rainfall was supplemented with overhead irrigation as needed. No fertilizer was applied to plots and recommended practices were followed for weed and insect control. Soil samples collected from the trial site before planting showed that in Field_2012 nitrate level in the soil was on average 36 mg kg^{-1}, whereas in Field_2013 it was 2.4 mg kg^{-1}. Before planting, seed was inoculated with commercial *Rhizobium* 'Nodulator' (Becker Underwood, Ames, IA, USA) with an undisclosed strain at the rate suggested on the package. Common bean has been grown on this site for many years and there is adequate native *Rhizobium*. In both seasons, the ADP was planted in a randomized complete block design with two replications. Each genotype was planted in two-row plots of 4.75 m long each and inter-row spacing of 0.50 m. The two

Andean no-nod mutants were included in the planting as checks and also for computations of amounts of N fixed. At 35 days after planting, chlorophyll content was measured on three plants in each plot using SPAD meter and then the average value was calculated. Days to flowering were recorded on all entries in both years. In Field_2012 at flowering, two plants were sampled from each plot by digging with a shovel and carefully removed the soil. These two plants were separated into shoot and roots. A visual nodulation score of 0–6 based on the number of nodules was recorded for each genotype in both years. A nodule score of 0 represented no nodules while 6 was for high nodulation. Roots were discarded while the shoots were oven-dried at 60 °C for 72 h and then weighed. Shoot was ground using Christy-Turner Lab Mill to pass through a 1 mm screen. About 5 mg of ground tissue was shipped to University of California, Davis, Stable Isotope Facility for ^{15}N natural abundance and total N analyses.

In the current study, the primary method used to evaluate SNF was estimating %Ndfa and Ndfa in shoot biomass at flowering in both GH and field experiments. However, from the plant breeding standpoint of enhancing SNF, an ideal genotype is one that not only fixes adequate N, and stores it in leaves and stems but one that is also efficient in partitioning or remobilizing fixed N from the stems, leaves, and pod walls to the seed, which is the economic yield. To identify genotypes with enhanced N fixation, and efficiency in partitioning and remobilizing the fixed N to the seed, we measured total N and ^{15}N natural abundance in the seed, which were used to estimate %Ndfa and Ndfa. Seed harvested from Field_2013 was ground to powder and sent to UC Davis Stable Isotope Facility for total N and ^{15}N natural abundance analyses. We focused on Field_2013 experiment because our interest was to determine N in the plant derived from fixation under low soil N levels. The data on %Ndfa and Ndfa in seed was also used in association analyses to determine whether genomic regions associated with %Ndfa and Ndfa in shoot biomass at flowering would co-localize with %Ndfa and Ndfa estimated using seed.

GH and field estimation of N fixed

The amount of N fixed or Ndfa by a single plant grown in GH experiments was estimated as total shoot biomass at flowering multiplied by the %N in shoot biomass, minus the total N of a non-fixing mutant. Although we used N-free nutrient solution to grow plants, we could not assume that the entire N in the plants came from fixation as the seed contains approximately 4 % N that sustains the plant prior to fixing N. It was on this premise that we subtracted the N of non-fixing mutants from that of fixing genotypes to have a more accurate estimate of N fixed under GH conditions.In field experiments the amount of Ndfa in the shoot biomass at flowering and in seed were estimated using the ^{15}N natural abundance method (Giller 2001; Shearer and Kohl 1986). This isotopic method has been reported to give more accurate estimates of Ndfa under field conditions (Peoples et al. 2009b). In this method the proportion of total N in the plant that was derived from fixation (%Ndfa) and N from fixation (Ndfa) were estimated using the following two equations from Giller (2001):

$$\%\text{Ndfa} = \left(\delta^{15}\text{N}_{\text{reference plant}} - \delta^{15}\text{N}_{\text{fixing plant}}\right) / \left(\delta^{15}\text{N}_{\text{reference plant}} - B\right) \quad (1)$$

$$\text{Ndfa} = \text{N}_{\text{total}} \times \%\text{Ndfa} \quad (2)$$

For Eq. 1: %Ndfa is percentage of N in the shoot biomass at flowering or in seed that is derived from atmosphere, i.e., N fixed, hereafter referred to as %Ndfa_Shoot or %Ndfa_Seed, respectively; $\delta^{15}\text{N}_{\text{reference plant}}$ is the ^{15}N in the no-nod, non-fixing plant (we used an average of two no-nods); $\delta^{15}\text{N}_{\text{fixing plant}}$ is the ^{15}N in the fixing plant; '*B*' is the δ^{15}N of the same N fixing plant when grown in N-free GH conditions. For estimating %Ndfa using shoot biomass at flowering each genotype had its own '*B*' value derived from GH evaluation of the ADP. This '*B*' was also used to estimate %Ndfa using seed. For Eq. 2: Ndfa is the N amount in the shoot biomass at flowering or N amount in the seed that is derived from atmosphere, i.e., fixed N, hereafter referred to as Ndfa_Shoot and Ndfa_Seed, respectively; N_{total} is the total N in the fixing plant that includes both fixed N and mineral N from the soil; %Ndfa is the N percent derived from atmosphere computed in Eq. 1. In the case of Ndfa_Seed, N_{total} refers to the total N in the seed yield per hectare that was computed from plot seed yield in Field_2013.

Phenotypic data analyses

Statistical analyses for field data were conducted using mixed models in SAS 9.3 (SAS Institute 2011). Assumption for normally distributed residuals required for analysis of variance (ANOVA) and SNP–trait association test was checked for all traits measured. Normality tests were conducted on the combined residuals of all treatments for each trait using PROC UNIVARIATE and traits that were not normally distributed were transformed. Normality test results indicated that all traits except seed N percentage were not normally distributed. Therefore, all traits except seed N percentage were transformed using natural logarithmic transformation for use in ANOVA and GWAS analyses. All the trait means are reported in their original values. An ANOVA using PROC MIXED was conducted on all the traits based on the following statistical model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(j)} + \mathcal{E}_{ijk},$$

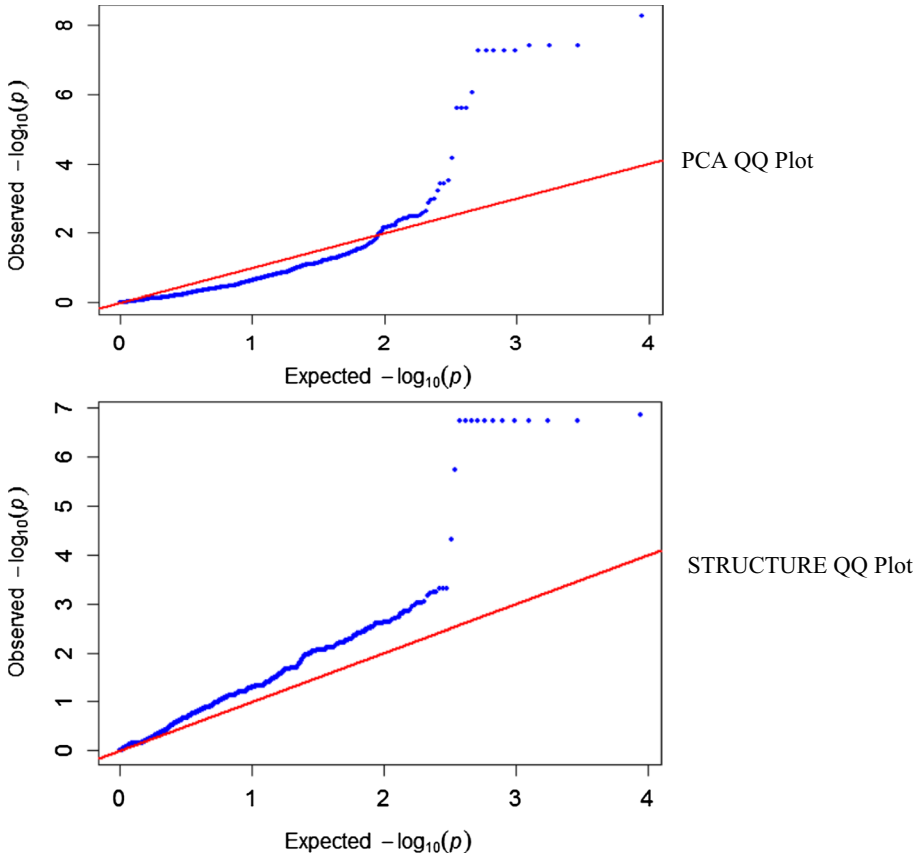

Fig. 1 The quantile–quantile (QQ plots) plots for seed nitrogen percentage comparing the effectiveness of using principal component analyses (PCA) and STRUCTURE software to control population structure in association tests using mixed linear model

where Y_{ijk} is the response variable, e.g., Ndfa, with genotype i in the environment j, replication k within environment; α_i was the fixed effect of the genotype i; β_j was the random effect of the environment j; $\alpha\beta$ was the random effect of the interaction between genotype i and environment j; γ was the random effect of a replication with environment j; ε was the random error term, which was assumed to be normally distributed with mean = 0. Genetic correlation analyses were conducted on selected traits using multivariate restricted maximum likelihood estimation with SAS PROC MIXED as described in Holland (2006).

Genotyping

DNA was collected as described in Cichy et al. (2015). DNA samples were genotyped using an Illumina BARCBean6K_3 BeadChip with 5398 SNPs (Hyten et al. 2010) in the Soybean Genomics and Improvement USDA Laboratory (USDA-ARS, Beltsville Agricultural Research Center) in MD, USA. The SNP genotyping was conducted on the Illumina platform by following the Infinium HD Assay Ultra Protocol (Illumina Inc.). The Infinium II assay protocol includes the procedures to make, incubate, and fragment amplified DNA, prepare the bead assay, hybridize samples to the BARCBean6K_3 BeadChip, extend and stain samples, and image the bead assay. The SNP alleles were called using the GenomeStudio Genotyping Module v1.8.4 (Illumina, Inc.). The data were manually adjusted for allele calls.

Population structure analysis and marker–trait association tests

After filtering for low-quality and monomorphic SNPs, 5326 SNPs were retained. These were filtered further for minor allele frequency (MAF $\geq$ 0.05) and a final total of 4623 SNPs were used in population structure, kinship, and association analyses. GWAS results can be confounded by population stratification. The two popular methods for detecting population stratification in association panels are principal component analysis (PCA) (Price et al. 2006) and subpopulation clustering using STRUCTURE (Pritchard et al. 2000). To decide on the best method to use in our study, we compared the effectiveness of PCA and STRUCTURE based on quantile–quantile (QQ) plots from association tests for all traits using a Mixed Linear Model implemented in TASSEL 5.0 (Bradbury et al. 2007). Principal component analysis (PCA) was implemented in software program EIGENSTRAT (Price et al. 2006). As illustrated in the QQ plots for seed N percentage (Fig. 1), both PCA

and STRUCTURE were effective in controlling for population structure. In both methods, there was a near-agreement of plots of expected and observed p values with the $X = Y$ line until a sharp curve towards the end representing what may be true associations. The trend was similar for all traits. Based on these results, we chose PCA for assessing the population structure in the panel and account for it in association tests for all traits reported in this study. To correct for cryptic relatedness in the panel, kinship matrix (K) was included in our association analyses. The kinship matrix was calculated using scaled identity by descent method implemented in TASSEL 5.0 (Bradbury et al. 2007). To determine the SNP–trait associations, we used a Mixed Linear Model (MLM) (Zhang et al. 2010) implemented in software program TASSEL 5.0. The following MLM equation was used:

$$Y = X\alpha + P\beta + K\mu + \varepsilon,$$

where Y the phenotype of a genotype; X was the fixed effect of the SNP; P was the fixed effect of population structure (from PCA matrix); K was the random effect of relative kinship, i.e., cryptic relatedness among genotypes (from kinship matrix); ε was the error term, which was assumed to be normally distributed with mean = 0. To estimate the proportion of phenotypic variation accounted for by a significant SNP, we used the R^2 computed in TASSEL. We used the conservative Bonferroni correction to control for error (false positives) associated with multiple tests. The Bonferroni-corrected threshold P value of 1.1×10^{-5} was calculated for 4623 SNP–trait association tests for each trait at $\alpha = 0.05$.

Candidate gene identification

We used Jbrowse on Phytozome v10 (Goodstein et al. 2012) to browse the common bean genome version 1.0 (Schmutz et al. 2014), to gain insights into positional candidate genes associated with significant SNPs. A gene was considered a candidate gene if it contained a significant SNP (Bonferroni corrected $P = 1.1 \times 10^{-5}$) or if there was a significant SNP within the immediate genomic region (±20 kb). The focus was on the most significant SNPs (peak SNPs with strongest signal) on each chromosome. We also conducted LD analyses in TASSEL 5.0 (Bradbury et al. 2007) to determine the strength of LD between the most significant SNP and its immediate surrounding significant SNPs to be more confident that they were tagging the same candidate gene. In addition, the gene was considered a candidate gene if it coded for a protein whose role or possible role in SNF or related traits had been established or proposed. If there was no functional annotation on Phytozome v10, we did a BLAST search (Zhang et al. 2000) using the genomic sequence as a query against the *Arabidopsis thaliana* genome on TAIR (Rhee et al. 2003) and soybean genome on NCBI.

Results

Population structure

The PCA indicated the presence of a population structure among genotypes in the ADP. The first, second, third and fourth principal components accounted for 41.3, 7.1, 4.3 and 2.6 % of the genotypic variability in the ADP, respectively. A plot of PC1 against PC2 revealed the existence of two clusters (Fig. 2). The smaller cluster comprising 22 genotypes. Fourteen of these were landraces from East Africa, five were cultivars from North America, two were cultivars from the Caribbean and one was a cultivar from South America. The larger cluster had 239 genotypes comprising landraces, elite lines and cultivars from several geographic regions. To account for this population structure, we used the first four PCs that together explained over 55.3 % of the genotypic variation in the ADP in MLM for association tests. The variation explained by subsequent PC's after the fourth PC only had marginal incremental values hence the decision to use only the first four PCs.

Greenhouse experiments

Highly significant differences ($P < 0.001$) were observed among ADP genotypes in GH_2012 and GH_2014 for chlorophyll content, nodule dry weight, shoot biomass, N % in shoot biomass and Ndfa (Table 1). The means of these five traits were slightly higher in GH_2014 than GH_2012. The frequency distribution graphs for Ndfa_Shoot in GH_2012 and GH_2014 showed a continuous distribution that is typical of a quantitative trait (Fig. 3). There were several significant genetic correlations among traits measured in the GH. In GH_2012, Ndfa significantly correlated with chlorophyll content ($r = 0.49$; $P < 0.05$), shoot biomass ($r = 0.98$; $P < 0.05$), and nodule dry weight ($r = 0.8$; $P < 0.05$).

Field experiments

There were highly significant differences ($P < 0.001$) among ADP genotypes evaluated in both Field_2012 and Field_2013 for chlorophyll content, nodule score, shoot biomass at flowering, and N % in shoot biomass. Genotype differences in N % in the seed were also significant in Field_2013. The means for chlorophyll content and shoot biomass were higher in Field_2012 than Field_2013 while the mean for nodule score was higher, in Field_2013 than Field_2012 (Table 1). In both Field_2012 and Field_2013, highly significant differences among

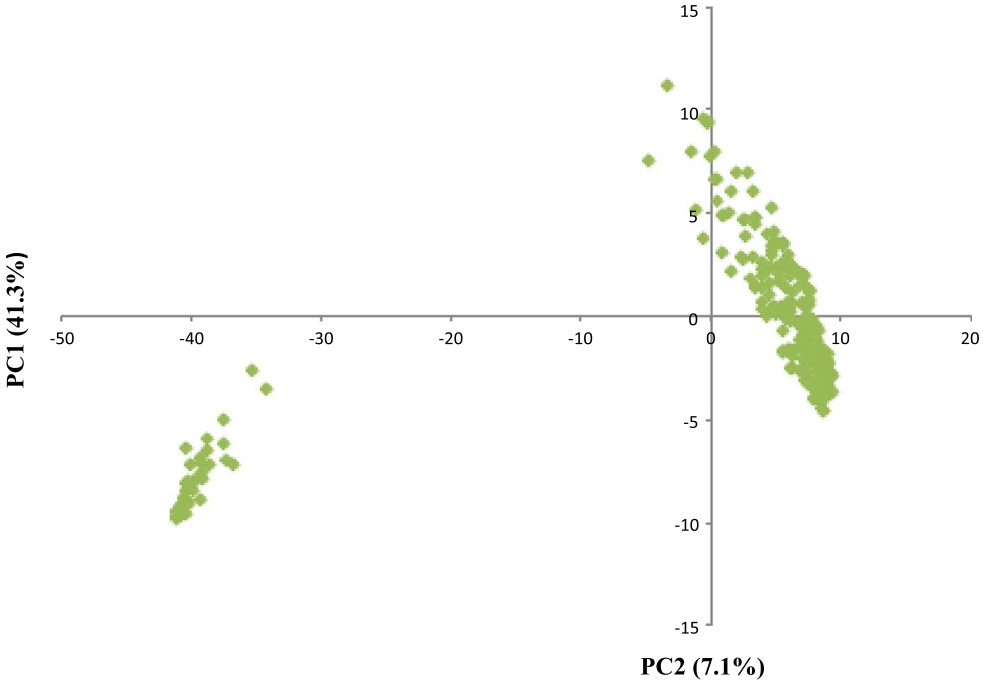

Fig. 2 Principal component analysis (PCA) plot of PC1 against PC2 illustrating the population structure comprising two major sub-groups in the Andean Diversity Panel

genotypes were observed for %Ndfa_Shoot and Ndfa_Shoot. In Field_2012, %Ndfa_Shoot ranged from 0.8 to 41.4 % with an average of 12.4 %, and exhibited a narrower range and smaller average than in Field_2013 where %Ndfa_Shoot ranged from 1.7 to 88.5 % with an average of 37.3 % (Table 1). This represented a nearly fourfold increase in Ndfa_Shoot average between Field_2012 and Field_2013. In Field_2013, there was a strong correlation ($r = 0.9$; $P < 0.05$) between Ndfa_Shoot and shoot biomass but in Field_2012 this correlation was not significant. Correlation between Ndfa_Shoot and chlorophyll content was significant in Field_2013 but not in Field_2012. Genotype by year interactions were highly significant ($P < 0.001$) for all six traits recorded in Field_2012 and Field_2013. The best performing genotype for Ndfa_Shoot in Field_2013 was ADP631, a Canadian light red kidney cultivar, OAC Inferno whose %Ndfa_Shoot estimate was 88.5 %. However, in Field_2012 the %Ndfa_Shoot for ADP631 was estimated at 7 %.

Significant differences for %Ndfa_Seed and Ndfa_Seed were observed among genotypes in Field_2013. The %Ndfa_Seed ranged from 3.6 to 98.2 % with an average of 45.5 %. The Ndfa_Seed ranged from 1.4 to 98.6 kg ha^{-1} with an average of 29.5 kg ha^{-1}. The Ndfa_Seed frequency distribution graph follows a pattern consistent with that for a quantitative trait (Fig. 3). There was a significant ($r = 0.30$; $P < 0.05$) correlation between %Ndfa_Shoot and %Ndfa_Seed. The 10 genotypes that performed better in both %Ndfa_Seed and Ndfa_Seed and %Ndfa_Shoot are shown in Table 2. These 10 genotypes had %Ndfa_Seed and Ndfa_Seed greater than 70 % and 55 kg ha^{-1}, respectively, and could be considered superior in both N fixation and partitioning of fixed N to the seed.

Marker–trait associations

Chlorophyll content

In GH_2012, two SNPs both on chromosome Pv09 were significantly associated with chlorophyll content. The most significant SNP (ss715647747; $P = 6.1 \times 10^{-06}$) explained about 7 % of the variation in chlorophyll content (Table 3). In GH_2014 three SNPs, all on Pv09 were significantly associated with chlorophyll content, and the most significant SNP (ss715648916; $P = 3.1 \times 10^{-6}$) explained about 8 % of the genetic variability. The two most significant SNPs in GH_2012 and GH_2014 (ss715647747 and ss715648916) were in strong linkage disequilibrium (LD) ($r^2 = 0.98$; $D' = 1$). The most significant SNP (ss715647747) on Pv09 in GH_2012 was also significant in GH_2014.

In Field_2012, only one significant SNP for chlorophyll content was identified on Pv01 that explained about 9 % of variation (Table 3). Significant SNPs for chlorophyll content were identified in Field_2013 on Pv09. In Field_2013, the most significant SNP was ss715648916 and explained about 7 % the chlorophyll content variation (Table 3). This SNP was also the most significant for chlorophyll content in GH_2014. Some of the significant SNPs for chlorophyll content on Pv09 were also significant for other traits in GH and field experiments. Significant SNP ss715648916 for chlorophyll content in GH_2014 was also significant for

Table 1 Means and ranges for traits associated with Symbiotic Nitrogen Fixation in Andean Diversity Panel of 259 common bean genotypes grown in the GH in 2012 and 2014 at Michigan State University, East Lansing, MI and in the Field at Montcalm Research Farm, Entrican, MI in 2012 and 2013

Trait	Experiment	Mean	Min.	Max.
Chlorophyll content (SPAD)	GH_2012	29.8 ± 0.2	20.6	39.8
	GH_2014	35.1 ± 0.2	25.1	45.8
	Field_2012	35.5 ± 0.1	27.0	44.9
	Field_2013	32.2 ± 0.1	22.2	40.8
Nodule dry wt./plant (mg)	GH_2012	117 ± 2.0	37	272
	GH_2014	140 ± 3.0	49	285
Nodule score (0–6 scale)	Field_2012	2.9 ± 0.1	0.5	6.0
	Field_2013	4.1 ± 0.1	1.0	6.0
Shoot biomass/plant (g)	GH_2012	3.1 ± 0	1.3	6.8
	GH_2014	3.6 ± 0.1	1.1	8.3
	Field_2012	20.0 ± 0.2	8.7	38.4
	Field_2013	10.1 ± 0.1	5.1	17.0
N % in shoot biomass	GH_2012	3.2 ± 0	2.2	4.5
	GH_2014	2.8 ± 0	2.1	3.7
	Field_2012	3.1 ± 0	2.3	4.1
	Field_2013	3.0 ± 0	1.7	3.9
N % in seed	Field_2013	3.9 ± 0	3.1	4.7
%Ndfa_Shoot	Field_2012	12.4 ± 0.6	0.8	41.4
	Field_2013	37.3 ± 0.8	1.7	88.5
%Ndfa_Seed	Field_2013	45.5 ± 1.2	3.6	98.7
Ndfa_Shoot/plant (mg)	GH_2012	59 ± 1	10	112
	GH_2014	62 ± 1	9	130
	Field_2012	71 ± 4	2	273
	Field_2013	123 ± 4	6	523
Ndfa_Seed (kg ha^{-1})	Field_2013	29.5 ± 2.7	2.9	92.3

Ndfa nitrogen derived from atmosphere, *GH_2012* evaluations in the GH in 2012, *GH_2014* evaluations in the GH in 2014, *Field_2012* evaluations in the field in 2012, *Field_2013* evaluations in the field in 2013, ±SE the mean; Max and Min represent the range a trait

shoot biomass in GH_2014. Another SNP (ss715647747) significant for chlorophyll content in GH_2013 was also significant for shoot biomass in GH_2013. Significant SNPs for chlorophyll content on Pv09 were consistently identified in two GH experiments and in Field_2013. In some cases the significant SNPs for chlorophyll content in GH_2014 were same as those in Field_2013 (Table 3).

Nodulation

Nodulation was evaluated as nodule dry weight in the GH, and as nodule score in the field experiments. Two SNPs were significant for nodulation in Field_2013, and both were located on Pv09 (Table 3). The most significant SNP (ss715648787; $P = 1.1 \times 10^{-6}$) explained about 12 % of the variability in nodule scores in Field_2013. This SNP was also significant for chlorophyll content (GH_2014 and Field_2013), shoot biomass (GH_2014), N % in shoot biomass (GH_2014 and Field_2013), N % in seed (Field_2013), %Ndfa_Shoot (Field_2013), Ndfa_Shoot (Field_2013 and GH_2014), and Ndfa_Seed (Table 3). No significant SNPs for nodulation were identified in Field_2012 or the two GH experiments.

Shoot biomass

Significant SNPs for shoot biomass were identified in both GH and field experiments. In GH_2012, 11 SNPs on Pv01, Pv03, Pv07, and Pv09 were significant. The most significant SNP was on Pv09, and explained about 11 % of variation in shoot biomass (Table 3). In GH_2014, four SNPs on Pv09 were significant. The most significant ($P = 1.3 \times 10^{-08}$) SNP was ss715648916, and explained about 13 % of variability in shoot biomass in GH_2014. This SNP was also significant for chlorophyll content, nodulation, N % in shoot biomass, %Ndfa and Ndfa in GH, and field experiments (Table 3). In Field_2012, five SNPs on Pv07 and Pv08 were significant, the most significant being on Pv07 that explained about 11 % of the variation in shoot biomass. In Field_2013, three significant SNPs, on Pv04 were significant. The most significant ($P = 6.4 \times 10^{-08}$) SNP in Field_2013 explained about 17 % of variation shoot biomass.

N percentage in biomass

Significant SNPs for N % in shoot biomass were identified in GH_2014 and Field_2013. In GH_2014, three SNPs on Pv03 and ten SNPs on Pv09 were significant. The most significant (3.7×10^{-09}) SNP in GH_2014 was ss715648916 on Pv09 and explained about 15 % of variation (Table 3; Fig. 4). In Field_2013, two SNPs on Pv09 were significant. The most significant SNP was the same one detected in GH_2014 but with a lower R^2 value of 10 %. This SNP (ss715648916) on Pv09 that was consistently significant in both GH and field experiments for N % in shoot biomass was also significant for N % in the seed (Fig. 4). In addition, ss715648916 was significant for chlorophyll content, nodulation, shoot biomass, %Ndfa_Shoot, Ndfa_Shoot, Ndfa_Seed in GH and field experiments (Table 3). No significant SNPs were identified for N % in shoot biomass in GH_2012 and Field_2012.

N percentage in seed

A total of 17 SNPs were significantly associated with N percentage in the seed. Sixteen SNPs were on Pv09 and one SNP was on Pv03 (Fig. 1). The most significant SNP (ss715648916; $P = 1.1 \times 10^{-9}$) was on Pv09 that explained about 17 % of variation in N percentage in the seed (Table 3). This SNP was also significant for N

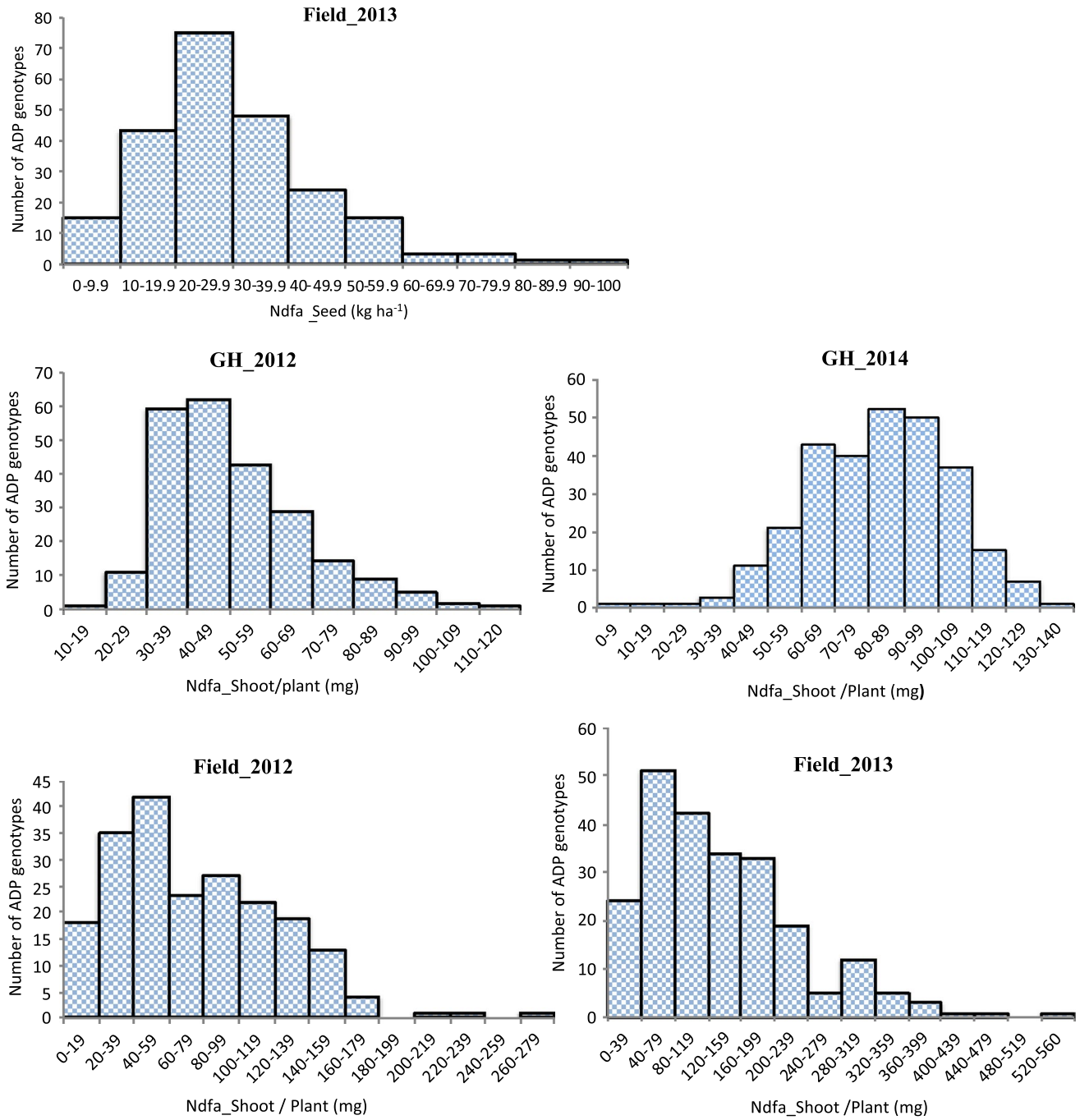

Fig. 3 Frequency distribution graphs for nitrogen derived from atmosphere in the seed (Ndfa_Seed) for Field_2013, and nitrogen derived from atmosphere in the shoot at flowering (Ndfa_Shoot) of the Andean diversity panel genotypes evaluated in the greenhouse (GH) and field

percentage in the shoot biomass at flowering in (GH_2012, GH_2013 and Field_2013), Ndfa (GH_2012), Ndfa_Shoot (Field_2013), and Ndfa_Seed (Field_2013).

%Ndfa in shoot biomass at flowering in field experiments

In Field_2013, significant SNPs for %Ndfa_Shoot were identified on Pv02, Pv03, Pv07, Pv09, Pv10 and Pv11 (Table 3). The most significant (ss715646392; $P = 2.9 \times 10^{-13}$) SNP was on Pv03 and explained about 22 % of variation in Field_2013. The most significant for %Ndfa_Shoot on Pv09 (ss715648916) that explained about 19 % of variation was also significant for chlorophyll content, shoot biomass, N % in shoot biomass and Ndfa in GH and field experiments (Table 3). There were no significant SNPs for %Ndfa_Shoot in Field_2012.

Table 2 Ten genotypes identified as superior in percentage of N derived from atmosphere (%Ndfa) and amounts of N in seed derived from atmosphere (Ndfa) from the Andean Diversity Panel and two non-nodulating mutants grown in the Field at Montcalm Research Farm, Entrican, MI in 2013

ID	Cultivar name	Country (region)	Seed color	DTM	%Ndfa_Shoot	%Ndfa_Seed	Ndfa_Seed (kg ha^{-1})	Seed yield (kg ha^{-1})
Ten ADP Genotypes								
ADP001	Rozi Koko	Tanzania (Africa)	Red mottled	94	78.2	84.1	89.5	2094
ADP280	G14440	Spain (Europe)	White	95	77.5	86.8	57.0	1904
ADP303	G17913	Hungary (Europe)	Biege	87	67.2	81.9	69.2	1594
ADP437	PC-50	Dominican (Caribbean)	Red mottled	93	80.9	84.9	64.9	1958
ADP483	PI209815	Kenya (Africa)	Yellow	93	61.2	76.7	60.0	1980
ADP601	Camelot	US (N. America)	DRK	86	51.0	94.2	61.3	1679
ADP631	OAC Inferno	Canada (N. America)	LRK	95	88.5	98.2	72.4	2253
ADP644	Fox Fire	US (N. America)	LRK	77	73.4	98.7	92.3	2570
ADP680	Clouseau	US (N. America)	LRK	82	65.4	84.2	89.6	2938
ADP684	Majesty	Canada (N. America)	DRK	85	68.2	78.1	56.4	1814
Experimental checks (non-nodulating mutants)								
G51493A	NA	–	Yellow Mottled	57	0	0	0	450
G51396A	NA	–	DRK	56	0	0	0	424
$LSD_{0.05}$ (for ADP genotypes)				4.4	32	33	29	893

ADP Andean Diversity Panel Identity, *DTM* days to maturity, *%Ndfa* percentage of N derived from atmosphere in the shoot biomass at flowering, *%Ndfa_Seed* percentage of N in the seed derived from atmosphere, *Ndfa_Seed* amount of N (kg ha^{-1}) in the seed derived from the atmosphere, *LRL* light red kidney, *DRK* dark red kidney, *LSD* least significant difference

Ndfa in shoot biomass at flowering in GH and field experiments

Significant SNPs for Ndfa were identified in both GH and field experiments (Fig. 5). In GH_2012, a total of 12 SNPs on Pv03, Pv07 and Pv09 were significant for Ndfa. The highest number (nine) of significant SNPs was on Pv07. The most significant SNP was on Pv09, and explained about 13 % of variability in Ndfa (Table 3). Most of the significant SNPs for Ndfa in GH_2012 were also significant for shoot biomass in GH_2012. In GH_2014, one SNP on Pv02, ten SNPs on Pv03 and 12 SNPs on Pv09 were significant for Ndfa. The most significant SNP (ss715648916; $P = 3.4 \times 10^{-13}$) was on Pv09, and explained about 20 % of variation in Ndfa in GH.

In Field_2013, a total of 25 SNPs on Pv02, Pv03, Pv07, Pv09, Pv10 and Pv11 were significant for Ndfa_Shoot. The most significant SNP was on Pv03 and explained about 23 % of Ndfa_Shoot variation in Field_2013 (Table 3). Most significant SNPs for Ndfa_Shoot in Field_2013 were consistently significant for Ndfa in GH_2012 and GH_2014 (Fig. 5). There were no significant SNPs for Ndfa_Shoot identified in Field_2012.

Ndfa and %Ndfa in seed for Field_2013

A total of 11 SNPs, five on Pv03 and six on Pv09 were significant for Ndfa_Seed in Field_2013 (Fig. 3). The most significant SNPs on Pv03 (ss715646392) and Pv09 (ss715648916) explained about 9 and 11 %, respectively, of Ndfa_Seed variability in Field_2013. These two most significant SNPs for Ndfa_Seed on Pv03 and Pv09 were also the most significant for Ndfa_Shoot on Pv03 and Pv09 in GH_2014 and Field_2013 (Table 3). In addition, ss715648916 on Pv09 was also significant for nodulation (Fig. 6), chlorophyll content, N percentage in shoot biomass at flowering and N percentage in seed in Field_2013. No significant SNPs for %Ndfa_Seed were identified.

Allelic effects of significant SNPs on Ndfa_Shoot

Using Ndfa_Shoot data from Field_2013, we assessed the allelic effects on Ndfa_Shoot of significant SNPs located on Pv03, Pv07, and Pv09. The most significant SNP on Pv03 ss715646392 had C as its minor allele (0.05; Table 3), and T as the major allele, which had a major effect on Ndfa_Shoot. The Ndfa_Shoot for homozygous TT and CC were 125 and 94 mg N per plant, respectively. The most significant SNP on Pv07 was ss715646473. The minor allele for this SNP was G (MAF = 0.06; Table 3) while A was the major allele and was the allele that had a major effect on Ndfa_Shoot in Field_2013. At this SNP, the homozygous AA and GG genotypes for Ndfa_Shoot were 120 and 92 mg N per plant, respectively. The most significant SNP on Pv09 was ss715648916. The minor allele

Table 3 Most significant SNP and candidate genes on relevant *Phaseolus vulgaris* chromosomes for SNF and related traits of the Andean Diversity Panel common bean genotypes evaluated in the GH at Michigan State University, East Lansing, MI in 2012 and 2014, and in the Field at Montcalm Research Farm, Entrican, MI in 2012 and 2013

Trait	Experiment	Chr.	SNP	Position	MAF	*P* value	R^2	Candidate gene/annotation
Chlorophyll Content	GH_2012	Pv09	ss715647747	26,619,346	0.10	6.1E−06	0.07	–
	GH_2014	Pv09	ss715648916	20,055,067	0.09	3.1E−06	0.08	*Phvul.009G136200*: LRR-RLK
	Field_2012	Pv01	ss715639380	36,919,960	0.12	7.8E−06	0.09	–
	Field_2013	Pv01	ss715641865	14,396,817	0.06	1.4E−10	0.15	–
	Field_2013	Pv09	ss715648916	20,055,067	0.09	1.0E−06	0.07	*Phvul.009G136200*: LRR-RLK
Nodule Score	Field_2013	Pv09	ss715648787	20,055,067	0.09	1.1E−06	0.12	*Phvul.009G136200*: LRR-RLK
Shoot Biomass	GH_2012	Pv01	ss715646315	48,116,724	0.17	1.0E−05	0.08	–
	GH_2012	Pv03	ss715645580	50,004,386	0.05	1.1E−05	0.10	–
	GH_2012	Pv07	ss715646458	4,252,888	0.09	8.2E−06	0.11	–
	GH_2012	Pv09	ss715647197	34,101,880	0.11	6.4E−06	0.11	*Phvul.009G231000*: Calmodulin
	GH_2014	Pv09	ss715648916	20,055,067	0.09	1.3E−08	0.13	*Phvul.009G136200*: LRR-RLK
	Field_2012	Pv07	ss715639237	42,895,691	0.12	2.3E−06	0.11	–
	Field_2012	Pv08	ss715647448	4,201,160	0.05	5.8E−08	0.15	–
	Field_2013	Pv04	ss715647346	45,251,507	0.14	6.4E−08	0.17	–
N % in Shoot Biomass	GH_2014	Pv03	ss715639320	47,948,032	0.12	1.8E−06	0.11	–
	GH_2014	Pv09	ss715648916	20,055,067	0.09	3.7E−09	0.15	*Phvul.009G136200*: LRR-RLK
	Field_2013	Pv09	ss715648916	20,055,067	0.09	3.2E−06	0.10	*Phvul.009G136200*: LRR-RLK
N % in Seed	Field_2013	Pv02	ss715639746	49,033,652	0.30	1.4E−06	0.12	–
	Field_2013	Pv03	ss715646392	1,178,905	0.05	1.0E−05	0.09	–
	Field_2013	Pv09	ss715648916	20,055,067	0.09	1.1E−09	0.17	*Phvul.009G136200*: LRR-RLK
%Ndfa_Shoot	Field_2013	Pv02	ss715649646	39,149,364	0.10	1.6E−06	0.09	–
	Field_2013	Pv03	ss715646392	1,178,905	0.05	2.9E−13	0.22	–
	Field_2013	Pv07	ss715646473	4,048,349	0.06	1.5E−07	0.17	–
	Field_2013	Pv09	ss715648916	20,055,067	0.09	6.4E−10	0.19	*Phvul.009G136200*: LRR-RLK
	Field_2013	Pv10	ss715650111	25,088,744	0.10	1.2E−06	0.11	–
	Field_2013	Pv11	ss715649573	42,485,910	0.20	1.6E−06	0.12	–
Ndfa	GH_2012	Pv03	ss715645580	50,004,386	0.05	3.1E−06	0.11	–
	GH_2012	Pv07	ss715646473	4,048,349	0.06	1.8E−06	0.12	*Phvul.007G050500*: LRR-RLK
	GH_2012	Pv09	ss715647197	34,101,880	0.11	8.4E−07	0.13	*Phvul.009G231000*: Calmodulin
	GH_2014	Pv02	ss715643723	25,332,620	0.16	7.2E−06	0.09	
	GH_2014	Pv03	ss715639320	47,948,032	0.12	2.1E−08	0.12	–
	GH_2014	Pv09	ss715648916	20,055,067	0.09	3.4E−13	0.20	*Phvul.009G136200*: LRR-RLK
Ndfa_Shoot	Field_2013	Pv02	ss715649646	39,149,364	0.09	1.0E−05	0.08	–
	Field_2013	Pv03	ss715646392	1,178,905	0.05	5.2E−13	0.23	–
	Field_2013	Pv07	ss715646473	4,048,349	0.06	1.4E−06	0.12	–
	Field_2013	Pv09	ss715648916	20,055,067	0.09	1.4E−09	0.14	*Phvul.009G136200*: LRR-RLK
	Field_2013	Pv10	ss715650111	25,088,744	0.10	6.2E−06	0.09	–
	Field_2013	Pv11	ss715649610	48,038,510	0.09	1.1E−05	0.08	–
Ndfa_Seed	Field_2013	Pv03	ss715646392	1,178,905	0.05	5.2E−06	0.11	–
	Field_2013	Pv09	ss715648916	20,055,067	0.09	8.9E−06	0.09	*Phvul.009G136200*: LRR-RLK

MAF minor allele frequency, *Ndfa* N derived from atmosphere, *GH_2012* evaluations in the GH in 2012, *GH_2014* evaluations in the GH in 2014, *Field_2012* evaluations in the field in 2012, *Field_2013* evaluations in the field in 2013, *SNP* SNP code, *E* exponent of the *P* value; R^2 is phenotypic variation explained by the SNP, *LRR-RLK* leucine-rich repeat receptor-like kinase

at this SNP was C (MAF = 0.09; Table 3), and the major allele was T. The minor allele C had a major effect on Ndfa_Shoot. At this SNP, homozygous CC genotypes fixed about 163 mg per plant of N compared to 112 mg N per plant for TT genotypes. Genotypes possessing alleles with major effects did not come from a single geographic region

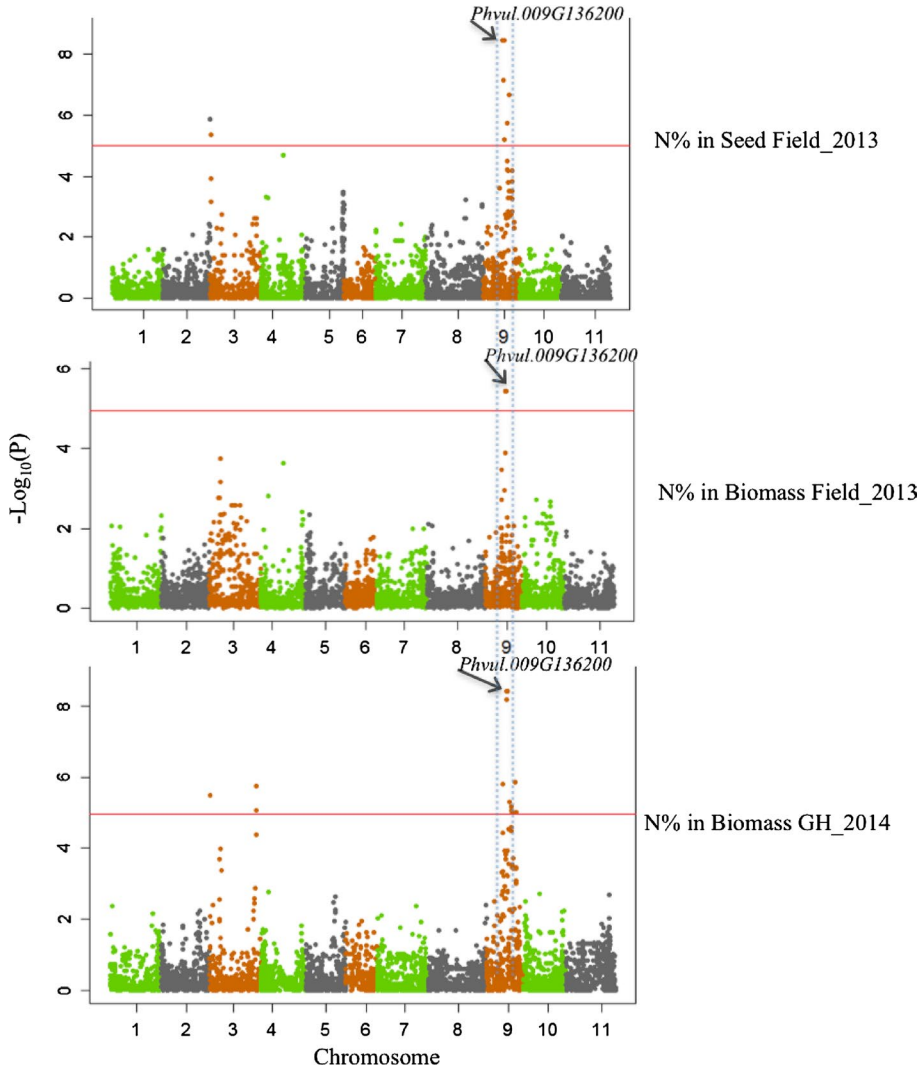

Fig. 4 Manhattan plots of association tests using MLM for N % in shoot biomass (GH_2014 and Field_2013) and N % in seed (Field_2013). A candidate gene for most significant SNP on Pv09 is also shown. The *red solid horizontal line* is the Bonferroni adjusted *P* value (1.1×10^{-05}). The *dotted gray vertical lines* are to show significant SNPs that were consistently significant for N % in shoot biomass and seed (color figure online)

or market class. We assessed the effect on Ndfa_Shoot of having all the three major effect alleles of ss715646392 (Pv03), ss715646473 (Pv07) and ss715648916 (Pv09) occurring simultaneously in a single genotype. The OAC Inferno cultivar (ADP631) was the only genotype in the ADP that had major effect alleles at all three most significant SNP loci for Ndfa_Shoot. In addition, this genotype also carried major effect allele at ss715647197, which was the most significant SNP in GH_2012.

Discussion

The genetic enhancement of SNF in common bean requires adequate genetic variability for the trait, and an understanding of the genetic basis of this variability would also foster breeding strategies that deploy marker technology. In this study, we investigated the variability of SNF and related traits in the Andean Diversity Panel of common bean. We explored the genetic basis of this variability, using a genome-wide association approach. We observed significant differences among genotypes, and wide phenotypic ranges for Ndfa_Shoot and Ndfa_Seed measured in GH and field experiments. The high averages for %Ndfa_Shoot (37.3 %) and %Ndfa_Seed (45.5 %) in Field_2013 are comparable to estimates from previous studies (Graham et al. 2003; Hardarson et al. 1993; Unkovich and Pate 2000). In Field_2013, ten genotypes in the panel had both %Ndfa_Shoot and %Ndfa_Seed values higher than 50 and 70 %, respectively (Table 2), which was higher than most previous estimates (Giller 2001; Hardarson et al. 1993; Tsai et al. 1993; Unkovich and Pate 2000; van Kessel and Hartley 2000). Common bean has been considered poor in SNF when compared to other grain legumes such as soybean. Reports indicate that soybean can be grown without supplemental N fertilizer and still produce competitive seed yields as most genotypes fix over 70 %Ndfa (Giller 2001). Results of the current study show that there are common

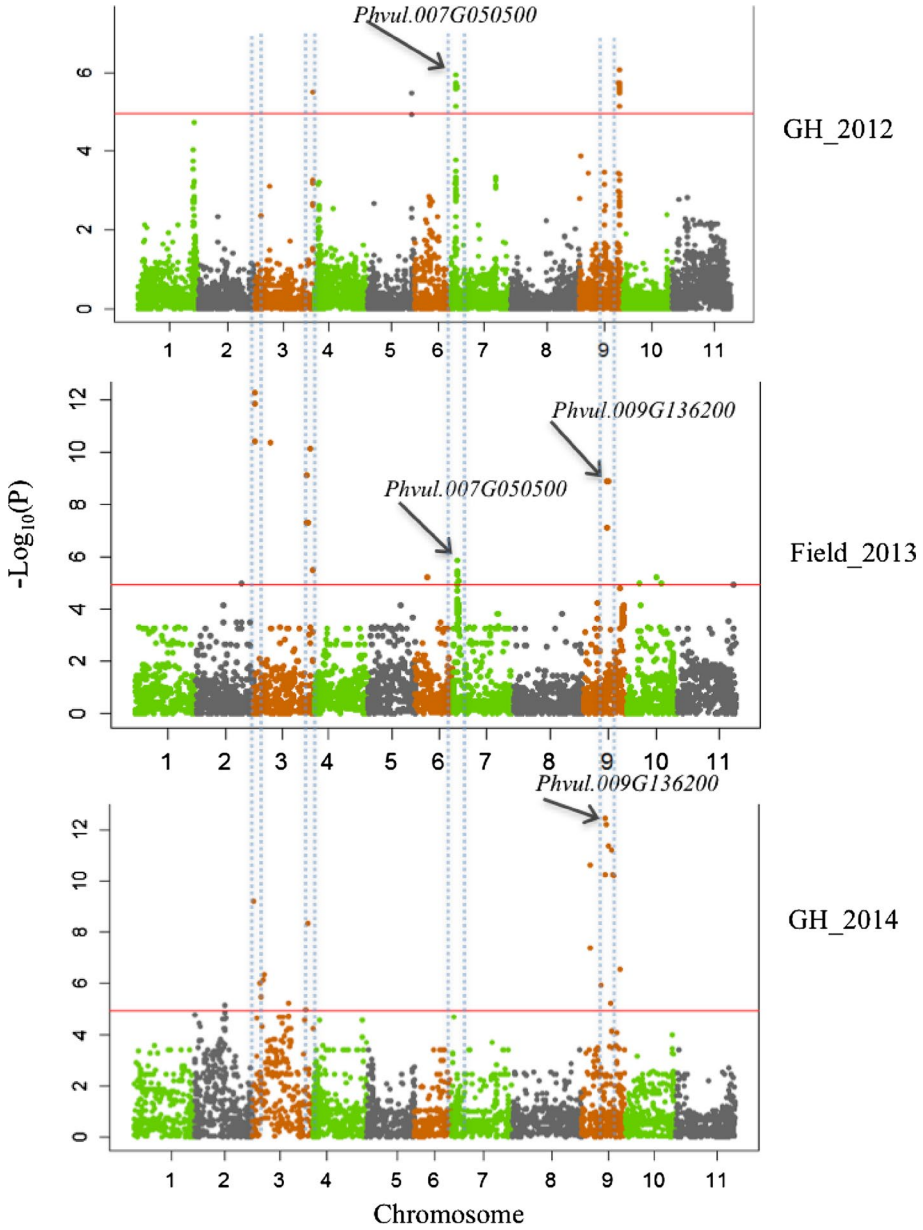

Fig. 5 Manhattan plots of association tests using MLM and candidate genes for amount of N derived from atmosphere (Ndfa) using the ADP grown in greenhouse (GH) and field. The *red solid horizontal line* is the Bonferroni adjusted *P* value (1.1×10^{-05}). The *dotted gray vertical lines* are to show significant SNPs that were consistently identified in GH_2012, GH_2014 and Field_2013 (color figure online)

bean genotypes within the Andean gene pool with competitive %Ndfa values that rival those of soybean, and could be grown without supplemental N fertilizer. In addition, this study has provided evidence of adequate genetic variability within the Andean gene pool to support genetic improvement of Andean beans for enhanced N fixation. The ten genotypes identified in Table 2 were not only superior in %Ndfa_Shoot and %Ndfa_Seed, but also in partitioning and remobilizing fixed N to the economic yield (i.e., seed). These ten genotypes were from different geographic regions of Africa, North America, and Europe and could potentially be used as germplasm in breeding for enhanced SNF. In addition different market classes were represented in this class of genotypes with enhanced SNF. This is advantageous from a plant breeding perspective as breeding programs from Africa or the Americas can choose genotypes adapted to local growing environments. Breeding for enhanced SNF within local market classes and maturity classes will increase the prospects of recovering progenies with desirable agronomic traits, provided selection is practiced in low-N soils. Although Ndfa_Seed does not capture all the variability for SNF, focusing selection on high Ndfa_Seed would be easier to integrate into most breeding programs as seed is generally harvested and does not necessitate additional measurement of traits at flowering.

A significant correlation ($r = 0.7$, $P < 0.05$) between flowering and Ndfa_Shoot was observed in Field_2013. In general, genotypes that flowered later had higher Ndfa_Shoot

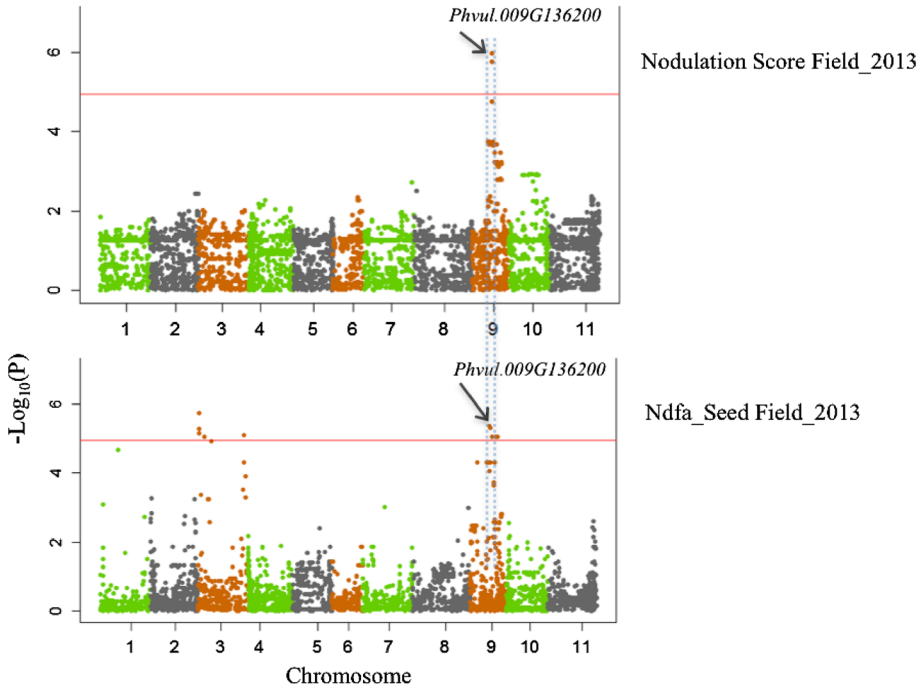

Fig. 6 Manhattan plots of association tests using MLM, and candidate gene for nodulation and amount of N in seed derived from atmosphere (Ndfa_Seed) identified using the ADP grown in the field in 2013. The *red solid horizontal line* is the Bonferroni adjusted *P* value (1.1×10^{-05}). The *dotted gray vertical lines* are to show SNPs that were consistently significant for nodulation and Ndfa_Seed in Field_2013 (color figure online)

values than those that flowered earlier. This result was expected because an early maturing genotype does not have sufficient time to accumulate sufficient above ground biomass that could serve as an adequate source and sink for photo-assimilates and fixed N, respectively. In addition, the period of active N fixation before the on-set of nodule senescence in early maturing genotypes is shorter resulting in lower amounts of N fixed. Delayed flowering has long been known to lead to a significant amount of N fixed in legumes (Graham 1981). One study in soybean suggested that a delay in flowering of 9 days would double seasonal N fixation (Hardy and Havelka 1976). This association complicates breeding for enhanced SNF in environments where growing seasons are short, and emphasis must be placed on earliness.

Chlorophyll content and shoot biomass were higher in Field_2012 than Field_2013, whereas nodule score, %Ndfa_Shoot and Ndfa_Shoot were higher in Field_2013 than Field_2012 (Table 1). This anomaly could be attributed to differences in soil N at the time of planting in these 2 years. In Field_2012, mineral N (nitrate) in the soil was 36 mg kg^{-1} compared to 2.4 mg kg^{-1} in Field_2013. Higher soil N suppresses nodulation while lower soil N enhances nodulation and SNF. Differences in the significance of correlations between shoot biomass and Ndfa_Shoot in Field_2012 and Field_2013 could also be due to differences in soil N in these 2 years. Under high soil N, most of the N required for shoot biomass production would be coming from soil N while in low soil N SNF would be the major source of N. The correlation between shoot biomass and Ndfa_Shoot is weakened when evaluations for SNF are conducted on high N soils, which has implications from a breeding perspective. Because of the expensive nature of SNF tests, routine selection for SNF is rarely conducted with common bean. Identification of traits indirectly related to SNF that can be measured cost effectively would be valuable. Shoot biomass fits this requirement, and has been used in previous studies to indirectly select for SNF, and as a proxy trait to identify QTL for SNF in soybean (Santos et al. 2013). However, indirect selections for SNF using shoot biomass would only be effective when conducted under low soil N. This also applies to using chlorophyll content since field measurements with a SPAD meter are fast and inexpensive, making it desirable as a phenotyping tool in breeding for enhanced SNF. Likewise, its effectiveness for use as an indirect trait for selecting for SNF would only be effective if plants are evaluated in a low-N site.

The total amount of N fixed by the plant is a product of biomass and N %. To maximize on the amount of N fixed by the plant both factors should be high. Genotypes that derive most of their N from the atmosphere (high %Ndfa) but have lower amount of shoot biomass would result in lower total N fixed. In this study, we observed genotypes that had higher %Ndfa_Shoot but only had modest amounts of Ndfa_Shoot because they produced less biomass. Genotypes with both high %Ndfa_Shoot and shoot biomass had the highest total N fixed. This relationship is consistent with prior knowledge that in general bush types fix lower amount of N than the climbing beans despite the %Ndfa being higher in some bush types than in climbing beans (Graham and Rosas 1977; Graham 1981). From a breeding perspective, however, a preferred genotype would be one that not only fixes adequate N, but also partitions and remobilizes the fixed N to the seed.

In Field_2013, the average %Ndfa_Shoot was 37 %, which was lower than 45 % for Ndfa_Seed (Table 1). The ^{15}N natural abundance method used in this study measures %Ndfa and Ndfa in a time-integrated manner. Therefore, this difference was expected, and it represents the amount of N that was fixed from flowering up to the time when the nodules senesced. The magnitude of this difference would depend on how long the nodules can continue actively fixing N after the on-set of plant reproductive phase given the competing needs for photo-assimilates by the nodules and seed filling. There are suggestions in literature that selecting genotypes whose nodules senescence late can be an avenue for maximizing the amount of Ndfa between flowering and physiological maturity (Giller 2001).

Marker–trait associations

In the current study, significant SNPs for nodulation were identified on Pv09 in Field_2013. The most significant SNP on Pv09 was also consistently associated with Ndfa_Shoot and related traits in both GH and field experiments. Previous studies in common bean that used bi-parental mapping populations have reported QTL for nodulation. Tsai et al. (1998) reported QTL for nodule number on Pv02, Pv04, Pv05 and Pv09 using BAT93 × Jalo EEP558 population of RILs evaluated under high N. Nodule number and nodule dry weight have previously been used as proxies for SNF in studies aimed at identifying QTL for SNF in common bean and soybean (Santos et al. 2013). We did not identify QTL for nodule dry weight in the two GH experiments. However, in the Field_2013, where we used a nodule score as a quick and less labor-intensive method than nodule dry weight, we identified significant SNPs for nodulation. In addition, significant SNPs for nodulation (nodule score) co-localized with significant SNPs for Ndfa_Shoot and Ndfa_Seed in Field_2013. These results demonstrated that nodule dry weight may not be a useful proxy trait to identify QTL for SNF in GH studies, but in field experiments a less labor-intensive nodule score is an effective proxy trait.

In the current study, several SNPs on Pv03, Pv07, and Pv09 were consistently significant for Ndfa_Shoot in both GH and field experiments. We explored the effects of alleles at the most significant loci for Ndfa_Shoot, i.e., ss715646392, ss715646473 and ss715648916 on Pv03, Pv07 and Pv09, respectively. We were particularly interested in genotypes that carried major effect alleles at all these three loci. We identified OAC Inferno (ADP631) as the only genotype in the panel that carried beneficial alleles at all three SNP loci. Interestingly, OAC Inferno had the highest Ndfa_Shoot (88.5 %) and second highest Ndfa_Seed (98.2 %) in Field_2013. This result though involving a single genotype may suggest the additive effects of these major alleles at significant SNP loci for Ndfa_Shoot. Combining these major alleles in the same background during breeding could provide phenotypes with enhanced SNF. However, combining these alleles through conventional selection would be challenging. The most effective and efficient way would be through marker-assisted selection using markers that could tag these alleles.

We also explored differences or similarities of association tests results for Ndfa measured at flowering using entire shoot biomass (Ndfa_Shoot) and Ndfa measured using seed (Ndfa_Seed). There were more significant SNPs, on more chromosomes that were associated with Ndfa_Shoot than Ndfa_Seed. In addition, when R^2 values of consistently significant SNPs for Ndfa_Shoot and Ndfa_Seed in Field_2013 were compared, R^2 for Ndfa_Shoot were larger than Ndfa_Seed. This trend is best illustrated by ss715646392, the most significant SNP for both Ndfa_Shoot and Ndfa_Seed in Field_2013 on Pv09, where R^2 was reduced from 23 % for Ndfa_Shoot to 11 % for Ndfa_Seed. In the case of the most significant SNP ss715648916 for Ndfa_Shoot and Ndfa_Seed on Pv09, the R^2 value decreased from 14 % for Ndfa_Shoot to 9 % for Ndfa_Seed. The reduction in the number of significant markers and the R^2 values when Ndfa_Shoot is compared to Ndfa_Seed may be attributed to the confounding effect of genotypic differences in remobilization and partitioning efficiency of fixed N to the seed (Ndfa_Seed). The Ndfa_Shoot biomass would not be confounded by genotypes differences in remobilization or partitioning since the entire above ground shoot biomass was used to estimate Ndfa_Shoot. Therefore, a correlation between Ndfa_Shoot and genotype would be expected to be stronger than correlation between Ndfa_Seed and genotype. This association could possibly have resulted in the identification of more significant SNPs with larger effects on Ndfa_Shoot than Ndfa_Seed.

Co-localization of significant SNPs for Ndfa_Shoot and Ndfa_Seed was observed on Pv03 and Pv09 for Field_2013. In addition, a significant correlation between Ndfa_Shoot and Ndfa_Seed was detected. Ndfa in seed only accounts for fixed N in seed, which is underlain by several physiological processes controlling partitioning and remobilization of N to the seed, and does not account for N in the rest of plant biomass. Therefore, we were intrigued by the co-localization of significant SNPs for Ndfa in shoot biomass at flowering, which includes the entire above ground biomass. This co-localization of significant SNPs for Ndfa_Shoot and Ndfa_Seed that were derived from different tissues and growth stages provided further support for important roles of genomic regions on Pv03 and Pv09 in controlling Ndfa. This co-localization suggests that at the time of physiological maturity the seed is the major sink of fixed N and most of the N in other plant parts, i.e., leaves, stems and pod walls is remobilized to the seed.

In this study, we identified a significant genetic correlation between days to flowering and Ndfa_Shoot ($r = 0.7$, $P < 0.05$). We explored whether the QTL for Ndfa_Shoot and Ndfa_Seed identified in this study co-localized with the flowering QTL on Pv01, previously identified by Kamfwa et al. (2015). None of the QTL for Ndfa identified in the current study co-localized with the flowering QTL on Pv01. Similarly, none of QTL for Ndfa identified in the current study co-localized with genomic region on Pv01 were the *fin* (*PvTFL1y*) gene that controls determinacy is located (Kwak et al. 2008; Repinski et al. 2012) and recently validated in the ADP (Cichy et al. 2015). In addition, correlation between Ndfa and determinacy was weak ($r = 0.2$, $P < 0.05$). These results suggest that the genetic basis of N fixation in the ADP was not influenced by flowering or determinacy despite the strong correlations between Ndfa and flowering. The QTL for Ndfa_Shoot and Ndfa_Seed co-localized with QTL for seed yield on Pv03 identified by Kamfwa et al. (2015). Since most of the N in the seed produced under low N is derived from SNF, genotypes superior in SNF are likely to produce higher seed yield than genotypes with low SNF potential on a low-N soil. This result provides further support for recommendations by Bliss (1993) that selection based on high seed yield produced under low N is effective for genetic enhancement of SNF in common bean.

Previous studies on the genetic architecture of SNF in common bean are scarce. Ramaekers et al. (2013) is the only published study in common bean that identified QTL for Ndfa. In that study QTL were identified for Ndfa on Pv04 and Pv10 using an intergene pool RIL population of G2333 x G19839 evaluated in the GH. When this population was evaluated in the field, QTL for Ndfa were identified on Pv01 and Pv10. Differences in marker platforms make it difficult to determine whether significant SNPs on Pv01 and Pv10 in the current study co-localize with the QTL identified by Ramaekers et al. (2013). We identified more QTL for Ndfa_Shoot since more alleles for Ndfa likely exist in the diverse ADP than the number of alleles segregating in the bi-parental population used by Ramaekers et al. (2013). The level of N available in field studies clearly effects the detection of QTL that control SNF based on lack of results from Field_2012 when N levels were high. The relatively high soil N levels (90 mg kg^{-1} N) available in the field where Ramaekers et al. (2013) evaluated the RIL population could have had a confounding effect on the expression of genes for Ndfa, resulting in fewer QTLs identified. In this study we identified several significant SNPs for Ndfa_Shoot in GH_2012, GH_2014 and Field_2013 on seven chromosomes, and the variation explained by individual significant SNPs ranged 8–23 %. Given the limitation of the size of the association panel, our study was underpowered to identify QTL with smaller effects. Therefore, we could have missed QTL with smaller effects. Larger association panels with greater marker density would help identify these QTL with smaller effects. Knowing whether the QTL identified in the Andean germplasm in the current study are the same in the Middle-American gene pool would be useful for breeders.

Candidate genes associated with significant SNPs

One of the advantages of GWAS over QTL mapping that uses bi-parental mapping populations is the ability to identify positional candidate genes, which results from enhanced mapping resolution. In this study, we identified three candidate genes for BNF and related traits. The first candidate gene was *Phvul.009G136200* on Pv09 that codes for leucine-rich repeat receptor-like protein kinase (LRR-RLK). This gene was 12.7 kb downstream of ss715648916, which was consistently significant for Ndfa_Shoot (GH_2014 and Field_2013) and Ndfa_Seed (Field_2013) (Figs. 4, 5, 6). In addition, ss715648916 was significant for nodulation (Field_2013), chlorophyll content (Field_2013), shoot biomass (GH_2014), N percentage in shoot biomass (Field_2013 and GH_2014), and N % in the seed (Field_2013). LRR-RLK's have been reported to play a critical role in signal transduction required for nodule formation (Sanchez-Lopez et al. 2012; Stracke et al. 2002). The Rhizobium releases the lipochitooligosaccharides (Nod factors) that are perceived by the LRR domain of the LRR-RLK. This results in the formation of signaling complex and subsequent downstream responses that include the formation of infection thread and nodules (Stracke et al. 2002). A second candidate gene *Phvul.007G050500* on Pv07 also encodes an LRR-RLK. The SNP (ss715646473) associated with this gene was consistently significant for Ndfa_Shoot in GH_2012 and Field_2013 (Table 3; Fig. 5). This SNP was located in the exon of *Phvul.007G050500* and is part of the LRR domain for signal perception. Three genes in the immediate upstream and two genes in the downstream region of *Phvul.007G050500* were identified also as LRR-RLK. Sanchez-Lopez et al. (2011) demonstrated the role of LRR-RLK's in nodule development in common bean. For example, the knockdown expression through RNAi of an LRR-RLK gene called *PvSymRK* in common bean resulted in the formation of scarce and defective nodules (Sanchez-Lopez et al. 2011). It is plausible that the three LRR-RLK candidate genes we have identified in the current study are among many other genes with a role in nodule development and nitrogen fixation in common bean as SNF is an integrated process occurring over a longer time period. The identification of four genes encoding LRR-RLK as candidate genes for Ndfa demonstrates that early events in the infection process may play a key role in determining the amount of N fixed by the plant.

The other candidate gene identified on Pv09 was *Phvul.009G231000* (Table 3) that was associated with ss715647197, the most significant SNP for Ndfa in GH_2012. *Phvul.009G231000* was 1.1 kb upstream of ss715647197. Because there was no functional annotation for this gene on Phytozome, a BLAST search revealed the highest hits in *A. thaliana* (TAIR) and *Medicago truncatula* (NCBI) were for genes that code for calmodulin, which are calcium-transporting proteins (Mitra et al. 2004). Following the perception of Nod factors by the legumes, there is a spike in the levels of free calcium in the cytoplasm of cells for roots hairs (Riely et al. 2004). Calcium spiking is reported to be an essential component of the signaling cascade required in nodule development (Levy et al. 2004). The nodulation signaling pathway has been reported to contain calcium-activated kinases. Alfalfa mutants defective in calcium-spike response do not nodulate (Ehrhardt et al. 1996). The influx of calcium to the root hair is reported to cause depolarization and subsequent curling of the root hair that precedes the formation of an infection thread and nodules. The flux of calcium in the root hair cells is mediated by calcium-binding proteins called calmodulin (Riely et al. 2004; Stacey et al. 2006). It is plausible that the candidate gene *Phvul.009G231000* that has high sequence similarity to calmodulin genes in *A. thaliana* played a significant role in calcium spikes and subsequent root hair morphological changes required for nodule formation. Further functional genomics studies are required to confirm the roles of the identified candidate genes in SNF.

Conclusions

In this study, we explored the genetic architecture of SNF and related traits in common bean. The enhanced mapping resolution from GWAS resulted in the identification of several significant SNPs and candidate genes for SNF and related traits. Once the identified QTL in this study are validated in different populations and genetic backgrounds, they could potentially be used in marker-assisted breeding to accelerate the genetic improvement of SNF in common bean.

Author contribution statement KK, KAC, JDK conceived and designed the experiments. KK performed the experiments, and analyzed the data. KK, KAC, JDK wrote the paper. All authors have read and approved the final version of the manuscript.

Acknowledgments Research was supported by the Feed the Future Innovation Lab for Collaborative Research on Grain Legumes by the Bureau for Economic Growth, Agriculture, and Trade, US Agency for International Development, under the terms of Cooperative Agreement No. EDH-A-00-07-00005-00; and the US Department of Agriculture, Agricultural Research Service. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the US Agency for International Development or the US Government. We also thank Dr. Zixang Wen for his helpful comments on some aspects of data analyses.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Ethical standards All experiments described in this manuscript comply with the current US laws in which they were performed.

References

Akibode CS, Maredia M (2012) Global and regional trends in production, trade and consumption of food legume crops. Staff Paper 2012–10. Department of Agricultural, Food and Resource Economics, Michigan State University

Beebe S (2012) Common bean breeding in the tropics. Plant Breed Rev 36:357–426

Bliss FA (1993) Breeding common bean for improved biological nitrogen fixation. Plant Soil 152:71–79

Bliss F, Pereira P, Araujo R (1989) Registration of five high nitrogen fixing common bean germplasm lines. Crop Sci 29:240–241

Boddey RM, Alves BJR, de Henrique B, Soares L, Jantalia CP, Urquiaga S (2009) Biological nitrogen fixation and the mitigation of greenhouse gas emissions. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 387–413

Bradbury P, Zhang D, Kroon T, Casstevens Y, Ramdoss Y, Buckler E (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Broughton WJ, Dilworth MJ (1970) Plant nutrient solutions. In: Somasegaran P, Hoben HJ (eds) Methods in legume-Rhizobium technology handbook for rhizobia niftal project. Univ of Hawaii, Hawaii, pp 245–249

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. Plant Soil 252:55–128

Buttery BR, Park SJ, Pv Berkum (1997) Effects of common bean (*Phaseolus vulgaris* L.) cultivar and Rhizobium strain on plant growth, seed yield and nitrogen content. Can J Plant Sci 77:347–351

Cichy KA, Porch TG, Beaver JS, Cregan P, Fourie D, Glahn RP, Grusak MA, Kamfwa K, Katuuramu DN, McClean P, Mndolwa E, Nchimbi-Msolla S, Pastor-Corrales MA, Miklas PN (2015) A *Phaseolus vulgaris* diversity panel for Andean bean improvement. Crop Sci. doi:10.2135/cropsci2014.09.0653

Ehrhardt DW, Wais R, Long SR (1996) Calcium spiking in plant root hairs responding to Rhizobium nodulation signals. Cell 85:673–681

Elizondo Barron J, Pasini RJ, Davis DW, Stuthman DD, Graham PH (1999) Response to selection for seed yield and nitrogen (N_2) fixation in common bean (*Phaseolus vulgaris* L.). Field Crops Res 62:119–128

Gage DJ (2009) Nodule development in legumes. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 1–24

Giller KE (2001) Nitrogen fixation in tropical cropping systems, 2nd edn. CABI, New York

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:1178–1186

Graham PH (1981) Some problems of nodulation and symbiotic nitrogen fixation in *Phaseolus vulgaris* L.: a review. Field Crops Res 4:93–112

Graham PH (2009) Soil biology with an emphasis on symbiotic nitrogen fixation. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 171–209

Graham P, Rosas J (1977) Growth and development of indeterminate bush and climbing cultivars of *Phaseolus vulgaris* L. inoculated with Rhizobium. J Agric Sci 88:503–508

Graham P, Rosas J, Estevez de Jensen C, Peralta E, Tlusty B, Acosta-Gallegos J, Arraes Pereira P (2003) Addressing edaphic constraints to bean production: the bean/cowpea CRSP project in perspective. Field Crops Res 82:179–192

Hardarson G, Bliss FA, Cigales-Rivero M, Henson RA, Kipe-Nolt JA, Longeri L, Manrique A, Pena-Cabriales J, Pereira PAA, Sanabria C, Tsai SM (1993) Genotypic variation in biological nitrogen fixation by common bean. Plant Soil 521:59–70

Hardy RWF, Havelka UD (1976) Photosynthate as a major factor limiting nitrogen fixation by field grown legumes with emphasis on soybeans. In: Nutman PS (ed) Symbiotic nitrogen fixation in plants. Cambridge University Press, London, pp 421–439

Herridge DF, Redden RJ (1999) Evaluation of genotypes of navy and culinary bean (*Phaseolus vulgaris* L.) selected for superior growth and nitrogen fixation. Aust J Exp Agric 39:975–980

Holland JB (2006) Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation with SAS Proc MIXED. Crop Sci 46:642–654

Hyten DL, Song Q, Fickus EW, Quigley CV, Lim JS, Choi IY, Hwang EY, Pastor-Corrales MA, Cregan PB (2010) High throughput SNP discovery and assay development in common bean. BMC Genom 11:475

Jensen E, Hauggaard-Nielsen H (2003) How can increased use of biological N_2 fixation in agriculture benefit the environment? Plant Soil 252:177–186

Kamfwa K, Cichy KA, Kelly JD (2015) Genome-wide association study of agronomic traits in common bean. Plant Genome. doi:10.3835/plantgenome2014.09.0059

Kwak M, Velasco D, Gepts P (2008) Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). J Hered 99:283–291

Levy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet EP, Ane JM, Lauber E, Bisseling T, Denarie J, Rosenberg C, Debelle F (2004) A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science 303:1361–1364

Mafongoya PL, Mpepereki S, Mudyazhezha S (2009) The importance of biological nitrogen fixation in cropping systems in nonindustrialized nations. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 329–348

Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GE, Long SR (2004) A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. Proc Natl Acad Sci USA 101:4701–4705

Nodari RO, Tsai SM, Guzmán P, Gilbertson RL, Gepts P (1993) Toward an integrated linkage map of common bean. III. Mapping genetic factors controlling host-bacteria interactions. Genetics 134:341–350

Peoples MB, Hauggaard-Nielsen H, Jensen ES (2009a) The potential environmental benefits and risks derived from legumes in rotations. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 349–385

Peoples MB, Unkovich MJ, Herridge DF (2009b) Measuring symbiotic nitrogen fixation by legumes. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 125–170

Pereira PAA, Miranda BD, Attewell JR, Kmiecik KA, Bliss FA (1993) Selection for increased nodule number in common bean (*Phaseolus vulgaris* L.). Plant Soil 148:203–209

Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Ramaekers L, Galeano CH, Garzon N, Vanderleyden J, Blair MW (2013) Identifying quantitative trait loci for symbiotic nitrogen fixation capacity and related traits in common bean. Mol Breed 31:163–180

Repinski SL, Kwak M, Gepts P (2012) The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. Theor Appl Genet 124:1539–1547

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res 31:224–228

Riely BK, Ané J-M, Penmetsa RV, Cook DR (2004) Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. Curr Opin Plant Biol 7:408–413

Sanchez-Lopez R, Jauregui D, Nava N, Alvarado-Affantranger X, Montiel J, Santana O, Sanchez F, Quinto C (2011) Down-regulation of SymRK correlates with a deficiency in vascular bundle development in *Phaseolus vulgaris* nodules. Plant, Cell Environ 34:2109–2121

Sanchez-Lopez R, Jauregui D, Quinto C (2012) SymRK and the nodule vascular system: an underground connection. Plant Signal Behav 7:691–693

Santos MA, Geraldi IO, Garcia AAF, Bortolatto N, Schiavon A, Hungria M (2013) Mapping of QTLs associated with biological nitrogen fixation traits in soybean. Hereditas 150:17–25

SAS Institute (2011) SAS version 9.3. SAS Institute Inc, Cary

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MMS, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet 46:707–713

Shearer G, Kohl D (1986) N_2 fixation in field settings: estimations based on natural ^{15}N abundance. Funct Plant Biol 13:699–756

Singh SP, Gutiérrez JA (1984) Geographical distribution of the DL1 and DL2 genes causing hybrid dwarfism in *Phaseolus vulgaris* L., their association with seed size, and their significance to breeding. Euphytica 33:337–345

Souza AA, Boscariol RL, Moon DH, Camargo LE, Tsai SM (2000) Effects of Phaseolus vulgaris QTL in controlling host-bacteria interactions under two levels of nitrogen fertilization. Genetics Mol Biol 23:155–161

Stacey G, Libault M, Brechenmacher L, Wan J, May GD (2006) Genetics and functional genomics of legume nodulation. Curr Opin Plant Biol 9:110–121

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417:959–962

Strodtman KN, Emerich DW (2009) Nodule metabolism. In: Emerich DW, Krishnan HB (eds) Nitrogen fixation in crop production. Crop Science Society of America, Madison, pp 95–124

Tsai SM, Da Silva PM, Cabezas WL, Bonetti R (1993) Variability in nitrogen fixation of common bean (*Phaseolus vulgaris* L.) intercropped with maize. Plant Soil 152:93–101

Tsai S, Nodari R, Moon D, Camargo L, Vencovsky R, Gepts P (1998) QTL mapping for nodule number and common bacterial blight in *Phaseolus vulgaris* L. Plant Soil 204:135–145

Uddling J, Gelang-Alfredsson J, Piikki K, Pleijel H (2007) Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. Photosynth Res 91:37–46

Unkovich MJ, Pate JS (2000) An appraisal of recent field measurements of symbiotic N_2 fixation by annual legumes. Field Crops Res 65:211–228

van Kessel C, Hartley C (2000) Agricultural management of grain legumes: has it led to an increase in nitrogen fixation? Field Crops Res 65:165–181

Vance CP (2001) Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. Plant Physiol 127:390–397

Vandemark GJ, Brick MA, Osorno JM, Kelly JD, Urrea CA (2014) Edible grain legumes. In: Smith S, Diers B, Specht J, Carver B (eds) Yield gains in major US field crops. CSSA Spec. Publ. 33. Crop Science Society of America, Madison, pp 87–124

Vincent JM (1970) A manual for practical study of root nodule bacteria. IBP Handbook No 15. Blackwell, Oxford

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214

Zhang Z, Ersoz E, Lai C, Todhunter R, Tiwari H, Gore M, Bradbury P, Yu J, Arnett D, Ordovas J (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360