

## ORIGINAL ARTICLE

# Improvement in nitrogen fixation capacity could be part of the domestication process in soybean

N Muñoz<sup>1,2,3</sup>, X Qi<sup>1</sup>, M-W Li<sup>1</sup>, M Xie<sup>1</sup>, Y Gao<sup>1</sup>, M-Y Cheung<sup>1</sup>, F-L Wong<sup>1</sup> and H-M Lam<sup>1</sup>

Biological nitrogen fixation (BNF) in soybeans is a complex process involving the interplay between the plant host and the symbiotic rhizobia. As nitrogen supply has a crucial role in growth and development, higher nitrogen fixation capacity would be important to achieve bigger plants and larger seeds, which were important selection criteria during plant domestication by humans. To test this hypothesis, we monitored the nitrogen fixation-related performance in 31 cultivated and 17 wild soybeans after inoculation with the slow-growing *Bradyrhizobium diazoefficiens* sp. nov. USDA110 and the fast-growing *Sinorhizobium (Ensifer) fredii* CCBAU45436. Our results showed that, in general, cultivated soybeans gave better performance in BNF. Electron microscopic studies indicated that there was an exceptionally high accumulation of poly- $\beta$ -hydroxybutyrate bodies in bacteroids in the nodules of all wild soybeans tested, suggesting that the C/N balance in wild soybeans may not be optimized for nitrogen fixation. Furthermore, we identified new quantitative trait loci (QTLs) for total ureides and total nodule fresh weight by employing a recombinant inbred population composed of descendants from a cross between a cultivated and a wild parent. Using nucleotide diversity ( $\theta\pi$ ), divergence index ( $F_{st}$ ) and distribution of fixed single-nucleotide polymorphisms as parameters, we found that some regions in the total ureides QTL on chromosome 17 and the total nodule fresh weight QTL on chromosome 12 exhibited very low diversity among cultivated soybeans, suggesting that these were traits specially selected during the domestication and breeding process.

*Heredity* (2016) **117**, 84–93; doi:10.1038/hdy.2016.27; published online 27 April 2016

## INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is one of the major global crops that also has a unique role in sustainable agriculture, due to its ability to fix atmospheric nitrogen through symbiotic interactions with rhizobia in the soil (Keyser and Li, 1992). There are two major groups of rhizobia that inhabit the root nodules of soybean: the slow-growing *Bradyrhizobium* and the fast-growing *Sinorhizobium* (Keyser and Li, 1992). In USA, Brazil and Argentina, slow-growing rhizobia, particularly the species *Bradyrhizobium japonicum* and *Bradyrhizobium elkanii*, are used as inoculants on soybean fields (Hungria *et al.*, 1998). On the other hand, the soils in China are dominated by fast-growing rhizobia such as *Sinorhizobium (Ensifer) fredii* (Tian *et al.*, 2012).

After rhizobial cells have encountered their compatible soybean partners, a series of developmental events in the infected roots occur, which lead to the formation of nodules, a special organ where biological nitrogen fixation (BNF) takes place (Stougaard, 2000). The primary metabolic product of BNF is ammonia. In soybean, ammonia is subsequently converted into ureides for long-range transport. Ureides will then provide organic nitrogen resources for the building of nitrogen-containing biomolecules (Shelp and Ireland, 1985).

During the domestication of soybean, the size and weight of seeds have been enhanced by human selection (Zhou *et al.*, 2015).

There may have been multiple ways to achieve these results. To support the increase in dry matter in the seed, which is a major sink of metabolites, corresponding increases in the carbon and nitrogen supplies are essential. Although an increased photosynthetic capacity could have provided more carbon (Richards, 2000), we hypothesize that the capacity of BNF was also enhanced during the domestication of soybean.

Several factors contribute to the capacity of BNF, including the development of nodules and the accumulation of ureides. Previous studies identified a non-nodulating and a super-nodulating mutant (Carroll *et al.*, 1985, 1986). The soybean mutant that cannot form nodules does not carry out BNF (Mathews *et al.*, 1992). On the other hand, super-nodulating mutants do not exhibit enhanced BNF and sometimes even exhibit a suppressed growth (Takahashi *et al.*, 1995). One possible reason is that the efficiencies of carbon and nitrogen metabolisms may not be able to catch up with the excessive nodule numbers. Genetic studies have been carried out to identify the quantitative trait loci (QTLs) associated with BNF in soybean. For example, there are QTLs identified on chromosomes 6, 9, 13 and 19 that are associated with the ureides content (the QTL on chromosome 19 accounting for up to 31% of the variance) (Hwang *et al.*, 2013). One QTL for the specific nodule weight (total nodule dry weight/number of nodules) was identified on

<sup>1</sup>Centre for Soybean Research of the Partner State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR; <sup>2</sup>Centro de Investigaciones Agropecuarias-INTA, Instituto de Fisiología y Recursos Genéticos Vegetales, Córdoba, Argentina and <sup>3</sup>Cátedra de Fisiología Vegetal, Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de Córdoba, Córdoba, Argentina

Correspondence: Professor H-M Lam, Centre for Soybean Research of the Partner State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR.

E-mail: honming@cuhk.edu.hk

Received 2 October 2015; revised 14 February 2016; accepted 16 March 2016; published online 27 April 2016

chromosome 20 (Santos *et al.*, 2013) and one for total nodule weight was identified on chromosome 18 (Hwang *et al.*, 2014).

Most of the previous genetic studies on soybean BNF have employed only cultivated soybeans as the sources of genetic materials (Hwang *et al.*, 2013; Santos *et al.*, 2013). Therefore, the results may not reflect the process of domestication. An investigation using wild germplasms which exhibit a much higher genetic diversity may fill this gap. The idea of using wild germplasms to study BNF has been underexplored despite some earlier discussions (Zahran, 2001; Provorov and Tikhonovich, 2003).

Advances in genetics and genomics of crop plants and their wild relatives have expanded our understanding of complex traits, leading to the discovery of new genes. Hence, we and other groups have characterized and compared the genomic variations of wild and cultivated soybeans by next-generation sequencing, contributing to our knowledge of the diversity and population structure of soybeans (Lam *et al.*, 2010; Li *et al.*, 2014; Zhou *et al.*, 2015). To identify QTLs controlling important traits of soybean, we also constructed a recombinant inbred (RI) population from one wild and one cultivated soybean accession, and used this RI population to clone a salt tolerance causal gene from a salt tolerance major QTL (Qi *et al.*, 2014). In this work, we made further use of this RI population to identify novel QTLs for total nodule fresh weight and ureides contents. The possible involvement of these QTLs in the domestication process is discussed.

## MATERIALS AND METHODS

### Plant material and bacterial strains

Thirty-one previously re-sequenced soybean germplasms (17 wild and 14 cultivated without pedigree relationship) (Lam *et al.*, 2010), and seventeen additional cultivated soybeans from various origins were used in this study (Supplementary Table S1). The non-nodulating soybean nod139 (Carroll *et al.*, 1986) was used as the negative control for BNF. Ninety-six RI lines from a biparental population were adopted for QTL analyses (Qi *et al.*, 2014). Seeds were disinfected with chlorine gas for 16 h and then germinated in plastic trays with moistened sterile vermiculite for 3 days until the radicles reached a length of 3–5 cm. These pre-germinated seeds were transferred to 210 ml plastic cups (one plant per cup) containing moistened sterile vermiculite (with 50 ml sterile water). An additional 50 ml of B and D nutrient solution (Broughton and Dilworth, 1971) without any nitrogen source was added to each cup. Immediately after transferring each plant to a cup, these pre-germinated seeds were inoculated with 1 ml of bacterial culture per plant. Two soybean nodulating rhizobia *Bradyrhizobium diazoefficiens* sp. nov. USDA110 (originally classified as *Bradyrhizobium japonicum* group Ia) (Delamuta *et al.*, 2013) and *S. (E.) fredii* CCBAU45436 (Tian *et al.*, 2012) were cultured in yeast extract mannitol medium at 28 °C with constant agitation (180 r.p.m.) for 5 days and 24 h, respectively, until they reached a density of  $3 \times 10^9$  cells ml<sup>-1</sup>.

### Morphological and biological nitrogen fixation traits

Soybean plants were grown under controlled environmental conditions: 25 °C and a 16:8 light:dark photoperiod. Plants at V3 stage (4 weeks post inoculation) with second trifoliolate leaves fully expanded were removed from the cups. The complete root system was washed with deionized water (Elix Water Purification System, Millipore, Darmstadt, Germany). Nodules were removed from each plant for nodule number and fresh weight quantification. Roots and shoots were separated and the fresh weight of each part was recorded for each plant. Each whole plant was dried inside a paper envelope in the oven at 80 °C for 24 h, and the plant's total dry weight was recorded. Eight to nine plants were evaluated on two independent replicates, which involved a total of approximately 3000 plants. The replicates have been performed under completely randomized design.

The dried plant material was ground to a fine powder using a sample mill (Cyclotec 1093, Foss, Hillerød, Denmark). For ureides extraction, 50 mg of fine powder per plant was boiled for 20 min in 25 mM potassium phosphate buffer

(pH 6). Samples were centrifuged at 10 000 g for 15 min, and the clear supernatant containing ureides was used for the colorimetric determination following a published method (Vogels and Van Der Drift, 1970). In brief, three 100- $\mu$ l aliquots of each sample were collected in three separate tubes for the measurement of endogenous glyoxylate, allantoic acid-derived glyoxylate and allantoin-derived glyoxylate. The content of allantoic acid and allantoin was converted into glyoxylate by acid and alkaline/acid hydrolysis, respectively (Vogels and Van Der Drift, 1970). Finally, the total glyoxylate content was measured by a colorimetric reaction where the glyoxylate is converted into glycolic acid-phenylhydrazone and then oxidized by ferricyanide in the presence of concentrated acid and phenylhydrazine to give a pink color. Absorbance at 520 nm by samples was determined using Synergy H1 Multi-Mode Reader (BioTek Instruments, Inc., Winooski, VT, USA) on 96-well plates, based on a standard curve with known concentrations of glyoxylate. By subtracting the amount of endogenous glyoxylate from the total amount of glyoxylate in each sample subjected to alkaline/acid or acid hydrolysis, the concentration of allantoin or allantoic acid could then be calculated. To check the accuracy of the ureides determination method, non-nodulating soybean plants nod139 were grown alongside the experimental plants as a negative control.

For total nitrogen content, 5–10 mg of dried and milled samples was weighed in tin capsules using an ultra-precision PerkinElmer AD-4 autobalance (PerkinElmer, Waltham, MA, USA). Nitrogen content was then determined by mass spectrometry using the stable isotope facility service of the University of California, Davis (USA).

### Nodule ultrastructure

The nodule ultrastructures of wild and cultivated soybeans were examined by high-pressure freezing (Studer *et al.*, 1992) of nodules at 28 days post inoculation using the High Pressure Freezer System Leica EM PACT2 (Leica Microsystems, Buffalo Grove, IL, USA) followed by observation using a transmission electron microscope Hitachi H-7650 (Hitachi High Technologies America, Inc., Schaumburg, IL, USA).

### RI population and QTL analyses

The RI population obtained by crossing the cultivated *Glycine max* (C08) with the wild-type *Glycine soja* (W05) was adopted from a previous study (Qi *et al.*, 2014). In our previous research (Qi *et al.*, 2014), a core panel of 96 RI lines was re-sequenced at  $\sim 1 \times$  depth to generate high-quality single-nucleotide polymorphism (SNP) information. An array of high-quality SNPs was grouped into bin marker. There are 2757 bin markers distributed on the 20 chromosomes. A bin map of 2992.0 cM in the Kosambi mapping function was generated using the bin markers. QTL analyses were performed using QTLCartographer (<http://statgen.ncsu.edu/qlcart/>). In brief, QTLs were identified using interval mapping and composite interval mapping with a 10-cM scanning window and a 0.5-cM walking step. Log-of-odds (LOD) cutoff was determined with 1000 times permutation with  $P < 0.05$ . The QTL boundary was identified with a 1.5 drop of LOD score peak. Two replicates were performed with nine plants per line per replicate.

### Genomic analyses of the QTL regions

In our previous study (Zhao *et al.*, 2015), sequencing reads data from 47 cultivated soybean accessions and 25 wild soybean accessions were retrieved from 4 studies (Kim *et al.*, 2010; Lam *et al.*, 2010; Li *et al.*, 2013; Chung *et al.*, 2014), of which 17 wild and 14 cultivated soybean accessions from Lam *et al.* (2010) were also used in this study. After mapping to the Williams 82 genome version Gmax\_189, SNPs were called for each soybean accession. Pairwise divergence ( $\theta$ ) within the wild soybean population or the cultivated soybean population and the divergence index ( $F_s$ ) between these two populations were calculated based on this SNP information by using a Perl script with a sliding window size of 20Kb and a step window size of 2Kb. If all the soybean accessions in the wild population or the cultivated population have the same genotype at an SNP locus, then this SNP was designated as a fixed SNP in the corresponding population. In this study, fixed SNPs retrieved from Zhao *et al.* (2015) were further filtered to remove low-confidence data. To be specific, fixed SNP loci with a missing genotype in more than 7 soybean accessions or with a minor allele frequency lower than 3 were removed.

### Statistical analyses

Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp, NY, USA). All the parameters analyzed on this work have passed the Shapiro–Wilk test for normal data distribution except in the case of ureides content. Given the non-normal distribution, non-parametric tests were used for the statistical analyses involving ureides contents.

## RESULTS

### A strong correlation between nitrogen accumulation and nodule development in soybean

A pilot experiment using a non-nodulating soybean mutant was performed (Carroll *et al.*, 1986). The non-nodulating plants contained only a trace amount of ureides which were the by-products of purine degradation, and these plants clearly exhibited symptoms of nitrogen deficiency (Supplementary Figure S1). Therefore, the much higher levels of ureides detected in the nodulating plants can be regarded as the specific product of nitrogen fixation under our experimental set-up.

Several physiological parameters related to nitrogen fixation were measured on a per-plant basis for the 17 wild and 31 cultivated soybean accessions: total plant fresh weight, total root fresh weight, total plant dry weight, total nodule fresh weight, nodule number, total nitrogen and total ureides (Supplementary Table S2). Our results indicated that in general, there are significant correlations between the development of nodules (that is, total nodule fresh weight and nodule number) and general plant growth (that is, total plant fresh and dry weights and total root fresh weight), and the accumulation of ureides and total nitrogen (Table 1).

As expected, total nitrogen per plant was highly correlated with total fresh weight, total dry weight and root fresh weight since nitrogen supports plant growth and development. This was independent of the soybean accessions or types of rhizobia used (Table 1). However, the correlation between total ureides vs total fresh weight, total dry weight and root fresh weight was not statistically significant under some conditions, for example, when wild soybeans were inoculated with CCBAU45436 or when cultivated soybeans were inoculated with USDA110 (Table 1).

Nodule number and total nodule fresh weight are two measurable parameters reflecting the extent of nodule development. Total nodule fresh weight was closely correlated with both total nitrogen and total ureides, independent of soybean accessions and types of rhizobia used (Table 1). On the other hand, nodule number also correlated with both total nitrogen and total ureides, except when wild soybeans were inoculated with CCBAU45436 (Table 1).

### Cultivated soybeans exhibited higher nitrogen fixation capacity than wild soybeans

In general, when inoculated with the same rhizobium, cultivated soybeans as a whole out-performed wild soybeans in most parameters tested, including total fresh weight, total root fresh weight, total dry weight, total nodule fresh weight, nodule number, total nitrogen and total ureides accumulation (Figure 1). This observation indicates that domestication and human selections might have led to improved overall growth in soybean that could be a result of enhanced nitrogen fixation.

In the Americas, it is recommended to use slow-growing rhizobia as the inoculants since they are believed to be more efficient in nitrogen fixation than fast-growing rhizobia. Considering the 31 cultivated soybean accessions that we have studied, we do not observe a significant difference in total fresh weight, root fresh weight or total

dry weight when inoculated with either USDA110 or CCBAU45436 (Figures 1a–c). However, on average, cultivated soybeans inoculated with USDA110 exhibited higher total nodule fresh weight and nodule number, and accumulated more total nitrogen and ureides, when compared with those inoculated with CCBAU45436 (Figures 1d–g).

Despite a seemingly better overall nitrogen fixation performance when USDA110 was used (Figure 1), 17 of the 24 cultivated soybean accessions originating from China exhibited equal performance in terms of total nitrogen per plant when inoculated with either USDA110 or CCBAU45436 (Supplementary Table S2). This suggests that the fast-growing rhizobia can also be used as inoculants when suitable soybean partners are identified.

For wild soybeans, on average, inoculation with USDA110 gave higher nodule number and total ureides, but no statistically significant differences were found for total fresh weight, total dry weight, total root fresh weight, total nodule fresh weight and total nitrogen (Figure 1). When the measurements were compared at the level of individual lines, the results were even more diverse and the advantage of using USDA110 was not clear (Supplementary Table S2; Figure 2). Therefore, the wild soybeans are of very different genetic backgrounds that may respond differently toward USDA110 and CCBAU45436. During domestication and subsequent human selection processes, some cultivated lines giving better performance toward slow-growing rhizobia may have been selected, especially in the Americas.

In symbiosis, while the bacteroids contribute to the nitrogenase activity, the host supplies the carbon to fuel nitrogen fixation. The efficiency of nitrogen accumulation is dependent on the optimized balance between nitrogen and carbon metabolisms.

To further explain the differences between wild and cultivated soybeans in the capacity of nitrogen fixation, we examined the ultrastructure of nodules. Since data are already available for cultivated soybeans in the literature (Cohen *et al.*, 1986; Studer *et al.*, 1992), we only confirmed the previous findings with one cultivated accession C08 (Figures 3a–d). In the same experiment, we examined the ultrastructure of nodules in four different wild soybean accessions inoculated with USDA110 or CCBAU45436 (Figures 3e–h and Supplementary Figure S2).

The ultrastructure of nodules from C08 was similar to those reported previously for both types of rhizobia (Figures 3a–d). Interestingly, there was an exceptionally higher accumulation of poly- $\beta$ -hydroxybutyrate bodies (PHBs) in bacteroids (Figures 3e–h; Supplementary Figure S2) in the nodules of all wild soybean accessions, compared with cultivated ones.

### Mapping QTLs of biological nitrogen fixation using an RI population

Our physiological analyses above suggested that cultivated soybeans may have accumulated important genetic determinants that enable the plants to perform more efficient nitrogen fixation. To explore the possible genetic determinants involved, we performed QTL analyses on an RI population originated from a cross between a wild (W05) and a cultivated (C08) soybean accession. In this study, we showed that, compared with W05, C08 exhibited a higher total fresh weight, total root fresh weight, total dry weight, total nodule fresh weight, total nitrogen and total ureides under the same experimental settings described above (Supplementary Table S2). As inoculation with USDA110 exhibited more consistent differences in these parameters, we used this rhizobium strain for our subsequent population studies.

We quantified the phenotypes of the 96 core RI lines (containing two parental lines), including total plant fresh weight, total root fresh weight, total nodule fresh weight and total ureides. The two parents of

**Table 1** Correlations between indicators of symbiotic performance on wild and cultivated soybean populations upon inoculations with *S. (E.) fredii* CCBAU45436 or *B. diazoefficiens* sp. nov. USDA110

Treatment	Trait 1	Trait 2	Correlation coefficient	P-value
Wild soybean with <i>S. (E.) fredii</i> (CCBAU45436)	<sup>a</sup> Total nitrogen vs	Nodule number	0.230	0.374
		Total nodule fresh weight	0.751	0.001
		Total plant fresh weight	0.925	<0.001
		Total plant dry weight	0.893	<0.001
		Total root fresh weight	0.857	<0.001
	<sup>b</sup> Total ureides vs	Nodule number	-0.348	0.171
		Total nodule fresh weight	0.414	0.098
		Total plant fresh weight	0.404	0.107
		Total plant dry weight	0.306	0.232
		Total root fresh weight	0.483	0.050
Wild soybean with <i>B. diazoefficiens</i> sp. nov. (USDA110)	<sup>a</sup> Total nitrogen vs	Nodule number	0.702	0.002
		Total nodule fresh weight	0.926	<0.001
		Total plant fresh weight	0.979	<0.001
		Total plant dry weight	0.938	<0.001
		Total root fresh weight	0.939	<0.001
	<sup>b</sup> Total ureides vs	Nodule number	0.564	0.018
		Total nodule fresh weight	0.855	<0.001
		Total plant fresh weight	0.885	<0.001
		Total plant dry weight	0.836	<0.001
		Total root fresh weight	0.836	<0.001
Cultivated soybean with <i>S. (E.) fredii</i> (CCBAU45436)	<sup>a</sup> Total nitrogen vs	Nodule number	0.419	0.019
		Total nodule fresh weight	0.446	<0.001
		Total plant fresh weight	0.779	<0.001
		Total plant dry weight	0.754	<0.001
		Total root fresh weight	0.597	<0.001
	<sup>b</sup> Total ureides vs	Nodule number	0.396	0.027
		Total nodule fresh weight	0.483	0.006
		Total plant fresh weight	0.319	0.080
		Total plant dry weight	0.377	0.037
		Total root fresh weight	0.169	0.364
Cultivated soybean with <i>B. diazoefficiens</i> sp. nov. (USDA110)	<sup>a</sup> Total nitrogen vs	Nodule number	0.383	0.033
		Total nodule fresh weight	0.726	<0.001
		Total plant fresh weight	0.850	<0.001
		Total plant dry weight	0.871	<0.001
		Total root fresh weight	0.517	0.003
	<sup>b</sup> Total ureides vs	Nodule number	0.424	0.017
		Total nodule fresh weight	0.453	0.010
		Total plant fresh weight	0.315	0.084
		Total plant dry weight	0.248	0.179
		Total root fresh weight	0.053	0.776

Correlations analyses were performed on 17 wild and 31 cultivated soybean accessions using the average value per line from 2 replicates of 8–9 plants each. Correlation is significant at the  $P < 0.05$  level.

<sup>a</sup>Pearson's correlation.

<sup>b</sup>Spearman's rho non-parametric correlation.

this population showed significant differences in these parameters (Supplementary Figure S3), consistent with the results in germplasm studies. The dependent variables exhibited normal distributions except for total ureides (Supplementary Figure S4), which showed an exponential distribution. To normalize the data distribution for total ureides, the data were transformed by taking the natural logarithm (Ln) of the original data. QTL analyses using both the transformed and untransformed data gave similar results (see below).

Using composite interval mapping, we identified several QTLs (Figure 4; Table 2). Figure 4 shows the peaks of the LOD score curve for each QTL. Table 2 shows the LOD cutoff, physical length and genetic distances for each characterized QTL. Only QTLs that could be identified in two independent experiments were further analyzed.

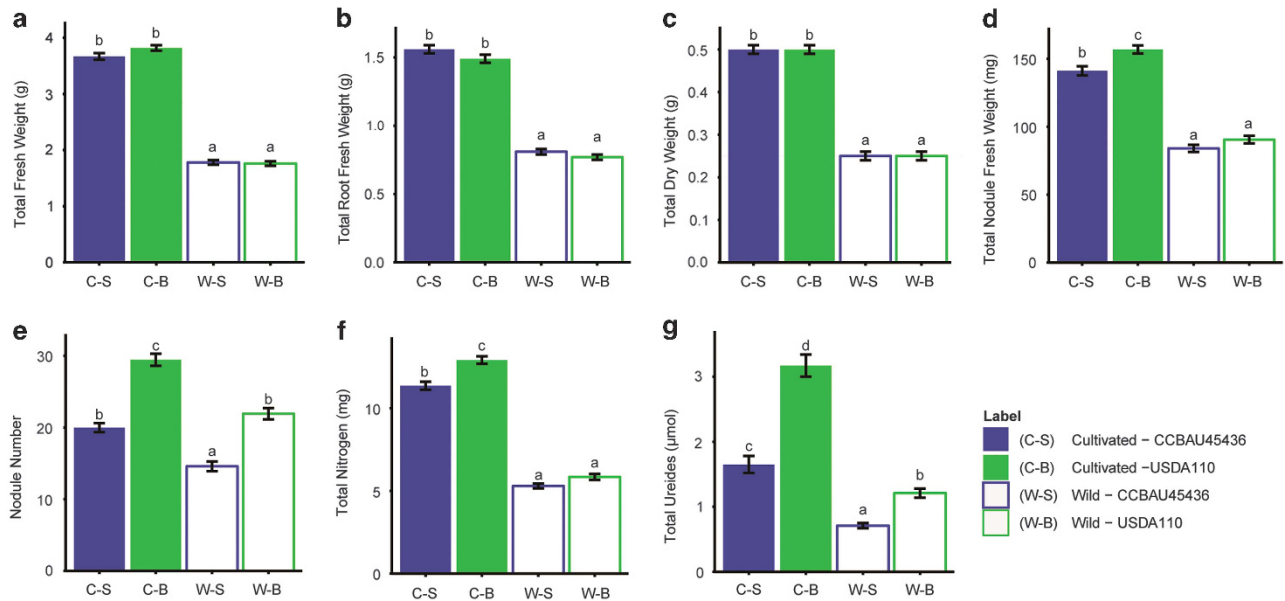
The span of each QTL was reported by combining the genomic regions of the two replicates.

In the QTL study, we identified one high-confidence QTL for root fresh weight on chromosome 8 (explaining 10.4% of the total variance) and two high-confidence QTLs for total plant fresh weight on chromosome 7 (explaining 12.8% of the total variance) and chromosome 18 (explaining 13.1% of the total variance), respectively.

The focus of this research is to identify new loci that are related to nitrogen fixation. Here, we found one high-confidence QTL for total ureides and two high-confidence QTLs for total nodule fresh weight.

The novel QTL for total ureides identified in this study is located on chromosome 17 (explaining 11.6% of the total variance), which is





**Figure 1** Overall symbiotic performance of cultivated and wild soybeans. Thirty-one cultivated and seventeen wild soybean accessions were inoculated with either *B. diazoefficiens* sp. nov. USDA110 or *S. (E.) fredii* CCBAU45436. Average data per plant for cultivated and wild soybean populations were compared: (a) total fresh weight, (b) total root fresh weight, (c) total dry weight, (d) total nodule fresh weight, (e) nodule number per plant, (f) total nitrogen and (g) total ureides. Solid bars: cultivated soybean; open bars: wild soybean; green bars: plants inoculated with USDA110; blue bars: plants inoculated with CCBAU45436. For total fresh weight, total root fresh weight, total dry weight, total nodule fresh weight, nodule number and total nitrogen per plant, different letters indicate a significant difference between groups by one-way ANOVA followed by *post hoc* Tukey test ( $P < 0.05$ ). For total ureides, different letters indicate a significant difference between groups using Mann–Whitney non-parametric test between pairs of independent samples ( $P < 0.05$  with Bonferroni correction). Each bar represents the average value of 31 cultivated soybean or 17 wild soybean accessions. Error bar:  $\pm$  s.e.m.

different from the previously identified QTLs that are located on chromosomes 6, 9, 13 and 19 (Hwang *et al.*, 2013).

We also identified one novel QTL for total nodule fresh weight on chromosome 12, and a second QTL on chromosome 18 which is located in close proximity to a total nodule fresh weight QTL reported previously (Hwang *et al.*, 2014), explaining 7.92 and 19.83% of the total variance, respectively.

The total nodule fresh weight QTL on chromosome 18 overlaps with a major QTL for total plant fresh weight. High total nodule fresh weight may indicate higher nitrogen fixation capacity that results in an increase in total plant fresh weight. Conversely, plants with higher total plant fresh weight could support better nodule development.

To provide more insights into the QTLs identified in this work, we investigated the genomic regions reported for these QTLs (Figure 5). We compared the nucleotide diversity ( $\theta\pi$ ), the divergence index ( $F_{st}$ ) and the distribution of fixed SNP between cultivated (47 accessions) and wild (25 accessions) soybeans. Such comparisons could point to the possible involvement of specific genomic regions within these QTLs during domestication.

The root fresh weight QTL on chromosome 8 identified in this study overlaps with previously reported QTLs for domestication (Brensha *et al.*, 2012) and root architecture (Manavalan *et al.*, 2015). Consistent with the involvement of this QTL in soybean domestication, we found that the  $\theta\pi$  value of the wild soybean population is much higher than that of the cultivated soybean population, over an extended genomic region in this QTL (Figure 5a). Consistent with this observation, the  $F_{st}$  values (black dots) in the QTL region are generally higher than the average  $F_{st}$  value of the entire chromosome 8 (purple line) (Figure 5a).

On the other hand, when we examined the QTL for total nodule fresh weight on chromosome 18 (which overlaps with a QTL for total

plant fresh weight), the  $\theta\pi$  values of the cultivated and wild soybean populations were very much alike and the  $F_{st}$  values in this genomic region were very low (Figure 5b), indicating that this QTL may not have a strong relationship with domestication.

Another QTL for total nodule fresh weight is located on chromosome 12. Some regions within this QTL exhibited higher  $\theta\pi$  values in the wild soybean population than in the cultivated soybean population (Figure 5c). Interestingly, these regions also have higher frequencies of fixed SNPs (Figure 5c).

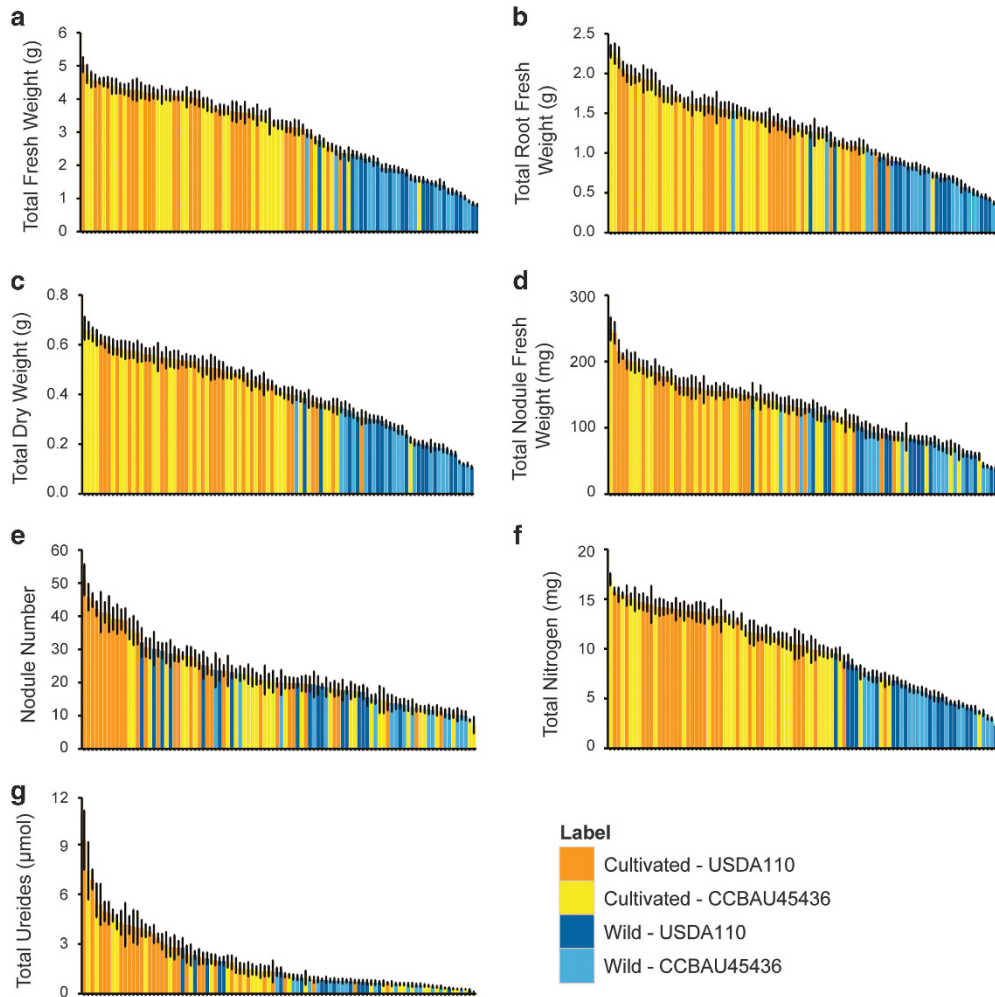
Similarly, when examining the genomic region of the QTL for total ureides located on chromosome 17, the region between 14.75 and 16.00 Mb exhibited higher  $\theta\pi$  values in the wild soybean population than in the cultivated population, which were accompanied with higher  $F_{st}$  values (Figure 5d). Higher frequencies of fixed SNPs were also observed in these regions in cultivated soybeans (Figure 5d).

Together, these results suggest that improvements in total nodule fresh weight and ureides accumulation are associated with the domestication process.

## DISCUSSION

In this study, we used a combination of soybean germplasm (both cultivated and wild) and a bi-parental RI population to test the hypothesis that improvement in nitrogen fixation capacity was the result of domestication and human selection. It should be noted that parameters such as initial seed weight, plant growth rate and photosynthetic efficiency of each accession could affect the physiological measurements. Hence, we have used a large number of accessions to reduce the statistical errors due to these differences.

The first piece of supporting evidence comes from a comprehensive physiological analysis of 17 wild and 31 cultivated soybean accessions



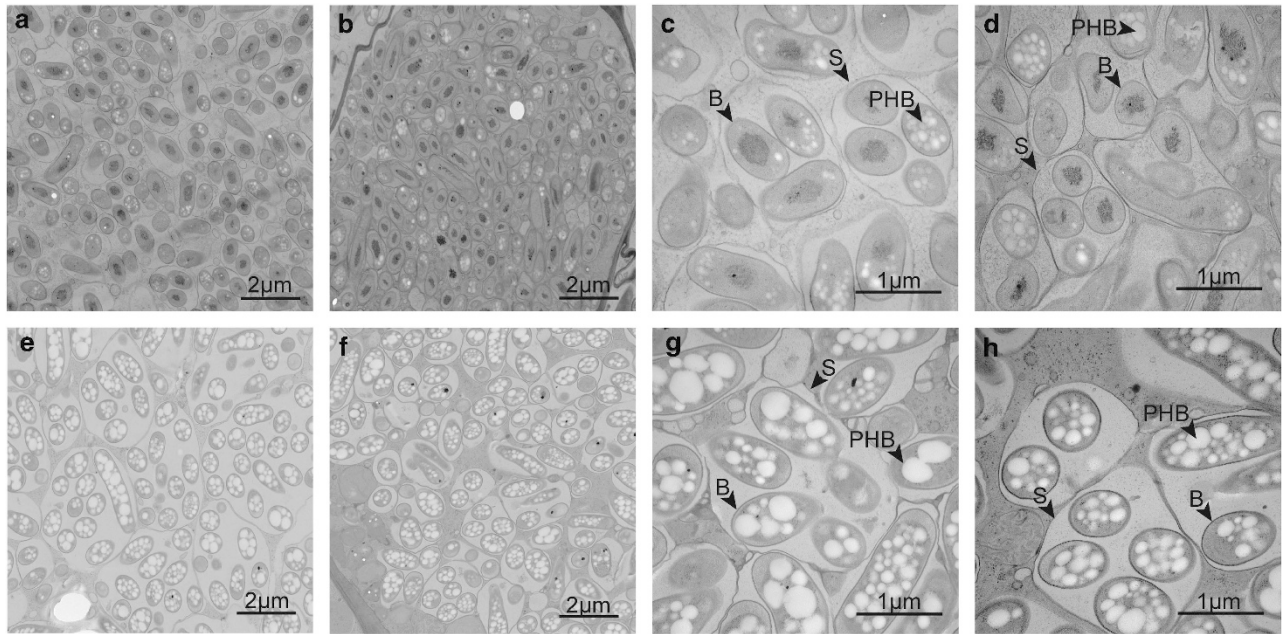
**Figure 2** Performance of individual soybean lines on nitrogen fixation-related traits. Thirty-one cultivated and seventeen wild soybean accessions were inoculated with either *B. diazoefficiens* sp. nov. USDA110 or *S. (E.) fredii* CCBAU45436. Average data per plant for individual lines were plotted in the following order: (a) total fresh weight, (b) total root fresh weight, (c) total dry weight, (d) total nodule fresh weight, (e) nodule number, (f) total nitrogen and (g) total ureides. Each bar represents the average value of 4–16 plants. Error bar:  $\pm$  s.e.m.

that were inoculated with two different types of rhizobia. Irrespective of the rhizobial strains used, we demonstrated a general trend that the nitrogen fixation capacity of cultivated soybeans out-performed wild soybeans (Figures 1 and 2).

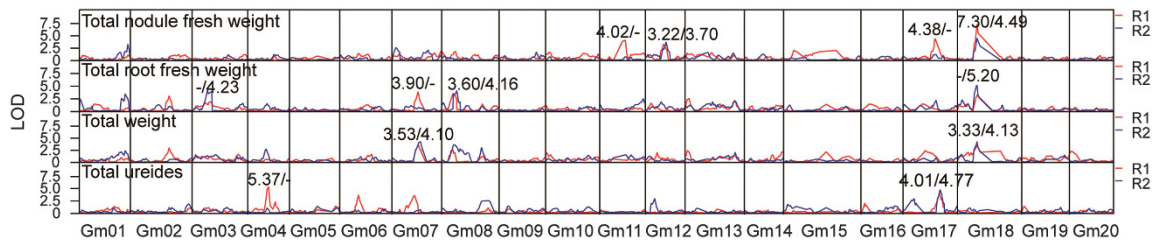
In connection to this, we also observed a prominent accumulation of PHBs in wild soybeans (Figure 3), suggesting that there may be an imbalance in carbon and nitrogen metabolisms in wild soybeans that could affect the capacity of nitrogen fixation. PHB accumulation has been associated with a release of TCA on bacteroids when high amounts of dicarboxylates are fuelling the nitrogen fixation (Trainer and Charles, 2006). Likewise, it has been suggested that this accumulation could confer adaptive advantages to rhizobia over other soil bacteria (Trainer and Charles, 2006). Although the role of PHB accumulation during symbiosis is not fully understood, our results show that the accumulation by the same rhizobium type could be highly conditioned by the host background. The strong accumulation of PHBs in wild soybean suggests that the process of nitrogen fixation in wild soybeans is not limited by the carbon supply. On the other hand, the C/N balance in wild soybeans might not be optimized for nitrogen fixation.

The performances of a slow-growing *Bradyrhizobium* USDA110 and a fast-growing *Sinorhizobium* CCBAU45436 were compared. We found that when wild soybeans were inoculated, the response to USDA110 and CCBAU45436 was very diverse and without a general pattern (Figure 2; Table 1). For cultivated soybeans, while it is expected that the performance of USDA110 would be better than CCBAU45436 when soybean cultivars from America were used (DuTeau *et al.*, 1986; Cregan and Keyser, 1988), the advantage of USDA110 is less prominent when Chinese soybean cultivars were employed (Supplementary Table S2). This result suggests that the advantage of using the slow-growing *Bradyrhizobium* may be due to the initial germplasm introduction and the subsequent breeding of soybean cultivars in the Americas. On the other hand, as a large proportion of fields under soybean cultivation in China is of high pH, a condition to which *Sinorhizobium* is better adapted, selections might have taken place so that some Chinese cultivars could use *Sinorhizobium* efficiently.

The advantages of the slow-growing *Bradyrhizobium* over the fast-growing *Sinorhizobium* were demonstrated in both cultivated and wild soybeans only in reference to total ureides and nodule number



**Figure 3** Electron microscopic images of the nodules of cultivated and wild soybeans inoculated with either *B. diazoefficiens* sp. nov. USDA110 or *S. (E.) fredii* CCBAU45436. (a, c) Cultivated soybean CO8 inoculated with USDA110. (b, d) CO8 inoculated with CCBAU45436. (e, g) Wild soybean W05 inoculated with USDA110. (f, h) W05 inoculated with CCBAU45436. Samples were collected and fixed 28 days post inoculation. B, Bacteroid; S, Symbiosome.



**Figure 4** LOD score distributions of nitrogen fixation-related traits across 20 soybean chromosomes. Peak LOD scores higher than the LOD cutoff values calculated by permutation test (1000 times,  $P < 0.05$ ) were indicated. '-' indicates the absence of a statistically significant LOD score in one replicate. R1 (red line): biological replicate 1; R2 (blue line): biological replicate 2. Gm1: soybean chromosome 1; Gm2: soybean chromosome 2, and so on.

**Table 2** QTLs identified for biological nitrogen fixation-related traits

Trait	LOD cutoff <sup>a</sup>	Chr. no.	Var (%)	QTL position <sup>b</sup>					
				Start position	End position	Physical length (kb)	Genetic start	Genetic end	Genetic distance (cM)
Total nodule fresh weight (g)	3.71	12	9.47	8 313 537	11 700 096	3387	52.559	59.822	7.263
Total root fresh weight (g)	3.76	8	10.25	12 018 056	13 254 281	1236	56.188	57.864	1.676
Total plant fresh weight (g)	3.68	7	12.86	4 951 994	7 287 837	2336	32.906	44.173	11.267
Ureides (μmol per plant)	3.65	17	13.26	16 166 636	17 934 015	1767	76.179	83.671	7.492
		18	11.75	12 018 056	12 815 362	797	56.188	57.304	1.116
		17	13.26	14 233 456	18 442 425	4209	104.561	111.434	6.873

Abbreviations: Chr. No., chromosome number; LOD, log-of-odds; QTL, quantitative trait locus; Var, variance.

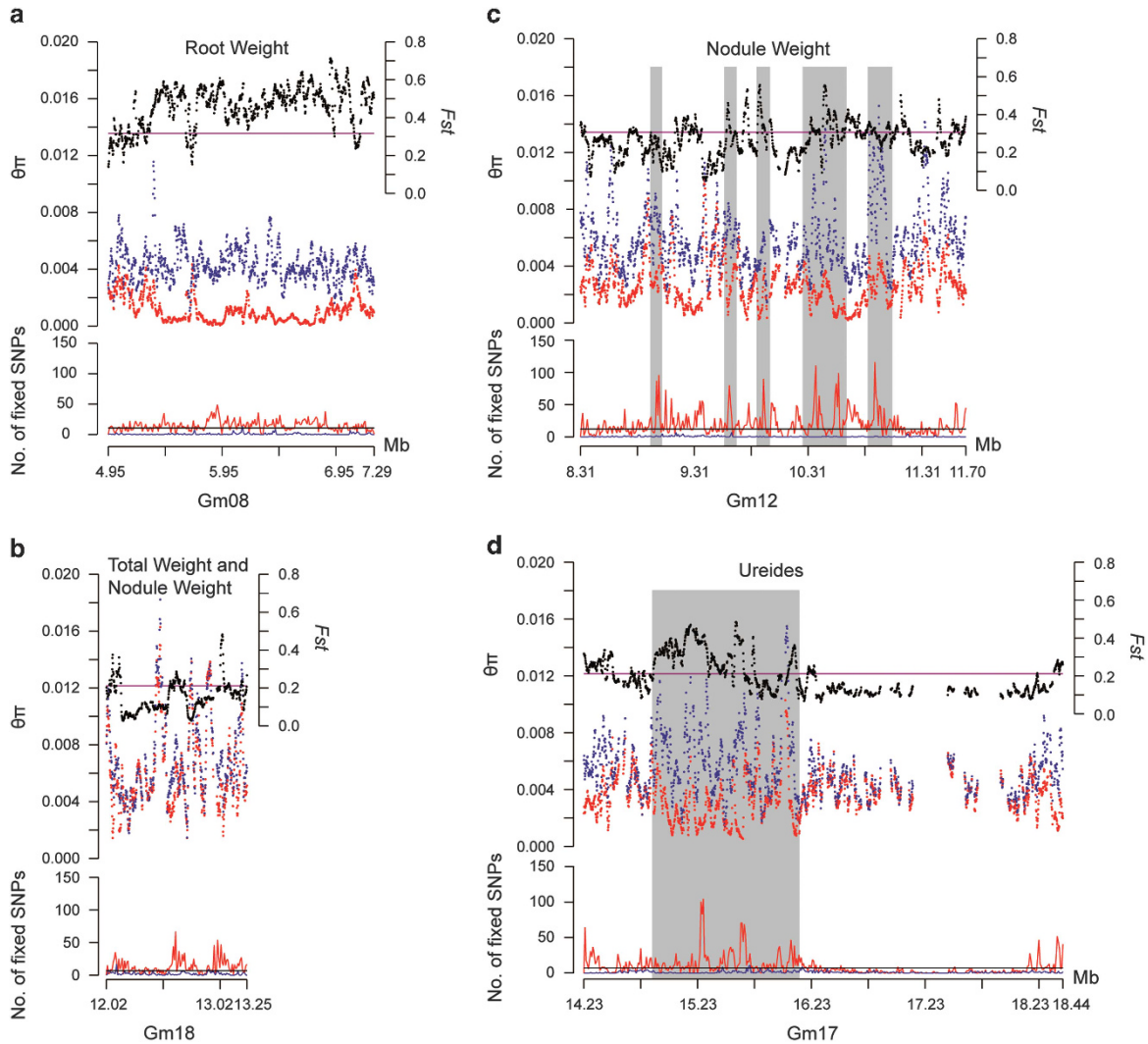
<sup>a</sup>LOD score cutoff of QTLs was determined by permutation tests (1000 times;  $P < 0.05$ ).

<sup>b</sup>Start position and end position are based on Williams 82 genome version Gmax\_189.

(Figure 2). However, higher ureides accumulation due to the inoculation of USDA110 could not be translated into higher total nitrogen or total fresh weight (Figure 2). One possible explanation is that the re-assimilation of ammonia stored in the form of ureides

could be a key limiting step as it is also very energy consuming (Todd *et al.*, 2006). Another parameter in which USDA110 showed consistently better performance than CCBAU45436 was nodule number. However, the study of hyper-nodulating mutant indicated that an





**Figure 5** Plots of divergence index ( $F_{st}$ ), nucleotide diversity ( $\theta\pi$ ) and fixed SNPs within the QTL regions. (a) QTL region of root fresh weight on chromosome 8. (b) QTL region of total plant fresh weight and total nodule fresh weight on chromosome 18. Gray boxes highlighted regions with high nucleotide diversity and large number of fixed nucleotides. (c) QTL region of total nodule fresh weight on chromosome 12. (d) QTL region of ureides content on chromosome 17. Gray box highlighted region with above-average  $F_{st}$  values, high nucleotide diversity and large number of fixed nucleotides. Black dots:  $F_{st}$  values between wild and cultivated soybeans. Purple line: average  $F_{st}$  value of the chromosome. Red dots and blue dots indicate the  $\theta\pi$  of cultivated soybeans and wild soybeans, respectively. Red lines and blue lines indicate the number of fixed SNPs of cultivated soybeans and wild soybeans, respectively. Black lines: average number of fixed SNPs in a chromosome of cultivated soybeans.

increase in nodule number may not always translate into higher grain yield (Song *et al.*, 1995).

To search for genomic clues to support the association between nitrogen fixation capacity and domestication, we employed an RI population to identify new QTLs for total ureides and total nodule fresh weight (Figure 4; Table 2). Only USDA110 was used as the inoculant. Since the genetic population was composed of descendants from a cross between a wild and a cultivated soybean (Qi *et al.*, 2014), the new QTLs may contain genomic information related to domestication. Ureides are metabolic intermediates of nitrogen fixation in soybean and are commonly used as parameters to measure nitrogen fixation (Herridge and Peoples, 1990). The importance of total nodule fresh weight as an indicator of symbiotic performance has been discussed in previous reports (Döbereiner, 1966; Hungria and Bohrer, 2000).

The QTLs mapped in this study account for ~10–25% of the genetic variance of the traits in question. It is not uncommon that

some QTLs can only explain a certain level (<20%) of genetic variations (Kearsey and Farquhar, 1998). For traits that are controlled by multiple loci, each locus may only account for a small variance.

As an attempt to assess domestication in the genome, we employed several parameters, including  $\theta\pi$ ,  $F_{st}$  and fixed SNPs. These indicators were used previously to trace the genomic footprints of domestication and human selection (Zhao *et al.*, 2015). Within the total nodule fresh weight QTL on chromosome 12 and the total ureides QTL on chromosome 17, we identified regions exhibiting high values of  $\theta\pi$  and  $F_{st}$ , and high frequencies of fixed SNPs (Figure 5). These pieces of evidence, although not in themselves a strong proof, do suggest that these genomic regions might have undergone substantial selection during domestication. However, low  $F_{st}$  values and high  $\theta\pi$  are observed in both wild and cultivated soybeans within the total plant weight and total nodule fresh weight QTLs on chromosome 18 (Figure 5b), suggesting that this region may not have been affected by domestication. On the other hand, this QTL region could be



important for subsequent soybean improvements through the breeding of cultivars. It is possible that the differences only occur among some soybean accessions, instead of representing a general difference between wild and cultivated soybeans. In this scenario, there may not be a strong indication based on  $F_{st}$  or  $\theta\pi$  values. This implied that unlike region on chromosome 12, QTL region on chromosome 18 may have escaped from human selection.

Identification of the causal genes in the novel QTLs related to nitrogen fixation capacity identified in this work (one QTL for total ureides and two QTLs for total nodule fresh weight) will require further functional studies. Nonetheless, when taking a quick look at these QTLs, some genes of potential interest could be identified. For example, in the total ureides QTL on chromosome 17, a gene (Glyma17g15190) encoding 5'-phosphoribosylglycinamide formyltransferase, a key enzyme of the *de novo* purine biosynthesis pathway, was found. The purine biosynthetic pathway provides the precursors for the biosynthesis of ureides (Schubert, 1986). Moreover, a gene (Glyma17g19660) encoding a nicotinate phosphoribosyltransferase was also found. This enzyme is required for pyridine nucleotide recycling (Hashida *et al.*, 2009), an essential process to maintain the high energy demand in nodules (Schubert, 1986).

In the total nodule fresh weight QTL on chromosome 12, a gene encoding succinate dehydrogenase 3-2 (Glyma12g10590) was identified. Succinate dehydrogenase catalyzes the oxidation of succinate to fumarate, which is later converted into malate by fumarase and serves as the carbon source for the symbiotic bacteroid.

In the total nodule fresh weight QTL on chromosome 18, a sucrose phosphate synthase gene (SPS: Glyma18g12890) was also found. SPS catalyzes the formation of sucrose-6-phosphate from UDP-glucose and fructose-6-phosphate, and this enzyme has a putative role in nodule carbon partitioning to different metabolic pathways (Aleman *et al.*, 2010).

In summary, by making use of wild germplasms and a unique RI population, we provided physiological and genomic data to support the notion that an improvement in nitrogen fixation capacity formed part of the domestication process of soybean. In addition, we identified new QTLs for total ureides and total nodule fresh weight, and genes related to nitrogen fixation could be identified therein.

## DATA ARCHIVING

Data archiving Genotypes and phenotype measurements for 96 RILs available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.vp6fs>.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work was supported by grants from Hong Kong Research Grants Council (General Research Fund: 468610; Area of Excellence: AoE/M-05/12) and Lo Kwee-Seong Biomedical Research Fund to H-ML. NM was supported by a fellowship from Instituto Nacional de Tecnología Agropecuaria (INTA), Argentina. Dr Changfu Tian (China Agricultural University, Beijing, PRC) kindly provided the rhizobium strain CCB4U45436. We also thank Mr Freddie Kwok and Ms Iris Tong for their technical assistance, Biologist Ignacio Lescano for the advice on ureides determination, and Ms Jee-Yan Chu for copy-editing the manuscript.

Aleman L, Ortega JL, Martinez-Grimes M, Seger M, Holguin FO, Uribe DJ *et al.* (2010). Nodule-enhanced expression of a sucrose phosphate synthase gene member (*MsSPSA*)

has a role in carbon and nitrogen metabolism in the nodules of alfalfa (*Medicago sativa* L.). *Planta* **231**: 233–244.

Brensha WB, Kantartzi SK, Meksem K, Grier I, Robert L, Barakat *et al.* (2012). Genetic analysis of root and shoot traits in the 'Essex' by 'Forrest' recombinant inbred line (RIL) population of soybean [*Glycine max* (L.) Merr.]. *J Plant Genome Sci* **1**: 1–9.

Broughton W, Dilworth M (1971). Control of leghaemoglobin synthesis in snake beans. *Biochem J* **125**: 1075–1080.

Carroll BJ, McNeil DL, Gresshoff PM (1985). A supernodulation and nitrate-tolerant symbiotic (*nts*) soybean mutant. *Plant Physiol* **78**: 34–40.

Carroll BJ, McNeil DL, Gresshoff PM (1986). Mutagenesis of soybean (*Glycine max* (L.) Merr.) and the isolation of non-nodulating mutants. *Plant Sci* **47**: 109–114.

Chung WH, Jeong N, Kim J, Lee WK, Lee YG, Lee SH *et al.* (2014). Population structure and domestication revealed by high-depth resequencing of Korean cultivated and wild soybean genomes. *DNA Res* **21**: 153–167.

Cohen HP, Sarath G, Lee K, Wagner FW (1986). Soybean root nodule ultrastructure during dark-induced stress and recovery. *Protoplasma* **132**: 69–75.

Cregan P, Keyser H (1988). Influence of *Glycine* spp. on competitiveness of *Bradyrhizobium japonicum* and *Rhizobium fredii*. *Appl Environ Microbiol* **54**: 803–808.

Döbereiner J (1966). Evaluation of nitrogen fixation in legumes by the regression of total plant nitrogen with nodule weight. *Nature* **210**: 850–852.

Delamata JRM, Ribeiro RA, Ormeño-Orrillo E, Melo IS, Martínez-Romero E, Hungria M (2013). Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group la strains as *Bradyrhizobium diazoefficiens* sp. nov. *Int J Syst Evol Microbiol* **63**: 3342–3351.

DuTeau N, Palmer R, Atherly A (1986). Fast-growing *Rhizobium fredii* are poor nitrogen-fixing symbionts of soybean. *Crop Sci* **26**: 884–889.

Hashida S, Takahashi H, Uchimiya H (2009). The role of NAD biosynthesis in plant development and stress responses. *Ann Bot* **103**: 819–824.

Herridge DF, Peoples MB (1990). Ureide assay for measuring nitrogen fixation by nodulated soybean calibrated by 15 N methods. *Plant Physiol* **93**: 495–503.

Hungria M, Boddey LH, Santos MA, Vargas MAT (1998). Nitrogen fixation capacity and nodule occupancy by *Bradyrhizobium japonicum* and *B. elkanii* strains. *Biol Fert Soils* **27**: 393–399.

Hungria M, Bohrer T (2000). Variability of nodulation and dinitrogen fixation capacity among soybean cultivars. *Biol Fert Soils* **31**: 45–52.

Hwang S, King CA, Davies MK, Ray JD, Cregan PB, Purcell LC (2013). QTL analysis of shoot ureide and nitrogen concentrations in soybean [*Glycine max* (L.) Merr.]. *Crop Sci* **53**: 2421–2433.

Hwang S, Ray JD, Cregan PB, King CA, Davies MK, Purcell LC (2014). Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (*Glycine max* L.(Merr.)). *Euphytica* **195**: 419–434.

Kearsey MJ, Farquhar AGL (1998). QTL analysis in plants; where are we now? *Heredity* **80**: 137–142.

Keyser HH, Li F (1992). Potential for increasing biological nitrogen fixation in soybean. *Plant Soil* **141**: 119–135.

Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY *et al.* (2010). Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc Natl Acad Sci USA* **107**: 22032–22037.

Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL *et al.* (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet* **43**: 387–387.

Li YH, Zhao SC, Ma JX, Li D, Yan L, Li J *et al.* (2013). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* **14**: 579.

Li YH, Zhou GY, Ma JX, Jiang WK, Jin LG, Zhang ZH *et al.* (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotechnol* **32**: 1045–1052.

Manavalan LP, Prince SJ, Musket TA, Chaky J, Deshmukh R, Vuong TD *et al.* (2015). Identification of novel QTL governing root architectural traits in an interspecific soybean population. *PLoS One* **10**: e0120490.

Mathews A, Carroll BJ, Gresshoff PM (1992). Studies on the root control of non-nodulation and plant-growth of non-nodulating mutants and a supernodulating mutant of soybean (*Glycine max* (L.) Merr.). *Plant Sci* **83**: 35–43.

Provorov NA, Tikhonovich IA (2003). Genetic resources for improving nitrogen fixation in legume-rhizobia symbiosis. *Genet Resour Crop Evol* **50**: 89–99.

Qi X, Li M-W, Xie M, Liu X, Ni M, Shao G *et al.* (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat Commun* **5**: 4340.

Richards RA (2000). Selectable traits to increase crop photosynthesis and yield of grain crops. *J Exp Bot* **51**: 447–458.

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

Schubert KR (1986). Products of biological nitrogen fixation in higher plants: synthesis, transport, and metabolism. *Annu Rev Plant Biol* **37**: 539–574.

Shelp BJ, Ireland RJ (1985). Ureide metabolism in leaves of nitrogen-fixing soybean plants. *Plant Physiol* **77**: 779–783.

Song L, Carroll B, Gresshoff P, Herridge D (1995). Field assessment of supernodulating genotypes of soybean for yield, N<sub>2</sub> fixation and benefit to subsequent crops. *Soil Biol Biochem* **27**: 563–569.

Stougaard J (2000). Regulators and regulation of legume root nodule development. *Plant Physiol* **124**: 531–540.

Studer D, Hennecke H, Müller M (1992). High-pressure freezing of soybean nodules leads to an improved preservation of ultrastructure. *Planta* **188**: 155–163.

- Takahashi M, Kokubun M, Akao S (1995). Characterization of nitrogen assimilation in a supernodulating soybean mutant En6500. *Soil Sci Plant Nutr* **41**: 567–575.
- Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF *et al.* (2012). Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc Natl Acad Sci USA* **109**: 8629–8634.
- Todd CD, Tipton PA, Blevins DG, Piedras P, Pineda M, Polacco JC (2006). Update on ureide degradation in legumes. *J Exp Bot* **57**: 5–12.
- Trainer MA, Charles TC (2006). The role of PHB metabolism in the symbiosis of rhizobia with legumes. *Appl Microbiol Biotechnol* **71**: 377–386.
- Vogels G, Van Der Drift C (1970). Differential analyses of glyoxylate derivatives. *Anal Biochem* **33**: 143–157.
- Zahrán HH (2001). Rhizobia from wild legumes: diversity, taxonomy, ecology, nitrogen fixation and biotechnology. *J Biotechnol* **91**: 143–153.
- Zhao S, Zheng F, He W, Wu H, Pan S, Lam H-M (2015). Impacts of nucleotide fixation during soybean domestication and improvement. *BMC Plant Biol* **15**: 81.
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W *et al.* (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* **33**: 408–414.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)