

RESEARCH ARTICLE

Open Access

# Detecting SNPs underlying domestication-related traits in soybean

Ying-Hui Li<sup>1</sup>, Jochen C Reif<sup>2</sup>, Scott A Jackson<sup>3</sup>, Yan-Song Ma<sup>1,4</sup>, Ru-Zhen Chang<sup>1</sup> and Li-Juan Qiu<sup>1\*</sup>

## Abstract

**Background:** Cultivated soybean (*Glycine max*) experienced a severe genetic bottleneck during its domestication and a further loss in diversity during its subsequent selection. Here, a panel of 65 wild (*G. soja*) and 353 cultivated accessions was genotyped at 552 single-nucleotide polymorphism loci to search for signals of selection during and after domestication.

**Results:** The wild and cultivated populations were well differentiated from one another. Application of the  $F_{st}$  outlier test revealed 64 loci showing evidence for selection. Of these, 35 related to selection during domestication, while the other 29 likely gradually became monomorphic as a result of prolonged selection during post domestication. Two of the SNP locus outliers were associated with testa color.

**Conclusions:** Identifying genes controlling domestication-related traits is important for maintaining the diversity of crops. SNP locus outliers detected by a combined forward genetics and population genetics approach can provide markers with utility for the conservation of wild accessions and for trait improvement in the cultivated gene pool.

**Keywords:** Soybean, Interspecific differentiation, Outliers, Selection, Testa color

## Background

The domestication of plants has been a key driver of the development of human civilization [1,2]. The necessary changes to plant phenotype and physiology have been brought about by a process of selection at key so-called “domestication” genes [1]. Both top-down and bottom-up approaches have been taken to identify the genomic regions most clearly affected by domestication and selection [3]. The former aims to isolate the genes or quantitative trait loci (QTL) responsible for a given phenotype, and has been successful in identifying a number of major effect genes in rice [4-9], maize [10,11] and wheat [12,13]. Bottom-up approaches apply population genetics strategies in which the focus is to uncover genes showing evidence for selection, followed up by attempting to link these genes to relevant phenotypes using a bioinformatics or a reverse genetics approach. Evidence for

selective sweeps has been discovered in maize, rice, wheat, soybean and sunflower [2,14-19].

Bottom-up approaches can be based on either a whole-genome re-sequencing program [19-23] or by concentrating on a pre-selected set of candidate genes [2,24-26]. The whole-genome approach is powerful, but scale-up requires a major investment. Typically, the number of accessions targeted for re-sequencing is less than 35 [19-23], a figure which reduces the detection power and simultaneously increases the risk of false positives [27,28]. In addition, many of the SNPs identified by re-sequencing do not in reality signal selection, but rather are the outcome of “genetic hitchhiking” [29,30]. To overcome this problem, the “outlier scan” test has been elaborated; this widely exploited test permits the screening of a large number of accessions [31,32]. To date, “ $F_{st}$  outliers” diagnostic of selection have been informative in several plant species, including sunflower [33], maize [16,34], white spruce [26], and other conifers [35].

In soybean (*Glycine max* (L.) Merr.), the traits most closely associated with domestication are a marked increase in the size of the inflorescences and in grain yield per plant, and an enhanced level of apical dominance. Other traits that likely have been subjected to prolonged

\* Correspondence: qiujuan@caas.cn

<sup>1</sup>The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/Key Lab of Germplasm Utilization (MOA), Institute of Crop Science, Chinese Academy of Agricultural Sciences, 100081 Beijing, P.R. China

Full list of author information is available at the end of the article

selection are the loss of testa color and increased resistance against a range of pathogens and insects. As in most crops, the effect of domestication and subsequent anthropogenic selection pressure has been to gradually reduce the genetic diversity remaining in the pool of cultivated materials. Cultivated soybean was likely domesticated from *G. soja* Sieb. & Zucc [36]. The existence of a genetic bottleneck has been established through an analysis of allelic diversity at both microsatellite and SNP loci as well as within genic sequences [37-39]. Here, an attempt was made to apply SNP genotyping to a panel of both cultivated and wild accessions to identify signals of selection, with a particular focus on testa color. The analyses reveal that combining a population genetics with a forward genetic approach provides an effective method to identify sequences that underlie an agronomic trait.

## Methods

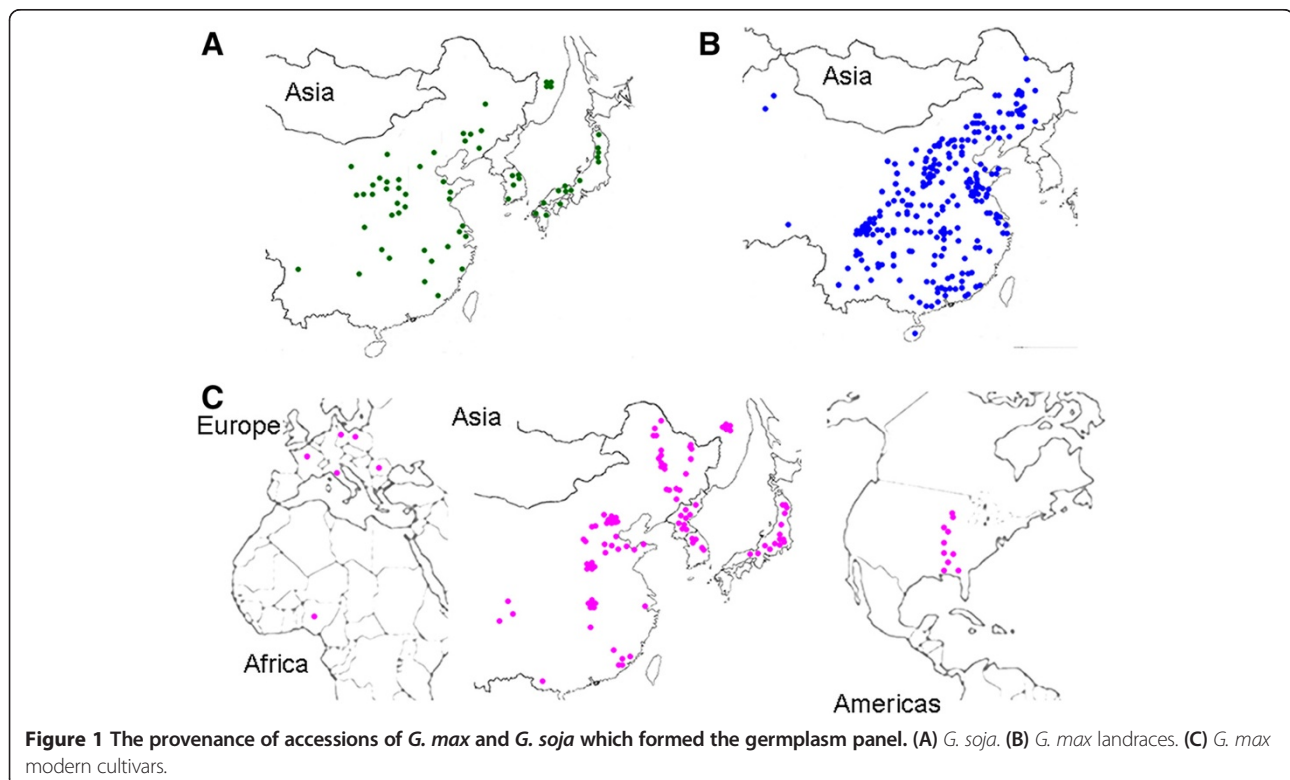
### Plant material

The germplasm panel comprised 65 accessions of *G. soja* and 353 of *G. max*. The provenance of the former included locations within the proposed area where soybean was domesticated (Figure 1). The cultivated population comprised 238 landraces and 115 modern cultivars. A diversity analysis of all of the wild accessions, 233 of the landraces and 65 of the modern cultivars has been reported elsewhere [37]. The additional 55 landraces and

modern cultivars originated from 12 countries and were included to broaden the level of geographic representation (Additional file 1).

### Data acquisition

The allelic constitution of 363 of the 418 accessions at 554 SNP loci has been published previously [37], and these were supplemented by equivalent data for 552 of the 554 loci with respect to the 55 added accessions (Additional file 2); the data were obtained using the Illumina GoldenGate platform [40]. The GenCall and GenTrain score thresholds were set at, respectively, 80% and 0.6, as described elsewhere [37]. On average, each accession harbored 2.1% missing data (range of 0–13.9%). According to the soybean reference genome (<http://www.phytozome.net>) [41], the 552 SNP loci are dispersed throughout the genome, with 505 (91.5%) residing within genic DNA. About 38% of the genic SNPs lie within coding sequence, and 137 of the alleles at these loci are non-synonymous. Testa color scores for the Chinese germplasm were recovered from the Chinese soybean germplasm catalog and various other sources [42-44], while the remainder were obtained from the Germplasm Resources Information Network (USDA) database (<http://www.ars-grin.gov/npgs/>). Testa color was considered as a qualitative trait, with five possible states: yellow, black, brown, cyan and double, following the conventional system [45].



### Analyses of molecular diversity

Summary statistics, including the proportion of heterozygosity in the population, Nei's indices of gene diversity and the frequency of major alleles were computed using Powermarker v3.25 [46]. A phylogenetic tree was generated, based on a neighbor-joining analysis of shared-allele distances [47] implemented in Powermarker v3.25, and this was visualized using a routine within the MEGA v4 software package [48]. Population structure was analyzed using a Bayesian Markov Chain Monte Carlo approach implemented in the software package STRUCTURE v2.1 [49]. The admixture and independent allele frequency models were adopted, testing  $K$  values between 1 and 10. Five runs were performed for each value of  $K$ , without using previous population information. The burn-in time and replication number were consistently set to 100,000.

### Identifying signals of selection

Since the low density of SNP markers (one SNP per 2 Mbp genomic region) limited the utilization of a window-sized approach to detect loci carrying a signature of selection, the summary statistic approach *f*dist2 [50,51] was adopted to identify SNP locus outliers. The focus was on divergence at domestication loci, so pairwise comparisons between wild accessions and landraces, and between wild accessions and modern cultivars were made. A neutral distribution of  $F_{st}$  with 50,000 interactions at the 99% confidence level was assumed, and the significance level was set at 95%.

## Results

### Population structure and genetic differentiation

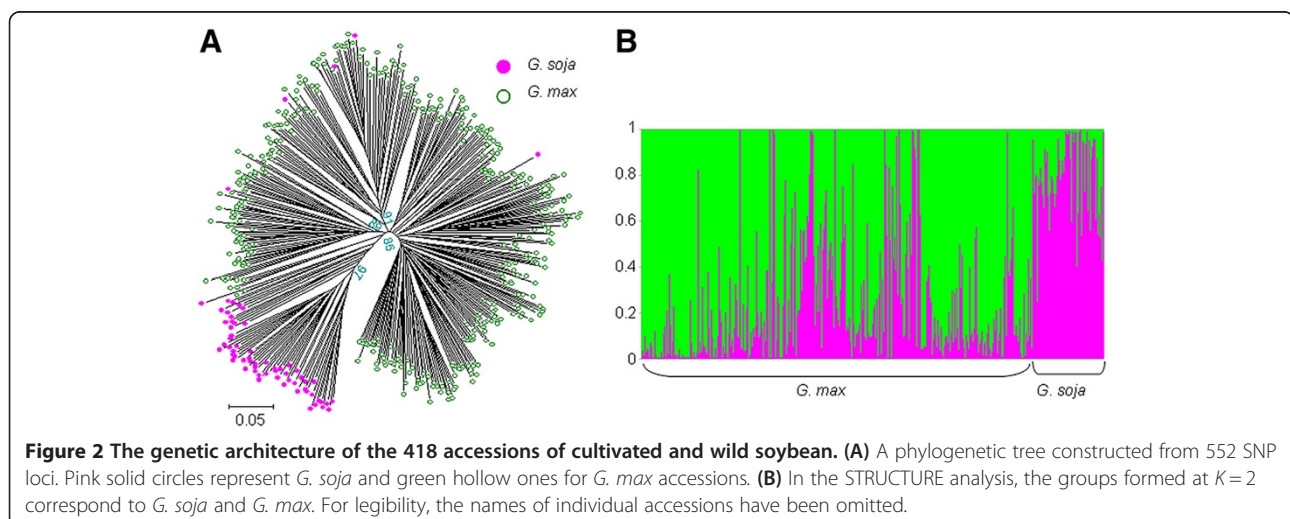
The addition of 55 accessions to the germplasm panel resulted in a slightly higher estimate of the extent of genetic diversity (Additional file 2) compared to that reported previously [39]. The population structure obtained

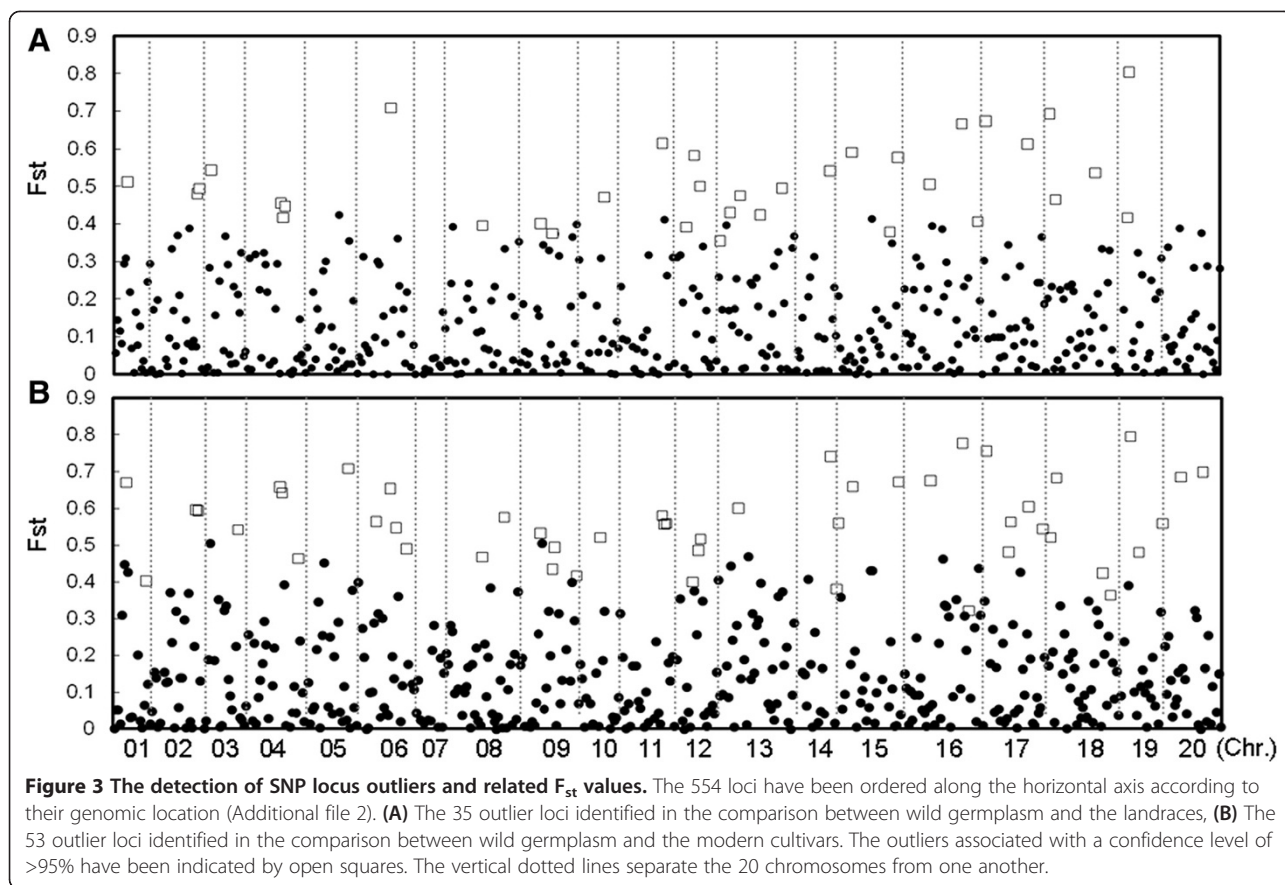
was consistent with a discontinuity between the wild and cultivated clusters ( $K = 2$ ), but there was evidence for introgression from wild to cultivated germplasm (Figure 2A, B). Based on mean pairwise  $F_{st}$  values, the wild, landrace and modern cultivar subsets were judged to be genetically distinct ( $p < 0.05$ ). The extent of the differentiation was greatest between the wild and modern cultivar subsets ( $F_{st} = 0.162$ ), and least between the landraces and modern cultivars ( $F_{st} = 0.047$ ).

### Detection of domestication genes

The presence of signatures of selection during domestication was inferred by comparing the allelic status at the 552 SNP loci between the wild and the landrace subgroups ("W-LC" comparison). In all, 6.3% of the loci were identified as SNP locus outliers at the 95% confidence level (Figure 3A). The  $F_{st}$  values of the SNP locus outliers ranged from 0.36-0.80 and were 3.3-7.4 fold higher than the mean  $F_{st}$  value taken over the full set of loci (0.11). Applying the same test to the comparison between the wild accessions and the modern cultivars ("W-MC") revealed 9.6% of the loci to be SNP locus outliers (Figure 3B). In all, nearly 70% (24/35) of the W-LC outliers were also outliers in W-MC. The major alleles of the wild population in these outlier loci changed to minor alleles in the populations of landrace or modern cultivars (Figure 4A). The strongest signal of selection was associated with the locus BARC-022029-04261, at which the major allele was represented in 82.0% of the wild accessions, but just 1.1% in the landraces and 1.7% in the modern cultivars.

In addition to the 24 shared SNP locus outliers, there were 11 W-LC- and 29 W-MC-specific ones. The frequencies of the major allele at most of the W-MC-specific loci decreased step-wise from wild accessions to landraces to modern cultivars (Figure 4B), indicating





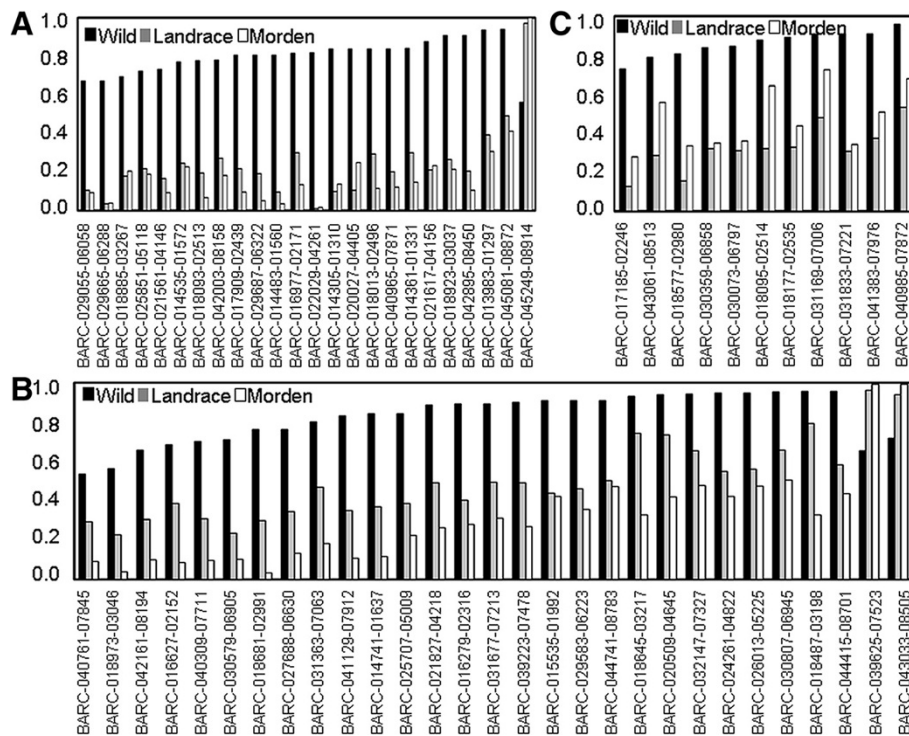
that these loci may be linked to genes/QTL subjected to prolonged selection during post domestication. The 11 W-LC-specific SNP locus outliers may represent domestication genes not subjected to selection during post domestication.

The genomic location of each of the 64 SNP locus outliers was obtained by querying the cv. Williams 82 whole genome sequence (<http://www.phytozome.net>) with the sequences flanking the SNP. This analysis identified regions on 19 of the 20 chromosomes (chromosome 7 had no hits). Chromosome 13 (Gm13) harbored the largest number of SNP locus outliers (six) (Additional file 3). Nine of the outliers mapped to intergenic regions. Of the 55 genic outliers, one was located in 5'-UTR, 25 in 3'-UTRs, 14 in introns and 15 (23.4%) within coding sequences. Among the latter, 13 were non-synonymous. Based on GO analysis, gene function was assignable to 38 of the W-LC + MC sequences harboring an SNP locus outlier [52]. Eleven of the genes, including five of the 13 genes harboring a non-synonymous SNP, were associated with the abiotic stress response (Additional file 3).

#### Association analysis for testa color

Most domesticated soybean materials are yellow-seeded, while black testa types predominate in wild accessions

(Additional file 4). In the present germplasm panel, 47 of the 65 wild accessions were black-seeded, and 237 of the 352 domesticated ones were yellow-seeded. The yellow testa trait was more frequent in the set of modern cultivars than in the set of landraces (Additional file 4). A comparison of SNP genotype with testa color across the full set of 418 accessions identified ten SNP loci potentially linked to the trait (Additional file 5). Of these, eight were W-LC + MC SNP locus outliers and two were W-MC-specific outliers. An analysis of the distribution of testa color and SNP locus outlier allele within the three populations (wild accessions, landraces and modern cultivars) is given in Additional file 6. Allele frequencies at nine of the ten loci (the exception was BARC-045249-08914) were correlated with testa color in the wild accession and modern cultivar populations, but the correlation was only retained for two of the loci (BARC-018681-02991 and BARC-018093-02513) when all three populations were considered (Figure 5A). With respect to BARC-018681-02991, 85.1% of the black testa wild accessions and 61.1% of the black testa landraces harbored the A allele, while 81.5% of the yellow testa landraces and 98.0% of the yellow testa modern cultivars carried the G allele. Similarly at BARC-018093-02513, 85.1% of the black testa wild accessions and 58.3% of the

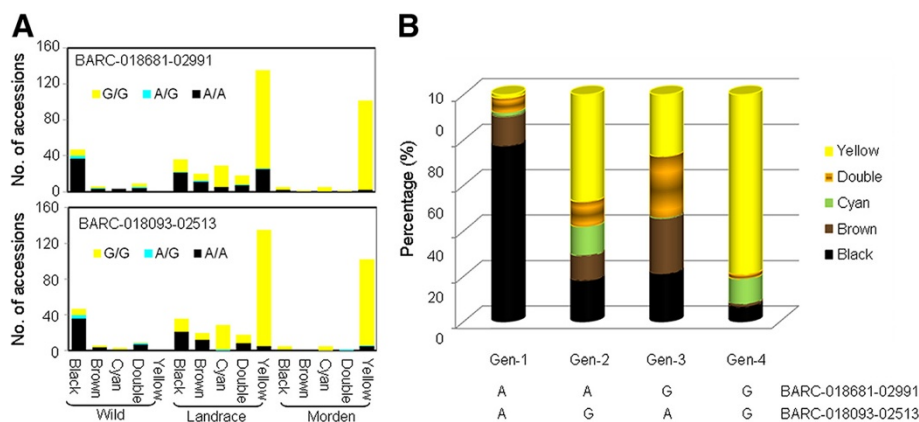


**Figure 4 Allele frequencies at SNP locus outliers.** (A) The 24 outliers in common between wild accessions vs landraces and wild accessions vs modern cultivars. (B) The 29 outliers specific to the wild accessions vs modern cultivars comparison. (C) The 11 outliers specific to the wild accessions vs landraces comparison.

black testa landraces harbored the A allele, while 96.3% of yellow testa landraces and 95.1% of the yellow testa modern cultivars carried the G allele. Four genotype combinations were recognized: Gen-1 (BARC-018681-02991 A, BARC-018093-02513 A), Gen-2 (AG), Gen-3 (GA), and Gen-4 (GG). The association between genotype combination and testa color was somewhat stronger than those based on a single locus: 77.0% of Gen-1 accessions were black seeded, while 79.3% of Gen-4 ones were yellow seeded (Figure 5B).

### Discussion

The strong selection pressure applied particularly during crop domestication and later subsequent genetic improvement has greatly narrowed the genetic base of cultivated types [23,53]. The current analysis identified 64 SNP locus outliers at which there was a significant difference ( $P < 0.05$ ) in diversity between wild and cultivated soybean populations, but failed to establish any clear distinction between landraces and modern cultivars, consistent with the conclusion drawn in previous



**Figure 5 SNP locus outliers associated with testa color.** (A) The relationship between SNP genotype and testa color in wild accessions, the landraces and the modern cultivars. (B) The frequency of the four haplotype classes based on SNPs within BARC-018681-02991 and BARC-018093-02513.

diversity studies that wild germplasm has become more strongly differentiated from landraces than landraces have become from modern cultivars [19,37,38]. A pedigree analysis has established that landraces have provided more than 76% of the nuclear genome carried by 1,300 Chinese modern cultivars released over the period 1923–2005 [54].

Even though selection signals have been identified in a number of genes, it is uncertain whether they are in reality identifying the presence of a domestication-associated genetic bottleneck [55] as opposed to reflecting the long term outcome of genetic improvement [56]. SSR markers in the vicinity of QTL underlying traits of agronomic importance tend to show a stronger level of genetic differentiation between wild and cultivated types than those unlinked to a known QTL [57]. When the location of the present SNP set was aligned with domestication-related QTL, it was established that six SNP locus outliers were linked to a domestication QTL, controlling the traits such as twining habit, maturity time, flower color, seed weight, protein content and resistance to soybean cyst nematode [58–62] (Additional file 3). In addition, eight of the SNP locus outliers are located around 1 Mbp distant from a QTL mapped in a population bred from a cross between a wild accession and a cultivated line [59–62]. Thus, it is likely that several of the SNP locus outliers identified here will have contributed to the phenotypic differentiation between wild and cultivated soybeans. SNPs BARC-025897-05144, BARC-031461-07098 and BARC-022043-04271 used in this study are located around 1 Mbp distant from an isoflavone synthase (IFS) gene (*IFS2*, *Glyma13g24200*), which controls isoflavone accumulation and is most expressed in the developing seed in soybean [63–65]. As isoflavone was not subject to selection during domestication, we used these SNPs to evaluate whether there exists a big change to detect false positive outliers. We observed for none of the relevant SNPs significant outliers, which suggests that our study is only marginally afflicted with an inflated rate of false-positives.

Testa color in soybean is controlled by five genes, namely *I*, *T*, *W1*, *R* and *O* [66]. A screen of 170 cultivated and 102 wild accessions based on sequence variation within the testa color-associated genes encoding flavonoid 3'-hydroxylase (*F3'H*) and flavonoid 3',5'-hydroxylase (*F3'5'H*) has shown that the joint allele constitution was more predictive of testa color than was the allelic state at either one of the two genes on its own [66]. Here, two SNP locus outliers (BARC-018681-02991 and BARC-018093-02513) were associated with testa color. One of the resulting four genotype combinations (Gen-4) was carried by 79.3% of the yellow testa accessions, a slightly lower proportion than was associated with a differently constituted haplotype [66]; at the same time, 77.4%

of the black-seeded accessions carried Gen-1, a rather higher proportion than was recorded for the differently constituted haplotype [66].

Some of the SNP locus outliers represent potential markers for other aspects of morphological differentiation between *G. max* and *G. soja*. Cultivated soybean plants are shorter and more compact than wild soybean plants, characteristics which better fit the requirements of modern soybean production systems. One of the outliers (BARC-040965-07871) mapped within the 3' UTR of *Glyma15g41130*, a gene which encodes a SAUR-like auxin-responsive protein family, and which is linked to a QTL controlling plant height [67]. Members of this gene family have been associated with the determination of flowering time and the regulation of growth and plant architecture [20].

## Conclusions

Genetic variation is the *sine qua non* for crop improvement. The domestication of soybean and the subsequent prolonged period of selection have resulted in a major loss in its genetic diversity. An overly narrow genetic base compromises the potential for achieving continuing gains from selection, underlining the importance of germplasm conservation, particularly of wild forms. At the same time, the identification of which genes were involved in domestication is required to recognize novel genes/alleles likely to contribute to soybean improvement. The SNP locus outliers identified here should not only aid in elaborating rational strategies for the conservation of wild germplasm [68], but may well also provide a source of markers suitable for the application of molecular breeding aimed at broadening the genetic base of soybean.

## Additional files

**Additional file 1:** The provenance of the 55 additional soybean accessions.

**Additional file 2:** Population statistics associated with the 552 SNP loci used to genotype the 418 soybean accessions.

**Additional file 3:** Properties of the 64 SNP locus outliers in soybean.

**Additional file 4:** The distribution of testa color among the wild accessions, landraces and modern cultivars.

**Additional file 5:** The joint distribution of testa color and allelic status at SNP locus outliers associated with soybean testa color.

**Additional file 6:** The joint distribution of testa color and allele at ten SNP locus outliers among the wild accessions, landraces and modern cultivars.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

LJQ, RZC and YHL were responsible for the research concept; YHL and YSM carried out the research and analyzed the data. YHL, JCR, SAJ and LJQ wrote the manuscript. All authors have read and approved the final manuscript.

## Acknowledgments

This research was financed by the State Key Basic Research and Development Plan of China (973) (No. 2010CB125903), an International Science and Technology Cooperation and Exchanges Project (No. 2008DFA30550) and a Science and Technology Project for the Supervisor of Excellent Doctoral Thesis in Beijing (No. YB20108210101). We thank the Chinese National Soybean GeneBank for supplying seed of *G. max* and *G. soja* accessions collected in China, and Dr. R. Nelson (University of Illinois, Urbana, IL) for the gift of seed of the additional accessions.

## Author details

<sup>1</sup>The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/Key Lab of Germplasm Utilization (MOA), Institute of Crop Science, Chinese Academy of Agricultural Sciences, 100081 Beijing, P.R. China. <sup>2</sup>Department of Cytogenetics and Genome Analysis, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany. <sup>3</sup>Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA. <sup>4</sup>Soybean Research Institute, Heilongjiang Academy of Agricultural Sciences, 150086 Harbin, China.

Received: 9 March 2014 Accepted: 18 September 2014

Published online: 26 September 2014

## References

1. Doebley J, Gaut B, Smith B: **The molecular genetics of crop domestication.** *Cell* 2006, **127**(7):1309–1321.
2. Wright S, Bi I, Schroeder S, Yamasaki M, Doebley J, McMullen M, Gaut B: **The effects of artificial selection on the maize genome.** *Science* 2005, **308**(5726):1310–1314.
3. Ross-Ibarra J, Morrell PL, Gaut BS: **Plant domestication, a unique opportunity to identify the genetic basis of adaptation.** *Proc Natl Acad Sci U S A* 2007, **104**(Suppl 1):8641–8648.
4. Izawa T: **Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice.** *J Exp Bot* 2007, **58**(12):3091–3097.
5. Jin J, Huang W, Gao J, Yang J, Shi M, Zhu M, Luo D, Lin H: **Genetic control of rice plant architecture under domestication.** *Nat Genet* 2008, **40**(11):1365–1369.
6. Konishi S, Izawa T, Lin S, Ebana K, Fukuta Y, Sasaki T, Yano M: **An SNP caused loss of seed shattering during rice domestication.** *Science* 2006, **312**(5778):1392–1396.
7. Li C, Zhou A, Sang T: **Rice domestication by reducing shattering.** *Science* 2006, **311**(5769):1936–1939.
8. Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, Fu Y, Cai H, Wang X, Xie D, Sun C: **Control of a key transition from prostrate to erect growth in rice domestication.** *Nat Genet* 2008, **200**:1361–1364.
9. Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H, Ma H, Zhang G, He Z: **Control of rice grain-filling and yield by a gene with a potential signature of domestication.** *Nature* 2008, **200**:1370–1374.
10. Doebley J, Stec A, Gustus C: ***Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance.** *Genetics* 1995, **141**(1):333–346.
11. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, Bombliès-Yant K, Lukens L, Doebley J: **The origin of the naked grains of maize.** *Nature* 2005, **436**(7051):714–719.
12. Simons K, Fellers J, Trick H, Zhang Z, Tai Y, Gill B, Faris J: **Molecular characterization of the major wheat domestication gene *Q*.** *Genetics* 2006, **172**(1):547–555.
13. Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J: **A *NAC* gene regulating senescence improves grain protein, zinc, and iron content in wheat.** *Science* 2006, **314**(5803):1298–1301.
14. Clark R, Linton E, Messing J, Doebley J: **Pattern of diversity in the genomic region near the maize domestication gene *tb1*.** *Proc Natl Acad Sci U S A* 2004, **101**(3):700–707.
15. Olsen K, Caicedo A, Polato N, McClung A, McCouch S, Purugganan M: **Selection under domestication: evidence for a sweep in the rice *Waxy* genomic region.** *Genetics* 2006, **173**(2):975–983.
16. Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J: **Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication.** *Proc Natl Acad Sci U S A* 2002, **99**(15):9650–9655.
17. Yamasaki M, Tenaillon M, Vroh Bi I, Schroeder S, Sanchez-Villeda H, Doebley J, Gaut B, McMullen M: **A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement.** *Plant Cell* 2005, **17**(11):2859–2872.
18. Kane NC, Rieseberg LH: **Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*.** *Genetics* 2007, **175**(4):1823–1834.
19. Li YH, Zhao SC, Ma JX, Li D, Yan L, Li J, Qi XT, Guo XS, Zhang L, He WM, Chang RZ, Liang QS, Guo Y, Ye C, Wang XB, Tao Y, Guan RX, Wang JY, Liu YL, Jin LG, Zhang XQ, Liu ZX, Zhang LJ, Chen J, Wang KJ, Nielsen R, Li RQ, Chen PY, Li WB, Reif J, et al: **Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing.** *BMC Genomics* 2013, **14**:579.
20. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W: **Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes.** *Nat Biotechnol* 2012, **30**:105–111.
21. Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B: **Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection.** *Nat Genet* 2010, **42**(12):1053–1059.
22. Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, Cheng T, Jiang T, Becquet C, Xu X, Liu C, Zha X, Fan W, Lin Y, Shen Y, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu H, Yu C, Zhang G, Li J, Cao J, Liu S: **Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*).** *Science* 2009, **326**(5951):433–436.
23. Hufford MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppeler SM, Lai JS, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang GY, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J: **Comparative population genomics of maize domestication and improvement.** *Nat Genet* 2012, **44**(7):808–811.
24. Chapman MA, Burke JM: **Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower.** *Theor Appl Genet* 2012, **125**(5):897–907.
25. Blackman BK, Rasmussen DA, Strasburg JL, Raduski AR, Burke JM, Knapp SJ, Michaels SD, Rieseberg LH: **Contributions of flowering time genes to sunflower domestication and improvement.** *Genetics* 2011, **187**(1):271–287.
26. Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J: **Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce.** *Mol Ecol* 2008, **17**(16):3599–3613.
27. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, Bentley DR, Chakravarti A: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**(7319):1061–1073.
28. Thornton KR, Jensen JD: **Controlling the false-positive rate in multilocus genome scans for selection.** *Genetics* 2007, **175**(2):737–750.
29. Dykhuizen DE: **Experimental studies of natural selection in bacteria.** *Ann Rev Ecol Syst* 1990, **21**:373–398.
30. Barton NH: **Genetic hitchhiking.** *Philos Trans R Soc Lond B Biol Sci* 2000, **355**(1403):1553–1562.
31. Helyar SJ, Hemmer-Ansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, Nielsen E: **Application of SNPs for population genomics of nonmodel organisms: new opportunities and challenges.** *Mol Ecol Resour* 2011, **11**(s1):123–136.
32. Bouret V, Kent MP, Primmer CR, Vasemañgi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L, Lien S: **SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*).** *Mol Ecol* 2013, **22**(3):532–551.
33. Chapman M, Pashley C, Wenzler J, Hvala J, Tang S, Knapp S, Burke J: **A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*).** *Plant Cell* 2008, **20**(11):2931–2945.
34. Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley J: **An analysis of genetic diversity across the maize genome using microsatellites.** *Genetics* 2005, **169**(3):1617–1630.

35. Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T: **Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica***. *Heredity* 2012, **109**:349–360.
36. Hymowitz T, Newell C: **Taxonomy, speciation, domestication, dissemination, germplasm resources, and variation in the genus *Glycine***. In *Advances in Legume Science*. Edited by Summerfield RJ, Bunting AH. Kew, Richmond, Surrey: Royal Botanical Gardens; 1980:251–264.
37. Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ: **Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci**. *New Phytol* 2010, **188**:242–253.
38. Hyten D, Song Q, Zhu Y, Choi I, Nelson R, Costa J, Specht J, Shoemaker R, Cregan P: **Impacts of genetic bottlenecks on soybean genome diversity**. *Proc Natl Acad Sci U S A* 2006, **103**(45):16666–16671.
39. Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y: **A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences**. *Ann Bot* 2010, **106**(3):505–514.
40. Shen R, Fan J, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Wickham Garcia E, McBride C: **High-throughput SNP genotyping on universal bead arrays**. *Mutat Res* 2005, **573**(1–2):70–82.
41. Schmutz J, Cannon S, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May G, Yu Y, Sakurai T, Umezawa T, Bhattacharyya M, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L: **Genome sequence of the paleopolyploid soybean (*Glycine max* (L.) Merr.)**. *Nature* 2010, **463**:178–183.
42. Chang RZ, Sun JY: *Catalogues of Chinese Soybean Germplasm and Resources: Continuation I*. Beijing: China Agricultural Press; 1991.
43. Chang RZ, Sun JY, Qiu LJ, Chen YW: *Catalogues of Chinese Soybean Germplasm and Resources: Continuation II*. Beijing: China Agricultural Press; 1996.
44. Wang GX: *Catalogues of Chinese Soybean Germplasm and Resources*. Beijing: China Agricultural Press; 1982.
45. Qiu LJ, Chang RZ, Liu ZX, Guan RX, Li YH: *Descriptors and Data Standard for Soybean (*Glycine* spp.)*. Beijing: Chinese Agricultural Press; 2006.
46. Liu K, Muse SV: **PowerMarker: an integrated analysis environment for genetic marker analysis**. *Bioinformatics* 2005, **21**(9):2128–2129.
47. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**(4):406–425.
48. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**(8):1596–1599.
49. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies**. *Genetics* 2003, **164**(4):1567–1587.
50. Beaumont MA, Balding DJ: **Identifying adaptive genetic divergence among populations from genome scans**. *Mol Ecol* 2004, **13**(4):969–980.
51. Beaumont MA, Nichols RA: **Evaluating loci for use in the genetic analysis of population structure**. *Proc Biol Sci* 1996, **263**(1377):1619–1626.
52. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C: **The Gene Ontology (GO) database and informatics resource**. *Nucleic Acids Res* 2004, **32**(Database issue):D258–D261.
53. Tanksley SD, McCouch SR: **Seed banks and molecular maps: unlocking genetic potential from the wild**. *Science* 1997, **277**(5329):1063–1066.
54. Xiong DJ, Zhao TJ, Gai JY: **Parental analysis of soybean cultivars released in China**. *Sci Agri Sin* 2008, **41**(9):2589–2598.
55. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS: **Investigation of the bottleneck leading to the domestication of maize**. *Proc Natl Acad Sci U S A* 1998, **95**(8):4441–4446.
56. Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS: **Selection versus demography: a multilocus investigation of the domestication process in maize**. *Mol Biol Evol* 2004, **21**(7):1214–1225.
57. Jun TH, Van K, Kim MY, Kwak M, Lee SH: **Uncovering signatures of selection in the soybean genome using SSR diversity near QTLs of agronomic importance**. *Genes Genome* 2011, **33**(4):391–397.
58. Bolon Y-T, Joseph B, Cannon SB, Graham MA, Diers BW, Farmer AD, May GD, Muehlbauer GJ, Specht JE, Tu ZJ, Weeks N, Xu W, Shoemaker R, Vance C: **Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean**. *BMC Plant Biol* 2010, **10**(1):41.
59. Kabelka EA, Carlson SR, Diers BW: **Localization of two loci that confer resistance to soybean cyst nematode from PI 468916**. *Crop Sci* 2005, **45**(6):2473–2481.
60. Liu B, Fujita T, Yan ZH, Sakamoto S, Xu D, Abe J: **QTL mapping of domestication-related traits in soybean (*Glycine max*)**. *Ann Bot* 2007, **100**(5):1027–1038.
61. Wang D, Graef GL, Procopiuk AM, Diers BW: **Identification of putative QTL that underlie yield in interspecific soybean backcross populations**. *Theor Appl Genet* 2004, **108**(3):458–467.
62. Yang K, Jeong N, Moon JK, Lee YH, Lee SH, Kim HM, Hwang CH, Back K, Palmer RG, Jeong SC: **Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean**. *J Hered* 2010, **101**(6):757–768.
63. Dhaubhadel S, McCarvey BD, Williams R, Gijzen M: **Isoflavonoid biosynthesis and accumulation in developing soybean seeds**. *Plant Mol Biol* 2003, **53**(6):733–743.
64. Gutierrez-Gonzalez JJ, Wu X, Gillman JD, Lee J-D, Zhong R, Yu O, Shannon G, Ellersieck M, Nguyen HT, Slepser DA: **Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds**. *BMC Plant Biol* 2010, **10**(1):105.
65. Gutierrez-Gonzalez JJ, Guttikonda SK, Tran L-SP, Aldrich DL, Zhong R, Yu O, Nguyen HT, Slepser DA: **Differential expression of isoflavone biosynthetic genes in soybean during water deficits**. *Plant Cell Physiol* 2010, **51**(6):936–948.
66. Guo Y, Qiu LJ: **Allele-specific marker development and selection efficiencies for both flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase genes in soybean subgenus *soja***. *Theor Appl Genet* 2013, **126**(6):1445–1455.
67. Kassem M, Shultz J, Meksem K, Cho Y, Wood A, Iqbal M, Lightfoot D: **An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits**. *Theor Appl Genet* 2006, **113**(6):1015–1026.
68. Freamo H, O'Eilly P, Berg PR, Lien S, Boulding EG: **Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs**. *Mol Ecol Resour* 2011, **11**(s1):254–267.

doi:10.1186/s12870-014-0251-1

Cite this article as: Li et al.: Detecting SNPs underlying domestication-related traits in soybean. *BMC Plant Biology* 2014 **14**:251.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

