View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Springer - Publisher Connector

Theor Appl Genet (2016) 129:1725–1738
DOI 10.1007/s00122-016-2735-x

ORIGINAL ARTICLE

Fast neutron-induced structural rearrangements at a soybean *NAP1* locus result in *gnarled* trichomes

Benjamin W. Campbell[1] · Anna N. Hofstad[1] · Suma Sreekanta[1] · Fengli Fu[1] · Thomas J. Y. Kono[1] · Jamie A. O'Rourke[2] · Carroll P. Vance[1] · Gary J. Muehlbauer[1,3] · Robert M. Stupar[1]

Received: 29 February 2016 / Accepted: 28 May 2016 / Published online: 9 June 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Key message **Three adjacent and distinct sequence rearrangements were identified at a NAP1 locus in a soybean mutant. Genetic dissection and validation revealed the function of this gene in soybean trichome development.**

Abstract A soybean (*Glycine max* (L.) Merr.) *gnarled* trichome mutant, exhibiting stunted trichomes compared to wild-type, was identified in a fast neutron mutant population. Genetic mapping using whole genome sequencing-based bulked segregant analysis identified a 26.6 megabase interval on chromosome 20 that co-segregated with the phenotype. Comparative genomic hybridization analysis of the mutant indicated that the chromosome 20 interval included a small structural variant within the coding region of a soybean ortholog (Glyma.20G019300) of Arabidopsis *Nck-Associated Protein 1* (*NAP1*), a regulator of actin nucleation during trichome morphogenesis. Sequence analysis of the candidate allele revealed multiple rearrangements within the coding region, including two deletions (approximately 1–2 kb each), a translocation, and an inversion. Further analyses revealed that the mutant allele perfectly co-segregated with the phenotype, and a wild-type soybean *NAP1* transgene functionally complemented an Arabidopsis *nap1* mutant. In addition, mapping and exon sequencing of *NAP1* in a spontaneous soybean *gnarled* trichome mutant (T31) identified a frame shift mutation resulting in a truncation of the coding region. These data indicate that the soybean *NAP1* gene is essential for proper trichome development and show the utility of the soybean fast neutron population for forward genetic approaches for identifying genes.

Communicated by D. A Lightfoot.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-016-2735-x) contains supplementary material, which is available to authorized users.

✉ Robert M. Stupar
stup0004@umn.edu

[1] Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108, USA

[2] USDA-ARS, Corn Insects and Crop Genetics Research, Iowa State University, Ames, IA 50011, USA

[3] Department of Plant Biology, University of Minnesota, Saint Paul, MN 55108, USA

Introduction

The plant trichome is an elongated epidermal cell that undergoes cell enlargement away from the plant surface. Trichomes develop on the surfaces of leaves, stems, petioles, and some reproductive organs (Vermeer and Peterson 1979; Nyman 1993). Trichomes perform many biological functions, including plant defense against insect predation, where they can affect larval growth and insect preferences (Levin 1973; Robbins et al. 1979; Hulburt et al. 2004), and adaptation of desert plants to drought conditions by increasing the leaf reflectance, which helps to moderate leaf temperatures without requiring increased transpiration (Ehleringer and Mooney 1978). Economically, cotton seed trichomes compose the fibers that make cotton a valuable commodity. Thus, an understanding of the genetic control of trichome development impacts both agricultural and economic productivity.

There has been considerable interest in understanding the genetic controls that underlie trichome formation, particularly in model plant systems. In Arabidopsis, several

genes required for trichome formation have been characterized. *GLABRA1* (*GL1*) and *TRANSPARENT TESTA GLABRA1* (*TTG1*) are important for trichome initiation (Oppenheimer et al. 1991; Walker et al. 1999). *GLABRA2* (*GL2*) controls normal trichome morphogenesis (Rerie et al. 1994), and *TRIPTYCHON* (*TRY*) and *CAPRICE* control the spacing pattern of trichomes across the leaf surface (Schellmann et al. 2002). *ZWICHEL* (*ZWI*), *CONSTITUTIVE PATHOGENE RESPONSE5 (CPR5)*, *TRANSPARENT TESTA GLABRA2* (*TTG2*), and *KAKTUS (KAK)* affect trichome branching (Oppenheimer et al. 1997; Kirik et al. 2001; Johnson 2002; El Refy et al. 2003; Downes et al. 2003), and mutations in *CROOKED* (*CRK*) and *GNARLED* (*GRL*) cause distorted trichomes (Mathur et al. 2003; Deeks et al. 2004; El-Assal et al. 2004). However, to our knowledge, the only previously isolated soybean gene affecting any trichome trait is the T allele, which has pleiotropic effects on the color of the trichome, hilum, and seed-coat (Woodworth 1921; Zabala and Vodkin 2003).

Several soybean mutagenesis platforms have been developed for functional characterization of soybean genes (reviewed by Campbell and Stupar 2016). These platforms include chemical mutagenesis (Cooper et al. 2008; Gillman et al. 2014), transposon tagging (Palmer et al. 2008a, b; Mathieu et al. 2009; Hancock et al. 2011; Cui et al. 2013; Raval et al. 2013), and irradiation mutagenesis (Men et al. 2002; Bolon et al. 2011; Gillman et al. 2014). Chemical mutagenesis causes single nucleotide polymorphisms (SNPs), and transposon tagging results in insertion mutants (Cooper et al. 2008; Palmer et al. 2008a, b; Mathieu et al. 2009; Hancock et al. 2011; Cui et al. 2013; Raval et al. 2013; Gillman et al. 2014). In contrast to the limited mutation types caused by chemical and transposon mutagenesis, irradiation mutagenesis has been reported to induce a wide variety of mutation types, including structural rearrangements (e.g. deletions, duplications, translocations, and inversions) of varying sizes, and SNPs (Bolon et al. 2011, 2014; Belfield et al. 2012).

Previously, forward genetic approaches to identify causative mutations induced through mutagenesis were often slow and typically required initial coarse mapping followed by one or more rounds of fine-mapping to positively identify a causative variant. Genetic methods, such as bulked segregant analysis (BSA) (Michelmore et al. 1991), have been developed to facilitate the coarse mapping of qualitative traits; however, fine-mapping based on the phenotyping and genotyping of individuals from large populations is arduous. The advent of new molecular technologies has rapidly decreased the time required to physically and genetically map potential causative polymorphisms. For example, irradiation-induced mutations can be detected using array Comparative Genomic Hybridization (aCGH) or genome resequencing. The aCGH approach is useful for quickly identifying sufficiently large (≥2 kb) deletions and duplications, as has been demonstrated in a variety of plant species such as *Arabidopsis thaliana* (Gong et al. 2004), clementine (*Citrus clementina* Hort. Ex Tan. Cv. Clemenules) (Ríos et al. 2008), rice (*Oryza sativa*) (Bart et al. 2010), and soybean (Bolon et al. 2011, 2014). Furthermore, sequencing-based genotyping of BSA samples can be used to decrease the time required to identify chromosomal loci that co-segregate with qualitative traits. This concept has been demonstrated using a range of different sequencing approaches (including sequencing of RNA, whole-genomes, and exomes) and has been demonstrated in several plant species, such as tomato *Solanum lycopersicum* (Illa-Berenguer et al. 2015), maize (Liu et al. 2012; Haase et al. 2015), rice (Takagi et al. 2013; Yang et al. 2013), barley (Mascher et al. 2014), and Arabidopsis (James et al. 2013; Zhang et al. 2014).

We were intrigued to test whether a combination of aCGH and whole genome sequencing-based bulked segregant analysis (WGS-BSA) could facilitate the rapid cloning of the causative gene(s) from an irradiated soybean mutant. In this study, we report the identification of a causative mutation underlying a previously identified soybean fast neutron mutant that exhibits *gnarled* trichomes. A combination of aCGH and WGS-BSA was used to identify a *Nck-Associated Protein 1* (*NAP1*) candidate gene for this trait, and subsequent genetic and molecular analyses confirmed the essential function of this gene in trichome development.

Materials and methods

Populations and phenotyping

A *gnarled* trichome mutant, R55C01 (Soybase.org mutant FN0175501), was identified in a soybean fast neutron mutant population developed at the University of Minnesota using the soybean line 'M92-220' which was derived from the variety 'MN1302' (Orf and Denny 2004; Bolon et al. 2011). This mutant was crossed to the wild-type accession 'Noir 1' [subline Noir 1-SGC-01 (McHale et al. 2012)] to generate a segregating mapping population. The F_1 hybrid and the subsequent segregating F_2 and F_3 individuals were grown in the greenhouse and visually phenotyped.

Detection of structural variants using comparative genomic hybridization microarrays

The aCGH array was designed using the first version of the soybean reference cv. 'Williams 82' genome sequence Glyma.Wm82.a1.v1.1 (Bernard and Cremeens 1988;

Schmutz et al. 2010). The array was composed of unique sequence probes (50–60mers) spaced across the genome at an interval typically ranging from 0.5 to 1.1 kb. The methods used for the labeling and the aCGH analysis were conducted according to the methods described in previous studies (Haun et al. 2011; Bolon et al. 2011, 2014; Anderson et al. 2014), using 'M92-220' as the reference sample for the array. Genomic DNA was isolated from leaf tissue using the Qiagen DNeasy kit, and 500 ng of genome DNA from each line was used for the labeling reaction. The mutant DNA was labeled using Cy3 dye and 'M92-220' reference sample was labeled with Cy5 dye. The labels were incorporated using the 3′–5′ exo-Klenow fragment from DNA polymerase I. The labeled DNA was quantified and hybridized for 72 h at 42° C to the 700k feature NimbleGen aCGH array. The methods used for array scanning and data analyses have been previously described (Bolon et al. 2011).

Sequencing of R55C01, 'Noir 1', and F_2 bulks

Fifty F_2 individuals with wild-type trichomes and 50 F_2 individuals with mutant trichomes were chosen from the 'Noir 1' × R55C01 population to compose the two mapping bulks. Genomic DNA of both bulks, the mutant R55C01, and the wild-type 'Noir 1' individual was extracted from leaf tissue using a Qiagen DNeasy kit. DNA samples were submitted to the University of Minnesota Genomics Center (UMGC) for sequencing on an Illumina HiSeq 2000 producing 101 bp paired-end reads with the goal of achieving an average sequencing coverage of 30×. Scythe (https://github.com/vsbuffalo/scythe) was used to remove adapter sequences from the 3′ ends of reads, with a 5 % prior on contamination rate. Sickle (https://github.com/najoshi/sickle) was then used to remove bases with a Phred quality below 20. Cleaned reads were aligned to the updated soybean reference genome assembly Glyma.Wm82.a2.v1 (Song et al. 2016) using BWA-MEM version 0.7.5a (Li 2013). Mismatch penalties and alignment reporting parameters were adjusted to report alignments with approximately 1 % mismatch from the reference. Alignments were sorted, de-duplicated, and labeled with read groups using Picard Tools version 1.107 (http://broadinstitute.github.io/picard/). Alignments were then re-aligned around potential insertion/deletion polymorphisms using the Genome Analysis Tool Kit (GATK) version 3.1-1. To minimize computational time while obtaining genome-wide coverage, we called SNP variants at the genomic positions previously identified for the SoySNP50K genotyping platform (Song et al. 2013) rather than all possible SNPs. Variants were called using the GATK UnifiedGenotyper (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). A custom Python script called VCF_MAF.py (available at https://github.com/TomJKono/Misc_Utils) was then used to estimate allele frequencies and calculate read depths in each bulk.

Whole genome sequencing-based bulked segregant analysis (WGS-BSA)

The bulk allele frequencies were initially calculated as the Glyma.Wm82.a2.v1 reference or the alternate state using custom PERL script for the SoySNP50K positions (Song et al. 2013, 2016). To utilize these data for mapping, the allele frequencies were converted to allele frequencies of the wild-type 'Noir 1' parent at each position in both bulks based on the alternate or reference allele state of 'Noir 1' at each SNP position from the 'Noir 1' sequence data. Non-polymorphic SNPs (having allele frequencies in both bulks of greater than or equal to 0.9 or less than or equal to 0.1), SNPs with missing data in either bulk, and SNPs with read counts less than ten in either bulk were removed from the dataset. The allele frequencies were plotted and graphically analyzed for spreads in allele frequency.

Genotyping and phenotyping of segregating F_3 individuals

R55C01, 'Noir 1', and F_3 families derived from different F_2 individuals were planted in the greenhouse and visually phenotyped for the presence of wild-type or *gnarled* trichomes. Genomic DNA was extracted from a single F_3 individual from each homozygous family and from one mutant and one wild-type plant from each segregating F_3 family. The genomic sequence of R55C01 was utilized to design PCR primers (Supplemental Table 1) that amplified distinct amplicons for the wild-type and mutant alleles.

Validation of chromosome rearrangements in *GmNAP1*

PCR was used to validate the chromosome rearrangements identified by whole genome sequencing at the locus of the candidate gene, *Glycine max NAP1* (*GmNAP1*). PCR primers were designed using the genomic sequence of R55C01 (Supplemental Table 2).

RNA sequencing of R55C01 and 'M92-220' root, seed, and leaf tissue

R55C01 and 'M92-220' seeds were imbibed in sterile water for 48 h before being transferred to pots containing quartz sand, with four seedlings planted per pot and later thinned to one plant. Plants were placed in a growth chamber at 28 °C, oscillating between 16 h of light and 8 h of dark. Each pot was watered daily with 500 mL of nutrient solution (O'Rourke et al. 2014). At the V2 stage, emerging

trifoliates and total root tissues were harvested and immediately immersed in liquid nitrogen from three biological replicates. Developing seeds were harvested at seed stage 0 (10 mg) from three additional biological replicates. RNA was extracted from leaf, root, and developing seeds using the Qiagen RNeasy kit and submitted to the UMGC, where samples were sequenced on an Illumina HiSeq2000. Illumina library preparation, clustering and sequencing reagents were used throughout the process, following the manufacturer's recommendations. Samples were sequenced as 50 bp paired-end reads with an insert size of 200 bp. On average, each sample generated 26 million paired-end reads. Read quality was confirmed using FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the reference genome version Glyma.Wm82.a1.v1 (Schmutz et al. 2010) using TopHat2 (Kim et al. 2013). Reads mapping to genic regions were identified using a combination of SAMtools and HTseq (Li et al. 2009; Anders et al. 2014). Differential gene expression and exon expression analyses were performed in R (R Development Core Team 2006) using DESeq (Anders and Huber 2010) and EdgeR (Robinson et al. 2010), respectively.

Mapping the *p2* introgression interval

The similarity of the R55C01 trichome mutant phenotype to the phenotypic descriptions and images of the *p2* trichomes of line T31 (PI548159) suggested that the *p2* allele was caused by a mutation affecting the same gene or pathway as the R55C01 mutant (Stewart and Wentz 1926; Bernard and Singh 1969; Singh et al. 1971; Healy et al. 2005). Bernard et al. (1991) backcrossed the *p2* allele into the cv. 'Harosoy' (PI548573) (Weiss and Stevenson 1955) and cv. 'Clark' (PI548533) (Johnson 1958) backgrounds to generate two 'Harosoy' *p2* backcross lines (PI547713 and PI547743) and three 'Clark' *p2* backcross lines (PI547449, PI547565, and PI547566). The SoySNP50K data for the three parents and five *p2* backcross lines (Song et al. 2015) were obtained from SoyBase (http://soybase.org). SNPs that were not polymorphic between mutant line T31 and wild-type lines 'Clark' and 'Harosoy' were removed. The genome was then scanned for SNPs at which T31 and all five *p2* backcross lines shared the same allele.

Sequencing the *GmNAP1* gene in T31

Seed of T31 (PI548159) was obtained from USDA Germplasm Resources Information Network, and T31 genomic DNA was extracted from leaf tissue. The candidate gene for this study, Glyma.20G019300, has a length of 22.5 kb, not including the promoter sequence, and the predicted transcript length is 4.8 kb (Song et al. 2016). PCR primers were designed to amplify the 5′UTR, 3′UTR, and all 23 exons including all splice site junctions (Supplemental Table 3). Reactions were PCR purified using a QIAquick PCR Purification Kit (Qiagen) and sequenced at the UMGC. Sequences were aligned to the sequence of gene model Glyma.20G019300 based on the reference sequence Glyma.Wm82.a2.v1.1 (Song et al. 2016).

Transgenic complementation of Arabidopsis *glr-4* using *GmNAP1*

The *GmNAP1* construct was designed in a pMDC123 backbone and consisted of the native promoter amplified from 'Williams 82' driving the synthesized *GmNAP1* cDNA (Piscataway, NJ) followed by a NOS terminator. The pMDC123 vector used also contained a 35S promoter driving a BAR herbicide resistance gene.

The Arabidopsis *nap1* mutant, *grl-4* (El-Assal et al. 2004), was transformed with the *GmNAP1* construct using the floral dip method (Clough and Bent 1998), and T_1 seeds were planted in a flat containing standard potting mix. Ten and seventeen days after germination the flat was sprayed with a 0.01 % solution of glufosinate and resistant plants were then transferred to individual pots. Leaf tissue was collected from each T_1 individual. PCR primers that amplified across the junction between the promoter and first exon of Glyma.20G019300 were used to test for the presence of *GmNAP1* construct in 'Williams 82', *grl-4*, and 20 T_1 individuals with wild-type trichomes (Supplemental Table 4).

Data availability

The aCGH data for mutant R55C01 (also known as FN0175501) can be found in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/gds) accession number GSM1402716. All DNA and RNA sequence data can be found in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) repository. Whole genome DNA sequence data for R55C01 and M92-220 can be found in SRP036841 (accession numbers SRX467193 and SRX467183, respectively). The RNA-seq data for R55C01 and M92-220 can be found in SRP074365 (accession numbers SRX1742547-SRX1742555 and SRX1742565-SRX1742573, respectively.)

Results

Identification and mapping of the *gnarled* trichome mutant

Several morphological and developmental mutants were discovered during the visual phenotypic screening of the

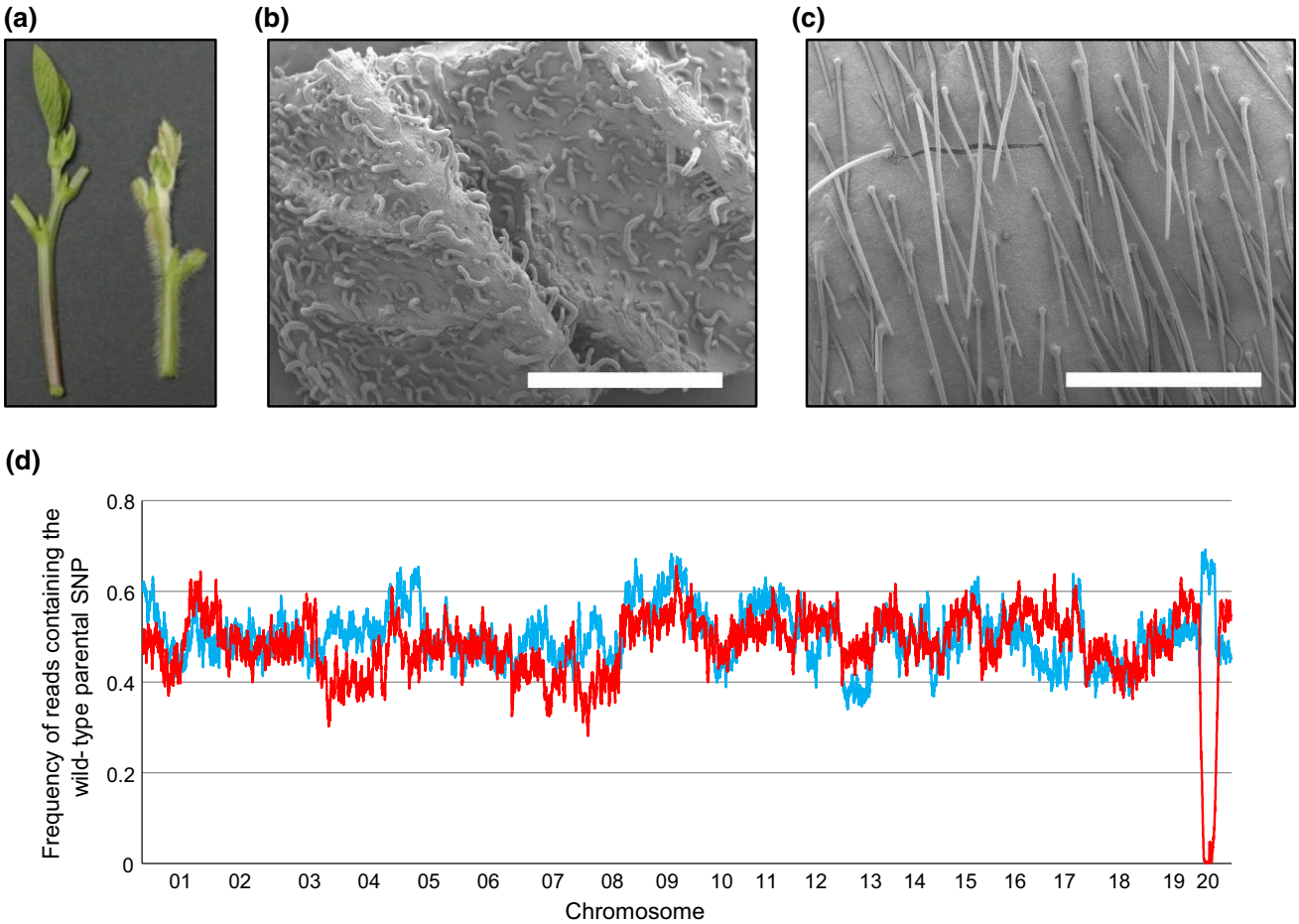

Fig. 1 Phenotype and coarse genetic mapping of the *gnarled* trichome mutant. **a** Phenotypes of mutant (*left*) and wild-type (*right*) shoots. SEM leaf trichome images of the *gnarled* mutant R55C01 **(b)** and the wild-type line 'M92-220' **(c)**. The mutant trichomes are wide, short, flaccid, and lay on the surface of the leaf or stem, as compared to the wild-type trichomes (narrow, long, straight, and project outward from the leaf or stem). *Scale bars* in **b** and **c** are 1 mm. **d** BSA-WGS allele frequencies were calculated for F_2 bulked samples that consisted of 50 mutant (*red lines*) and 50 wild-type (*blue line*) individuals. The allele frequencies were calculated as the proportion of reads containing the wild-type parental SNP (i.e. SNPs that match the wild-type parent 'Noir 1') at each position for over 16,000 polymorphic SoySNP50K positions. The allelic frequencies are shown as the average value across a 21 SNP sliding window. The obvious spread in allele frequencies indicates that the causative locus is located on chromosome 20 (color figure online)

soybean fast neutron population generated at the University of Minnesota (Bolon et al. 2011). Mutant R55C01 (Soybase mutant FN0175501) was identified as a short trichome mutant (SOY:0001804) (Fig. 1a). Scanning electron microscope images of the leaves from wild-type and mutant plants indicated that the mutant has a *gnarled* trichome phenotype (Fig. 1b, c), which is characterized by trichomes that are swollen, twisted and reduced in length (Szymanski et al. 1999; Deeks et al. 2004; El-Assal et al. 2004). The *gnarled* mutant trichomes are shorter, exhibit wide, flaccid shafts (SOY:0001720), lay on the surface of the leaf or stem (SOY:0001977), and have round, blunt tips (SOY:0001722).

aCGH was conducted on an M5 mutant plant to identify potential causative mutations [this plant was assigned the identification number FN0175501.x2.02.01.M5 in Bolon et al. (2011)]. The aCGH results did not identify any duplications but did identify two deletions in the mutant genome: a putative 26 kb deletion on chromosome 5 and a putative ~2 kb deletion on chromosome 20. However, the aCGH method is not sensitive enough to detect some types of rearrangements (inversions and translocations), small deletions, and small duplications, which may underlie the mutant phenotype.

Genetic mapping was conducted to identify the genomic interval co-segregating with the trichome phenotype. The mutant was outcrossed to the wild-type accession 'Noir 1' to generate the mapping population. The F_1 plant had wild-type trichomes, and the F_2 population segregated in a 3:1 wild-type to mutant ratio (144 to 53; Chi-squared p

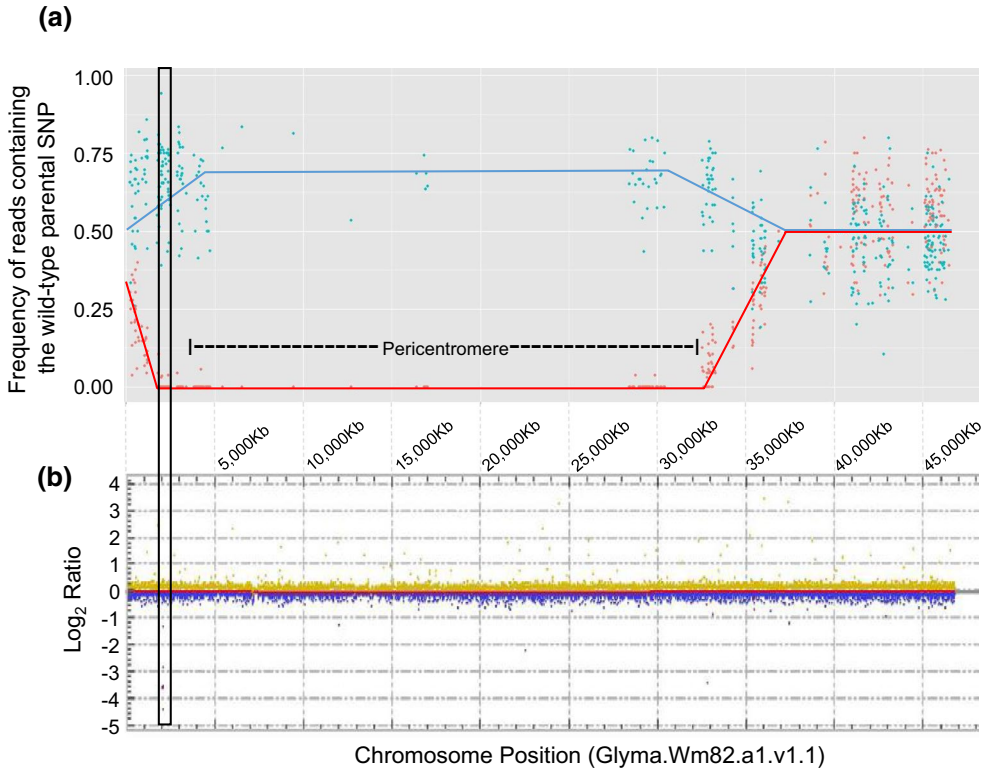

Fig. 2 Genetic mapping of the *gnarled* mutant and physical mapping of the deletion on chromosome 20. Coincidental mapping of the **a** WGS-BSA mapping interval with **b** a deletion detected by array Comparative Genomic Hybridization (aCGH). In **a**, blue data points indicate the 'Noir 1' SNP frequency at each marker position in the wild-type bulk; red data points indicate the 'Noir 1' SNP frequency at each marker position in the mutant bulk. *Blue* and *red lines*, respectively, are drawn in **a** to assist in visualizing the separation in the bulk allele frequencies. A *vertical rectangle* shared between **a** and **b** identifies the chromosome region containing the *GmNAP1* gene. In **b**, probes below the 0.0 $\log_2$ value indicates the absence of mutant DNA (i.e. a putative deleted segment) (color figure online)

value = 0.537 for one locus), indicating that the trichome mutant phenotype was recessive and was caused by a mutation at a single locus.

WGS-BSA was conducted using a bulked sample of F_2 mutant segregants and a bulked sample of F_2 wild-type segregants. The allele frequencies were calculated with respect to the 'Noir 1' SNP state such that SNPs closely linked to the causative locus would exhibit a mutant bulk allele frequency of approximately zero and a wild-type bulk frequency of approximately 0.66 (as two out of every three wild-type plants would be expect to be heterozygous at the causative locus). The allele frequencies at each SNP position were visualized as the proportion of reads derived from 'Noir 1' and averaged across a 21 SNP sliding window (Fig. 1d). Chromosomes 1 through 19 did not show any major divergence in allele frequencies between the bulks. However, the chromosome 20 allele frequencies exhibited the expected divergence for the causative locus (Fig. 1d).

A detailed analysis of chromosome 20 showed that the mutant bulk had an average 'Noir 1' allele frequency of zero for a 26.6 Mb interval spanning the pericentromere on chromosome 20, between positions 1.67 and 28.3 Mb (Fig. 2a). This 26.6 Mb interval only encompasses 65 gene models, as the region is mostly heterochromatic. The mutant bulk's average 'Noir 1' allele frequency of zero indicates that the mutant parent markers in the 26.6 Mb interval co-segregated with the F_2 mutant phenotype. For the same interval, the wild-type bulk had the excepted average 'Noir 1' allele frequency of 0.66. The 26.6 Mb interval overlapped with one structural variant detected by the aCGH experiment, the approximately 2 kb deletion (Fig. 2b). This deletion, presumably generated by the fast neutron mutagenesis, was located within a single gene model, Glyma.20G019300. The nearest ortholog to this gene in Arabidopsis is *NAP1*, or *Nck-Associated Protein 1* (gene model AT2G35110), and is involved in the actin cytoskeleton formation (Deeks et al. 2004; El-Assal et al. 2004). The soybean gene model Glyma.20G019300 (named Glyma20g02370 in previous genome releases) has high amino acid similarity to this Arabidopsis *NAP1* ortholog (87.4 %). No other Arabidopsis gene model showed high sequence similarity to this soybean gene. Furthermore, the Arabidopsis *nap1* trichome mutant (*grl-4*) has a phenotype with swollen, twisted, and shorter trichomes, similar to the soybean mutant phenotype

observed in R55C01 (Deeks et al. 2004; El-Assal et al. 2004).

To validate the mapping results, co-segregation between the trichome phenotype and the candidate mutation was tested on segregating F_3 progeny. A co-dominant PCR marker was designed using three primers to amplify unique bands for the mutant and wild-type alleles (Supplemental Table 1, Supplemental Fig. 1A). The phenotypes of 50 F_3 individuals, representing different $F_{2:3}$ families, perfectly co-segregated with their expected genotypic classes (mutant: mt/mt or wild-type: wt/wt or wt/mt) (Supplemental Fig. 1). Altogether, these data and prior information about the *NAP1* gene indicated that Glyma.20G019300 was the leading candidate gene for the soybean *gnarled* phenotype.

Complex rearrangements detected in Glyma.20G019300

The aCGH data detected an approximately 2 kb deletion in the Glyma.20G019300 candidate gene, but was not able to resolve the fine structure of this event. Therefore, whole genome resequencing was conducted to resolve the specific breakpoints of this deletion. Surprisingly, the resequencing data revealed a much more complex structure to this locus than anticipated (Fig. 3).

Glyma.20G019300 consists of 23 exons stretched across 22,550 nucleotides (including exons, introns, and untranslated regions) that encode a protein with 1388 amino acids. Sequencing of the mutated allele resolved two distinct deletions (2021 and 1224 bp, respectively) nearby one another. The 1224 bp deletion was not originally detected by aCGH due to the deletion of only a single probe, but occurred upstream of the aCGH identified mutation. In addition, a 1138 bp segment separating these deletions was found in an inverted orientation. This inverted fragment consisted of the thirteenth exon and part of the twelfth and thirteenth introns. This fragment inverted and fused to a sequence 7 bp upstream of the fourteenth exon, forming a novel junction between positions Gm20:2,010,290 and Gm20:2,007,928. Finally, an unresolved chromosome rearrangement, possibly an intra-chromosomal translocation, was identified adjacent to the inversion–deletion junction. The Gm20:2,009,152 side of the inverted fragment was connected to sequence found nearly 15 Mb downstream in the reference genome, at Gm20:16,920,485. The first half of the gene was interrupted at Gm20:2,012,311 and was found to be connected to Gm20:16,939,673. A novel 22 bp sequence was also found in the junction. PCR primers were designed to span these junctions, and the amplicons confirmed the presence of the three novel junctions found by whole genome resequencing (Fig. 3).

The putative intra-chromosomal translocation remains unresolved. Orientation of the sequences at the junction Gm20:2,009,152 to Gm20:16,920,485 and the junction Gm20:2,012,311 to Gm20:16,939,673 do not suggest that a single contiguous piece was inserted into Glyma.20G019300. Rather, the orientations of the sequences suggest that additional chromosome rearrangements have likely occurred.

RNA sequencing transcription analysis of the *gnarled* mutant

Complex rearrangements may alter transcription of the genes at and nearby the disrupted locus. To test this, RNA-seq analysis was conducted on leaf, seed, and root tissues from wild-type ('M92-220') and *gnarled* mutant individuals (R55C01). Approximately 95 % of the RNA-Seq reads were mapped to the soybean genome, and roughly 89 % of the read-pairs were mapped concordantly. Under normal conditions in wild-type plants, Glyma.20G019300 has been observed to be transcribed in all previously examined tissues (Libault et al. 2010; Severin et al. 2010a). Similarly, this gene was observed to be transcribed in all three tissues tested in this study; however, the *gnarled* mutant exhibited altered relative transcript abundances among the 23 exons (Supplemental Fig. 2).

Exons one through ten all exhibited similar transcription levels in mutant and wild-type plants, with some minor exon-specific fluctuations among the different tissues and genotypes, while exon 11 appeared to show higher relative transcription levels in the mutant (Supplemental Fig. 2). Exons 12, 13, and 14, however, exhibited essentially no transcription in the *gnarled* mutant (Supplemental Fig. 2), while the wild-type individuals exhibited transcription levels similar to the other exons of this gene. As described in the previous section, exon 12 resides within a deleted interval in the *gnarled* mutant, and an interval encompassing exon 13 is inverted relative to wild-type. These structural rearrangements may be expected to eliminate transcription in these intervals. Exon 14, while present and in proper orientation in the mutant, is directly adjacent to the exon 13 inversion and also appeared to be transcriptionally silent. The downstream exons, 15–23, exhibited fewer relative transcripts in the *gnarled* mutant relative to wild-type (Supplemental Fig. 2). Furthermore, it is worth noting that the transposition event that separates exons 11 and 13 in the *gnarled* mutant is of unknown size and sequence composition (Fig. 3). Therefore, it is possible that the reads observed from exons 1–11 and exons 15–23 are completely independent transcripts in the *gnarled* mutant line. Furthermore, gene model Glyma.07G221000, the nearest paralog to Glyma.20G019300, did not exhibit any exon-level transcript differences between the mutant and wild-type

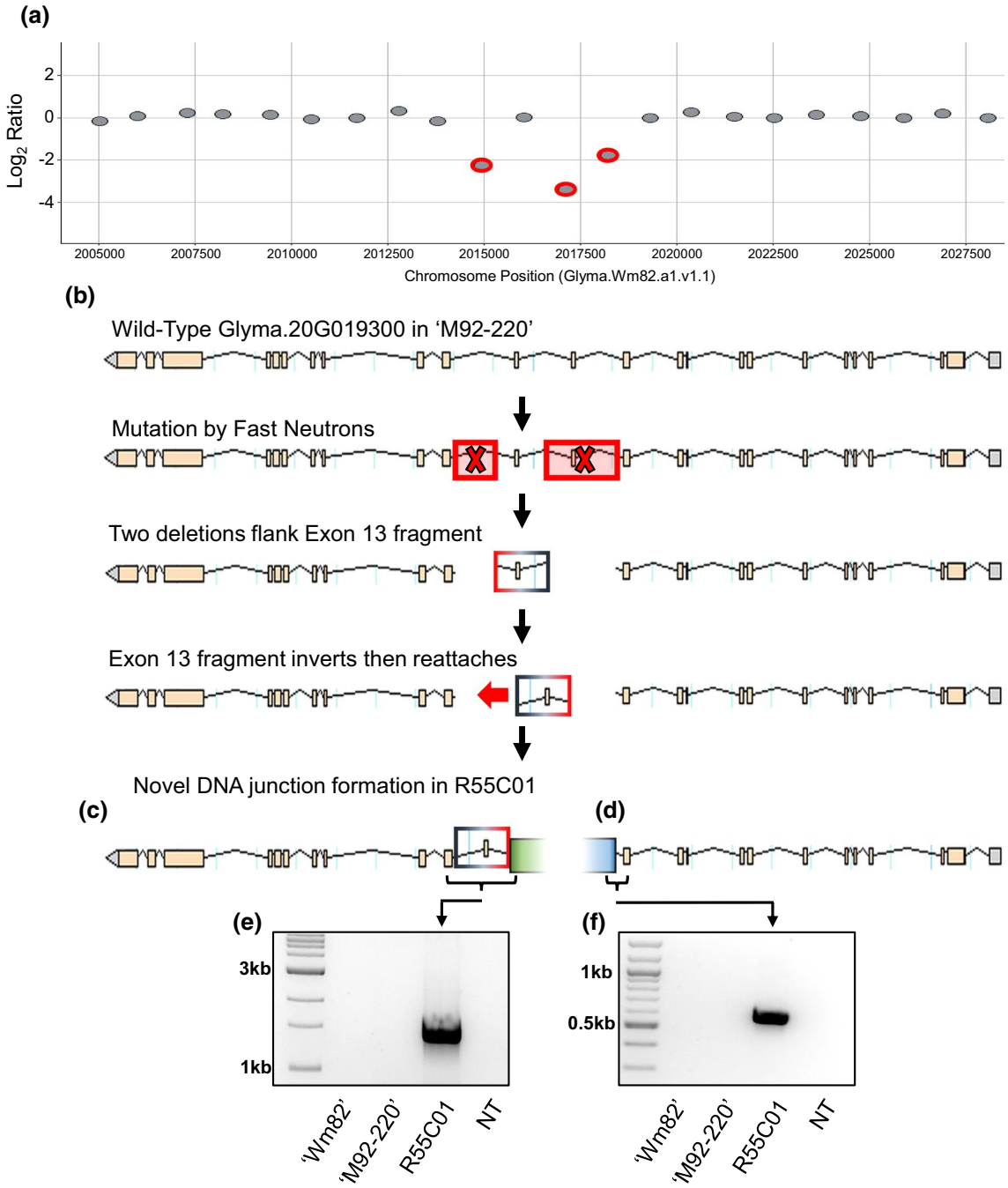

Fig. 3 Mutations in the candidate gene demonstrate the complexity of mutations that can occur by fast neutron mutagenesis. **a** aCGH report depicting two deletions in *GmNAP1* indicated by probes with corrected $\log_2$ ratios of less than −2. aCGH array was designed using version 1 assembly (Glyma.Wm82.a1.v1.1), thus v1 positions are listed. **b** Wild-type Glyma.20G019300 and fast neutron mutations to the gene. **c** The inverted 13th exon connected to the second half of the gene forms a novel junction between Gm20:2,010,290 and Gm20:2,007,928 (positions are according to the version 2 genome assembly, Glyma.Wm82.a2.v1). The Gm20:2,009,152 side of the inverted fragment is connected to sequence found at Gm20:16,920,485. **d** The first half of the gene is interrupted at Gm20:2,012,311 and is connected to sequence found at Gm20:16,939,673. A novel 22 bp sequence was found in the junction. PCR amplification was used to confirm the novel DNA junctions created by fast neutron mutagenesis. **e** A 1.4 kb fragment spanning across two novel junctions created in the second half of the gene. **f** A 605 bp fragment spanning the novel junction created in the first half of the gene. For **e** and **f**, the samples tested were (*left* to *right*): 'Williams 82', 'M92-220', R55C01, and a no template control. The orientations of the sequences at the junctions do not suggest that a single contiguous piece was inserted into Glyma.20G019300, and the extent of chromosomal rearrangements that occurred on chromosome 20 is unclear at this time

plants. This offers additional evidence that the transcriptional changes observed in Glyma.20G019300 are a consequence of the structural rearrangement per se, rather than RNA-interference or other post-transcriptional silencing mechanism.

Transcriptional alterations were also observed genome-wide between the *gnarled* mutant compared to wild-type, including 2299 genes differentially transcribed in at least one of the three tissue types. The genes differentially transcribed in leaf tissues included those involved in processes such as lipoxygenase activity (GO:0016165) and photosynthetic processes (GO:0009769 and GO:0016168), which are pathways that have been linked to trichome development (Schilmiller et al. 2010; Yan et al. 2012). However, homologs for genes previously demonstrated to be involved in trichome developmental processes, including monosaccharide and lignin biosynthesis (Marks et al. 2009), were not observed to be differentially transcribed between the *gnarled* mutant and wild-type in any tissue (data not shown). Among the 65 gene models located within the 26.6 Mb interval mapped by WGS-BSA, Glyma.20G019300 was the only differentially transcribed gene. None of the remaining 64 gene models exhibited differential transcription among any of the three tissue types.

Complementation of *Atnap1* using *GmNAP1*

Due to the high similarity of both the *nap1* gene sequences and phenotypes, complementation of an Arabidopsis *nap1* mutant with *GmNAP1* could be used to validate the function of Glyma.20G019300. A construct consisting of 2 kb of the soybean *NAP1* promoter driving the soybean *NAP1* cDNA and a D35S promoter driving the BAR herbicide resistance gene was transformed into the Arabidopsis *nap1 gnarled* mutant (*grl-4*). Twenty T_1 individuals were recovered that displayed wild-type trichomes, were resistant to glufosinate, and tested positive for the *GmNAP1* transgene based on PCR analyses (Supplemental Fig. 3). The functional complementation of the Arabidopsis *nap1* mutant indicates that *GmNAP1* is important for trichome formation and the two orthologs share functional homology.

Identification of a spontaneous *NAP1* soybean mutant T31 (*p2*)

A search for historic soybean mutants with *gnarled* trichomes led to the identification of the mutant line T31 (PI548159) (Stewart and Wentz 1926; Bernard and Singh 1969; Singh et al. 1971; Healy et al. 2005). T31's recessive *p2* mutant trichome allele was initially described as 'puberulent' but more closely resembles the *gnarled* phenotype. The *p2* allele was previously backcrossed into the cv. 'Harosoy' (PI548573) and into cv. 'Clark' (PI548533) to generate several advanced backcross lines (Weiss and Stevenson 1955; Johnson 1958; Bernard et al. 1991). There are two *p2* backcross (BC_6) near-isogenic lines (PI547713 and PI547743) in 'Harosoy', and there are three *p2* backcross (one BC_6 and two BC_7) near-isogenic lines (PI547449, PI547565, and PI547566) in 'Clark' (Bernard et al. 1991). The SoySNP50K chip data (Song et al. 2015) obtained from SoyBase (http://soybase.org) were used to identify a genomic interval shared by T31 and the five *p2* backcross lines.

A single genomic interval shared between T31 and the five *p2* backcross lines was located on chromosome 20 (Fig. 4). At position Gm20:1,742,275 (ss715636805), all five *p2* backcross lines carried the T31 allele, and at position Gm20:2,053,056 (ss715636914) three of the five *p2* backcross lines contained the T31 allele and two lines (PI547449 and PI547565) had missing genotypes. PI547565 had the T31 allele for the polymorphic SNP at Gm20:2,148,735 (ss71563945), adjacent to the missing genotype at Gm20:2,053,056, suggesting that the genotyping score at Gm20:2,053,056 would likely match T31. PI547449 had either missing data or heterozygous calls at all polymorphic SNP positions downstream of position Gm20:1,742,275 until Gm20:2,353,994 which had the 'Clark' allele. The observed heterozygous genotype calls are likely due to heterogeneity found between sampled individuals in a line rather than to residual heterozygosity within a specific individual of a line. The narrow 566 kb *p2* introgression interval, marked by the resumption of the recurrent parent haplotypes upstream at Gm20:1,582,950 (ss715636740) and downstream at Gm20:2,148,735 (ss715636945) contains the *GmNAP1* gene Glyma.20G019300. The inclusion of Glyma.20G019300 in the *p2* introgression interval and the similarity of the R55C01 and T31 trichome phenotypes suggested that *p2* could be caused by a mutation in Glyma.20G019300.

Sequencing the exons of Glyma.20G019300 from T31 identified a single base pair deletion in the 22nd exon (Fig. 5). The resulting frame shift mutation and early stop codon resulted in the mutation or loss of 202 amino acids (14.5 % of the gene). Sequence analysis of the locus in 25 wild-type diverse soybean accessions (McHale et al. 2012) confirmed that the single base pair deletion is unique to T31.

Discussion

The present study combined aCGH and WGS-BSA methods to identify fast neutron-induced chromosome rearrangements in Glyma.20G019300 that caused a *gnarled* trichome phenotype in the soybean mutant R55C01. This gene is the soybean ortholog of the well-characterized

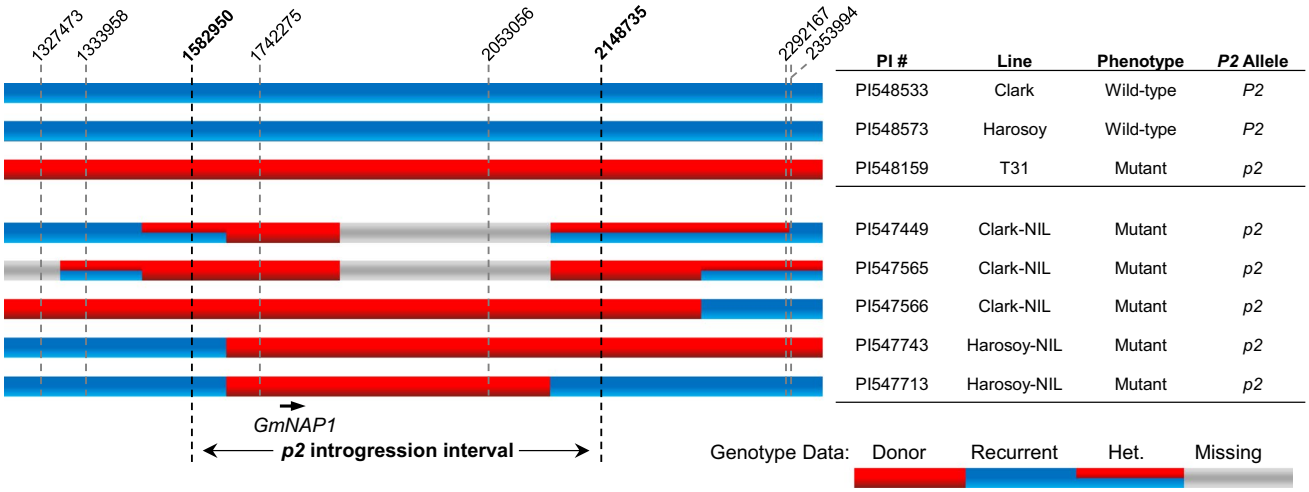

Fig. 4 *p2* introgression interval identified on chromosome 20. Positions are given for polymorphic SoySNP50K markers in the genomic region. The *p2* allele, from the donor line T31, was previously backcrossed into the recurrent parents 'Clark' and 'Harosoy' to generate five Near Isogenic Lines (NILs) as part of the Soybean Isoline Collection. The recurrent parent genotypes are color coded in *blue*, and the donor parent genotypes are color coded in *red*, heterozygous (Het.) genotypes are color coded in *blue* and *red*, and missing genotypes are color coded in *gray*. Heterozygous scores are likely due to heterogeneity in the NIL. Examination of the five *p2* NILs' genotypes identified a single introgression interval (566 kb) in the genome in which all five lines shared the donor parent genotype (Gm20:1,582,950–2,148,735). This interval contains the *GmNAP1* gene (Gm20:1,999,216–2,021,765) (color figure online)

```
Wm82 Gm07 39610713  ACATGCTAGTGACAGAATGGATCCTGAATTATCAAGTC 39610676
Wm82 Gm20  2000396  ACATGCTAGTGAGGGAATGGATCCCGAATTATCAAGTC  2000359
T31                 ACATGCTAGTGAGGGAATG-ATCCCGAATTATCAAGTC …(45bp)… TGA

Archer              ...................G..................
Minsoy              ...................G..................
Noir1               ...................G..................
Missoy              ...................G..................
Parker              ...................G..................
PI468922            ...................G..................
Bert                ...................G..................
A.K.(Harrow)        ...................G..................
Captial             ...................G..................
CNS                 ...................G..................
Dunfield            ...................G..................
Illini              ...................G..................
Jackson             ...................G..................
Lincoln             ...................G..................
Forrest             ...................G..................
Lee                 ...................G..................
Odgen               ...................G..................
Perry               ...................G..................
Richland            ...................G..................
Roanoke             ...................G..................
S-100               ...................G..................
Haberlandt          ...................G..................
Liu yue bao         ...................G..................
Shirome             ...................G..................
Zontanorukon        ...................G..................
```

Fig. 5 Sequence comparison of T31, 'Williams 82' (Wm82), and 25 diverse wild-type lines for the interval flanking the frame shift deletion found in T31. Sequencing of Glyma.20g019300 (*NAP1*) exon 22 in 25 diverse wild-type lines identified that the single base pair deletion is unique to T31. Sequence differences between the two soybean *NAP1* paralogs (highlighted in *gray*) make it possible to differentiate the chromosome 20 (Gm20) and chromosome 7 (Gm07) sequences. The given sequence positions are for genome assembly version 2 (Glyma.Wm82.a2.v1). T31's 1 bp deletion is highlighted in *gray*. The resulting frame shift caused an early stop codon starting 64 bases downstream from the single base deletion. The early stop codon is shown in bold in the T31 downstream sequence

Arabidopsis *NAP1* gene (Deeks et al. 2004; El-Assal et al. 2004). The morphological development of trichomes is guided by the actin cytoskeleton and thus proper actin nucleation is critical for proper trichome formation (Beilstein and Szymanski 2004). The *NAP1* gene was identified as a component of the SCAR/WAVE protein complex that activates the ARP2/3 complex involved in actin nucleation, and mutations in this gene have been shown to result in the *gnarled* trichome phenotype of Arabidopsis. Arabidopsis and soybean trichomes are morphologically distinct, with Arabidopsis exhibiting branched trichomes and soybean displaying unbranched trichomes. Despite such differences in form, mutations to the *nap1* gene in both species cause similar perturbations to trichome development and demonstrate that the two species share this key component in the actin polymerization pathway that underlies trichome morphology. This result suggests that other genes and pathways controlling trichome morphology are also likely to be conserved between the two species, which may facilitate the identification of other genes that are critical for soybean trichome function and development.

Several previous studies have described soybean trichome morphological mutants (Nagai and Saito 1923; Piper and Morse 1923; Stewart and Wentz 1926; Owen 1927; Johnson and Hollowell 1935; Ting 1946; Williams 1950; Bernard and Singh 1969; Bernard 1975; Healy et al. 2005). However, to our knowledge, no soybean study has cloned the underlying causative variant of a soybean trichome morphology mutant. The functional complementation of an Arabidopsis *nap1 gnarled* mutant (*grl-4*) by whole plant transformation with *GmNAP1* validated the soybean *GmNAP1* gene function in the present study. Additionally, this study further validated the function of *GmNAP1* by identifying a second mutant allele of *GmNAP1*, the *p2* trichome mutant locus in line T31 (PI548159). This phenotype is the result of a single base pair deletion in the 22nd exon of *NAP1* (Fig. 5).

The T31 mutant is part of the USDA Soybean Isoline Collection and several previous studies have used mutants from this collection for trait mapping and candidate gene validation (Muehlbauer et al. 1991; Thompson et al. 1997; Cober and Voldeng 2001; Jeong et al. 2002; Molnar et al. 2003; Watanabe et al. 2009; Cober et al. 2010; Severin et al. 2010b; Peiffer et al. 2012). The recent genotyping of this collection with the SoySNP50K chip (Song et al. 2015) facilitated the rapid mapping of the T31 mutant. This study demonstrates the ability to leverage the valuable Isoline Collection historic mutant resource to validate a candidate gene. Efforts are currently underway to map additional mutant alleles from the Isoline Collection. To our knowledge, this is the first study to combine the SoySNP50K chip data with the Isoline Collection's historic phenotypic data to validate a candidate gene by identifying a second mutant allele.

The 26.6 Mb mapping interval identified by WGS-BSA in this study was wider than expected, but the size of the interval was likely inflated by low regional recombination rates. The mapping interval had reasonable resolution on the distal side of the candidate gene with the mapping interval starting approximately 329 kb from *GmNAP1*. However, on the proximal side of *GmNAP1*, the mapping interval extended approximately 26.3 Mb from *GmNAP1* to the other arm of the chromosome. Due to the position of *GmNAP1* near the edge of the heterochromatic region, it was likely that repressed recombination on the proximal side of the gene expanded the mapping interval significantly. A recent mapping study also identified suppressed recombination in this region on chromosome 20 (Li et al. 2014). It is likely that the repressed recombination in this region led to the large mapping interval. Furthermore, it is probable that subsequent mapping studies using WGS-BSA in soybean will have smaller mapping intervals for regions of the genome with higher recombination rates.

Combining the aCGH data with the WGS-BSA mapping interval led to the identification of a single candidate gene, despite the large mapping interval. The only aCGH-detected mutation in the mapping interval was the approximately 2 kb deletion in Glyma.20G019300. Further examination at this locus identified additional mutations that were not detected by aCGH. These additional mutations include a second deletion, an inversion, and two novel junctions which suggest additional chromosome rearrangements occurred on chromosome 20. The complexity of fast neutron-induced mutations identified within this single gene was unexpected and further challenges the common assumption that fast neutron mutagenesis results in simple deletions (see Bolon et al. 2014 for additional evidence).

This study has demonstrated the effective combination of WGS-BSA and aCGH to identify a candidate fast neutron-induced mutation from a reasonable sized F_2 mapping population. The combination of technologies demonstrated the ability to save significant cost and time by identifying the causative variant with only one round of BSA mapping, and without the need for additional fine-mapping.

Author contribution statement BWC, CPV, GJM, and RMS designed the research; BWC, ANH, SS, TJYK, FF, and JAO performed the research; BWC, TJYK and JAO analyzed the data; and BWC, JAO and RMS wrote the manuscript

Acknowledgments The authors are grateful to Dan Szymanski for providing seed for the Arabidopsis *grl-4* mutant, Yung-Tsi Bolon and Jeff Roessler for early identification of the soybean *gnarled* mutant, Bruna Bucciarelli for assisting with the RNA-seq experiments, and both Junqi Liu and Jean-Michel Michno for their advice regarding the *GmNAP1* construct creation. The authors acknowledge the Minnesota

Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. URL: http://www.msi.umn.edu. Microscopy imaging and analysis was performed at the University Imaging Centers, University of Minnesota. The authors would also like to thank the Minnesota Soybean Research and Promotion Council, United Soybean Board, and the National Science Foundation for their help and support. We would also like to thank the National Science Foundation for supporting BWC through the Graduate Research Fellowship Program (NSF-GRFP).

Compliance with ethical standards

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

Ethical standards The authors declare that the experiments comply with the current laws of the countries in which the experiments were performed

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Anders S, Pyl PT, Huber W (2014) HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169

Anderson JE, Kantar MB, Kono TY, Fu F, Stec AO, Song Q, Cregan PB, Specht JE, Diers BW, Cannon SB et al (2014) A roadmap for functional structural variants in the soybean genome. G3 (Bethesda) 4:1307–1318

Bart RS, Chern M, Vega-Sánchez ME, Canlas P, Ronald PC (2010) Rice Snl6, a cinnamoyl-CoA reductase-like gene family member, is required for NH1-mediated immunity to Xanthomonas oryzae pv. oryzae. PLoS Genet 6:e1001123

Beilstein M, Szymanski DB (2004) Cytoskeletal requirements during Arabidopsis tichome development. In: Hussey P (ed) The plant cytoskeleton in cell differentiation and development. Blackwell, Oxford, UK, pp 265–289

Belfield EJ, Gan X, Mithani A, Brown C, Jiang C, Franklin K, Alvey E, Wibowo A, Jung M, Bailey K et al (2012) Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. Genome Res 22:1306–1315

Bernard RL (1975) The inheritance of appressed pubescence. Soybean Genet Newsl 2:34–36

Bernard RL, Cremeens CR (1988) Registration of "Williams 82" Soybean. Crop Sci 28:1027

Bernard RL, Singh BB (1969) Inheritance of Pubescence Type in Soybeans: glabrous, curly, dense, sparse, and puberulent. Crop Sci 9:192–197

Bernard RL, Nelson RL, Cremeens CR (1991) USDA soybean genetic collection: isoline collection. Soybean Genet Newsl 18:27–57

Bolon YT, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ et al (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in Soybean. Plant Physiol 156:240–253

Bolon YT, Stec AO, Michno JM, Roessler J, Bhaskar PB, Ries L, Dobbels AA, Campbell BW, Young NP, Anderson JE et al (2014) Genome resilience and prevalence of segmental duplications following fast neutron irradiation of Soybean. Genetics 198:967–981

Campbell BW, Stupar RM (2016) Soybean mutant and germplasm resources: current status and future prospects. Curr Protoc Plant Biol 1:307–327. doi:10.1002/cppb.20015

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16:735–743

Cober ER, Voldeng HD (2001) A new Soybean maturity and photoperiod-sensitivity locus linked to E1 and T. Crop Sci 41:698–701

Cober ER, Molnar SJ, Charette M, Voldeng HD (2010) A new locus for early maturity in Soybean. Crop Sci 50:524–527

Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N et al (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol 8:9

Cui Y, Barampuram S, Stacey MG, Hancock CN, Findley S, Mathieu M, Zhang Z, Parrott WA, Stacey G (2013) Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. Plant Physiol 161:36–47

Deeks MJ, Kaloriti D, Davies B, Malhó R, Hussey PJ (2004) Arabidopsis NAP1 is essential for Arp2/3-dependent trichome morphogenesis. Curr Biol 14:1410–1414

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498

Development Core Team R (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0

Downes BP, Stupar RM, Gingerich DJ, Vierstra RD (2003) The HECT ubiquitin-protein ligase (UPL) family in Arabidopsis: UPL3 has a specific role in trichome development. Plant J 35:729–742

Ehleringer JR, Mooney HA (1978) Leaf hairs: effects on physiological activity and adaptive value to a desert shrub. Oecologia 37:183–200

El Refy A, Perazza D, Zekraoui L, Valay JG, Bechtold N, Brown S, Hülskamp M, Herzog M, Bonneville JM (2003) The Arabidopsis KAKTUS gene encodes a HECT protein and controls the number of endoreduplication cycles. Mol Genet Genom 270:403–414

Gillman JD, Stacey MG, Cui Y, Berg HR, Stacey G (2014) Deletions of the SACPD-C locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol 14:143

Gong J, Waner DA, Horie T, Li SL, Horie R, Abid KB, Schroeder JI (2004) Microarray-based rapid cloning of an ion accumulation deletion mutant in Arabidopsis thaliana. Proc Natl Acad Sci (USA) 101:15404–15409

Haase NJ, Beissinger T, Hirsch CN, Vaillancourt B, Deshpande S, Barry K, Buell CR, Kaeppler SM, de Leon N (2015) Shared genomic regions between derivatives of a large segregating population of maize identified using bulked segregant analysis sequencing and traditional linkage analysis. G3 (Bethesda) 5:1593–1602

Hancock CN, Zhang F, Floyd K, Richardson AO, LaFayette P, Tucker D, Wessler SR, Parrott WA (2011) The rice miniature inverted repeat transposable element mPing is an effective insertional mutagen in Soybean. Plant Physiol 157:552–562

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM (2011) The composition and origins of genomic variation among

individuals of the soybean reference cultivar Williams 82. Plant Physiol 155:645–655

Healy RA, Horner HT, Bailey TB, Palmer RG (2005) A microscopic study of trichomes on gynoecia of normal and tetraploid clark cultivars of glycine max and seven near-isogenic lines. Int J Plant Sci 166:415–425

Hulburt DJ, Boerma HR, All JN (2004) Effect of pubescence tip on soybean resistance to lepidopteran insects. J Econ Entomol 97:621–627

Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015) Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theor Appl Genet 128:1329–1342

James G, Patel V, Nordström KJ, Klasen JR, Salomé PA, Weigel D, Schneeberger K (2013) User guide for mapping-by-sequencing in Arabidopsis. Genome Biol 14:R61

Jeong SC, Kristipati S, Hayes AJ, Maughan PJ, Noffsinger SL, Gunduz I, Buss GR, Maroof MA (2002) Genetic and sequence analysis of markers tightly linked to the soybean mosaic virus resistance gene, Rsv3. Crop Sci 42:265–270

Johnson HW (1958) Registration of soybean varieties. Agron J 50:690–691

Johnson CS (2002) TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of arabidopsis, encodes a WRKY transcription factor. Plant Cell Online 14:1359–1375

Johnson HW, Hollowell EA (1935) Pubescent and glabrous characters of soybeans as related to resistance to injury by the potato leafhopper. J Agric Res 51:371–381

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36

Kirik V, Bouyer D, Schöbinger U, Bechtold N, Herzog M, Bonneville JM, Hülskamp M (2001) CPR5 is involved in cell proliferation and cell death control and encodes a novel transmembrane protein. Curr Biol 11:1891–1895

Levin DA (1973) The role of trichomes in plant defense. Q Rev Biol 48:3–15

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Li YH, Liu YL, Reif JC, Liu ZX, Liu B, Mette MF, Chang RZ, Qiu LJ (2014) Biparental resequencing coupled with SNP genotyping of a segregating population offers insights into the landscape of recombination and fixed genomic regions in elite soybean. G3 (Bethesda) 4:553–560

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G (2010) An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. Plant J 63:86–99

Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One 7:e36406

Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA (2009) Transcriptome analysis of Arabidopsis wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. Mol Plant 2:803–822

Mascher M, Jost M, Kuon JE, Himmelbach A, Aßfalg A, Beier S, Scholz U, Graner A, Stein N (2014) Mapping-by-sequencing accelerates forward genetics in barley. Genome Biol 15:R78

Mathieu M, Winters EK, Kong F, Wan J, Wang S, Eckert H, Luth D, Paz M, Donovan C, Zhang Z et al (2009) Establishment of a soybean (Glycine max Merr. L) transposon-based mutagenesis repository. Planta 229:279–289

Mathur J, Mathur N, Kirik V, Kernebeck B, Srinivas BP, Hülskamp M (2003) Arabidopsis CROOKED encodes for the smallest subunit of the ARP2/3 complex and controls cell shape by region specific fine F-actin formation. Development 130:3137–3146

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159:1295–1308

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Men AE, Laniya TS, Searle IR, Iturbe-Ormaetxe I, Gresshoff I, Jiang Q, Carroll BJ, Gresshoff PM (2002) Fast neutron mutagenesis of soybean (Glycine soja L.) produces a supernodulating mutant containing a large deletion in linkage group H. Genome Lett 1:147–155

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci (USA) 88:9828–9832

Molnar SJ, Rai S, Charette M, Cober ER (2003) Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. Genome 46:1024–1036

Muehlbauer GJ, Staswick PE, Specht JE, Graef GL, Shoemaker RC, Keim P (1991) RFLP mapping using near-isogenic lines in the soybean [Glycine max (L.) Merr]. Theor Appl Genet 81:189–198

Nagai I, Saito S (1923) Linked factors in soybeans. Japan Jour Bot 1:121–136

Nyman Y (1993) The pollen-collecting hairs of Campanula (Campanulaceae). I. Morphological variation and the retractive mechanism. Am J Bot 80:1427–1436

O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX et al (2014) An RNA-Seq based gene expression atlas of the common bean. BMC Genom 15:866

Oppenheimer DG, Herman PL, Sivakumaran S, Esch J, Marks MD (1991) A myb gene required for leaf trichome differentiation in Arabidopsis is expressed in stipules. Cell 67:483–493

Oppenheimer DG, Pollock MA, Vacik J, Szymanski DB, Ericson B, Feldmann K, Marks MD (1997) Essential role of a kinesin-like protein in Arabidopsis trichome morphogenesis. Proc Natl Acad Sci (USA) 94:6261–6266

Orf JH, Denny RL (2004) Registration of "MN1302" Soybean. Crop Sci 44:693

Owen FV (1927) Inheritance studies in soybeans. II. Glabrousness, color of pubescence, time of maturity, and linkage relations. Genetics 12:519–529

Palmer RG, Sandhu D, Curran K, Bhattacharyya MK (2008a) Molecular mapping of 36 soybean male-sterile, female-sterile mutants. Theor Appl Genet 117:711–719

Palmer RG, Zhang L, Huang ZP, Xu M (2008b) Allelism and molecular mapping of soybean necrotic root mutants. Genome 51:243–250

Peiffer GA, King KE, Severin AJ, May GD, Cianzio SR, Lin SF, Lauter NC, Shoemaker RC (2012) Identification of candidate genes underlying an iron efficiency quantitative trait locus in soybean. Plant Physiol 158:1745–1754

Piper CV, Morse WJ (1923) The Soybean. McGraw-Hill Book Co, New York

Raval J, Baumbach J, Ollhoff AR, Pudake RN, Palmer RG, Bhattacharyya MK, Sandhu D (2013) A candidate male-fertility female-fertility gene tagged by the soybean endogenous transposon, Tgm9. Funct Integr Genom 13:67–73

Rerie WG, Feldmann KA, Marks MD (1994) The GLABRA2 gene encodes a homeo domain protein required for normal trichome development in Arabidopsis. Genes Dev 8:1388–1399

Ríos G, Naranjo MA, Iglesias DJ, Ruiz-Rivero O, Geraud M, Usach A, Talón M (2008) Characterization of hemizygous deletions in citrus using array-comparative genomic hybridization and microsynteny comparisons with the poplar genome. BMC Genom 9:381

Robbins JC, Daugherty DM, Hatchett JH (1979) Ovipositional and feeding preference of leafhoppers (Homoptera: Cicadellidae) on clark soybeans in relation to plant pubescence. J Kansas Entomol Soc 52:603–608

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Schellmann S, Schnittger A, Kirik V, Wada T, Okada K, Beermann A, Thumfahrt J, Jürgens G, Hülskamp M (2002) TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in Arabidopsis. EMBO J 21:5036–5046

Schilmiller AL, Miner DP, Larson M, McDowell E, Gang DR, Wilkerson C, Last RL (2010) Studies of a biochemical factory: tomato trichome deep expressed sequence tag sequencing and proteomics. Plant Physiol 153:1212–1223

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Sel-D El-Assal, Le J, Basu D, Mallery EL, Szymanski DB (2004) Arabidopsis GNARLED encodes a NAP125 homolog that positively regulates ARP2/3. Curr Biol 14:1405–1409

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE et al (2010a) RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biol 10:160

Severin AJ, Peiffer GA, Xu WW, Hyten DL, Bucciarelli B, O'Rourke JA, Bolon YT, Grant D, Farmer AD, May GD et al (2010b) An integrative approach to genomic introgression mapping. Plant Physiol 154:3–12

Singh BB, Hadley HH, Bernard RL (1971) Morphology of pubescence in soybeans and its relationship to plant Vigor1. Crop Sci 11:13–16

Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8:1–12

Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2015) Fingerprinting soybean germplasm and its utility in genomic research. G3 (Bethesda) 5:1999–2006

Song Q, Jenkins J, Jia G, Hyten DL, Pantalone V, Jackson SA, Schmutz J, Cregan PB (2016) Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genom 17:33

Stewart RT, Wentz JB (1926) A recessive glabrous character in soybeans. Agron J 18:997–1009

Szymanski DB, Marks MD, Wick SM (1999) Organized F-actin is essential for normal trichome morphogenesis in Arabidopsis. Plant Cell 11:2331–2347

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S et al (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183

Thompson JA, Bernard RL, Nelson RL (1997) A third allele at the soybean dt1 locus. Crop Sci 37:757–762

Ting CL (1946) Genetic studies on the wild and cultivated soybeans. Agron J 38:381–393

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinf 11:11.10.1–11.10.33

Vermeer J, Peterson RL (1979) Glandular trichomes on the inflorescence of *Chrysanthemum-morifolium* cv. Dramatic (Compositae). I. Development and morphology. Can J Bot 57:705–713

Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC (1999) The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in arabidopsis, encodes a WD40 repeat protein. Plant Cell 11:1337–1349

Watanabe S, Hideshima R, Xia Z, Tsubokura Y, Sato S, Nakamoto Y, Yamanaka N, Takahashi R, Ishimoto M, Anai T et al (2009) Map-based cloning of the gene associated with the soybean maturity locus E3. Genetics 182:1251–1262

Weiss MG, Stevenson TM (1955) Registration of soybean varieties, V. Agron J 47:541–543

Williams LF (1950) Structure and genetic characteristics of the soybean. In: Markley KS (ed) Soybeans Soybean Prod. Interscience Publishers, New York, pp 111–134

Woodworth CM (1921) Inheritance of cotyledon, seed-coat, hilum and pubescence colors in soy-beans. Genetics 6:487–553

Yan A, Pan J, An L, Gan Y, Feng H (2012) The response of trichome mutants to enhanced ultraviolet-B radiation in Arabidopsis thaliana. J Photochem Photobiol B 113:29–35

Yang Z, Huang D, Tang W, Zheng Y, Liang K, Cutler AJ, Wu W (2013) Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. PLoS One 8:e68433

Zabala G, Vodkin L (2003) Cloning of the pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. Genetics 163:295–309

Zhang XC, Millet Y, Ausubel FM, Borowsky M (2014) Next-Gen sequencing-based mapping and identification of ethyl methanesulfonate-induced mutations in *Arabidopsis thaliana*. Curr Protoc Mol Biol 108:7.18.1–7.18.16