

TILLING by Sequencing: A Successful Approach to Identify Rare Alleles in Soybean Populations

Rima Thapa ^{1,†}, Militza Carrero-Colón ², Katy M. Rainey ¹ and Karen Hudson ^{2,*}

¹ Department of Agronomy, Purdue University, 915 West State Street, West Lafayette, IN 47907, USA; rimathapa7@gmail.com (R.T.); krainey@purdue.edu (K.M.R.)

² USDA-ARS Crop Production and Pest Control Research Unit, 915 West State Street, West Lafayette, IN 47907, USA; militza.carrero-colon@usda.gov

* Correspondence: Karen.Hudson@USDA.gov; Tel.: +1-765-494-8057

† Current address: BASF Corporation, 925 County Road 378, Beaver Crossing, NE 68313, USA.

Received: 29 October 2019; Accepted: 28 November 2019; Published: 3 December 2019

Abstract: Soybean seeds produce valuable protein that is a major component of livestock feed. However, soybean seeds also contain the anti-nutritional raffinose family oligosaccharides (RFOs) raffinose and stachyose, which are not digestible by non-ruminant animals. This requires the proportion of soybean meal in the feed to be limited, or risk affecting animal growth rate or overall health. While reducing RFOs in soybean seed has been a goal of soybean breeding, efforts are constrained by low genetic variability for carbohydrate traits and the difficulty in identifying these within the soybean germplasm. We used reverse genetics Targeting Induced Local Lesions in Genomes (TILLING)-by-sequencing approach to identify a damaging polymorphism that results in a missense mutation in a conserved region of the *RAFFINOSE SYNTHASE3* gene. We demonstrate that this mutation, when combined as a double mutant with a previously characterized mutation in the *RAFFINOSE SYNTHASE2* gene, eliminates nearly 90% of the RFOs in soybean seed as a proportion of the total seeds carbohydrates, and results in increased levels of sucrose. This represents a proof of concept for TILLING by sequencing in soybean.

Keywords: soybean; raffinose; stachyose; TILLING; *Glycine max*; RFO

1. Introduction

Soybean provides 70% of the global supply of protein meal [1]. However, the soybean seed also contains carbohydrates that are considered antinutritional. Raffinose and stachyose family oligosaccharides (RFOs) are not easily digested by monogastric animals [2,3], and studies have suggested that lower levels of raffinose and stachyose contribute to faster livestock growth and to better overall health [4,5].

Overall, the identification of genetic variation in the carbohydrate biosynthesis pathway of soybean presents a good opportunity for optimizing meal composition for soybean consumers. There are several genes in the carbohydrate biosynthetic pathway that have previously been used to reduced raffinose and stachyose levels, and related genes whose functions are unexplored. The soybean genome contains four genes encoding raffinose synthase enzymes. *RAFFINOSE SYNTHASE2* (*RS2*), *RAFFINOSE SYNTHASE3* (*RS3*), and *RAFFINOSE SYNTHASE4* (*RS4*) share 60% amino acid identity with the *Arabidopsis* raffinose synthase enzyme, and soybean *RAFFINOSE SYNTHASE1* (*RS1*) is less similar [6]. A naturally-occurring polymorphism in the *RAFFINOSE SYNTHASE2* gene was discovered in the soybean accession PI 200508, a deletion of three nucleotides relative to the reference sequence that results in the loss of tryptophan 331 (W331-) in the protein sequence, and has the phenotypic effect of increasing the ratio of sucrose to RFOs in the seed [6]. Reverse genetics approaches also previously identified a missense mutation (T107I) in the *RS2* gene

that reduces levels of raffinose and stachyose and increases sucrose levels [7]. Both of these *rs2* mutants have reduced levels of raffinose (0.1%–0.2% raffinose as a percentage of total seed carbohydrates) relative to lines wild type for *RS2* (0.8%–1% raffinose) in replicated field tests [8]. Stachyose levels are also reduced in the *rs2* single mutant (0.5%–2% of total seed carbohydrates) contrasted with 4% of total carbohydrate in wild type lines, and moderate increases in total sucrose levels [8,9]. A biotechnological approach silencing the *RS2* gene in soybean seeds also achieved significant reductions in seed RFOs and an increase in sucrose [4]. Mutation in *RAFFINOSE SYNTHASE3* has been shown to reduce RFOs further when combined with a non-functional allele of *rs2*, resulting in ultra-low levels of raffinose and stachyose of less than 2% as a fraction of total seed carbohydrates [8–11]. While the decrease in RFO content is desirable for meal digestibility, the alteration of carbohydrate partitioning resulting in increased levels of sucrose, which contributes to the metabolizable energy in soybean meal is an added advantage [12,13]. RFOs have been demonstrated to have a role in desiccation and cold tolerance in plants and may contribute to seed vigor [14–19]. However, studies have shown that soybean *rs2* and *rs3* mutants germinate and emerge with normal efficiency [11,20]. Overall, variation for RFO content is limited in available soybean germplasm [21], and it remains an open question in the field of soybean improvement what impact a null mutation in *RAFFINOSE SYNTHASE* could have on seed carbohydrate levels, and which genetic combination optimally reduces RFOs across growing environments to optimize meal traits.

The *Cel I* endonuclease-based TILLING (Targeting Induced Local Lesions in Genomes) approach has previously been successfully implemented in soybean and resulted in the identification of mutants in seed composition, disease signaling, and other important agronomic traits [7,22–25]. The reverse-genetic method TILLING presents the ability to identify mutants without prior knowledge of the phenotype or the degree of phenotypic severity. Further benefits to TILLING are the capability to identify mutations that may be lethal or result in reduced viability in the homozygous state. Particularly for the soybean genome, where most genes have two highly similar homeologs, which are both functional. Another advantage that TILLING provides over phenotypic screens is a means to identify mutations in genes where deleterious effects can be masked by the presence of another gene that compensates for the loss of function. Additionally, the resulting mutations are non-transgenic variation and can be used in both conventional or transgenic breeding programs. However, there are several significant challenges: Each gene is different in its coding sequence and GC content and in turn the probability of creating an amino acid change or nonsense mutation as a result of a single base pair change. In many cases, DNA point mutations can be silent or create synonymous changes that are unlikely to be deleterious to the protein. Another challenge is the cost: While sequencing costs per base have declined precipitously in recent years, the effort of maintaining and extracting DNA from large populations, library construction, barcoding, and high-fidelity amplification, as well as the technical effort to perform these techniques, remain a cost impediment. Finally, soybean has undergone a relatively recent genome duplication, and homologous genes are, in some cases, identical at the coding level, presenting a challenge in detecting mutations in one specific member of the homologous pair [26]. In the case of genes involved in carbohydrate biosynthesis, a phenotypic screen of a large population is tedious and costly and requires HPLC (high performance liquid chromatography) analysis of individual samples. In the case of the *RAFFINOSE SYNTHASE* gene family, it is known that *RS2* and *RS3* both contribute to carbohydrate partitioning. We sought to obtain additional and more severe mutant alleles to further reduce raffinose and stachyose levels in soybean seeds. We pursued a reverse genetic strategy utilizing high throughput sequencing (TILLING-by-Sequencing, TbyS [27]) to identify additional loss-of-function mutations in the *RS2* and *RS3* genes. To reduce costs, we constructed only one-dimensional pools, which limited the number of libraries. After sequence analysis, we selected mutations that we expected to be deleterious. As a secondary screen, we designed PCR-based SNP (Single Nucleotide Polymorphism) markers to confirm the polymorphisms and determine which pool and subpool contained the mutant individual. This provided an alternative method of confirming the polymorphism, as well as a usable marker for downstream breeding approaches.

2. Materials and Methods

2.1. Plant Material and Growth Conditions

Mutagenesis of soybean (*Glycine max*) cultivar Williams-82 with N-nitroso-N-methylurea (NMU), as previously described [22,28]. M₁ and M₂ plants were grown in the field in West Lafayette, Indiana, and leaf disc punches were obtained from each M₂ plant. For validation of phenotype, crosses to a mutant containing the *rs2* W331-allele (line KB10-23#1677) were performed in the field, and F₁ plants were advanced and genotyped in the greenhouse. F₂ populations were planted in the field during the subsequent growing season, and individuals were genotyped to select double and single mutants and wild type siblings [29]. PCR based genotyping was performed using CAPs or dCAPS markers specific for *rs2* and *rs3* mutant alleles (listed in Supplemental Table S1).

2.1. DNA Extraction, Amplification, Sequencing, and Mutant Identification

DNA was extracted from 100 mg freeze-dried leaf tissue using the Omega EZ 96 Plant DNA kit (Omega Bio-tek, Norcross, GA, USA). Samples were quantified by visualization on agarose gels and digital densitometry with ImageQuant software (GE Healthcare) and diluted to a concentration of 9 ng/μL. These individual samples were diluted to a working concentration of 3–6 ng/μL in 8-sample pools. PCR primers (listed in Supplemental Table S1) were designed to amplify exons of the *RS2* (Glyma.06g179200) and *RS3* (Glyma.05g003900) genes. We chose to pursue amplification of exon 1 of the *RS2* and *RS3* gene for the possibility of the introduction of an early termination codon and because of the reliable and robust amplification observed from these primer sets. Amplification was performed (95 °C for 60 s, 5 cycles of 94 °C 30 s, 54 °C 20 s, 68 °C for 180 s, the 25 cycles of 94 °C 30 s, 56 °C 20 s, 68 °C for 180 s) on DNA sub-pools (8 samples) using the Advantage 2 high fidelity polymerase mix (Takara Biosciences). Subpool amplicons were combined into larger pools of 512 samples, purified using Ampure (Beckman Coulter, Indianapolis, IN, USA), quantified using fluorimetry on the Qubit 2.0 (Life Technologies, Carlsbad, CA, USA). In addition, 1.5 μg of each sample was sheared into 200 bp fragments, and Illumina libraries were constructed using standard protocols as described previously [29]. Three libraries were prepared for the *RS2* amplicon, and 3 libraries were prepared for the *RS3* amplicon. Each library included a non-overlapping set of M₂ individuals. The *RS2* and *RS3* amplicons were sequenced for a total of 1533 unique M₂ individuals. DNA from PI 200508, which carries multiple point mutations in the *RS2* gene was included as a sample in each *RS2* library to act as a positive control for mutation detection. Sequencing was performed on the Illumina HiSeq2500, to obtain paired-end reads of 150 bp.

2.3. Sequence Analysis

Illumina sequence assembly was performed with bowtie2 [30] to reference soybean genomic sequences corresponding to our targeted regions from the Williams 82 a2.v1 genomic sequence obtained from Phytozome [31]. The samtools pileup utility was used to identify sequence variants and determine the variant frequency, and output frequency was parsed from the alignment via the mpileup-tools suite (run mpileup.py and mpileup-parser-v2.py) obtained from (<http://comailab.genomecenter.ucdavis.edu/index.php/mpileup>, last visited 10/2019) as described [27]. Mpileup was used with the following parameters: -d 8000000-q 21-Q 21 (maximum depth per base in each input BAM (binary alignment map) file of 8 million, minimum base, and minimum mapping quality scores of 21). Base frequencies that substantially exceeded background values were targeted as potential induced variation (see Results section). For each predicted mutation identified in the parsed pileup file, a manual comparison with predicted coding sequence was then used to determine if the mutation caused a non-synonymous change in the coding sequence, and SIFT (Sorting Intolerant From Tolerant) was used to estimate the likelihood of a damaging mutation and further select targets for PCR-based mutation screening [32]. Only changes predicted to be both non-synonymous and damaging were pursued for isolation by the secondary screen.

2.4. SNP Assays

For each novel mutation we pursued, a positive control was generated in the form of a plasmid carrying the mutated sequence. First, *RS2* or *RS3* amplicons from wild type Williams-82 were inserted into the TOPO-TA plasmid (Invitrogen, Carlsbad, CA, USA) and validated by Sanger dye-terminator sequencing using standard M13 forward and reverse sequencing primers. Second, selected polymorphisms were introduced into the soybean sequence in the plasmid using the QuikChange Site-Directed Mutagenesis kit (Agilent, Santa Clara, CA, USA) following standard kit protocols. Cleaved amplified polymorphic sequence (CAPS) and derived CAPS (dCAPS) genotyping assays to discriminate the wild-type from the novel mutant sequence were designed using dCAPS Finder [33] and tested on the mutated plasmids, which became positive controls for pool screening. Amplification using mutation discriminating CAPS/dCAPs assays was carried out on pools of 8 M₂ samples, and finally on individual plants once a pool was found to carry a mutation of interest. These assays were also used to genotype F₁ and F₂ plants.

2.5. Carbohydrate Measurement

Crosses were performed to combine mutant *rs2* (W331-) and *rs3* genes. Single F₃ seeds produced in the field in the 2017 growing season were genotyped for the 2 genes and analyzed for carbohydrate composition by HPLC. In the 2018 growing season, F₃ plants were propagated, and bulk seed from single F₄ plants (5 of each genotype) was analyzed by HPLC. A detailed protocol for HPLC from soybean seed was previously described [10]. Statistical significance was calculated with a two-tailed, type 2 *t*-test between wild type and mutant values, with either 8 single seed samples (for 2017 data) or 5 single plant bulk samples (for 2018).

3. Results

3.1. Sequencing of *RS2* and *RS3* Amplicons

A fragment of the *RS2* gene corresponding to exon 1 was amplified from pooled mutant genomic DNA with a sample of DNA from the previously described *rs2* mutant (from the PI 200508 accession) [6] included as a spike-in control, and PCR amplicons were pooled for construction of Illumina libraries. From the sequence data, the approximate coverage of each fragment in each library was calculated. For *RS2* libraries 1, 2, and 3, we obtained 43 million, 53 million, and 45 million paired-end 100 bp reads, respectively, totaling 4.4–5.3 Gb of sequence in each, each with an overall alignment rate > 99%. For *RS3* libraries 1, 2, and 3, we obtained 42 million, 37 million, and 44 million paired ends reads, respectively, totaling 3.8–4.4 Gb each. Coverage was not even across the amplicons, but much higher at the ends of the fragments as expected (Supplemental Figure S1). We estimated that minimum coverage per base per plant (considering 512 plant samples per library) was 2000–3000. In the *RS2* sequence, we found a known polymorphism from PI 200508—several polymorphisms in the *RS2* gene have been described from this accession [6], including a G to C change that occurred 598 bp downstream of the translation start site (720 bp from the start of our amplification primer and the sequence obtained in our experiment) in an area of relatively deep coverage of 4.8 million to 5.9 million. This SNP was present at a frequency of 0.83%, 2.2%, and 0.37% (different fraction of representation for each library) roughly consistent with our estimate. Any single base in a diploid individual should be covered by 0.1% of reads. Small variations in sample concentration and amplification efficiency may be exaggerated through the TbyS procedure, which resulted in the observed range of abundance. However, we found many putative polymorphisms that occurred in approximately 0.1% of the reads at any given reference base and suspected that many might be a result of PCR error (Supplemental Figure S2). Thus, to identify new SNPs in the mutant population, an empirically determined frequency of polymorphism from our spike-in control was used as a starting point. After calling SNPs that appeared in the sequenced DNA, we prioritized screening SNPs that met two criteria: (1) An occurrence above a cutoff of 0.15% (above the noise level but somewhat lower than the expected frequency for a homozygous mutation); (2) that created non-

synonymous substitutions or nonsense mutations at a conserved position and were, therefore, predicted to be damaging for protein function. We preferred to select SNPs that resulted in a G to A or C to T change diagnostic of NMU-induced mutation (and less likely to be PCR error), and that was unique to one library (thus not a result of amplification of sequences from a homologous sequence in the soybean genome). However, we screened for both transitions and transversions, when the amino acid change was predicted to be damaging (including mutations near the splicing sites of the *RS3* gene). After successful positive control and genotyping assay design, we pursued five putative polymorphisms in the *RS2* gene, and five putative mutations in the *RS3* gene (Supplemental Table S2).

3.2. Polymorphism Validation

CAPS or dCAPS assays were designed to screen the pools for putative polymorphisms (Supplemental Table S1). We targeted five putative mutations in the *RS2* amplicon for secondary screening by dCAPS assay (Supplemental Table S2). It was determined that two of the identified SNPs (L180P and S194N) both originated in one plant sample, and we determined that this sample was, therefore, likely to be a contaminating the non-Williams-82 individual in the original M_1 population (Supplemental Figure S3). While we chose not to pursue this variation in *RS2*, this finding served as partial validation of the TbyS method. We were unable to validate the other sequence polymorphisms that we selected from the *RS2* amplicon in the population, and we hypothesized that they could be due to PCR-generated errors early in the amplification of the products used for sequencing or in library construction.

Five putative mutations in the *RS3* amplicon were targeted (Supplemental Table S2), including two sites that were flagged as putatively interfering with splicing. One pool demonstrated a positive band using the confirmatory dCAPS marker for a missense mutation in the *RS3* gene (Figure 1a). It was further determined that this polymorphism existed in a single plant sample (Figure 1b). PCR results suggested that the mutation was homozygous in the M_2 plant in the original population. A G to A polymorphism 223 base pairs from the translation start site created a glycine to glutamic acid change at position 75 in the predicted amino acid sequence. The glycine 75 residue was conserved across the soybean raffinose synthase family and in *Arabidopsis* raffinose synthase enzymes (Figure 1c). We will refer to this mutation as *rs3G75E*. The dCAPS primers and the G to A polymorphism in the mutant introduced a restriction site for the XmnI enzyme, and we used this assay (Figure 1d) developed for screening the pools to identify individuals for phenotypic screening. We were unable to validate and isolate individuals with the other polymorphisms that we observed in the *RS3* amplicon sequence.

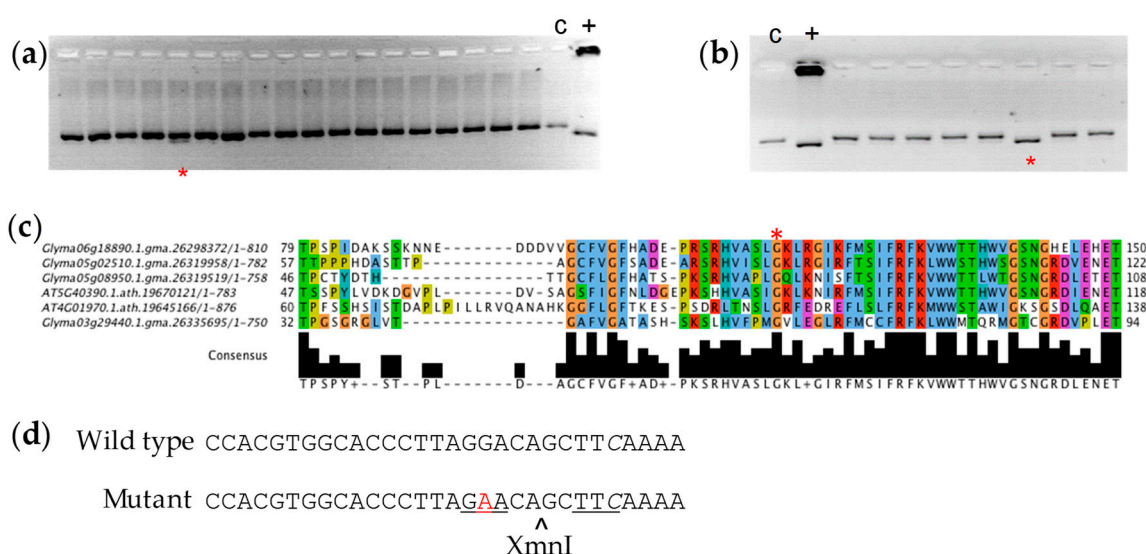


Figure 1. Polymorphism identified in the *RAFFINOSE SYNTHASE3* gene. The presence of multiple bands in (a) subpool PCR and (b) individual sample PCR (asterisk) after agarose gel electrophoresis indicates single nucleotide polymorphism in *RS3*. c = Wild type genomic DNA control, + = positive control for mutation (see Methods). (c) Mutation in *RS3* affects conserved glycine residue at position 75 in the amino acid sequence. (d) DNA sequence surrounding the G to A polymorphism (red) in the *rs3^{G75E}* mutant. The italicized C is where the mismatch in the genotyping primer introduces the *XmnI* site (underlined) that is cut in the mutant but not the wild-type sequence.

3.3. Low RFO Phenotype in *rs2rs3* Double Mutants

Based on previous studies with *rs3* mutants, we suspected that it would be difficult to observe a phenotype in the *rs3* single mutant, confirming the expectation on initial analyses for raffinose and stachyose content that showed no effect in the single *rs3^{G75E}* mutant. We expected that the *rs3* mutation would, however, enhance the phenotype of *rs2^{W331-}* mutants. We crossed *rs3^{G75E}* to *rs2^{W331-}*, and 11 individuals (to control for variation in the genetic backgrounds of the two parents) were selected from the *F*₂ generation of the cross using PCR-based genotyping markers. The seed was harvested from individual *rs2* and *rs3* single mutants, as well as the *rs2rs3* double mutant and wild type siblings. Analysis of seed composition by HPLC in Figure 2 for two field seasons showed that the seeds wild-type for both *RS2* and *RS3* have ~4% sucrose, 1% raffinose, and 3% stachyose. Mutation in *RS3* alone has no significant effect on carbohydrate content. As previously observed, the mutation in *RS2* reduced raffinose content to 0.2%, and stachyose levels to less than 1% with a modest increase in sucrose levels. As observed in other studies [10], galactinol was undetectable in lines with the wild type allele of *RS2*. A combination of the *rs2* and *rs3* mutations reduced residual stachyose content, and sucrose increased to 5%. The reductions in raffinose and stachyose were statistically significant in the *rs2rs3* double mutant in both years.

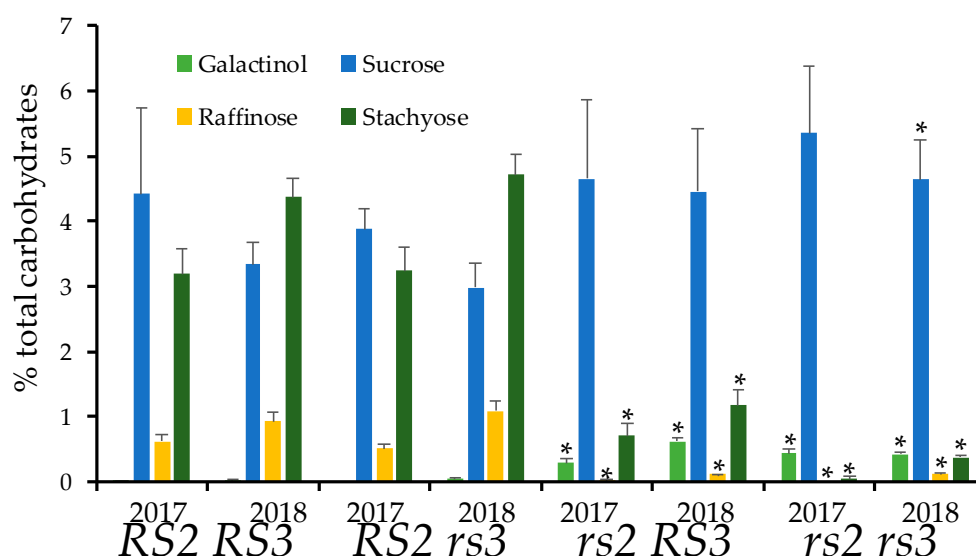


Figure 2. Reduction of raffinose family oligosaccharides (RFOs) in *rs2 rs3* double mutant. Single plants (5 per genotype) were harvested in 2017 and 2018 and assayed for carbohydrate content. Error bars indicate standard deviation, asterisk indicates statistical significance (two-tailed, type 2 t-test between wild type and the double or single mutant) at *p* = 0.05.

4. Discussion

Lower levels of raffinose and stachyose, along with increased levels of sucrose in soybean seed, have been demonstrated to improve the early growth of animals fed on soybean meal rations [4]. While mutation in *RS3* alone does not make a significant difference to raffinose and stachyose levels

in the soybean seed, we have shown that in combination with *rs2_{W331-}*, *rs3_{G75E}* can dramatically reduce levels of raffinose and stachyose, while increasing sucrose levels. The G75E mutation results in a reduction in raffinose and stachyose similar to that of a previously isolated *rs3* mutant allele [10], a variant of the *RS3* sequence that contains two distinct missense mutations that reduce RFOs in combination with a mutant *rs2* allele [11]. Recently, an additional haplotype variant in *rs3* has been demonstrated to have a similar additive phenotype when in combination with *rs2*. While there are distinguishing polymorphisms outside the *rs3* open reading frame in this accession, the mechanism of action of this allele is not understood [34]. As reported here, *rs3_{G75E}* is an alternative source that can be used for further non-transgenic improvement in the soybean meal profile, and in the presence of *rs2_{W331-}* reduces raffinose and stachyose to levels comparable to these other ultra-low RFO genotypes. Further studies using the combination of *rs2_{W331-}* and *rs3_{G75E}* will be needed to identify lines where sucrose levels are increased in the double mutants. In this study, genetic loci from the parents controlling physiological maturity were not fixed, and it has been observed that maturity effects contribute significantly to sucrose levels [8]. Interestingly, the roles for *RAFFINOSE SYNTHASE1* and *RAFFINOSE SYNTHASE4* in soybean seed composition remain unknown, and it is untested whether the elimination of these enzymes could further improve the soybean seed carbohydrate profile.

In this case, the reverse genetic TbyS approach enabled the isolation of a new loss-of-function allele of *RAFFINOSE SYNTHASE3* in soybean. We believe that our approach of using CAPS or dCAPS markers to isolate and validate mutant lines represent perhaps the easiest and lowest cost method of this type. The *rs3_{G75E}* mutation has no effect on raffinose or stachyose content on its own, and, therefore, could not have been detected using a forward genetics-based phenotypic screen, unless the *rs2* mutation was present in the population before mutagenesis. An advantage to this method is cost: The sequencing-based method of TILLING does not require specific mutation-detection equipment beyond a high throughput sequencing instrument, such as those found in many core facilities. Pooling large numbers of samples into each library reduces the cost of sequencing and library construction. The CAPS/dCAPS screening method is fairly inexpensive and easy to implement in any basic molecular lab, which allows for the inexpensive design of assays for validation of potentially damaging polymorphisms while minimal effort is expended on silent mutations, and the SNP marker can be used effectively in population genotyping. As observed in other TbyS experiments, many sequence polymorphisms [35] were observed in the initial sequencing data, at frequencies that made it difficult to discriminate between false and true positives, and, notably, PCR errors include the same types of nucleotide substitutions as those induced by the mutagen [27,35–38]. Alternative multi-dimensional pooling strategies (or initial amplification from smaller pools) would circumvent this problem (although with additional costs), and as library construction becomes more routine and less costly, we hope that the TILLING-by-sequencing approach can be used in soybean to identify more mutations with less noise. We are currently applying this successful, straightforward, and inexpensive method to other gene targets.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Sequence coverage for *RS2* and *RS3* amplicons, Figure S2: Sequence polymorphisms in the *RS2* and *RS3* amplicon sequences, Figure S3: Detection of *RS2* polymorphisms, Table S1: Primer sequences, Table S2: Unique sequence polymorphisms identified in *RS2* and *RS3* amplicons.

Author Contributions: Conceptualization, K.M.R. and K.H.; data curation, R.T. and K.H.; formal analysis, R.T. and K.H.; funding acquisition, K.M.R.; investigation, R.T. and M.C.-C.; methodology, M.C.-C.; validation, M.C.-C.; writing—original draft, K.H.; writing—review and editing, R.T., M.C.-C., and K.M.R.

Funding: Funding for this project was provided by USDA-ARS (CRIS Project #5020-21000-007 “Genetic Enhancement of Seed Quality and Plant Health Traits, and Designing Soybeans with Improved Functionality”) and the United Soybean Board (USB Project #2020-152-0119 “Modifying Soluble Carbohydrates in Soybean Seed for Enhanced Nutritional Energy Meal”).

Acknowledgments: The authors are grateful to K. Bilyeu for the KB10-23#1677 line that contained the *rs2_{W331-}* deletion mutation and for carbohydrate analyses, and to P. SanMiguel and the Purdue Genomics Core Facility for library construction and sequencing services. The authors are grateful to Matthew Hudson for critical reading of the manuscript. Mention of trade names or commercial products in this publication is solely for the purpose

of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. www.soystats.com. Soystats.com. Available online: <http://www.soystats.com/2012/Default-frames.htm> (accessed on 19 October 2019).
2. Naczek, M.; Amarowicz, R.; Shahidi, F. α -Galactosides of Sucrose in Foods: Composition, Flatulence-Causing Effects, and Removal. In *Antinutrients and Phytochemicals in Food*; American Chemical Society: Washington, DC, USA, 1997; Volume 662, pp. 127–151.
3. Hartwig, E.E.; Kuo, T.M.; Kenty, M.M. Seed Protein and its Relationship to Soluble Sugars in Soybean. *Crop Sci.* **1997**, *37*, 770–773, doi:10.2135/cropsci1997.0011183X003700030013x.
4. Valentine, M.F.; De Tar, J.R.; Mookkan, M.; Firman, J.D.; Zhang, Z.J. Silencing of Soybean Raffinose Synthase Gene Reduced Raffinose Family Oligosaccharides and Increased True Metabolizable Energy of Poultry Feed. *Front. Plant Sci.* **2017**, *8*, 692, doi:10.3389/fpls.2017.00692.
5. Karr-Lilienthal, L.K.; Kadzere, C.T.; Grieshop, C.M.; Fahey, G.C. Chemical and nutritional properties of soybean carbohydrates as related to nonruminants: A review. *Livest. Prod. Sci.* **2005**, *97*, 1–12, doi:10.1016/j.livprodsci.2005.01.015.
6. Dierking, E.C.; Bilyeu, K.D. Association of a Soybean Raffinose Synthase Gene with Low Raffinose and Stachyose Seed Phenotype. *Plant Genome* **2008**, *1*, 135–145, doi:10.3835/plantgenome2008.06.0321.
7. Dierking, E.C.; Bilyeu, K.D. New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biol.* **2009**, *9*, doi:10.1186/1471-2229-9-89.
8. Jo, H.; Lorenz, A.J.; Rainey, K.M.; Shannon, J.G.; Chen, P.Y.; Bilyeu, K.D. Environmental Stability Study of Soybeans with Modified Carbohydrate Profiles in Maturity Groups 0 to V. *Crop Sci.* **2019**, *59*, 1531–1543, doi:10.2135/cropsci2018.09.0600.
9. Jo, H.; Lee, J.D.; Bilyeu, K.D. Environmental Stability of Carbohydrate Profiles in Different Soybean Genotypes. *Crop Sci.* **2018**, *58*, 773–782, doi:10.2135/cropsci2017.08.0497.
10. Bilyeu, K.D.; Wiebold, W.J. Environmental Stability of Seed Carbohydrate Profiles in Soybeans Containing Different Alleles of the Raffinose Synthase 2 (RS2) Gene. *J. Agric. Food Chem.* **2016**, *64*, 1071–1078, doi:10.1021/acs.jafc.5b04779.
11. Schillinger, J.A.; Dierking, E.C.; Bilyeu, K.D. Soybeans Having High Germination Rate and Ultra-Low Raffinose and Stachyose Content. U.S. Patent No 8471 107, 21 June 2013.
12. Hagely, K.B.; Palmquist, D.; Bilyeu, K.D. Classification of Distinct Seed Carbohydrate Profiles in Soybean. *J. Agr. Food Chem.* **2013**, *61*, 1105–1111, doi:10.1021/jf303985q.
13. Leske, K.L.; Akavanichan, O.; Cheng, T.K.; Coon, C.N. Effect of Ethanol Extract on Nitrogen-Corrected True Metabolizable Energy for Soybean-Meal with Broilers and Roosters. *Poult. Sci.* **1991**, *70*, 892–895, doi:DOI 10.3382/ps.0700892.
14. Wang, Z.; Zhu, Y.; Wang, L.L.; Liu, X.; Liu, Y.X.; Phillips, J.; Deng, X. A WRKY transcription factor participates in dehydration tolerance in *Boea hygrometrica* by binding to the W-box elements of the galactinol synthase (BhGolS1) promoter. *Planta* **2009**, *230*, 1155–1166, doi:10.1007/s00425-009-1014-3.
15. Koster, K.L.; Leopold, A.C. Sugars and Desiccation Tolerance in Seeds. *Plant Physiol.* **1988**, *88*, 829–832, doi:DOI 10.1104/pp.88.3.829.
16. Zuther, E.; Buchel, K.; Hundertmark, M.; Stitt, M.; Hinch, D.K.; Heyer, A.G. The role of raffinose in the cold acclimation response of *Arabidopsis thaliana*. *FEBS Lett.* **2004**, *576*, 169–173, doi:10.1016/j.febslet.2004.09.006.
17. Blochl, A.; Peterbauer, T.; Richter, A. Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. *J. Plant Physiol.* **2007**, *164*, 1093–1096, doi:10.1016/j.jplph.2006.10.010.
18. Li, T.; Zhang, Y.M.; Wang, D.; Liu, Y.; Dirk, L.M.A.; Goodman, J.; Downie, A.B.; Wang, J.M.; Wang, G.Y.; Zhao, T.Y. Regulation of Seed Vigor by Manipulation of Raffinose Family Oligosaccharides in Maize and *Arabidopsis thaliana*. *Mol. Plant* **2017**, *10*, 1540–1555, doi:10.1016/j.molp.2017.10.014.

19. Gangl, R.; Tenhaken, R. Raffinose Family Oligosaccharides Act As Galactose Storesin Seeds and Are Required for Rapid Germination of Arabidopsis in the Dark. *Front. Plant Sci.* **2016**, *7*, 1115, doi:10.3389/fpls.2016.01115.
20. Dierking, E.C.; Bilyeu, K.D. Raffinose and stachyose metabolism are not required for efficient soybean seed germination. *J. Plant Physiol.* **2009**, *166*, 1329–1335, doi:10.1016/j.jplph.2009.01.008.
21. La, T.; Large, E.; Taliercio, E.; Song, Q.; Gillman, J.D.; Xu, D.; Nguyen, H.T.; Shannon, G.; Scaboo, A. Characterization of Select Wild Soybean Accessions in the USDA Germplasm Collection for Seed Composition and Agronomic Traits. *Crop Sci.* **2019**, *59*, 233–251, doi:10.2135/cropsci2017.08.0514.
22. Cooper, J.L.; Till, B.J.; Laport, R.G.; Darlow, M.C.; Kleffner, J.M.; Jamai, A.; El-Mellouki, T.; Liu, S.; Ritchie, R.; Nielsen, N.; et al. TILLING to detect induced mutations in soybean. *BMC Plant Biol.* **2008**, *8*, 9, doi:10.1186/1471-2229-8-9.
23. Hoshino, T.; Watanabe, S.; Takagi, Y.; Anai, T. A novel GmFAD3-2a mutant allele developed through TILLING reduces alpha-linolenic acid content in soybean seed oil. *Breed. Sci.* **2014**, *64*, 371–377, doi:10.1270/jsbbs.64.371.
24. Lakhssassi, N.; Zhou, Z.; Liu, S.; Colantonio, V.; AbuGhazaleh, A.; Meksem, K. Characterization of the FAD2 Gene Family in Soybean Reveals the Limitations of Gel-Based TILLING in Genes with High Copy Number. *Front. Plant Sci.* **2017**, *8*, 324, doi:10.3389/fpls.2017.00324.
25. Hoshino, T.; Takagi, Y.; Anai, T. Novel GmFAD2-1b mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with GmFAD2-1a mutant alleles. *Breed. Sci.* **2010**, *60*, 419–425, doi:10.1270/jsbbs.60.419.
26. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463*, 178–183, doi:10.1038/nature08670.
27. Tsai, H.; Howell, T.; Nitcher, R.; Missirian, V.; Watson, B.; Ngo, K.J.; Lieberman, M.; Fass, J.; Uauy, C.; Tran, R.K.; et al. Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiol.* **2011**, *156*, 1257–1268, doi:10.1104/pp.110.169748.
28. Ritchie, R.; Wiley, H.; Szymanski, D.; Nielsen, N. Targeting Induced Local Lesions in Genomes—TILLING. In *Legume Crop Genomics*; Wilson, R.F., Stalker, H.T., Brummer, E.C., Eds.; AOCS Press: Champaign, IL, USA, 2004; pp. 194–203.
29. Thapa, R.; Carrero-Colón, M.; Addo-Quaye, C.; Held, J.; Dilkes, B.; Hudson, K.A. New Alleles of FAD3A Lower the Linolenic Acid Content of Soybean Seeds. *Crop Sci.* **2018**, *58*, 713–718, doi:10.2135/cropsci2017.08.0490.
30. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357, doi:10.1038/Nmeth.1923.
31. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186, doi:10.1093/nar/gkr944.
32. Sim, N.L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457, doi:10.1093/nar/gks539.
33. Neff, M.M.; Neff, J.D.; Chory, J.; Pepper, A.E. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: Experimental applications in Arabidopsis thaliana genetics. *Plant J.* **1998**, *14*, 387–392.
34. Hagely, K.B.; Jo, H.; Kim, J.-H.; Hudson, K.A.; Bilyeu, K. Molecular assisted breeding for improved carbohydrate profiles in soybean seed. *Theor. Appl. Genet.* **2019**, in press.
35. Guo, Y.; Abernathy, B.; Zeng, Y.; Ozias-Akins, P. TILLING by sequencing to identify induced mutations in stress resistance genes of peanut (*Arachis hypogaea*). *BMC Genom.* **2015**, *16*, 157, doi:10.1186/s12864-015-1348-0.
36. Kim, S.-I.; Tai, T.H. Identification of novel rice low phytic acid mutations via TILLING by sequencing. *Mol. Breed.* **2014**, *34*, 1717–1729, doi:10.1007/s11032-014-0127-y.
37. Potapov, V.; Ong, J.L. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS ONE* **2017**, *12*, e0169774, doi:10.1371/journal.pone.0169774.

38. Shagin, D.A.; Shagina, I.A.; Zaretsky, A.R.; Barsova, E.V.; Kelmanson, I.V.; Lukyanov, S.; Chudakov, D.M.; Shugay, M. A high-throughput assay for quantitative measurement of PCR errors. *Sci. Rep.* **2017**, *7*, 2718, doi:10.1038/s41598-017-02727-8.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).