Theoretical and Applied Genetics (2020) 133:665–676
https://doi.org/10.1007/s00122-019-03499-7

ORIGINAL ARTICLE

Genome-wide association analyses reveal the genetic basis of biomass accumulation under symbiotic nitrogen fixation in African soybean

Davoud Torkamaneh[1,2,3] · François-P. Chalifour[1] · Chantal J. Beauchamp[1] · Hesham Agrama[4,5] · Steve Boahen[4] · Halim Maaroufi[2] · Istvan Rajcan[3] · François Belzile[1,2]

Received: 29 January 2019 / Accepted: 30 November 2019 / Published online: 10 December 2019

Abstract

Key message **We explored the genetic basis of SNF-related traits through GWAS and identified 40 candidate genes. This study provides fundamental insights into SNF-related traits and will accelerate efforts for SNF breeding.**

Abstract Symbiotic nitrogen fixation (SNF) increases sustainability by supplying biological nitrogen for crops to enhance yields without damaging the ecosystem. A better understanding of this complex biological process is critical for addressing the triple challenges of food security, environmental degradation, and climate change. Soybean plants, the most important legume worldwide, can form a mutualistic interaction with specialized soil bacteria, bradyrhizobia, to fix atmospheric nitrogen. Here we report a comprehensive genome-wide association study of 11 SNF-related traits using 79K GBS-derived SNPs in 297 African soybeans. We identified 25 QTL regions encompassing 40 putative candidate genes for SNF-related traits including 20 genes with no prior known role in SNF. A line with a large deletion (164 kb), encompassing a QTL region containing a strong candidate gene (CASTOR), exhibited a marked decrease in SNF. This study performed on African soybean lines provides fundamental insights into SNF-related traits and yielded a rich catalog of candidate genes for SNF-related traits that might accelerate future efforts aimed at sustainable agriculture.

Communicated by Albrecht E. Melchinger.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-019-03499-7) contains supplementary material, which is available to authorized users.

✉ François Belzile
francois.belzile@fsaa.ulaval.ca

Davoud Torkamaneh
dtorkama@uoguelph.ca

François-P. Chalifour
Francois-P.Chalifour@fsaa.ulaval.ca

Chantal J. Beauchamp
Chantal.Beauchamp@fsaa.ulaval.ca

Hesham Agrama
Hesham.agrama@squ.edu.om

Steve Boahen
S.Boahen@cgiar.org

Halim Maaroufi
Halim.Maaroufi@ibis.ulaval.ca

Istvan Rajcan
irajcan@uoguelph.ca

[1] Département de Phytologie, Université Laval, Quebec City, QC, Canada

[2] Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Quebec City, QC, Canada

[3] Department of Plant Agriculture, University of Guelph, Guelph, ON, Canada

[4] International Institute for Tropical Agriculture (IITA), Ibadan, Nigeria

[5] Sultan Qaboos University, Muscat, Oman

Introduction

Nitrogen is an essential element for growth and reproduction in both plants and animals (Burns et al. 1997). It is a fundamental component of amino acids, nucleic acids, and many other organic and inorganic compounds (Nelson and Cox 2008). Despite the fact that nitrogen comprises about 80% of the earth's atmosphere (Weeks 1932), this atmospheric nitrogen is not the form needed by plants and animals. Thus, in crop production, synthetic N-fertilizers are provided as nitrogen sources (Sutton and Bleeker 2013). It is estimated that more than 50% of the world's food supply is directly dependent on the application of nitrogen fertilizer in agriculture (van Hameren et al. 2013). As nitrogen fertilizers are often expensive and are not accessible everywhere (e.g., Africa), this contributes to food insecurity by severely impeding crop productivity (Masson-Boivin et al. 2009). On the other hand, excess chemical nitrogen fertilization or runoff can lead to highly undesirable consequences. Agriculture is the main source (~65%) of nitrous oxide (N_2O), produced by the breakdown of nitrogen fertilizers and a potent contributor to global warming and climate change (Crutzen et al. 2007). It is 292 times more potent as a greenhouse gas than CO_2 and contributes to the formation of ground-level ozone and acid rain (Jensen et al. 2012). The excess of N-fertilizers applied in agriculture is one of the main sources of harmful algal blooms (Moore et al. 2011). Finally, drinking water contaminated with nitrogen can cause methemoglobinemia, or "blue-baby syndrome," which is potentially fatal to infants (Knobeloch et al. 2000).

To overcome this problem, legumes offer an attractive alternative: biological nitrogen fixation. Legumes are able to form a symbiotic relationship with specialized nitrogen-fixing soil bacteria (rhizobia) to convert atmospheric nitrogen (N_2) into forms usable by plants (Burns et al. 1997; Masson-Boivin et al. 2009; van Hameren et al. 2013). Symbiotic nitrogen fixation (SNF) is a complex biological system as it includes several different genes and pathways. This system starts with cross talk between bacteria and their host. The roots of legumes release flavonoids into the soil, and these attract rhizobia that will in turn produce signaling molecules (Nod factors) that alter root hair growth in such a way to wrap itself around the bacteria, eventually leading to bacterial colonization and formation of nodules. It is in these nodules that these bacteria will transform atmospheric nitrogen into ammonia, a form that can be assimilated by the legume host (Oldroyd 2013; Zipfel and Oldroyd 2017).

The precise phenotyping of SNF is extremely demanding in terms of labor, cost, and time. This, in turn, has limited the number and scope of SNF studies. Most of the advances in the area of genetic and molecular mechanisms underlying SNF have come from the study of *Medicago truncatula*, *Lotus japonicus*, and *Glycine max*. The molecular components of the pathways and genes involved in SNF were first identified via the study of symbiosis-defective mutants (Penmetsa and Cook 1997). Recent advances in sequencing technologies led to define SNF-related genes and pathways through transcriptomic analysis (Karmakar et al. 2019), host–microbe interaction (Steidinger and Bever 2016), proteomics (Larrainzar and Wienkoop 2017), and methylomics (Andrews and Andrews 2017). Even though many molecular components controlling SNF have now been discovered, very little information is available regarding the genetic and allelic diversity within the cultivated germplasm for a species such as soybean, the most widely grown oil crop in the world. Because of their large-scale cultivation, soybeans can biologically fix more than 17 Mt/Yr [15% of worldwide nitrogen demand (FAOSTAT)] and makes a net positive contribution to soil nitrogen content.

In this study, we used 297 African cultivated soybean lines to perform a genome-wide association study. This research material is important for two reasons: (1) this population consists of cultivars and advanced breeding lines developed for African growing conditions, and (2) African farmers cultivate soybean without coating the seed with rhizobial inoculum nor nitrogen fertilization; thus, yields will depend on efficient uptake of nitrogen via symbiosis. We performed a comprehensive GWAS analysis to identify the genomic regions controlling SNF-related traits in soybean. This study provides key insights into existing genetic variation that underlies SNF-related traits in cultivated soybeans.

Materials and methods

Plant materials and phenotyping

A set of 297 soybean genotypes representative of the soybean-breeding program at the International Institute of Tropical Agriculture (IITA, Ibadan, Nigeria) were obtained. The soybean lines were grown under the following conditions: 26 °C/24 °C (day/night) and a relative humidity of ~60% in a greenhouse at Université Laval in a two-factor experiment (2 × 2). Five seeds of each line were planted in 3.7-L pots (Classic 400 from Nursery Supplies Inc., Los Angeles, California, USA) with substrate that was composed of medium-sized vermiculite (Perlite Canada Inc., Montreal, QC, Canada) and Turface MVP grade (Profile Products LLC, Buffalo Grove, IL, USA). Pots were arranged according to a split-plot design where soybean lines were randomized within two blocks. The main factor was the N treatments [N-free nutrient solution for inoculated seeds with *Bradyrhizobium japonicum* (strain 532C) (Hume and Shelp 1990;

Soulemanov et al. 2002) and N-inclusive nutrient solution for non-inoculated seeds (Supplementary Table S1)], and the sub-factor was the soybean lines. This design was used to prevent cross-contamination between inoculated and fertilized treatments. Two plants were kept for each sample per pot. Forty-five days after sowing, roots were cleaned and separated from the aerial. Nodules were collected and counted. Roots, aerial parts, and the nodules were all dried at 65 °C for 72 h before measuring their dry weight. In total, 11 traits were measured: shoot dry weight in inoculated and fertilized samples (SDW_i and SDW_f), root dry weight in inoculated and fertilized samples (RDW_i and RDW_f), number of nodules (NN), nodule dry weight (NDW), NDW/SDW ratio, normalized NDW (nNDW) (obtained by dividing the NDW (gr per plant) by the SDW (gr per plant) × 100 (value expressed as a percentage), NDW/RDW ratio, SDW_i/SDW_f (SDW ratio), and RDW_i/RDW_f (RDW ratio) (Ryle et al. 1981; Sheng and Hunt 1991; Gan et al. 2004; Olivera et al. 2004; Vasileva and Ilieva 2007). The statistical analysis of traits data was conducted in SAS (SAS Institute 2013) using PROC GLM model as follows:

$$Y_{ijk} = \mu + b_i + \alpha_i + b\alpha_{ij} + \beta_k + \alpha\beta_{jk} + e_{ijk},$$

where μ is the overall mean, b_i, $b\alpha_{ij}$, and e_{ijk}, are random effects for blocks (replications), replication × cultivar interactions and errors with variances σ_b^2, $\sigma_{b\alpha}^2$, and σ_e^2, respectively, and α_j, β_k, and $\alpha\beta_{jk}$ are fixed effects for cultivars, N-treatment, and cultivar × N-treatment interactions, respectively (Supplementary Table S2).

Genotyping

DNA extraction was performed using the Qiagen DNeasy 96 Plant kit according to the manufacturer's protocol. SNP genotyping was performed using a GBS approach (with *Ape*K1 digestion) and libraries prepared following Elshire's protocol (Elshire et al. 2011; Sonah et al. 2013) at the Plateforme d'analyses génomiques [Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval (Quebec, QC, Canada)]. Single-end, 100-bp sequencing of multiplex GBS libraries (one 96-plex GBS library per sequencing lane) was performed on an Illumina HiSeq 2000 at the Genome Quebec Innovation Center, McGill University (Montreal, QC, Canada).

GBS data analysis

Illumina sequence reads (581 million 100-bp reads) were processed using the Fast-GBS pipeline (Torkamaneh et al. 2017). In brief, FASTQ files were demultiplexed based on barcodes sequences. Demultiplexed reads were trimmed and then mapped against the soybean reference genome (Williams 82 (Gmax_275_Wm82.a2.v1)) (Schmutz et al. 2010). About 92.1% of the reads were successfully mapped to the soybean reference genome. Then nucleotide variants were identified from mapped reads. Variants were removed if (1) they were multi-allele, (2) the overall read quality (QUAL) score was < 32, (3) the mapping quality (MQ) score was < 30, (4) read depth of was < 2, (5) missing data > 80%, and (6) heterozygosity > 50%. Missing data imputation was performed with BEAGLE v4.1 (Browning and Browning 2007) as described by Torkamaneh and Belzile (2015). Functional annotation of SNPs was performed with the help of SnpEff (Cingolani et al. 2012) (version 5) using the soybean reference genome (Williams82) (Schmutz et al. 2010) annotation (Gmax_275_Wm82.a2.v1.gene.gff3) downloaded from Phytozome (V12) (Goodstein et al. 2012).

Population genetic analysis

Population structure was estimated using a variational Bayesian inference implemented in fastStructure (Raj et al. 2014). Five runs were performed for each number of populations (K) set from 1 to 12. The most likely K value was determined by the log probability of the data (LnP(D)) and delta K (Evanno et al. 2005), based on the rate of change in LnP(D) between successive K values. Principal component analysis was performed using the P3D R package (Zhang et al. 2010). Linkage disequilibrium (LD) between SNPs on each chromosome was estimated with the r^2 option using PLINK (Purcell et al. 2007).

Genome-wide association analysis

GWAS analyses were performed using the rMVP package in R (https://github.com/XiaoleiLiuBio/rMVP) using mixed linear models (MLMs) (Yu et al. 2006) and fixed and random model circulating probability unification (FarmCPU) (Liu et al. 2016). Two different approaches for population structure were used: PCA (covariate P) and fastStructure (covariate Q) to capture population structure. Two kinship matrices provided estimates of the relatedness among individuals (covariates K and K^*) (Kang et al. 2008; Li et al. 2012). A total of seven different models, from a naïve model devoid of any correction for confounding to models incorporating a correction for both population structure (covariates P or Q) and/or kinship (covariates K or K^*), were tested. Models that took into account kinship and PCA ($P + K^*$) were found to provide the best fit based on the cumulative distribution of p values for SDW_i, SDW_f, NDW/RDW ratio, SDW ratio, and RDW ratio. The FDR-adjusted p value (q value) below 0.05 was used to establish a significance threshold (Wang et al. 2012a, b). Finally, the SNP-heritability estimated by GCTA-GREML method using all SNPs (Yang et al. 2010). The percentage of variation explained

(PVE) by the marker–trait association (R^2) was calculated as the difference between the R^2 of the model with and without the strongest associated SNP.

Synteny and candidate gene identification

First, we determined if QTLs identified in this work were located in macrosyntenic regions detected between the *Glycine max* and *Medicago truncatula* genomes as reported in Lee et al. (2017) and Young et al. (2011). For syntenic regions containing a soybean QTL, we determined whether symbiosis-related QTLs had been reported in these same regions in *M. truncatula* based on the GWAS analysis reported by Stanton-Geddes et al. (2013). We then designed a systematic analytical process to identify candidate genes for SNF-related traits. (1) We identified the genes residing within QTL regions, the extent of which was defined by markers in high LD ($r^2 > 0.7$) with the peak SNP. (2) Then we searched different plant databases (SoyBase, SoyKB, LIS, PlantDB, SoyGD, and agriGO) (Shultz et al. 2006; Exner et al. 2008; Du et al. 2010; Grant et al. 2010; Joshi et al. 2014) to keep genes labeled as being involved in SNF. (3) We kept the genes that were expressed in at least one tissue or organ related to SNF (e.g., root, root hairs, nodule, etc.), using the soybean transcriptome atlas dataset (Severin et al. 2010). (4) We identified genes that were differentially expressed in response to inoculation with *Bradyrhizobium japonicum* using the RNA-seq data from Libault et al. (2010). (5) Finally, we explored the resulting catalog of genes for known orthologues.

Fast-neutron mutant assay

We searched a FN-mutant database (Bolon et al. 2011; Grant et al. 2010) to identify lines carrying a deletion spanning QTL regions detected in this study. In this fashion, we identified five FN-mutant lines, of which, two had large deletions (~4 Mb) and two other lines were not fixed for the deletions spanning QTLs of interest. One mutant line (FN0186954) carried a small homozygous deletion (164 kb). This line and its wild-type counterpart (M92-220) were obtained from Dr. R. Stupar (University of Minnesota, USA). To characterize the resulting phenotype, mutant and wild-type lines were grown in the same greenhouse under the same conditions and treatments as described above for the GWAS analysis.

Data availability

The genotypic, phenotypic and GWAS data are publicly available at Figshare (https://figshare.com/projects/Genome-Wide_Association_Analyses_Reveal_the_Genetic_Basis_of_Symbiotic_Nitrogen_Fixation_in_African_Soybean/65885).

Results

Characterization of GWAS population

To comprehensively assess SNF, we selected a collection of 297 African soybean lines derived from an ongoing breeding program at the IITA. These lines are important because they were developed to efficiently uptake the nitrogen via symbiosis. The plants were grown in highly controlled greenhouse conditions under two treatments (inoculation with rhizobia and fertilization with nitrogen) (Supplementary Table S1) and 11 SNF-related traits were measured (see details in "Materials and methods" section). These traits showed a normal distribution without any significant skewness or kurtosis (Shapiro–Wilk test, p value = 0.32) (Fig. 1a, Supplementary Figure 1, and Supplementary Table S2). Generally, we observed very low coefficients of variation (CV), ranging from 1.2 to 2.3%, between different replications for all traits studied with the exception of NN (CV = 40.1%). Furthermore, significant correlations were observed between shoot dry weight (SDW) and NN (0.53*), SDW and nodule dry weight (NDW) (0.69*), root dry weight (RDW) and NN (0.53*), RDW and NDW (0.69*), NDW and NN (0.77**), and NDW/SDW ratio and NDW/RDW ratio (0.59*) (Supplementary Table S3). We observed a high variability among lines in biomass production under inoculation with rhizobia, but not under N fertilization, thus indicating variability in SNF efficiency among the collection. To best describe the SNF capacity, we measured the ratio between inoculated and fertilized samples for shoot dry weight and root dry weight (SDW and RDW ratio) for each accession.

To obtain genome-wide nucleotide variants, we performed genotyping-by-sequencing (GBS) of the 297 lines, and obtained 581 million 100-bp reads (mean of ~2M reads/sample). Sequencing reads were processed using the FastGBS analytical pipeline for variant calling. We obtained an initial catalog of 78,961 nucleotide variants (Supplementary Figure S2). After missing data imputation and removing nucleotide polymorphisms with a minor allele frequency (MAF) < 0.05, we generated a final dataset of 54,225 variants including 3792 indels for GWAS analysis. Of these variants, 12.5% were located in exons (Supplementary Figure S3); 46.4% were classified as silent, 52.3% as missense, and 1.3% as nonsense mutations. Principal component analysis (PCA), phylogenetic tree clustering and population structure analysis suggested that this population was composed of five subpopulations ($K = 5$), with most lines showing some admixture, but these did not form very tight and distinct subsets (Fig. 1b, c and Supplementary Figure S4). Genome-wide linkage disequilibrium (LD) analysis showed LD decay varied across different chromosomes from 260 kb to 1.6 Mb at $r^2 < 0.2$ and 189 kb to 743 kb at $r^2 < 0.3$. The mean LD

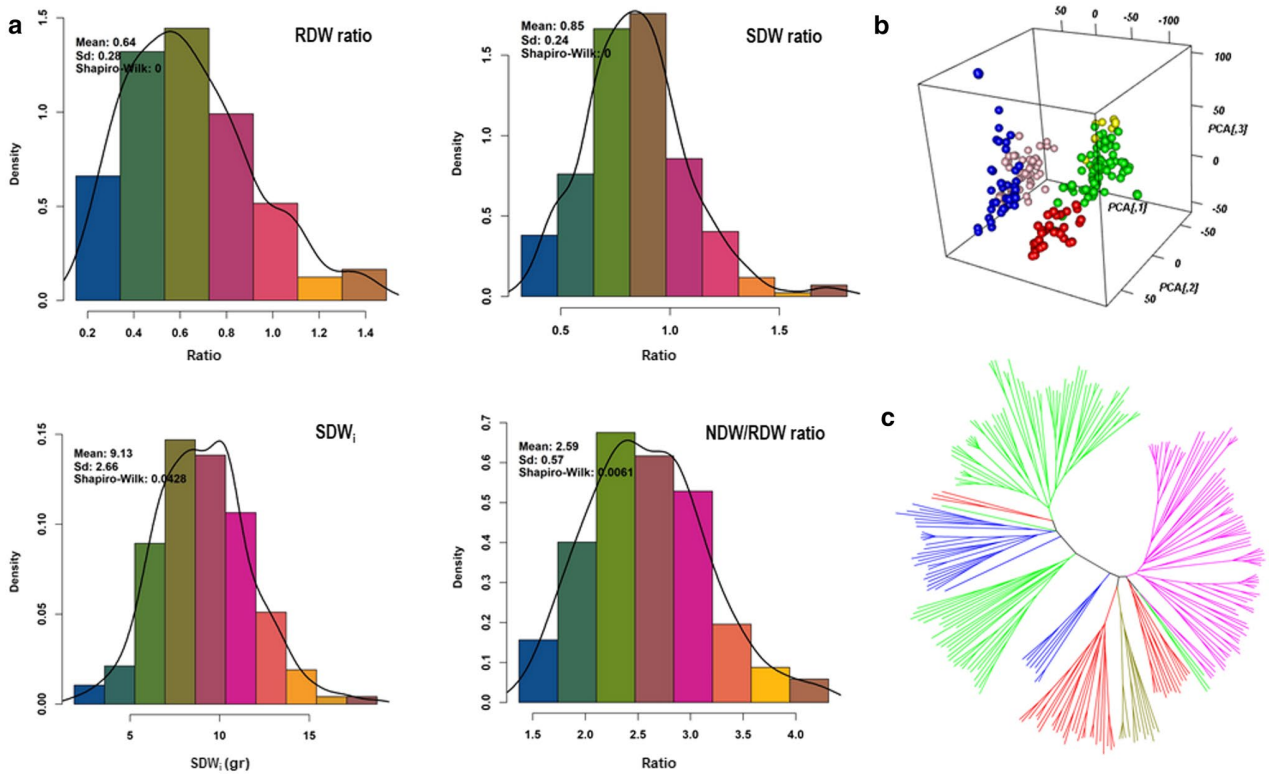

Fig. 1 Phenotypic diversity and genetic structure among this panel of 297 African soybean lines. **a** Histograms of phenotypic values for traits which were used for GWAS (SDW$_i$, NDW/RDW ratio, SDW ratio, and RDW ratio). **b** 3D PCA plot of the first three principal components. **c** Phylogenetic tree constructed using the neighbor-joining method. Colors represent different clusters based on fastSTRUCTURE analysis (Supplementary Figure S4) (color figure online)

decay for the entire genome was estimated as 550 kb at $r^2 < 0.3$, which is comparable to previous studies (Supplementary Figure S5).

GWAS on SNF-related traits in African soybean

GWAS analysis was performed using MLM and FarmCPU methods, and population structure (P) and cryptic relatedness (K*) were incorporated as covariates to reduce false-positive signals. Using this approach, 25 QTL regions ($-\log_{10} P \geq 5.2$ and q value ≤ 0.05) (Fig. 2, Table 1 and Supplementary Figure S6) were detected for four traits: RDW ratio, SDW ratio, NDW/RDW ratio, and SDW$_i$. No significant association was observed for NN (the most commonly measured phenotype in previous studies), presumably due to the high level of intra-line

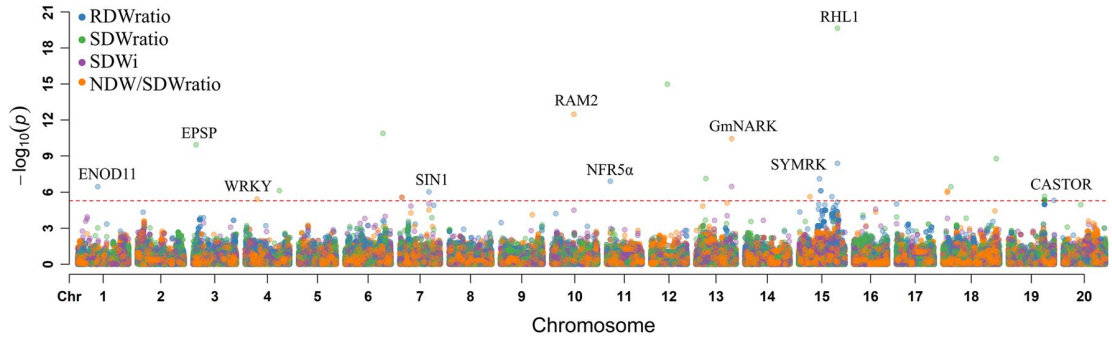

Fig. 2 GWAS for four SNF-related traits; RDW ratio, SDW ratio, SDW and NDW/RDW ratio, respectively, in the African soybean lines. The genome-wide significance threshold drawn at a FDR-adjusted p value (or q value) below 0.05

Table 1 Characteristics of QTLs showing as significant association with selected phenotypic traits related to SNF among a collection of African soybean accessions

Trait	QTL ID	Chr	MSS[a] position	Synteny[b]	Allele frequency		MSS *p* value	MSS *q* value	R^{2d}	SNP-heritability (%)
					Efficient	Inefficient				
RDW ratio	*qtl_01*	Chr01	22 027 329	Yes[c]	0.94	0.06	3.1E−07	2E−03	0.09	61.3
	qtl_02	Chr03	3 012 188	Yes[c]	0.11	0.89	1.3E−10	7E−04	0.05	
	qtl_03	Chr04	37 677 065	No	0.92	0.08	7.3E−07	3E−03	0.04	
	qtl_04	Chr06	41 634 280	No	0.09	0.91	1.3E−11	2E−05	0.06	
	qtl_05	Chr07	32 115 284	Yes	0.08	0.92	9.1E−07	3E−03	0.06	
	qtl_06	Chr07	37 270 973	Yes	0.12	0.88	1.3E−05	2E−02	0.08	
	qtl_07	Chr11	4 577 083	Yes[c]	0.61	0.39	6.1E−08	9E−04	0.09	
	qtl_08	Chr12	18 290 967	Yes	0.94	0.06	1.1E−15	1E−05	0.06	
	qtl_09	Chr13	11 049 293	No	0.91	0.09	7.2E−08	2E−03	0.04	
	qtl_10	Chr15	22 932 181	Yes[c]	0.07	0.93	8.6E−09	9E−04	0.08	
	qtl_11	Chr15	42 576 692	Yes	0.09	0.91	2.3E−20	1E−06	0.07	
	qtl_12	Chr17	499 223	Yes	0.08	0.92	8.2E−06	2E−02	0.05	
	qtl_13	Chr18	8 821 143	No	0.93	0.07	3.5E−07	1E−03	0.04	
	qtl_14	Chr18	58 648 473	Yes[c]	0.89	0.11	1.7E−09	1E−04	0.04	
	qtl_15	Chr19	50 562 657	Yes	0.92	0.08	4.8E−06	1E−02	0.08	
SDW ratio	*qtl_16*	Chr04	13 209 559	No	0.93	0.07	3.8E−06	1E−03	0.10	45.6
	qtl_17	Chr07	12 271 656	No	0.07	0.93	9.3E−06	1E−02	0.12	
	qtl_18	Chr07	32 115 284	Yes	0.08	0.92	1.9E−05	1E−02	0.09	
	qtl_19	Chr10	24 366 467	No	0.09	0.91	3.4E−13	2E−06	0.11	
	qtl_20	Chr13	39 223 028	No	0.94	0.06	3.6E−11	2E−05	0.16	
	qtl_21	Chr15	12 431 650	Yes[c]	0.68	0.32	2.3E−06	1E−03	0.07	
	qtl_22	Chr16	24 157 230	Yes[c]	0.93	0.07	1.8E−05	1E−02	0.05	
	qtl_23	Chr18	4 999 836	Yes	0.89	0.11	8.0E−07	3E−03	0.06	
SDW_i & SDW_f	*qtl_24*	Chr19	39 939 106	Yes[c]	0.63	0.37	6.5E−06	3E−02	11.1	26.6
NDW/RDW ratio	*qtl_25*	Chr07	2 419 122	No	0.88	0.12	2.6E−06	5E−02	14.6	29.9

[a]MSS, most significant SNP

[b]Located in macrosyntenic regions with *M. truncatula*

[c]Overlapped with symbiosis-related QTLs identified in *M. truncatula*

[d]R^2 was calculated by R^2 of model with SNP minus R^2 of model without SNP

variation for this trait (Supplementary Figure S7). Furthermore, GWAS analysis for SDW_i showed a strong association peak ($-\log10\ P \geq 6.2$ and q value ≤ 0.03) at the end of chromosome 19, while the same association was observed for SDW_f (Supplementary Figure S8). This suggests that this genomic region is important for plant biomass production in the presence of nitrogen, regardless of its sources (inoculation or fertilization). Of 25 QTLs detected in this study, 16 were located in macrosyntenic genomic regions identified between *G. max* and *M. truncatula*. Interestingly, in half of these cases (8 QTLs), potentially orthologous QTLs have been reported in *M. truncatula* (Table 1).

We found that ratios (RDW and SDW ratio), comparing the phenotype of a single line under contrasting conditions (inoculation vs. fertilization treatment), provided powerful tools to describe SNF capacity. As shown in Table 1, the vast majority of marker-trait associations were detected for the two traits showing the highest degree of heritability (SDW and RDW ratios; 45.6 and 61.3%, respectively), making these traits highly informative for the study of SNF. The other traits were eliminated due to their low heritability (< 10%). The GWAS analysis was performed using common genetic variants underlying SNF-related traits, but these variants explain only a portion of the heritability.

Identification of candidate genes for SNF-related traits

One of the aims of GWAS studies is to identify the causal genes or alleles that explain the portion of phenotypic variability that is genetically determined. To identify candidate genes, we followed a systematic approach. We identified 207 genes residing within the 25 QTL regions

Table 2 List of candidate genes for SNF-related traits identified in this study

Cat	Gene ID[a]	QTL ID	Gene name	Gene product	Function	Known legume orthologs[b]
Reported in soybean	Glyma.13g292300	*qtl_09*	GmNARK	LRR receptor kinase	AON[c]	LjHAR1, MtSUNN, PsSYM29
	Glyma.16g111200	*qtl_22*	CCaMK	Ca2 + /CaM-dependent kinase	CSP[d]	LjCCaMK, MtDMI3, PsSYM9
	Glyma.11g063100	*qtl_07*	NFR5	LysM receptor kinase	NF[e] receptor	LjNFR5, MtNFP, PsSYM10
	Glyma.07g204600	*qtl_06*	GmCLE	CLAVATA3/ESR-related	AON	LjCLE-RS1
Reported in other legumes	Glyma.07g030100	*qtl_25*	CRE	Cytokinin receptor	Nodule organogenesis	MtCRE1, LjLHK1
	Glyma.07g203100	*qtl_06*	ND	RHO GTPASE	Infection	LjROP
	Glyma.07g203300	*qtl_07*	SIN 1	GRAS transcription regulator	Infection, CCD[f]	LjSIN1, PsSIN1
	Glyma.15g203900	*qtl_10*	ND	Putative transcription factor	Infection, CCD	LjRHL1
	Glyma.15g197300	*qtl_10*	SYMRK	LRR receptor kinase	CSP	PsSYM19, LjSYMRK, MtDMI2
	Glyma.01g091600	*qtl_01*	Nodulin	Glutamate-ammonia ligases	Infection	ENOD11
	Glyma.07g175400	*qtl_05&18*	Annexin	Calcium-activated GTPase	IT[g] growth	MtAnn1
	Glyma.07g201700	*qtl_06*	ND	EIN	Regulation of IT growth	MtSICKLE
	Glyma.07g202900	*qtl_06*	ND	GTPase Rab11	Infection	MtRAB1F
	Glyma.17g005100	*qtl_12*	Ssh1p	Polyphosphoinositide protein	Nodule organogenesis	LjPLPs
	Glyma.17g005300	*qtl_12*	ND	Acetyl-CoA acetyl-transferase	Infection	LjLbAAT
	Glyma.04g115500	*qtl_16*	WRKY	Transcription regulator	Infection	MtWRAKY, LjWRAKY
	Glyma.10g107100	*qtl_19*	RAM2	Glycerol-3-phosphate shuttle	IT growth	MtRAM2, LjRAM2
	Glyma.15g150300	*qtl_21*	MatE	Transmembrane transport	Nodule organogenesis	MtMatE
	Glyma.18g089600	*qtl_14*	Syntaxin	Intracellular transporter	IT growth	MtSYP132
	Glyma.19g263500	*qtl_15*	CASTOR	Ca^{2+}-gated ion channel	CSP	LjCASTOR
Novel SNF-related candidates	Glyma.07g031100	*qtl_25*	ND	Phospholipase D	Infection	
	Glyma.07g030400	*qtl_25*	ND	Glycosyl hydrolases	CSP	
	Glyma.19g140800	*qtl_24*	ND	Ca^{2+}/CaM-dependent kinase	CSP	
	Glyma.13g308000	*qtl_20*	ND	Unknown protein	Nitrate transport	
	Glyma.07g117000	*qtl_17*	ND	ATP synthase	CSP	
	Glyma.07g202300	*qtl_05&18*	IFS	Isoflavone synthase	Infection	
	Glyma.07g202100	*qtl_05&18*	ND	Transcription factor	Nodule organogenesis	
	Glyma.15g231000	*qtl_11*	ND	Callose synthase	CCD	
	Glyma.15g230900	*qtl_11*	ND	LRR receptor kinase	AON	
	Glyma.15g231100	*qtl_11*	ND	LRR receptor kinase	AON	
	Glyma.15g231300	*qtl_11*	ND	LRR receptor kinase	AON	

Table 2 (continued)

Cat	Gene ID[a]	QTL ID	Gene name	Gene product	Function	Known legume orthologs[b]
	Glyma.15g197600	*qtl_10*	ND	LEM3 / CDC50	CCD	
	Glyma.15g194600	*qtl_10*	ND	Sugar transporter	IT growth	
	Glyma.15g210200	*qtl_10*	ND	LRR-RLK	Infection, NR[h]	
	Glyma.03g027400	*qtl_02*	EPSP	Catalytic activity	Infection	
	Glyma.04g157500	*qtl_03*	ND	Polygalacturonase	IT growth	
	Glyma.06g246600	*qtl_04*	ND	CTL transporter	Infection	
	Glyma.12g142400	*qtl_08*	ND	LLR	AON	
	Glyma.18g056800	*qtl_23*	ND	Cation transport	Infection	
	Glyma.07g202300	*qtl_06*	IFS	Isoflavone synthase	Infection	

[a]Putative and Candidate genes identified in QTL regions

[b]Lj, *Lotus japonicus*; Mt, *Medicago truncatula*; Ps, *Pisum sativum*

[c]Autoregulation of nodulation

[d]Common symbiosis pathway

[e]Nod factor

[f]Cortical cell division

[g]Infection thread

[h]Negative resistance

identified above. We then identified those that were annotated as being involved in nitrogen pathways and that were expressed in at least one tissue or organ related to SNF (e.g., root, root hair, nodule). We finally identified genes that were differentially expressed in response to inoculation with *Bradyrhizobium japonicum*. Using this comprehensive approach, we identified 40 candidate genes for SNF-related traits (Table 2). This approach allowed us to capture four well-characterized SNF genes in soybean [*GmNARK* (Kinkema and Gresshoff 2008), *CCaMK* (Harper and Harmon 2005), *NFR5α* (Indrasumunar et al. 2010) and *GmCLE* (Lim et al. 2011)] out of a total of seven described so far (Indrasumunar et al. 2010; Okazakia et al. 2013). For the three remaining known SNF genes, we observed no nucleotide variation (genetic variation) in the vicinity of these genes in this population. Furthermore, among our set of 40 candidate genes, we identified 16 homologs of well-studied SNF genes that have been characterized in other legumes. Finally, 20 other genes for which we found no prior link with SNF can be viewed as novel candidate genes.

Validation of a candidate SNF-related gene

To functionally assess whether these candidate regions do indeed partially govern SNF in soybean, we looked for deletions overlapping these QTLs. We investigated a collection of fast-neutron lines and found five mutant lines with a deletion overlapping with our QTL regions. Of these five lines, two lines carried very large deletions (> 2 Mb) and two lines were not fixed for these deletions. The last mutant line, FN0186954, carried a relatively small (164 kb) homozygous deletion (50.42–50.59 Mb) that spanned the RDW ratio QTL on the end of chromosome 19. This region contains the Glyma.19g263500 gene, an ortholog of the *Lotus japonicus* CASTOR gene which codes for an ion channel identified as essential for symbiosis in *L. japonicus* (Chen et al. 2009). Line FN0186954 and its wild-type counterpart (M92-220) were phenotyped in the same fashion as the lines of the association panel. When inoculated, FN0186954 exhibited a significantly lower SDW (6.3 vs. 9.6 grams; $p = 0.01$), RDW (1.8 vs. 3.6 g; $p = 0.02$), and NN (72 vs. 211; $p = 0.001$) compared to the wild-type control. When grown with chemical N fertilizer, FN0186954 showed a phenotype that did not differ from the wild type (Fig. 3), indicating that the mutant line was not lacking some major genes for growth, but rather had suffered the loss of a gene needed specifically for the plant to obtain nitrogen through SNF. This thus provides the first evidence suggesting that the soybean CASTOR gene is involved in SNF, as is the case in *L. japonicus*.

Discussion

We report here that, using a systematic approach, we uncovered 25 QTLs contributing to SNF-related traits within a collection of African soybean cultivars and breeding lines via a GWAS approach. This represents a great stride forward in at least two respects: (1) the large number and strong candidates for the involvement of many of these in SNF-related traits and (2) the loci and alleles uncovered in this work are relevant to varietal development as they were discovered in elite material. In prior work, very few QTL mapping analyses, mostly using the progeny of biparental crosses, have been performed in soybean for SNF or related traits (Santos et al. 2013; Dhanapal et al. 2016). These analyses not only were generally conducted on only inoculated samples, but were also phenotyped for a limited number of traits: typically, nodule number (NN), nodule dry weight (NDW), NDW/NN ratio and shoot dry weight (SDW) (Ramaekers et al. 2013; Santos et al. 2013; Dhanapal et al. 2016). Here, not only were QTLs defined for SNF-related traits, but we also carried out a systematic analysis to define candidate genes within each QTL region based on their annotation and expression, resulting in a set of 40 candidate genes.

One measure of the quality of such a catalog of candidate genes is certainly its ability to capture genes already known to contribute to SNF. In this work, of the seven genes previously shown through functional analysis to contribute to SNF in soybean, four were found among our catalog of candidate genes [*GmNARK* (Kinkema et al. 2008); *CCaMK* (Harper and Harmon 2005); *NFR5α* (Indrasumunar et al. 2010); and *GmCLE* (Lim et al. 2011)]. For the three other known genes [*NFR1* (Indrasumunar et al. 2010), *NIN* and *ENOD40* (Okazakia et al. 2013)], there may be absence of allelic variation within this set of elite lines, the frequency of alternate alleles may have been too low, or these may have escaped detection due to incomplete coverage of the genome with SNP markers. Further 16 candidate genes can also be thought of as excellent candidates based on the fact that their orthologues in other legumes have been demonstrated to play important roles in SNF. In the case of the ortholog of the *L. japonicus* CASTOR gene, a line homozygous for a 164-kb deletion spanning the soybean CASTOR gene exhibited highly significant reductions in both SDW and RDW under inoculation but not N fertilization, suggesting strongly that this gene also contributes to SNF in soybean. Taken together, this represents the first experimental evidence for the contribution of these 16 soybean orthologues in SNF and constitutes a major contribution to our knowledge of genes underlying this key trait in this crop. In summary, there is good evidence for the involvement in SNF of half of our catalog of candidate genes (20/40). The remaining 20 candidate genes, for which no prior evidence of their involvement in SNF has been reported, nonetheless, represent a novel set of genes of interest for which further study would be highly warranted.

In our view, a key to the success of this approach was the choice of phenotypes. Although nodule number, nodule weight, and plant dry weight have been used in the past (Ramaekers et al. 2013; Santos et al. 2013; Dhanapal et al. 2016), we found that these did not generate strong association signals. In the case of nodule number and weight, these traits were highly variable and did not prove very useful, similar to what has been reported by other researchers (Wang et al. 2012a, b; Lira et al. 2015). It is likely that nodule number or weight alone, without assessing the size and

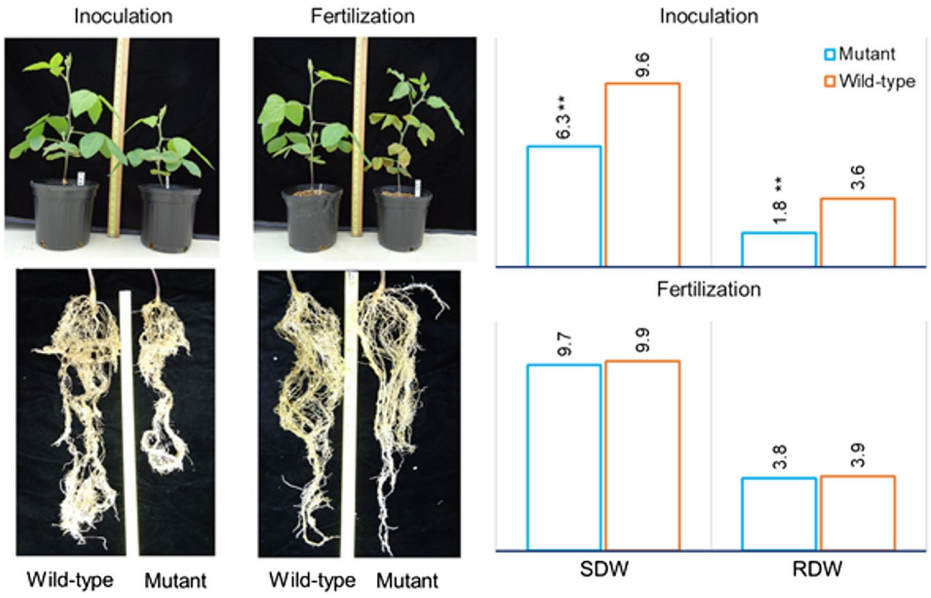

Fig. 3 Left, phenotypic diversity observed between FN0186954 (mutant line) and its wild-type counterpart for the GmCASTOR gene under two conditions (inoculated and fertilized). Right, histograms showing the phenotypic differences between the mutant and wild-type lines for SDW (gr) and RDW (gr) in response to inoculation and fertilization. Asterisks above a column indicate a significant difference (p value < 0.01) between the mutant line and its wild-type counterpart

nitrogen-fixing activity of these nodules, is a poor descriptor of SNF activity. Similarly, plant dry weight did not yield strong association signals. Alone, it may not directly relate to SNF as other traits such as plant architecture could confound the results. When the accumulation of dry weight (roots or shoots) of a line was measured under both N fertilization and rhizobial inoculation, the ratio between the two truly provided a very powerful phenotype for assessing SNF. Indeed, 24 of 25 QTL regions identified in our GWAS were identified based on such ratios. These likely better describe how well a particular line achieves its full growth potential (presumably that obtained under N fertilization) under conditions of rhizobial inoculation and therefore better describe the plant's ability to fix atmospheric nitrogen.

In the final case, a QTL on chromosome 19 was detected using both SDW data (under fertilized and inoculated conditions). The fact that phenotypic variation for SDW under N fertilization was observed suggests that the lines differed in their ability to convert the available N into biomass. The observed variation for SDW under inoculation likely captures both the ability of the plant to capture atmospheric N and to convert it to biomass. One can think of the SDW ratio as providing a measure of how well a line can capture atmospheric N, relative to the fertilizer check. Thus, because we detected no association signal for the SDW ratio on chromosome 19, we propose that the observed signal most likely reflects how efficiently a line was able to convert available N (whether obtained through fertilization or symbiosis) into biomass. Although it is of agronomic interest, this trait is not directly related to SNF.

We recognize at least two limitations to our study. In this study, only a single rhizobial strain was used for phenotyping of SNF-related traits. Although this strain [*Bradyrhizobium japonicum* (strain 532C) (Hume and Shelp 1990; Soulemanov et al. 2002)] is known to be a predominant and commercially available strain, this single strain cannot cover all possible types of host-genotype by strain-genotype interaction. Recent studies reported that different strains and mixed inoculants can show different interaction with a host (Burghardt et al. 2018). Such complex interactions will be missed from this study. Secondly, the sum of estimated PVEs for identified QTLs exceeds the estimated heritability. We assume that one possible cause of such differences is an overestimation of individual PVEs.

We conclude that the knowledge that comes out from this study will promote the addressing of the triple challenges of food security, environmental degradation, and climate change, under sustainable agricultural practices. We expect that genetic regions identified in this study will become a key tool in SNF breeding programs and also enable geneticists and molecular biologists to rapidly deploy this knowledge to dissect the genetic architecture of SNF-related traits.

Acknowledgements We thank the Global Affairs Canada, Canadian Field Crop Research Alliance, and International Institute of Tropical Agriculture (IITA). We thank Fausta Karaboneye for her help on greenhouse phenotyping. We also thank sequencing platform of Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, QC, Canada.

Author contribution statement DT and FB conceptualized the project. FC, CB, HA, and SB prepared the plant materials and performed sample selection. DT, FC, and CB conducted greenhouse phenotyping. HA and SB conducted field phenotyping. DT, IR, and HM formal analysis. DT and FB wrote the original draft. DT and FB wrote review and editing.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Andrews M, Andrews ME (2017) Specificity in legume-rhizobia symbioses. Int J Mol Sci 18(4):705. https://doi.org/10.3390/ijms18040705

Bolon Y-T et al (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol 156(1):240–253

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

Burghardt LT, Epstein B, Guhlin J, Nelson MS, Taylor MR, Young ND et al (2018) Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. Proc Natl Acad Sci USA 115(10):2425–2430

Burns IG et al (1997) How do nutrients drive growth? Plant Soil 196:321–325

Chen C et al (2009) Antiquity and function of CASTOR and POLLUX, the twin ion channel-encoding genes key to the evolution of root symbioses in plants. Plant Physiol 149(1):306–317

Cingolani P, Platts A, le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80–92

Crutzen PJ et al (2007) N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. Atmos Chem Phys Discuss 7:11191–11205

Dhanapal A et al (2016) Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. Plant Genome 8(3):1–15

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38(Web Server issue):W64–W70

Olivera M, Tejera RA, Iribarne C, Ocana A, Lluch C (2004) Growth, nitrogen fixation and ammonium assimilation in common bean (*Phaseolus vulgaris*): effect of phosphorus. Physiol Plant 121:498–505

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Exner V, Hirsch-Hoffmann M, Gruissem W, Hennig L (2008) PlantDB—a versatile database for managing plant research. Plant Methods 4:1

Gan Y, Stulen IV, Keulen H, Kuiper PJC (2004) Low concentrations of nitrate and ammonium stimulate nodulation and N_2 fixation while inhibiting specific nodulation (nodule DW g^{-1} root dry weight) and specific N_2 fixation (N_2 fixed g-1 root dry weight) in soybean. Plant Soil 258:281–292

Goodstein DM, Shu S, Howson R et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(Database issue):D1178–D1186

Grant D, Nelson RT, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucl Acids Res 38(suppl 1):D843–D846

Harper JF, Harmon A (2005) Plants, symbiosis and parasites: a calcium signalling connection. Nat Rev Mol Cell Biol 6:555–566

Hume DJ, Shelp BJ (1990) Superior performances on the Hup$^-$*Bradyrhizobium japonicum* strain 532C in Ontario soybean field trials. Can J Plant Sci 70:661–666

Indrasumunar A et al (2010) Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max* L. Merr.). Plant Cell Physiol 51(2):201–214

Jensen E et al (2012) Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. Agron Sustain Dev 32:329–364

Joshi T, Fitzpatrick MR, Chen S, Liu Y, Zhang H, Endacott RZ, Gaudiello EC, Stacey G, Nguyen HT, Xu D (2014) Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. Nucl Acids Res 42(D1):D1245–D1252

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Karmakar K, Kundu A, Rizvi AZ, Dubois E, Severac D, Czernic P, Cartieaux F, DasGupta M (2019) Transcriptomic analysis with the progress of symbiosis in 'crack-entry' legume arachis hypogaea highlights its contrast with 'infection thread' adapted legumes. Mol Plant Microbe Interact 32(3):271–285

Kinkema M, Gresshoff PM (2008) Investigation of downstream signals of the soybean autoregulation of nodulation receptor kinase GmNARK. Mol Plant Microbe Interact 10:1337–1348

Knobeloch L et al (2000) Blue babies and nitrate-contaminated well water. Environ Health Perspect 108:675–678

Larrainzar E, Wienkoop S (2017) A proteomic view on the role of legume symbiotic interactions. Frontiers Plant Sci 8:1267. https://doi.org/10.3389/fpls.2017.01267

Lee C, Yu D, Choi HK, Kim RW (2017) Reconstruction of a composite comparative map composed of ten legume genomes. Genes Genom 39(1):111–119

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N (2012) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45:43–50

Libault M et al (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiol 152:541–552

Lim CW, Lee YW, Hwang CH (2011) Soybean nodule-enhanced CLE peptides in roots act as signals in GmNARK-mediated nodulation suppression. Plant Cell Physiol 52(9):1613–1627

Lira MAJ, Nascimento LRS, Fracetto GGM (2015) Legume-rhizobia signal exchange: promiscuity and environmental effects. Front Microbiol 6:945

Liu X et al (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12(2):e1005957

Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol 17:458–466

Moore S et al (2011) Impacts of climate variability and future climate change on harmful algal blooms and human health. Proc Centers Oceans Hum Health Investig Meet 7:S4. https://doi.org/10.1186/1476-069X-7-S2-S4

Nelson DL, Cox MM (2008) Principles of biochemistry, 5th edn. W.H. Freeman and Company, New York, p 272

Okazakia S, Kanekob T, Satoc S, Saeki K (2013) Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. Proc Natl Acad Sci 110:17131–17136. https://doi.org/10.1073/pnas.1302360110

Oldroyd EDG (2013) Speak, friend, and enter: signaling systems that promote beneficial symbiotic associations in plants. Nat Rev Microbiol 11(4):252–263

Penmetsa RV, Cook DR (1997) A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. Science 275:527–530. https://doi.org/10.1126/science.275.5299.527

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al (2007) PLINK: a tool set for whole genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197(2):573–589

Ramaekers L et al (2013) Identifying quantitative trait loci for symbiotic nitrogen fixation capacity and related traits in common bean. Mol Breed 31:163–180

Ryle GJA, Arnott RA, Powell CE (1981) Distribution of dry weight between root and shoot in white clover dependent on N_2 fixation or utilizing abundant nitrate nitrogen. Plant Soil 60:29–39

SAS Institute (2013) The SAS system for Windows Release 9.4. SAS Institute, Cary

Santos MA et al (2013) Mapping of QTLs associated with biological nitrogen fixation traits in soybean. Hereditas. https://doi.org/10.1111/j.1601-5223.2013.02275.x

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278):178–183

Severin JA et al (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10:160

Sheng Q, Hunt LA (1991) Poids sec des pousses et des racines du blé, de la triticale et du seigle et teneur en eau du sol. Can J Plant Sci 71:41–49

Shultz JL, Kurunam D, Shopinski K et al (2006) The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of *Glycine max*. Nucleic Acids Res 34(Database issue):D758–D765

Sonah H, Bastien M, Iquira E, Tardivel A, Legare G et al (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE 8(1):e54603

Soulemanov A, Prithiviraj B, Carlson RW, Jeyaretnam B, Smith DL (2002) Isolation and characterization of the major nod factor of *Bradyrhizobium japonicum* strain 532C. Microbiol Res 157:25–28

Stanton-Geddes J, Paape T, Epstein B, Briskine R, Yoder J, Mudge J et al (2013) Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. PLoS ONE 8(5):e65688

Steidinger BS, Bever JD (2016) Host discrimination in modular mutualisms: a theoretical framework for meta-populations of mutualists and exploiters. Proc R Soc B 283:20152428. https://doi.org/10.1098/rspb.2015.2428

Sutton MA, Bleeker A (2013) Environmental science: the shape of nitrogen to come. Nature 494:435–437

Torkamaneh D, Belzile F (2015) Scanning and filling: ultra-dense SNP genotyping combining genotyping-by-sequencing, SNP array and whole-genome resequencing data. PLoS ONE 10(7):e0131533

Torkamaneh D et al (2017) Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. BMC Bioinform 18:5

van Hameren B, Hayashi S, Gresshoff PM, Ferguson BJ (2013) Advances in the identification of novel factors required in soybean nodulation, a process critical to sustainable agriculture and food security. J Plant Biol Soil Health 1(1):6

Vasileva V, Ilieva A (2007) Effect of pre-sowing treatment of seeds with insecticides on nodulating ability, nitrate reductase activity and plastid pigments content of Lucerne (*Medicago sativa* L.). Agron Res 5:87–92

Wang D et al (2012) Symbiosis specificity in the legume—rhizobial mutualism. Cell Microbiol 14(3):334–342

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y, Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci 196:125–131

Weeks ME (1932) The discovery of the elements. IV. Three important gases. J Chem Educ 9(2):215

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42(7):565–569. https://doi.org/10.1038/ng.608

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK et al (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480(7378):520–524

Yu JM et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang Z et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360