MOLECULAR APPROACHES TO IMPROVE SOYBEAN PLANT ARCHITECTURE AND SEED COMPOSITION

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

JEONGHWA KIM

Dr. Kristin Bilyeu, Dissertation Supervisor

December 2021

The undersigned, appointed by the Dean of the Graduate School,
have examined the dissertation titled:

MOLECULAR APPROACHES TO IMPROVE SOYBEAN PLANT ARCHITECTURE AND SEED COMPOSITION

presented by Jeonghwa Kim,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Kristin Bilyeu

Dr. Felix Fritschi

Dr. Andrew Scaboo

Dr. Trupti Joshi

DEDICATIONS

To my parents, Taeseop Kim and Hyeonju Lee, who have sacrificed their lives for me.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my appreciation to my supervisor, Dr. Kristin Bilyeu. Kristin is a person who taught me to grow intellectually, and look at the world in a broader and wiser perspective. Throughout my PhD journey, she has always believed in me and supported me not only academically but also personally. For me, she is not just a dissertation advisor, but also a role model. She has succeeded as a woman in science, as well as a human being; she is consistently nonjudgmental and positively minded. I am sincerely grateful that I was able to work with her for the past several years of my life.

In addition, I would like to thank my dissertation committee members, Dr. Felix Fritschi, Dr. Andrew Scaboo, and Dr. Trupti Joshi. With their advice and support, I walked the right path during my PhD and completed my dissertation. I am honored to have them on my dissertation committee.

I also would like to express my sincere appreciation to the two technicians in Dr. Bilyeu's lab: Christine Cole and Mark Johnston. For my dissertation research, Christine analyzed the fatty acid composition, conducted genotyping, and helped to measure phenotypes for plant architecture study. Mark also helped measure the phenotype data for the plant architecture study and aided my analysis of carbohydrate composition. Without their help and support, my dissertation research would not have been successfully completed; and I am profoundly grateful. Furthermore, the research findings that I present here would have been impossible to achieve without the work of former graduate students in Dr. Bilyeu lab, Anh Pham, Emily Dierking, Katherine Hagely, and Hyun Jo. I

really appreciate all their hard work and achievements for fulfilling seed composition projects.

Finally, there are so many people who I would like to personally express my appreciation. First, I would like to thank to Dr. Jeong-Dong Lee. He was my advisor during my masters at Kyungpook National University, and he gave me the opportunity to experience broader worlds academically. I deeply appreciate for all his advice, help, and support. Also, I would like to specially thank two of my lab-mates: Carrie Miranda and Nick Dietz. Thanks to these two people, I adjusted easily to graduate life, and also felt attached and involved. I really appreciate for their presence during my PhD journey. Furthermore, I would like to thank all of my 'non-academic' friends, Soobin Choi, Jiyoon Kim, Hyewon Lee, Bomi Kim, Dongho Lee, Taeseon Woo, Jiyung Hwang, and Yujin Jeong. I sincerely thank their love and support during my life in Columbia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACADEMIC ABSTRACT

Soybean is an important agronomic crop which has been widely used as a vegetable oil as well as a source of protein for animals and humans. Among various desired traits for genetic improvement to improve the value of soybean, yield and seed composition deserve attention considering its ever-increasing demand as well as where its overall value comes from. In terms of yield improvement, the current research suggested an alternative strategy to remodel the architecture of soybean plants for better adaptation, especially to high yield environments. From the field trials in three latitudinal environments in the US, tall determinate soybean types were found to change the overall plant architecture of soybean in both Midwest and Southern environments in the US by increasing plant height and node number and creating similar stem thickness and pod density at the stem tip compared to typical determinate types. The increase in pod-bearing nodes with lodging resistance has the potential to result in more yield, especially in high yield planting environments. Further agronomic merits need to be examined in diverse environments for its full utilization in soybean breeding. For the improvement of soybean seed composition, this research showed the feasibility of soybean variety development with an oil and meal value bundle. The results from eight environments indicated that the desired oil, the high oleic and low linolenic acid oil trait (HOLL; >70% oleic and <3% linolenic acid), and meal value traits, the reduced seed content of the raffinose family of oligosaccharides (RFO), can be successfully combined by genotype selection without interference. The modified seed composition traits will contribute to creating healthier, more oxidatively stable soybean oil and higher metabolizable energy soybean meal.

CHAPTER 1

Literature Review

Soybean [*Glycine max* (L.) Merr.], which refers to cultivated soybean, belongs to the family *Fabaceae*/*Leguminosae*, subfamily *Papilionoideae*, tribe *Phaseoleae*, subtribe *Glycininae*, genus *Glycine* Willd., and subgenus *Soja* (Moench) F.J. Herm. (Singh et al. 2007; Hymowitz 2008). The genus *Glycine* includes about 28 species and it is classified into two subgenera: *Glycine* and *Soja* (Asaf et al. 2017). The subgenera *Glycine* includes 23 wild perennial species which are native to Australia, while the subgenera *Soja* comprises the cultivated soybean and its annual wild counterpart, *Glycine soja* Sieb. & Zucc., which originated in East Asia (Palmer et al. 2016). The cultivated soybean, a self-pollinated diploid (2n=40) species became the basis of the development of modern soybean cultivars (Palmer and Hymowitz 2016).

History of soybean: origin, domestication, and dissemination

Modern cultivated soybean (*G. max*) is widely believed to be domesticated from *G. soja* ca. 6000 – 9000 years ago in East Asia (Carter et al., 2004; Kofsky et al., 2018). Historical literature stated several putative regions where domestication occurred, such as northeastern China, Yellow River Valley in central China, southern China, in a corridor from southwest to northeast China, or simultaneously at multiple regions but no broad consensus has yet been reached (Singh et al., 2007). The remaining mysteries in the history of domestication of soybean have been somewhat revealed by recent advances in whole-genome resequencing of wild and cultivated soybeans (Sedivy et al., 2017). Based on the results of molecular and genome-based studies as well as archeological findings, three contrasting hypotheses on the origin of modern soybean have recently been presented: single-origin hypothesis, multiple-origin hypothesis, and complex hypothesis

(Sedivy et al., 2017). The single-origin hypothesis states that domesticated soybeans were originated from *G. soja* through a single domestication event hat occurred between 5000 and 9000 years ago (Sedivy et al., 2017). Phylogenetic analysis using whole-genome SNPs of 302 wild, landrace, or improved soybeans supports the single-origin hypothesis by showing that all wild soybeans are clustered together in a same group and domesticated soybeans were derived from the single cluster of wild soybeans (Zhou et al., 2015). The multiple-origin hypothesis states *G. max* was diverged from *G. soja* from multiple independent domestication events in several regions between 5000-9000 years ago (Sedivy et al., 2017). Phylogenetic studies using chloroplast gene SNPs from the 302 resequenced accessions indicated that multiple maternal lines had been selected during domestication of soybean (Fang et al., 2016). Additionally, according to an analysis of chloroplast DNA variations from 326 wild and cultivated soybeans collected from several Asian countries, the cultivated accessions were found to have originated independently from multiple regions with different wild gene pools (Xu at al., 2002). Moreover, the genetic diversity and population structure analysis among various *G. max* landraces collected from China, Korea, and Japan showed that accessions from Korea and Japan were genetically distinct from those from China (Lee et al., 2011; Li and Nelson 2001). These findings, together with the fact that *G. soja* grows wild in East Asia including China, Korea, Japan, Taiwan, and far eastern Russia (Singh et al., 2007), support the multiple-origin hypothesis. In terms of complex hypothesis, domestication of soybean took place as a gradual process, including divergence of *G. soja* prior to the multiple domestication events (Sedivy et al., 2017). Divergence studies in soybean through genome sequence comparisons of *G. max* and *G. soja* suggested that divergence between

G. soja and *G. max* took place 0.27 or 0.8 million years ago, and the modern cultivated soybean was domesticated from a *G. soja*/*G. max* complex which is created during divergence procedures (Kim et al., 2010; Li et al., 2014). Since conflicting molecular and genomic evidence persist about domestication of soybean, further archeological, molecular, and genomic studies, especially about the *G. soja*/*G. max* complex relevant to the molecular bases of soybean domestication, are needed to shed light on the history of soybean domestication.

The history of the dissemination of soybean is only partially known due to the lack of historical records (Hymowitz, 2008). From about the 1st to 16th century A.D., soybean was disseminated through trade routes to many Asian countries, such as Japan, Indonesia, Malaysia, Myanmar, Vietnam, Thailand, Nepal, India, Philippines (Hymowitz, 2008). It was not until the 18th century that soybean began to be spread to Western countries. Historical records stated that soybean was first known to Europe in 1712 through a book of Engelbert Kaempfer, a German naturalist who lived in Japan for about two years in 1690s (Shurtleff et al., 2015). Later, soybeans were officially grown in botanical gardens in France in 1739, in England in 1790, and in Germany in 1794 (Singh et al., 2007; Shurtleff et al., 2015). The first dissemination of soybean to United States happened in 1765 by Samuel Bowen, a seaman of East India Company, who brought soybeans from China to Savannah, Georgia and planted them in Thunderbolt, Georgia, a few miles east of Savannah (Hymowitz and Harlan, 1983). However, it was not until the 20th century that soybean in the US, a crop first grown primarily for forage, hay, and green fertilizer, became an economically important crop (Singh et al., 2007). In 1917, Osborne and Mendel experimentally demonstrated that heated soybean meal has

nutritional benefits than raw soybean meal, which ultimately showed potential of soybean for human and animal consumption and raised the necessity of soybean processing industry (Hymowitz, 1990). In 1922, A.E Staley established the first soybean processing facilities in Decatur, Illinois, which became the foundation of soybean processing industries the Unites States (Singh et al., 2007). Presently, soybean is grown in ca. 119 million ha worldwide and continues to be recognized for its economic value as the world's largest oilseed crop as well as a nutritional source for animal protein feeding and human consumption (Grassini et al, 2021).

Yield-influencing factors

For a long time since the domestication of soybean, yield has been the major concern in soybean breeding efforts. The soybean seed yield is a multifactorial trait that is affected not only by the genetic potential that each soybean genotype has but also by several environmental factors (Li et al., 2020). Therefore, to maximize the yield potential of soybean, it is required to understand the multiple yield-influencing factors and apply the knowledge appropriately to breeding and cultivation systems.

Soybean is a photoperiod-sensitive short-day crop. Thus, different daytime lengths depending on the latitude of each production environment have a great impact on determining the production and adaptation of soybean in specific environments because it regulates the duration of vegetative and reproductive stages of soybean (Li et al. 2020). Significant yield reduction was observed when certain soybean cultivars were planted 2°N apart from their typical cultivation latitudes (Gai and Wang, 2001). Planting date affects the growth development, and yield of soybean since delayed planting is associated

with a reduction in the duration of the soybean growth stages (Hu and Wiatrak 2012). The production environments exposed to abiotic stresses, such as extremely temperature (below 12 °C or above 36 °C), as well as drought and flood conditions, can also result in the reduction of soybean yield (Li et al. 2017).

Soybean yield is a complex quantitative trait controlled by many genes along with environmental effects and interactions, and several genetic loci that govern soybean yield, such as seed size, number of seeds per pot, and other yield components have been identified (Hao et al. 2012; Sedivy et al. 2017). Potential approaches and perspectives to improve the yield potential of soybean, with the means of genetic improvement and plant breeding efforts, were well summarized by Liu et al. (2020). It was pointed out that the number of seeds per unit area and seed mass are the main factors that ultimately determine the soybean yield and several yield components affecting those need to be considered simultaneously. Specifically, soybeans with desired plant architecture showing appropriate plant height, shorter internode length, more internodes, few or no branches, moderate pod number per node, higher podding rate, a higher ratio of four seeds per pod, moderate 100-seed weight, smaller petiole angle, and shorter petiole were considered ideal for maximizing yield potential of soybean.

Stem termination types

Soybeans show a wide range in the abruptness of stem termination. The variations in stem termination have been named with several terms, such as pod-bearing habit (Woodworth, 1932), stem termination type (Bernard, 1972), or stem growth habit (Hartung 1981), but these basically refer to the same trait. The stem termination is a type

of post-flowering reproductive growth, which is associated with not only domestication and adaptation of soybean, but also other important characteristics such as flowering time and duration, node formation, maturity, water use efficiency, abiotic stress tolerance, and ultimately final yield of soybean (Bernard, 1972; Heatherly and Smith, 2004; Ping et al., 2014). Due to its agronomic significance, it is important to accurately classify stem termination type and appropriately use it in breeding programs targeting various production environments. However, the accurate characterization of stem termination type is sometimes difficult because it is influenced by other genetic and environmental factors (Liu et al., 2015). Therefore, it is necessary to understand the distinct morphological characteristics of each stem termination type, and to broaden our knowledge about how soybean accessions with each of stem termination type would respond differently under diverse production environments.

The distinct stem termination types primarily result from the differences in the timing at which soybeans terminate their apical stem growth. Based on the timing of the termination of apical stem growth, most soybean varieties are generally classified into three categories: indeterminate, determinate, and semi-determinate. The indeterminate varieties continue vegetative growth (i.e., production of new nodes with trifoliolate leaves) on apical meristems at the stem and branch apices for a long while after flowering until the beginning of seed fill (growth stage R5) (Bernard, 1972). Since the terminal growth continues as long as lateral growth, indeterminate varieties feature a tapered stem with little or no secondary lateral growth near the stem tip (Tian et al., 2010). In contrast, determinate soybean varieties abruptly halt vegetative growth at or soon after flowering begins, producing a thick stem due to the continuous lateral growth even after the

cessation of apical stem growth (Ping et al., 2014). In semi-determinate varieties, the stem tip ceases vegetative growth a few days earlier than the indeterminate ones, thus resulting in fewer main-stem nodes and somewhat shorter stem length than the indeterminate varieties (Hartung et al., 1981; Ping et al., 2014).

While the three categories of stem termination types have been generally accepted, an additional type, termed tall determinate, was introduced (Thompson et al., 1997). The tall determinate stem type was observed from two isolines, L91-8052 and L91-8060, generated from crosses between an indeterminate soybean cultivar 'Clark' and either 'Soysota' or 'Peking', respectively. The tall determinate isolines significantly delayed the timing of stem termination compared to typical determinate types and almost identically terminated their stem growth with semi-determinate types. Also, the isolines with tall determinate stem type showed similar plant height with those having semi-determinate stem type, which were ~70% of the mature plant height of the indeterminate parent, Clark; however, they showed significant differences in numbers of nodes and terminal leaflet areas compared to the semi-determinate types. Considering morphological characteristics at the top of the stem, the tall determinate isolines showed higher similarity with determinate types having a thick stem and long raceme

Genetic control of soybean stem termination

Classical genetic analyses identified two genes, *Dt1* and *Dt2*, regulating stem growth habit in soybean (Bernard 1972). In the *Dt1* genetic background, dominant *Dt2* genotypes produced semi-determinate phenotypes, and recessive *dt2* genotypes produced indeterminate phenotypes. In the recessive *dt1* genetic background, the phenotype was

determinate regardless of the alleles at *Dt2* locus because of an epistatic effect of the *dt1* allele on the expression of the *Dt2*. *Dt2* was completely dominant over *dt2*, while *Dt1* was partially dominant over *dt1* (Bernard, 1972; Ping et al., 2014). The *Dt1*/*dt1* heterozygote exhibited an intermediate or semi-determinate phenotype (Bernard, 1972).

The molecular basis of the genes regulating stem termination was elucidated. *Dt1* (designated as *GmTfl1* or *GmTFL1b*; *Glyma19g37890.1*) was found to be an ortholog of *Arabidopsis TERMINAL FLOWER1* (*TFL1*) (Tian et al., 2010; Liu et al., 2010). During early vegetative growth stages, the functional *Dt1* and recessive *dt1* alleles both function similarly and produce a similar level of transcripts in the shoot apical meristems (SAMs), which results in phenotypic similarity in stem growth in those development stages. But the recessive *dt1* allele abruptly loses the expression, concomitant with floral induction, which causes the determinate stem termination (Liu et al., 2010; Xu et al., 2013). The early loss of the expression in the recessive *dt1* background is considered to result from decreased binding affinity of *dt1* with interactors such as *GmFDLs*, bZIP transcription factors *FLOWERING LOCUS D* (*FD*) in soybean, possibly under competitive interaction *GmFTs*, soybean orthologs of *Arabidopsis FLOWERING LOCUS T* (*FT*) (Liu et al., 2010). In the case of the functional *Dt1* allele, their expression of *Dt1* was found to be also reduced after flowering even in the indeterminate plants, and the artificial down-regulation of the *Dt1* transcripts by VIGS showed suppression in the indeterminate terminal stem growth at the shoot apical meristems (SAMs) of the *Dt1* plants (Liu et al., 2010). Therefore, the fate of SAM in the functional *Dt1* background is considered to be determined by the quantity of the transcripts, just as in the recessive *dt1* background (Liu et al., 2010). The reduced expression of *Dt1* and the resulting suppression of the terminal

stem growth is found to be controlled by floral meristem (FM) identity genes, such as orthologues of *Arabidopsis AP1* and *LFY*, which are induced by the FT–FD complex (Sablowski, 2007; Yue et al. 2021). The expression of *Dt1* was also found to be influenced by daylength and the phytochrome A genes, *E3* and *E4* (Xu et al. 2013). Under short-day conditions, the expression of *Dt1* in indeterminate plants were very low in young seedlings regardless of the genotype at *E3* and *E4*; but, after those plants are transferred to long-day conditions, the expression was increased and maintained to various levels, depending on the genotypes at *E3* and LD light quality (Xu et al., 2013). Under long-day or natural daylength conditions, the expression of *Dt1* in the recessive *e3*/*e4* genotype background was almost not induced, and thus the stem growth of the indeterminate plants was terminated shortly after flowering under both natural daylength or long-day conditions (Xu et al., 2013).

Six independent missense mutations, hereafter called *dt1R62S*, *dt1L67Q*, *dt1P113L*, *dt1R130K*, *dt1H141R*, and *dt1R166W*, were identified while searching for allelic variations at the *Dt1* genic region (Tian et al., 2010; Liu et al., 2010; Yue et al., 2021). The majority of the six nonsynonymous mutations were found in a subset of *G. max* but rarely found in the *G. soja* accessions, which suggests that the artificial selection of the mutations took place during diversification of soybean (Tian et al., 2010). The specific functions of each missense mutation and potential differences in phenotypes resulting from each mutant allele need to be studied further. In 1997, before the six mutant alleles at the *Dt1* locus were identified, an additional allele, termed as *dt1-t*, was reported at the genic region, which was known to confer the alternative stem termination type, tall determinate (Thompson et al., 1997). The *dt1-t* allele was found to be allelic to

the *dt1* locus and independent of *Dt2* (Thompson et al., 1997). Since the initial characterization of the tall determinate genetic type, no further articles have been published to date, which thus calls for the necessity of further studies on the molecular basis of the allele.

In terms of *Dt2*, the second gene regulating stem termination type, it was shown to be a gain-of-function MADS domain factor gene which is classified into the *APETALA1/ SQUAMOSA* (*AP1/SQUA*) subfamily which include floral meristem (FM) identity genes *AP1*, *FUL*, and *CAL* in *Arabidopsis* (Ping et al. 2014). Quantitative differences in expression levels were observed between dominant *Dt2* and recessive *dt2*, and the increased expression of *Dt2* was found to downregulate functional *Dt1* in the shoot apical meristems (SAMs), which ultimately leads to the early conversion of the SAMs into reproductive inflorescences (Ping et al., 2014). Along with its function as a direct repressor of *Dt1*, *Dt2* was found to activate the putative floral integrator/identity genes, such as *GmSOC1*, *GmAP1*, and *GmFUL*, and thus promotes flowering in soybean (Zhang et al., 2019). A total of 37 SNPs in the non-coding region of *Dt2* were identified when it was cloned, but it is still required to clarify the causative mutations and further elucidate the molecular functions of the genes.

Practical uses of genes modifying plant architecture of soybean

The genes regulating stem termination types and timing of flowering and maturity can modify the overall plant architecture of soybean, and optimizing plant architecture appropriate to each of the various production environments is considered a

promising strategy to maximize the yield of soybean (Hartung et al., 1981; Yue et al., 2021).

Stem termination type has been a target of artificial selections for adaptation to different latitudinal environments. Studies about genotypic selection of stem growth habit showed that most soybean varieties in the northern part of China have the functional *Dt1* allele which confers the indeterminate stem termination type, while determinate types were predominant in the southern area (Tian et al. 2010; Liu et al. 2015). Similar tendencies on the distribution of stem termination types depending on different latitude were observed in the US production environments (Cooper, 1985; Hartung et al. 1981; Wilcox, 1987).

Historically, several breeding efforts have been made to genetically optimize plant architecture for its better yield potential and adaptation in diverse environments. During the late 1970s to early 1980s, soybean cultivars with semi-dwarf stature were released to overcome yield losses due to lodging (Cooper, 1985). These short stature soybeans were developed by introducing determinate (*dt1*) alleles into high-yielding indeterminate soybean backgrounds and found to improve yield specifically under the production environments with higher seeding rates (Cooper 1985; Beuerlein 1988). However, the semi-dwarf phenotype has limitations for their broader uses, since they require specific management practices, such as high seeding rates and narrow row spacing, and dense planting of soybeans has a much greater negative effect on the fertility of soybean (Cooper 1985; Liu et al. 2020).

The modification of soybean plant architecture for improved yield potential can also be achieved by manipulating gene expression regulators, such as *GmmiR156b*,

GmMYB14, and *GmWRI1b* (Bao et al 2019; Chen et al. 2021; Guo et al. 2020; Sun et al. 2019) Overexpression of *GmmiR156b* was found to have effects to create an increased number of long branches, nodes, and pods increased 100-seed weight, and thicker stem, and thus result in a 46%–63% increase in yield per plant (Sun et al. 2019). *GmMYB14*-overexpressing transgenic soybean plants showed semi-dwarf stature along with a decrease in internode length, leaf area, leaf petiole length, and leaf petiole angle, and improved yield in high density under field conditions (Chen et al. 2021). Transgenic soybean lines overexpressing *GmWRI1b* showed decreases in plant height and internode length, increases in numbers of node and branch, stem diameter, shoot dry weight, pod number per plant, and seed number per plant, thereby resulting in increases in yield per plant and yield per ha at three plant density levels under field conditions (Guo et al. 2020).

Economic importance of soybean

Soybean is one of the most important economic crops globally with its broader uses as an oilseed and a source of protein for animal and human consumptions. In 2020, about 339 million metric tons of soybean were produced worldwide, which represent 59% of world oilseed production (SoyStats, 2021). The United States is one of the largest soybean producers, along with Brazil, with approximately 29% of the total world soybean production in 2020 (SoyStats, 2021). The US soybean growers planted 83.1 million acres of soybeans and harvested 4.14 billion bushels of soybeans, and these were valued at $30.5 billion dollars (SoyStats, 2021). About 24.9 billion pounds of soybean oil was produced in the US in 2020 and 55% of the vegetable oil consumed in the US was

soybean oil. Soybean represents about 70% of the protein meal consumed globally and around 97% of total soybean meal production is used for animal feeding. And in relation to the consumption of soybean meal by livestock, poultry (60.8%) and swine (18.9%) account for the majority, and the rest of the soybean meal is used for beef, dairy, pet food, or aquaculture, among others. Since the uses of soybean seeds for soybean oil and meal account for most of its commercial value, the overall value of soybean comes from the vegetative oil and high protein meal for animal feeds.

Soybean oil and fatty acid composition

Soybean oil is mainly composed of triacylglycerols, containing three fatty acids attached to a glycerol backbone, along with non-glycerides fractions, including phytosterols, waxes, hydrocarbons, carotenoids, tocopherols, and phosphatides (De Sousa et al. 2014). The fatty acid compositions thus directly determine the overall quality of soybean oil. Soybean oil extracted from commodity soybeans consists of 11 % palmitic acid (16:0), 4 % stearic acid (18:0), 23 % oleic acid (18:1), 54 % linoleic acid (18:2), and 8 % linolenic acid (18:3) (Wilson 2004). Compared to other vegetable oils consumed globally, soybean oil features relatively high amounts of linoleic acid and linolenic acid, polyunsaturated fatty acids (PUFA). The linoleic and linolenic acids are essential fatty acids for humans which must be incorporated from the diet, and also precursors of other beneficial long-chain ω-3 PUFA, such as eicosapentaenoic acid (EPA, 20:5) and docosahexaenoic acid (DHA, 22:6). These fatty acids are known to have health benefits by preventing inflammation, cardiovascular diseases, and Alzheimer's disease, as well as promoting fetal development (Jo et al. 2020). Despite the health benefits of the PUFA

content, these fatty acids have been a target of reduction or removal when it comes to the uses of soybean oil, especially as cooking vegetable oil, since the high PUFA content causes low oxidative stability of soybean oil. The high level of PUFA in soybean oil is known to result in rancidity, rapid decrease in optimum flavor, and shortened storage time of manufactured food products (Warner and Fehr 2008). Therefore, soybean oil has typically undergone a chemical hydrogenation process. The hydrogenation process does improve overall oxidative stability of soybean oil by reducing amount of PUFA and increasing the more beneficial and oxidatively stable oleic acid oil content as a consequence; but at the same time it creates an unwanted by-product, 10–40 % *trans* fats, which have been regulated in foods due to their negative effects on health (Hu et al. 1997; FDA 2003; FDA 2015). Therefore, research has been directed towards improving soybean oil functionality without the hydrogenation process, through genetic improvements of soybean seed oil having increased oleic acid content at the expense of linoleic and linolenic acids. Several soybean germplasm lines having high oleic acid of over 80%, low linolenic acid of 1-3% of total fatty acid content, or both these two traits combined have been developed either by genetic engineering or molecular marker-assisted breeding approaches (Bilyeu et al., 2011; Brace et al., 2011; Buhr et al., 2002; Hoshino et al. 2010; Pham et al., 2010; Pham et al., 2011; Pham et al., 2012).

Genes for the modification of fatty acid profiles in soybean

Genes controlling oleic acid and PUFA content in soybean seed oil have been characterized (Bilyeu et al. 2003; Heppard et al. 1996; Schlueter et al. 2007; Schmutz et al. 2010). Two microsomal oleate fatty acid desaturase genes, *FAD2-1A*

(Glyma.10g278000) and *FAD2-1B* (Glyma.20g111000), were found to play a key role controlling the oleic acid level in developing soybean seed oil by converting oleic acid precursors to linoleic acid precursors (Anai et al., 2008; Dierking and Bilyeu, 2008). The *FAD2-1A* and *FAD2-1B* are homologous genes which are most closely related to one another with 99% identity in encoded amino acid sequence (Li et al., 2008). Several independent mutant alleles for *FAD2-1A* and independent mutant alleles for the *FAD2-1B* have been identified, and the following are relevant: *FAD2-1A* indel, a null allele derived from PI 603452; *FAD2-1A* S117N, a missense mutation derived from mutagenized Williams 82 line 17D; and *FAD2-1B* P137R, a missense mutation derived from PI 283327 (Dierking and Bilyeu 2009; Pham et al. 2010; Pham et al. 2011). The mutations identified in either *FAD2-1A* or *FAD2-1B* resulted in elevated oleic acid content in the range of 27–50 % of the seed oil (Anai et al. 2008; Combs et al. 2019; Dierking and Bilyeu 2009; Haun et al. 2014; Hoshino et al. 2010; Pham et al. 2010; Sweeney et al. 2017). The *FAD2-1A* indel allele caused a frameshift mutation that was a more severe mutation than the missense *FAD2-1A* S117N, and the indel allele was found to produce higher and more stable oleic acid content in soybean seed oil, presumably due to less residual enzymatic activity of *FAD2*.

There are three microsomal linoleate desaturase genes, *FAD3A* (Glyma.14g194300), *FAD3B* (Glyma.02g227200) and *FAD3C* (Glyma.18g062000), which are responsible for the conversion of linoleic acid precursors to linolenic acid precursors and thus control the linolenic acid content in soybean seed oil (Bilyeu et al. 2003; Bilyeu et al. 2005; Bilyeu et al. 2006; Heppard et al. 1996; Pham et al. 2011; Pham et al. 2012; Schlueter et al. 2007; Schmutz et al. 2010). Independent mutations have been

found for each of the three *FAD3* genes (Bilyeu et al., 2005; Bilyeu et al., 2006): *FAD3A* Splice site (G810A), a null allele derived from CX1512-44; and *FAD3C* G128E, a missense allele derived from CX1512-44 are relevant. Among the three *FAD3* genes, *FAD3A* was found to have a greater impact in modifying linolenic acid content in soybean seed than *FAD3B* and *FAD3C*, probably due to its higher expression in developing seeds (Bilyeu et al. 2003, 2005). A double mutation in both *FAD3A* and either *FAD3B* or *FAD3C* was shown to result in about 3% linolenic acid content, while triple mutations in *FAD3A*, *FAD3B*, and *FAD3C* lower the linolenic acid to just 1% of the seed oil.

For maximizing the functionality of soybean oil, four alleles responsible for the most dramatic increase in oleic acid and decrease in linolenic acid have been selected for the use in the molecular marker-driven breeding (Bilyeu et al. 2018a): *FAD2-1A* indel, *FAD2-1B* P137R, *FAD3A* Splice site (G810A), and *FAD3C* G128E. It has been confirmed that soybean lines having the combination of these four alleles can successfully produce the high oleic and low linolenic acid (HOLL) seed oil phenotype with over 80% oleic acid, 3–7% linoleic acid, and less than 3% linolenic acid (Hagely et al. 2021; Pham et al. 2011; Pham et al. 2012).

Soybean meal and anti-nutritional factors

Soybean is an important plant-based protein source in diets for animals and humans because of its high quantity and quality of protein for proper nutrition (Hagely et al., 2020). The dry weight of a typical soybean is comprised of about 20% oil, 40% protein, and 15% soluble carbohydrate (Openshaw and Hadley 1978). The whole soybean

seeds are subjected to processing procedure including cleaning, crushing, dehulling, flaking, and pre-press or solvent extraction, to extract oil fraction. The remaining flakes are used to produce soybean meal and diverse soybean protein products. As an animal protein feeding source, soybean is generally consumed in the form of the defatted soybean meal, which is the by-product of the processing for seed oil extraction (Banaszkiewicz 2011). Globally, about 69% of protein sources for animal feeds are derived from soybean meal, and the soybean meal is widely used in mixtures of animal feed for poultry, swine, cattle and fish. However, there are factors that adversely affect the efficiency of soybean meal by reducing digestibility and nutritional value. The negative effects are known to result from the presence of anti-nutritional factors in soybean seeds, such as protease inhibitors, soybean agglutinin, soluble carbohydrates, phytic acid and soyasaponins (Gilman et al., 2009; Hitz et al., 2002; Jo et al., 2018; Miao et al., 2018; Herman et al., 2016).

Soybeans contain high levels of protease inhibitors, such as Kunitz and Bowman-Birk trypsin inhibitors, which account for 6% of total seed protein (Orf and Hymowitz, 1979). Kunitz inhibitors are major inhibitors that specifically target trypsin and account for up to 91% of trypsin inhibitor activity in seeds (Deak et al., 2008). Bowman-Birk inhibitors are known to inhibit trypsin and chymotrypsin, although they account for a smaller fraction of total seed trypsin inhibitor activity (Deak et al., 2008). The activities of these trypsin inhibitors in the soybean meal restrict digestibility of protein and lipid in monogastric animals, and thus negatively affect their growth rate (Vagadia et al., 2017). Letcin is a sugar-specific protein which has the ability to bind to cell surfaces with a high degree of specificity towards the oligosaccharides and glycopeptides (Moreira et al.,

1991). Raw soybean contains lectin at a concentration of 10–20 g/kg; but lectins even in excess of 7 g/kg are known to be harmful for digestibility and unsuitable for consumption (George et al., 2008). Lectin molecules interact with specific glycoconjugate receptors, which results in agglutination of red blood cells in animals (George et al., 2008). Lectin binds to the intestinal epithelium, cause morphology changes in the intestine, and thus reduce the absorption of nutrients which ultimately cause reduction of growth rate or loss in body weight (Dersjant-Li 2002).

Phytic acid (phytate) is the storage form of phosphorus in seeds, and it accounts for up to 1-3% of soybean seed composition (Gilman et al. 2009; Redekar et al., 2015). The phytic acid is hardly digested by monogastric animals due to the lack of phytase enzyme (Bilyeu et al., 2008). The underutilized phytic acid in animal feeds can cause phosphorus pollution of groundwater and eutrophication of freshwater lakes and streams when it is excreted in manure (Hatten III et al., 2001; Gilman et al. 2009). Furthermore, phytic acid negatively affects the bioavailability of essential minerals due to chelation of the minerals (Bilyeu et al., 2008). Soyasaponins account for 0.6~6.2% of soybean seeds and are major components of secondary metabolites in soybean seeds (Park et al., 2016). Soyasaponins are divided into two groups, group A and 2,3-dihydro-2,5-dihy-droxy-6-methyl-4H-pyran-4-one (DDMP) saponins, and an acetylated terminal sugar at the C-22 position of group A saponins are known to cause bitterness and astringent aftertastes of soybean (Okubo et al., 1992). Also, soyasaponins are known to create bubbles during tofu production, which interfere with the processing procedure (Tsukamoto and Yoshiki 2005). Moreover, in some fish species, saponins play a crucial role in the onset of intestinal inflammation (Gu et al., 2016; Knudsen et al., 2008; Krogdahl et al., 2015).

These anti-nutritional factors necessitate additional processing of soybean: heat-treatment, solvent extraction, bioprocessing, and enzyme-treatment, which are energy- and cost-intensive (Zhou et al., 2016). These processes have not only the possibility to cause unintended consequences, such as reduced protein quality and digestibility, but also a downfall in that each type of process is specific to certain types of anti-nutritional factors, which results in incomplete removal of particular compounds (Berhow et al., 2020; Gilman et al., 2015).

Carbohydrates and RFOs in soybean

Carbohydrates in soybean seeds consist of non-starch polysaccharides, free sugars, and starch (Choct et al. 2010). The non-starch polysaccharides are mainly divided into insoluble fraction (mainly cellulose) and soluble fraction, including pectic polymers (Choct et al. 2010). These carbohydrates are known to play important roles in seed germination, seed desiccation tolerance and cold stress tolerance (Obendorf and Kosina 2011). For the whole carbohydrate composition, the majority consists of the non-starch polysaccharides, and the free sugars and starch represent about 10% and less than 1% of the total carbohydrates, respectively (Choct 1997; Macrae et al. 1993). The carbohydrate compositions are variable depending on genotype of each soybean and other affecting factors, such as production environments, weather conditions, harvest conditions, and post-harvest processing (Karr-Lilienthal et al., 2005).

When referring to soluble carbohydrates in soybean meal, it mainly refers to three components, sucrose, raffinose and stachyose. These soluble carbohydrates are closely associated with the functionality of soybean meal. Sucrose is fully digestible for

monogastric animals, thereby positively affecting metabolizable energy of soybean meal; but raffinose and stachyose, the Raffinose Family of Oligosaccharides (RFO), are indigestible in guts of monogastric animals including humans as well as poultry and swine, the major livestock who consume soybean meal. The poor digestion results from the lack of the relevant enzymes required, α-galactosidase, and thereby cause no weight gain, flatulence, diarrhea and other digestive distress (Wang et al., 2003). There are extra processing procedures, such as ethanol extraction, which have been employed to remove RFO in the soybean meal. But the processes were found to lead the reduction of sucrose at the same time with the reduction of RFO, thereby lowering the overall quality of soybean meal (Coon et al. 1990).

Genes for the modification of soluble carbohydrate profiles in soybean

Two genes responsible for raffinose biosynthesis have been identified (Dierking and Bilyeu, 2008; Dierking and Bilyeu, 2009; Hagely et al. 2013; Schillinger et al. 2013, 2018): *RS2* (Glyma.06g179200) and *RS3* (Glyma.05g003900). When functional, these genes generate raffinose synthase enzymes which belong to a group of hydrolase family enzymes which require two substates, galactinol and sucrose, and as a consequence produce the three-ring molecule raffinose, and a by-product, myo-inositol (Dierking and Bilyeu 2008). From an array of previous research, two independent variant alleles for each of the *RS2* and *RS3* have been identified and characterized (Dierking and Bilyeu 2008, 2009; Hagely et al. 2020; Jo et al. 2018; Thapa et al. 2019): *rs2W331*-, an indel in-frame allele with a three base pair nucleotide deletion normally encoding a highly conserved tryptophan at amino acid position 331 from the start codon of the *RS2* gene

derived from PI 200508; *rs2T107I*, a missense mutation of *RS2* derived by the TILLING method; *rs3snp6*, a SNP in the non-coding region found in a variety of released cultivars; and rs3G75, a missense mutation derived by the TILLING method. The *rs2W331-* was shown to have a greater impact compared to the *rs2T107I* allele in reducing RFO, and the *rs2W331-* genotype results in the Low RFO trait causing less than 2% seed RFO content (Hagely et al. 2013; Hagely et al. 2020; Jo et al. 2018; Jo et al. 2019). The two variant alleles of *RS3* were found to be responsible for Ultra-Low (UL) RFO trait when combined with *rs2W331-*, causing significant reduction in seed RFO contents up to less than 1% of the total soluble carbohydrates (Hagely et al. 2020; Jo et al. 2018; Jo et al. 2019). The impact of the *rs3snp6* and r*s3G75E* in the reduction of seed RFO contents was found to be similar, and also no significant differences were observed in galactinol and sucrose contents generated from soybeans having each of the mutant alleles (Hagely et al. 2020).

REFERENCES

Asaf, S., Khan, A.L., Aaqil Khan, M., Muhammad Imran, Q., Kang, S.M., Al-Hosni, K., Jeong, E.J., Lee, K.E. and Lee, I.J., Comparative analysis of complete plastid genomes from wild soybean (Glycine soja) and nine other Glycine species. PloS one, 2017. 12(8): p.e0182281.

Anai, T., Yamada, T., Hideshima, R., Kinoshita, T., Rahman, S.M. and Takagi, Y., Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by GmFAD2-1a. Breeding Science, 2008. 58(4): p.447-452.

Banaszkiewicz, T., Nutritional value of soybean meal. Soybean and nutrition, 2011. p.1-20.

Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., Qiu, D., Shan, Z., Yang, Z., Yuan, S. and Zhang, C., CRISPR/Cas9-mediated targeted mutagenesis of GmSPL9 genes alters plant architecture in soybean. BMC plant biology, 2019. 19(1): p.1-12.

Berhow, M.A., Singh, M., Bowman, M.J., Price, N.P., Vaughn, S.F. and Liu, S.X., Quantitative NIR determination of isoflavone and saponin content of ground soybeans. Food chemistry, 2020. 317, p.126373

Bernard, R.L., Two genes affecting stem termination in soybeans. Crop Science, 1972. 12(2): p.235-239.

Beuerlein, J.E., Yield of indeterminate and determinate semidwarf soybean for several planting dates, row spacings, and seeding rates. Journal of Production Agriculture, 1988. 1(4): p.300-303.

Bilyeu, K., Gillman, J.D. and LeRoy, A.R., Novel FAD3 mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Science, 2011. 51(1): p.259-264.

Bilyeu, K.D., Palavalli, L.A.V.A.N.Y.A., Sleper, D.A. and Beuselinck, P.R., Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science, 2003. 43(5): p.1833-1838.

Bilyeu, K.D., L. Palavalli, D.A. Sleper, and P.R. Beuselinck., Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science, 2005. 45(5): p.1830–1836.

Bilyeu, K., Palavalli, L., Sleper, D.A., and Beuselinck, P., Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science, 2006. 46(5): p. 1913-1918.

Bilyeu, K.D., Zeng, P., Coello, P., Zhang, Z.J., Krishnan, H.B., Bailey, A., Beuselinck, P.R. and Polacco, J.C., Quantitative conversion of phytate to inorganic phosphorus in soybean seeds expressing a bacterial phytase. Plant physiology, 2008. 146(2): p.468-477.

Bilyeu, K., Škrabišová, M., Allen, D., Rajcan, I., Palmquist, D.E., Gillen, A., Mian, R. and Jo, H., The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. Journal of the American Oil Chemists' Society, 2018a. 95(1): p.39-49.

Brace, R.C., 2010. Agronomic and seed traits of soybean lines with high-oleate concentration. Iowa State University.

Buhr, T., Sato, S., Ebrahim, F., Xing, A., Zhou, Y., Mathiesen, M., Schweiger, B., Kinney, A., Staswick, P. and Tom Clemente, Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. The Plant Journal, 2002. 30(2): p.155-163.

Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A. and Kong, F., 2017. Molecular mechanisms of flowering under long days and stem growth habit in soybean. Journal of Experimental Botany, 2017. 68(8): p.1873-1884.

Carter, T.E., Hymowitz, T. and Nelson, R.L., Biogeography, local adaptation, Vavilov, and genetic diversity in soybean. In Biological resources and migration (pp. 47-59). Springer, Berlin, Heidelberg., 2014.

Chen, L., Yang, H., Fang, Y., Guo, W., Chen, H., Zhang, X., Dai, W., Chen, S., Hao, Q., Yuan, S. and Zhang, C., Overexpression of GmMYB14 improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. Plant biotechnology journal, 2021. 19(4): p.702-716.

Combs, R. and Bilyeu, K., Novel alleles of fad2-1a induce high levels of oleic acid in soybean oil. Molecular Breeding, 2019. 39(6): p. 79.

Choct, M., Dersjant-Li, Y., McLeish, J. and Peisker, M., Soy oligosaccharides and soluble non-starch polysaccharides: a review of digestion, nutritive and anti-nutritive effects in pigs and poultry. Asian-Australasian Journal of Animal Sciences, 2010. 23(10): p.1386-1398.

Coon, C.N., Leske, K.L., Akavanichan, O. and Cheng, T.K., Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. Poultry Science, 1990. 69(5): p.787-793.

Cooper, R.L., Breeding semidwarf soybeans. Plant breeding reviews, 1985. 3: p.289-311.

Deak, N.A., Johnson, L.A., Lusas, E.W. and Rhee, K.C., 2008. Soy protein products, processing, and utilization. In Soybeans, 2008. (pp. 661-724). AOCS Press.

Dersjant-Li, Y., The use of soy protein in aquafeeds. Avances en Nutrición Acuícola. 2002.

Dierking, E.C. and Bilyeu, K.D., Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome, 2008. 1(2): p. 135-145.

Dierking, E.C. and Bilyeu, K.D., New sources of soybean seed meal and oil composition traits identified through TILLING. BMC plant biology, 2009. 9(1): p.1-11.

Fang, C., Ma, Y., Yuan, L., Wang, Z., Yang, R., Zhou, Z., Liu, T. and Tian, Z., Chloroplast DNA underwent independent selection from nuclear genes during soybean domestication and improvement. Journal of genetics and genomics= Yi chuan xue bao, 2016. 43(4): p.217-221.

FDA (2003) Food labeling: trans fatty acids in nutrition labeling, nutrient content claims, and health claims. Federal Register Rule Number 68 FR 41433 Session Number 21 CFR Part 101, 68:41433–41506

FDA (2015) Final determination regarding partially hydrogenated oils. Federal Register Rule Number 80 FR 34650, 80:34650– 34670

Gillman, J.D., Pantalone, V.R. and Bilyeu, K., 2009. The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome, 2009. 2(2): 179-190.

Gai, J.Y. and Wang, Y.S., A study on the varietal eco-regions of soybeans in China. Scientia Agricultura Sinica, 2001. 34(2): p.139-145.

George, M.A., Bhide, S.V., Thengane, R.J., Hosseini, G.H. and Manjaya, J.G., Identification of low lectin mutants in soybean. Plant breeding, 2008. 127(2): p.150-153.

Grassini, P., La Menza, N.C., Edreira, J.I.R., Monzón, J.P., Tenorio, F.A. and Specht, J.E., Soybean. In Crop Physiology Case Histories for Major Crops, 2021. (pp. 282-319). Academic Press.

Gu, M., Jia, Q., Zhang, Z., Bai, N., Xu, X. and Xu, B., Soya-saponins induce intestinal inflammation and barrier dysfunction in juvenile turbot (Scophthalmus maximus). Fish & shellfish immunology, 2018. 77: p.264-272

Guo, W., Chen, L., Chen, H., Yang, H., You, Q., Bao, A., Chen, S., Hao, Q., Huang, Y., Qiu, D. and Shan, Z., Overexpression of GmWRI1b in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions. Plant biotechnology journal, 2020. 18(8): p.1639-1641.

Guzman, P.S., Diers, B.W., Neece, D.J., St. Martin, S.K., LeRoy, A.R., Grau, C.R., Hughes, T.J. and Nelson, R.L., QTL associated with yield in three backcross-derived populations of soybean. Crop Science, 2007. 47(1): p.111-122.

Hagely, K.B., Palmquist, D. and Bilyeu, K.D., Classification of distinct seed carbohydrate profiles in soybean. Journal of agricultural and food chemistry, 2013. 61(5): p.1105-1111.

Hagely, K.B., Jo, H., Kim, J.-H., Hudson, K.A., and Bilyeu, K., Molecular-assisted breeding for improved carbohydrate profiles in soybean seed. Theoretical and Applied Genetics, 2020. 133(4): p. 1189-1200.

Hagely, K., Konda, A.R., Kim, J.-H., Cahoon, E.B., and Bilyeu, K., Molecular-assisted breeding for soybean with high oleic/low linolenic acid and elevated vitamin e in the seed oil. Molecular Breeding, 2021. 41(1): p. 3.

Hao, D., Cheng, H., Yin, Z., Cui, S., Zhang, D., Wang, H. and Yu, D., Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (Glycine max) landraces across multiple environments. Theoretical and Applied Genetics, 2012. 124(3): p.447-458.

Hatten III, L.F., Ingram, D.R. and Pittman, S.T., Effect of phytase on production parameters and nutrient availability in broilers and laying hens: a review. Journal of Applied Poultry Research, 2001. 10(3): p.274-278.

Hartung, R.C., Specht, J.E. and Williams, J.H., Modification of soybean plant architecture by genes for stem growth habit and maturity. Crop Science, 1981. 21(1): p.51-56.

Haun, W., Coffman, A., Clasen, B.M., Demorest, Z.L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A., Cedrone, F. and Mathis, L., Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant biotechnology journal, 2014. 12(7): p.934-940.

Heatherly, L.G. and Smith, J.R., Effect of soybean stem growth habit on height and node number after beginning bloom in the midsouthern USA. Crop science, 2004. 44(5): p.1855-1858.

Heppard, E.P., Kinney, A.J., Stecca, K.L. and Miao, G.H., Developmental and growth temperature regulation of two different microsomal [omega]-6 desaturase genes in soybeans. Plant Physiology, 1996.110(1): p.311-319.

Herman, E.M. and Schmidt, M.A., The potential for engineering enhanced functional-feed soybeans for sustainable aquaculture feed. Frontiers in plant science, 2016. 7: p.440.

Hitz, W.D., Carlson, T.J., Kerr, P.S. and Sebastian, S.A., Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant physiology, 2002. 128(2): p.650-660.

Hoshino, T., Takagi, Y. and 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. Breeding science, 2010. 60(4): p.419-425.

Hymowitz, T., On the domestication of the soybean. Economic botany, 1970. 24(4): p.408-421.

Hymowitz, T. and Harlan, J.R., Introduction of soybean to North America by Samuel Bowen in 1765. Economic Botany, 1983. 37(4): p.371-379.

Hymowitz, T., Soybeans: The success story. Advances in new crops, 1990. p.159-163.

Hymowitz, T., The history of the soybean. In Soybeans, 2008. (pp. 1-31). AOCS Press.

Hu, F.B., Stampfer, M.J., Manson, J.E., Rimm, E., Colditz, G.A., Rosner, B.A., Hennekens, C.H. and Willett, W.C., Dietary fat intake and the risk of coronary heart disease in women. New England journal of medicine, 1997. 337(21): p.1491-1499.

Hu, M. and Wiatrak, P., Effect of planting date on soybean growth, yield, and grain quality. Agronomy Journal, 2012. 104(3): p.785-790.

Jo, H., Lee, J.-D., and Bilyeu, K.D., Environmental stability of carbohydrate profiles in different soybean genotypes. Crop Science, 2018. 58(2): p. 773-782.

Jo, H., Lorenz, A.J., Rainey, K.M., Shannon, J.G., Chen, P., and Bilyeu, K.D., Environmental stability study of soybeans with modified carbohydrate profiles in maturity groups 0 to v. Crop Science, 2019. 59(4): p. 1531-1543.

Jo, H., Kim, M., Ali, L., Tayade, R., Jo, D., Le, D.T., Phommalth, S., Ha, B.K., Kang, S., Song, J.T. and Lee, J.D., Environmental stability of elevated α-linolenic acid derived from a wild soybean in three Asian countries. Agriculture, 2020. 10(3): p.70.

Karr-Lilienthal, L.K., Kadzere, C.T., Grieshop, C.M. and Fahey Jr, G.C., Chemical and nutritional properties of soybean carbohydrates as related to nonruminants: A review. Livestock Production Science, 2005. 97(1): p.1-12.

Kim, M.Y., Lee, S., Van, K., Kim, T.H., Jeong, S.C., Choi, I.Y., Kim, D.S., Lee, Y.S., Park, D., Ma, J. and Kim, W.Y., Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences, 2010. 107(51): p.22032-22037.

Kofsky, J., Zhang, H. and Song, B.H., The untapped genetic reservoir: the past, current, and future applications of the wild soybean (Glycine soja). Frontiers in plant science, 2018. 9: p.949.

Krogdahl, Å., Gajardo, K., Kortner, T.M., Penn, M., Gu, M., Berge, G.M. and Bakke, A.M., Soya saponins induce enteritis in Atlantic salmon (Salmo salar L.). Journal of agricultural and food chemistry, 2015. 63(15): p.3887-3902.

Knudsen, D., Jutfelt, F., Sundh, H., Sundell, K., Koppe, W. and Frøkiær, H., Dietary soya saponins increase gut permeability and play a key role in the onset of soyabean-induced enteritis in Atlantic salmon (Salmo salar L.). British Journal of Nutrition, 2008. 100(1): p.120-129.

Lee, J.D., Vuong, T.D., Moon, H., Yu, J.K., Nelson, R.L., Nguyen, H.T. and Shannon, J.G., Genetic diversity and population structure of Korean and Chinese soybean [Glycine max (L.) Merr.] accessions. Crop Science, 2011. 51(3): 1080-1088.

Li, Z. and Nelson, R.L., Genetic diversity among soybean accessions from three countries measured by RAPDs.Crop Science, 2001. 41(4): p.1337-1347.

Li, L., Wang, X., Gai, J. and Yu, D., Isolation and characterization of a seed-specific isoform of microsomal omega-6 fatty acid desaturase gene (FAD2-1B) from soybean: Full Length Research Article. DNA Sequence, 2008. 19(1): p.28-36.

Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L. and Zhang, S.S., De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nature biotechnology, 2014. 32(10): p.1045-1052.

Li, M.W., Xin, D., Gao, Y., Li, K.P., Fan, K., Muñoz, N.B., Yung, W.S. and Lam, H.M., Using genomic information to improve soybean adaptability to climate change. Journal of Experimental Botany, 2017. 68(8): p.1823-1834.

Li, M., Liu, Y., Wang, C., Yang, X., Li, D., Zhang, X., Xu, C., Zhang, Y., Li, W. and Zhao, L., Identification of traits contributing to high and stable yields in different soybean varieties across three Chinese latitudes. Frontiers in plant science, 2020. 10: p.1642.

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., Nagamatsu, A., Arai, M., Yamada, T., Kitamura, K. and Masuta, C., The soybean stem growth habit gene Dt1 is an ortholog of Arabidopsis TERMINAL FLOWER1. Plant Physiology, 2010. 153(1): p.198-210.

Liu, G., Zhao, L., Averitt, B.J., Liu, Y., Zhang, B., Chang, R., Ma, Y., Luan, X., Guan, R. and Qiu, L., Geographical distribution of GmTfl1 alleles in Chinese soybean varieties. The Crop Journal, 2015. 3(5): p.371-378.

Macrae, R., Robinson, R.K., and Sadler, M.J., Soya beans. In: (Ed. R. Macrae, R. K. Robinson and, M. J. Sadler), Encyclopaedia of Food Science, Food Technology, 1993.

Miao, S., Zhao, C., Zhu, J., Hu, J., Dong, X. and Sun, L., Dietary soybean meal affects intestinal homoeostasis by altering the microbiota, morphology and inflammatory cytokine gene expression in northern snakehead. Scientific reports, 2018. 8(1): p.1-10.

Moreira, R.D.A., Ainouz, I.L., Oliveira, J.T.A.D. and Cavada, B.S., Plant lectins, chemical and biological aspects. Memórias do Instituto Oswaldo Cruz, 1991. 86: p.211-218.

Nutrition. Academic Press; London. 1993:4215-4218.

Okubo, K., Iijima, M., Kobayashi, Y., Yoshikoshi, M., Uchida, T. and Kudou, S., Components responsible for the undesirable taste of soybean seeds. Bioscience, biotechnology, and biochemistry, 1992. 56(1): p.99-103.

Obendorf, R.L., Kosina, S.M. and Ng, T.B., Soluble carbohydrates in soybean. Soybean-Biochemistry, Chemistry and Physiology, InTech Open Access Publisher, Rijeka, Croatia, 2011. p.201-228.

Openshaw, S.J., and Hadley, H.H., Maternal effects on sugar content in soybean seeds. Crop Science, 1978. 18(4): p.581-584.

Orf, J.H. and Hymowitz, T.H., Inheritance of the Absence of the Kunitz Trypsin Inhibitor in Seed Protein of Soybeans 1. Crop Science, 1979. 19(1): p.107-109.

Palmer, R.G. and Hymowitz, T., 2016. Soybean: germplasm, breeding, and genetics.

Park, J., Kim, J.H., Krishnamurthy, P., Tsukamoto, C., Song, J.T., Chung, G., Shannon, J.G. and Lee, J.D., Characterization of a New Allele of the Saponin-Synthesizing Gene Sg-1 in Soybean. Crop Science, 2016. 56(1): p.385-391.

Pham, A.T., Lee, J.D., Shannon, J.G., and Bilyeu, K., Mutant alleles of fad2-1a and fad2-1b combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology, 2010. 10(1): p. 195.

Pham, A.T., Lee, J.D., Shannon, J., and Bilyeu, K., A novel fad2-1a allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. TAG Theoretical and Applied Genetics, 2011. 123(5): p. 793-802.

Pham, A.T., Shannon, J.G., and Bilyeu, K., Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. Theoretical and Applied Genetics, 2012. 125(3), 503-515.

Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., Sui, Y., Lin, F., Liu, X., Tang, Z. and Nguyen, H., Dt2 is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. The Plant Cell, 2014. 26(7): p.2831-2842.

Redekar, N.R., Biyashev, R.M., Jensen, R.V., Helm, R.F., Grabau, E.A. and Maroof, M.S., Genome-wide transcriptome analyses of developing seeds from low and normal phytic acid soybean lines. BMC genomics, 2015. 16(1): p.1-16.

Wang, K., Bu, T., Cheng, Q., Dong, L., Su, T., Chen, Z., Kong, F., Gong, Z., Liu, B. and Li, M., Two homologous LHY pairs negatively control soybean drought tolerance by repressing the abscisic acid responses. New Phytologist, 2021. 229(5): p.2660-2675.

Sablowski, R., Flowering and determinacy in Arabidopsis. Journal of experimental botany, 2007. 58(5): p.899-907.

Schillinger, J.A., Dierking, E.C. and Bilyeu, K.D., U.S. Patent 10,081,814. Soybeans having high germination rates and ultra-low raffinose and stachyose content. September 2018.

Schillinger, J.A., Dierking, E.C., and Bilyeu, K.D., U.S. Patent Number 8,471,107. Soybeans having high germination rates and ultra-low raffinose and stachyose content. June 2013.

Schillinger, J.A., Dierking, E.C. and Bilyeu, K.D., U.S. Patent 10,081,814. Soybeans having high germination rates and ultra-low raffinose and stachyose content. September 2018.

Schlueter, J.A., Vasylenko-Sanders, I.F., Deshpande, S., Yi, J., Siegfried, M., Roe, B.A., Schlueter, S.D., Scheffler, B.E. and Shoemaker, R.C., The FAD2 gene family of soybean: insights into the structural and functional divergence of a paleopolyploid genome. Crop Science, 2007. 47: p.S-14.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.-C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A., Genome sequence of the palaeopolyploid soybean. Nature, 2010. 463(7278): p.178-183.

Sedivy, E.J., Wu, F. and Hanzawa, Y., Soybean domestication: the origin, genetic architecture and molecular bases. New Phytologist, 2017. 214(2): p.539-553.

Shurtleff, W. and Aoyagi, A., History of soybeans and soyfoods in Germany (1712–2015). Lafayette, CA: Soyinfo Center., 2015.

Singh, R.J., Nelson, R.L. and Chung, G.H., Soybean (Glycine max (L.) Merr.). Genetic resources, chromosome engineering, and crop improvement. Oilseed crops, 2007. 4: p.13-50.

De Sousa, L.S., De Moura, C.V.R., De Oliveira, J.E. and De Moura, E.M., Use of natural antioxidants in soybean biodiesel. Fuel, 2014. 134: p.420-428.

SoyStats (2021) http://www.soystats.com.

Sun, Z., Su, C., Yun, J., Jiang, Q., Wang, L., Wang, Y., Cao, D., Zhao, F., Zhao, Q., Zhang, M. and Zhou, B., Genetic improvement of the shoot architecture and yield in soya bean plants via the manipulation of GmmiR156b. Plant biotechnology journal, 2019. 17(1): p.50-62.

Sweeney, D.W., Carrero-Colón, M. and Hudson, K.A., Characterization of new allelic combinations for high-oleic soybeans. Crop Science, 2017. 57(2): p.611-616.

Thapa, R., Carrero-Colón, M., Rainey, K.M. and Hudson, K., Tilling by sequencing: a successful approach to identify rare alleles in soybean populations. Genes, 2019. 10(12): p.1003.

Tsukamoto, C. and Yoshiki, Y., Soy saponin. In Soy in health and disease prevention, 2005. (pp. 155-172). CRC Press.

Thompson, J.A., Bernard, R.L. and Nelson, R.L., A third allele at the soybean dt1 locus. Crop science, 1997. 37(3): p.757-762.

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J.E., Nelson, R.L., McClean, P.E., Qiu, L. and Ma, J., Artificial selection for determinate growth habit in soybean. Proceedings of the National Academy of Sciences, 2010. 107(19): p.8563-8568.

Vagadia, B.H., Vanga, S.K. and Raghavan, V., Inactivation methods of soybean trypsin inhibitor–A review. Trends in Food Science & Technology, 2017. 64: p.115-125.

Wang, T.L., Domoney, C., Hedley, C.L., Casey, R. and Grusak, M.A., Can we improve the nutritional quality of legume seeds?. Plant Physiology, 2003. 131(3): p.886-891.

Warner, K. and Fehr, W., Mid-oleic/ultra low linolenic acid soybean oil: A healthful new alternative to hydrogenated oil for frying. Journal of the American Oil Chemists' Society, 2008. 85(10): p.945.

Wilcox, J.R. and Frankenberger, E.M., Indeterminate and Determinate Soybean Responses to Planting Date 1. Agronomy Journal, 1987. 79(6): p.1074-1078.

Wilson, R.F., 2004. Seed composition. Soybeans: improvement, production, and uses, 2004. 16: p.621-677.

Woodworth, C.M., 1933. Genetics of the soybean.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y. and Fang, C., Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature biotechnology, 2015. 33(4): p.408-414.

Zhou, Z., Ringø, E., Olsen, R.E. and Song, S.K., Dietary effects of soybean products on gut microbiota and immunity of aquatic animals: a review. Aquaculture Nutrition, 2018. 24(1): p.644-665.

Xu, D., Abe, J., Gai, J. and Shimamoto, Y., Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theoretical and Applied Genetics, 2002. 105(5): p.645-653.

Xu, M., Xu, Z., Liu, B., Kong, F., Tsubokura, Y., Watanabe, S., Xia, Z., Harada, K., Kanazawa, A., Yamada, T. and Abe, J., Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biology, 2013. 13(1): p.1-14.

Yue, L., Li, X., Fang, C., Chen, L., Yang, H., Yang, J., Chen, Z., Nan, H., Chen, L., Zhang, Y. and Li, H., FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. Journal of Integrative Plant Biology, 2021. 63(6): p.1004-1020.

Zhang, D., Wang, X., Li, S., Wang, C., Gosney, M.J., Mickelbart, M.V. and Ma, J., A post-domestication mutation, Dt2, triggers systemic modification of divergent and convergent pathways modulating multiple agronomic traits in soybean. Molecular plant, 2019. 12(10): p.1366-1382.

CHAPTER 2

Utilization of Plant Architecture Genes in Soybean to Positively Impact Adaptation to High Yield Environments

ABSTRACT

Optimization of plant architecture by modifying stem termination and timing of flowering and maturity of soybean is a promising strategy to improve its adaptability to specific production environments. Therefore, it is important to choose a proper stem termination type and to understand morphological differences between each stem termination type under various environmental conditions. Variations in abruptness of stem termination have been generally classified into three classical genetic types, indeterminate (*Dt1*), determinate (*dt1*), and semi-determinate (*Dt1 Dt2*). However, an additional stem termination type, termed tall determinate, and its genetic symbol, *dt1-t*, were introduced about 25 years ago. The tall determinate soybean lines show delayed cessation of apical stem growth and about 50% taller plant height than the typical determinate soybeans, even though the genetic control of the tall determinate phenotype was found to be allelic to the *dt1* locus. Despite the potential agronomic merits of the alternative stem termination type, knowledge about the tall determinate soybean remains limited. In this research, we clarified the molecular basis of the tall determinate stem termination type and examined potential agronomic merits of the alternative stem type under three different production environments in the US. Sequence analysis of the classical tall determinate soybean lines revealed that the *dt1-t* allele responsible for tall determinate stem architecture is caused by two of the identified missense alleles of *dt1*, *dt1R62S* and *dt1R130K*. Also, from the comparison among soybean accessions belonging to each of the genotype categories for stem termination type, soybean accessions with tall determinate alleles were found to have a high discrepancy rate in phenotyping. The tall determinate soybeans had taller plant height and a greater number of nodes with similar

stem diameter and similar pod density at the apical stem compared to typical determinate soybeans having *dt1R166W* allele in both Midwest and Southern environments in the US. The phenotype of increased pod-bearing nodes with lodging resistance has the potential to improve yield, especially grown in high yield environments. This study suggests an alternative strategy to remodel the shape of soybean plants which can possibly lead to yield improvement through the modification of soybean plant architecture.

INTRODUCTION

Soybean [*Glycine max* (L.) Merrill] is an important crop worldwide and improving its yield has always been a major concern for increasing the profits from this important crop. Since soybean was domesticated, there have been long-term efforts for improving yield, and the world soybean production increased approximately 13-fold from 1961 to 2017 (FAO stat). However, the yield increase was mainly due to an increase in planting area; the yield per unit area of soybean has not changed significantly over the past few decades (Liu et al., 2020). Unlike other major crops which increased yields per unit area by increasing planting densities with the modification of plant architecture, improving yields per unit area in soybean is not that simple due to its unique plant architecture. Since soybean is a typical legume which has leaves, inflorescences, and pods growing at each node, several agronomic traits, such as plant height, number of nodes, flowering time, and maturity, need to be considered simultaneously to optimize soybean plant architectures for improved yield potential.

Soybean plant architecture can be modified by adjusting stem termination and timing of flowering and maturity. Soybeans show a wide range in the abruptness of stem termination, and the variations primarily result from the differences in timing at which soybeans terminate their apical stem growth (Bernard 1972). Based on the timing, most soybean varieties are generally classified into three categories: indeterminate, determinate, and semi-determinate. The indeterminate soybean varieties continue vegetative growth on apical meristems at the stem and branch apices after flowering until the beginning of seed fill (Bernard 1972). Since the terminal growth continues as long as lateral growth, indeterminate varieties feature a tapered stem with little or no secondary

lateral growth near the stem tip (Tian et al., 2010). In contrast, determinate varieties abruptly halt vegetative growth at or soon after flowering begins, producing a thick stem due to the continuous lateral growth even after the cessation of apical stem growth (Ping et al., 2014). In semi-determinate varieties, the stem tip ceases vegetative growth several days earlier than the indeterminate ones, resulting in fewer main-stem nodes and somewhat shorter stem length than the indeterminate varieties (Hartung et al., 1981; Ping et al., 2014). Classical genetic analyses identified two genes regulating the stem growth habit in soybean, *Dt1* and *Dt2* (Bernard 1972). In the *Dt1* genetic background, *Dt2* genotypes have semi-determinate phenotypes, while the recessive gene *dt2* produces indeterminate varieties. However, in the *dt1* genetic background, the phenotype is determinate regardless of the alleles at the *Dt2* locus due to an epistatic effect of the recessive *dt1* allele over the *Dt2* alleles.

While the three categories of stem growth habit were generally accepted, Thompson et al (1997) introduced an additional type, termed tall determinate, and its genetic symbol, *dt1-t*, to describe a distinct phenotype they found. The tall determinate phenotype was observed from near-isogenic lines (hereafter referred to isolines) of an indeterminate soybean cultivar 'Clark' with either 'Soysota' or 'Peking' as a donor parent of L91-8052 and L91-8060, respectively. The authors found that the tall determinate (*dt1-t*) isolines significantly delayed the timing of stem termination compared to typical determinate (*dt1*) isolines and almost identically terminated their stem growth with semi-determinate (*Dt1 Dt2*) isolines, even though the genetic control of the tall determinate phenotype was later found to be allelic to the *dt1* locus and independent of *Dt2* (Thompson et al., 1997). The tall determinate isolines showed similar plant height to the

semi-determinate isolines, which were ~70% of the mature plant height of indeterminate Clark; however, the tall determinate isolines had significantly different numbers of final stem nodes and terminal leaflet areas compared to the semi-determinates. When considering the leaf and stem characteristics at the top of the plant, tall determinate isolines were similar with determinate isolines. Given that the stem termination type affects other important agronomic traits, such as flowering time, node formation, plant height, lodging resistance, and ultimately yield of soybean, more specific research about the distinct stem characteristics of the tall determinate type is required in order to characterize and utilize its potential agronomic merits, but no further articles have been published to date (Cao et al., 2016; Heatherly and Smith, 2004; Liu et al., 2010).

The molecular basis of the typical stem growth habits in soybean was elucidated. *Dt1* (*GmTfl1*; *Glyma19g37890.1*), the major gene affecting stem growth habit, was found to be a functionally conserved ortholog of *Arabidopsis thaliana TERMINAL FLOWER1* (*TFL1*), where the functional gene participates in forming indeterminate stems (Liu et al., 2010; Tian et al, 2010). Six independent missense mutations were identified while searching for allelic variations of the *Dt1* genic region (Liu et al., 2010; Tian et al, 2010; Yue et al., 2021). While four out of the six identified nonsynonymous mutations were suggested to cause the transition from indeterminate to determinate phenotype, the specific functions of each of the resulting alleles and possible differences in phenotypes conferred from each of these mutations have not been defined. *Dt2*, the second soybean gene regulating stem growth, was characterized as a dominant MADS domain factor gene classified into the *APETALA1/SQUAMOSA* (*AP1/SQUA*) subfamily (Ping et al., 2014). There were quantitative differences in expression level between dominant *Dt2* and

recessive *dt2*, and the increased expression of *Dt2* was found to downregulate functional *Dt1* in the shoot apical meristems (SAMs) to promote early conversion of the SAMs into reproductive inflorescences (Ping et al., 2014). Along with its function as a direct repressor of *Dt1*, *Dt2* was found to activate the putative floral integrator/identity genes, such as *GmSOC1*, *GmAP1*, and *GmFUL*, thereby promoting flowering in soybean (Zhang et al., 2019). A total of 37 SNPs in the non-coding region of *Dt2* were identified when it was cloned, but further efforts are still required to pinpoint causative mutations and elucidate molecular mechanisms responsible for the *Dt2* activity (Ping et al., 2014).

The genes regulating stem termination contribute diversity in phenotypes of soybean plant architecture, and each of these morphological characteristics has distinct advantages depending on diverse production environments. For instance, nearly all soybean varieties grown commercially in the northern US and in Canada (MG IV and earlier) are indeterminate types, due to their better adaptation to the shorter growing season at high latitudes, while most soybean varieties commercially grown in the southern US (MG V and later) are determinate (Bernard 1997; Heatherly and Elmore, 2004; Hartung et al., 1981). However, depending on environments, there would be agronomic merits which can be achieved by choosing an alternate stem termination type for soybean. In late planting and non-irrigated southern US environments, indeterminate lines showed better yield potential than determinate lines (Kilgore-Norquest and Sneller 2000). Moreover, semi-determinate soybean varieties have been developed in the past decade to be particularly used in high-yield, lodging-prone environments due to improved productivity in those environments resulting from short plant stature (Ping et al., 2014). Given that the distinct stem characteristics each soybean has are associated with its

production, a precise classification of stem termination types are required in order to choose a proper plant architecture suitable for the production environment (Liu et al., 2010; Liu et al., 2020).

Germplasm collections which contain phenotype data of stem growth habits as well as various other associated traits, such as plant height, lodging, flowering and maturity date, and yield, are publicly accessible from the USDA National Plant Germplasm System and Germplasm Resources Information Network (GRIN) database collection (www.ars-grin.gov). Soybean accessions having a wide range in the abruptness of stem termination were evaluated by two separate descriptors, termed stem termination score and stem termination type. In terms of the stem termination score, the evaluations were made at maturity with a number range from 1 (very determinate) to 5 (very indeterminate). The stem termination scores less than 2 and greater or equal to 2.5 are considered determinate and indeterminate, respectively and the score in between 2 and 2.5 is considered semi-determinate (Heatherly and Smith 2004). With regard to the stem termination type, soybean accessions were coded based on characteristics at the top of the stem with three categories: D (determinate; stem abruptly terminating), I (indeterminate; stem tapering gradually toward tip), and S (Semi-determinate; intermediate between determinate and indeterminate). Considering the measurements are somewhat subjective and possibly influenced by other factors, these simple classifications may lead to discrepancies between the designated stem growth habit and the actual underlying genotype of the evaluated soybean accessions.

Alteration of plant architecture by modifying genes that affect stem termination and timing of flowering and maturity is a promising strategy to improve yield potential of

soybean. However, soybeans show a wide range in the abruptness of stem termination and the variations are even further complicated to definitively categorize because of other genetic and environmental factors, particularly flowering time. Therefore, a precise classification of stem termination types, as well as a broad understanding of their responses under various environments in combination with other genes are required to choose a proper plant architecture suitable for the production environment (Liu et al., 2010; Liu et al., 2020). In particular, the potential agronomic merits of the new stem termination type, tall determinate, which has not been generally accepted and thus has not been widely studied need to be examined for its broader use in breeding programs. In this study, we clarified the molecular basis of the *dt1-t* alleles controlling the tall determinate stem termination type and examined the morphological characteristics and possible agronomic merits of the tall determinate stem growth habit in three different latitudinal environments ranging from MG III to VII in the US.

MATERIALS AND METHODS

Sequencing analysis at *Dt1* locus for the classical tall determinate soybean lines

A total of four classical tall determinate soybean lines, Peking, Soysota, L91-8060 and L91-8052, were used in the sequence analysis to clarify the molecular basis of the *dt1-t* alleles controlling the tall determinate stem termination type. Peking and Soysota are soybean cultivars, and the L91-8060 and L91-8052 are isolines with the soybean cultivar Clark as a recurrent parent and each of Peking and Soysota as a donor parent, respectively. Details about the development of the two isolines were described in Thompson et al. (1997). Since the tall determinate allele was found to be allelic to *dt1* (Thompson et al., 1997), the nucleotide polymorphisms in the coding region of *Dt1* (*Glyma.19g194300*; 45183357 – 45185175 Wm82.a2.v1) were examined (Table 2.1).

Whole genome resequenced data of the two tall determinate cultivars, Peking and Soysota, were publicly available (Zhou et al., 2015), so the nucleotide polymorphisms at the *Dt1* locus were investigated in the sequence data using positions on Wm82.a2.v1 chromosome 19 (Table 2.1). The data were downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (ftp-trace.ncbi.nlm.nih.gov/sra/reports/Metadata/). The data sets were processed with a custom Pgen workflow and analyzed on SNPViz v2.0 (Zeng et al., 2021). For the two tall determinate isolines, L91-8052 and L91-8060, Sanger sequencing of PCR amplified products was conducted for checking allele status at the locus. Template DNA was prepared from leaf presses from 1.2 mm washed FTA (Whatman, Clifton, NJ) card punches according to the manufacturer's instructions. Fully-developed trifoliate leaves were used to prepare the FTA card punches. The DNA templates were amplified by PCR

using the two primer sets which are targeting each of exon 1 and exon 4 region at the Dt1 locus where R62S and P113L, R130K, and R166W alleles are located, respectively: a symmetric mix of primers (Dt1upf1: 5'-CACACTCGATCTACCT-3', Dt1in1r: 5'-ACATACCGTGTGACCATG-3') with a product size of 485 bp targeting exon 1, and a symmetric mix of primers (Dt1in31f: 5'-CATGAGAGAGATCACTGAC -3', Dt1endr1; R: 5'-GCAAAACCAGCAGCTACTT-3') with a product size of 292 bp targeting exon 4 region at *Dt1* locus. The total volume of reactions was 50 µl containing templates DNA, primers, buffer (40 mM Tricine- KOH [pH 8.0] 16 mM KCl, 3.5 mM MgCl2, 3.75 µgml−1 BSA,), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). The cycling condition were 95°C for 3 min, 50 cycles of 95°C for 20 s, 60°C for 20 s and 72°C for 20 s, final extension at 72°C for 3 min, and ended at 4°C. The PCR products were determined by 1.2% agarose gel electrophoresis and then sequenced using each of the forward primers at the University of Missouri DNA Core Facility (https://dnacore.missouri.edu/). Sequencing results were analyzed using Chromas software.

Analysis on disparity in measuring stem termination types

To examine how historically observers have phenotyped stem termination types of soybean accessions, a set of 528 soybean accessions of which both phenotype and resequenced data are publicly accessible were selected for this analysis. The detailed information of the accessions is listed in (Supplementary Table 1 in Appendix). The phenotype data about stem termination type was downloaded from The USDA National Plant Germplasm System and Germplasm Resources Information Network (GRIN)

database collection (www.ars-grin.gov). The soybean accessions were evaluated with three categories, D (determinate), I (indeterminate), and S (semi-determinate), depending on the morphological characteristics at the stem tip. The publicly available resequenced data of the soybean accessions were analyzed from our curated accession panel of 775 soybean accessions with whole genome resequence data. The Soy775 resource is comprised of data sets from the USB-481 resequencing project and Zhou302 data set remapped to Wm82.a2.v1 and available on soykb.org. (Zhou et al., 2015; Valliyodan et al., 2016; Škrabišová, et al., personal communication 2021). The genotype categories of each soybean accession were assigned based on the allele status at the two genes conferring stem termination types, *Dt1* and *Dt2*. In terms of the *Dt1* gene, the allele status of each line was sorted depending on the presence of one of four of the most frequently found alleles, *dt1R62S*, *dt1P113L*, *dt1R130K*, and *dt1R166W*, out of the six previously identified missense mutations. For the *Dt2* allele, we first identified a highly associated SoySNP50K maker at the locus since no causative mutation of the gene has been found yet. The associated SNP marker, ss715632223 (Pos: 55642486 on chromosome 18) was selected from a GWAS analysis using SoySNP50K (Song et al., 2013) as genotype and the phenotype data downloaded from GRIN database. The genotype at the *Dt2* locus was assigned based on the nucleotide polymorphism at the genomic position of the marker: functional *Dt2* in presence of nucleotide G, and recessive *dt2* in presence of nucleotide A. The disparity rates of each genotype category were calculated based on percentage about how often each soybean within the genotype category evaluated as each of the three stem termination types.

Field experiment in three distinct maturity group environments in the United States

To evaluate morphological characteristics and potential agronomic merits of soybean lines with various plant architecture types, soybean lines with each of the targeted genotype combinations were planted in three different latitudinal environments (maturity groups [MG] III, V, and VI) in the US in 2019 and 2020. Agronomic traits were measured from the soybean plants produced each of the production environments. The details of each field experiment were summarized in (Supplementary Table 3 in Appendix).

Plant materials and population developments

The plant materials used for each of the three field trials are listed in Table 2.2. Since the field experiments aimed to evaluate potential agronomic merits of soybeans with various plant architecture genotypes, especially the genotypes not commonly introduced for soybean genetic improvements targeting the southern environments of the US MG V and VI, we first developed recombinant inbred line (RIL) populations having either the tall determinate or semi-determinate stem termination alleles, along with the functional *E1* maturity gene which is critical for southern US soybean production environments (Langewisch et al., 2017). Four experimental populations were generated for this study and the details of population development were summarized in Supplementary Table 2 in Appendix. Each population had one donor parent of the functional *E1* maturity gene and one donor parent of the target stem termination alleles, either tall determinate (Glyma.19g194300 R62S or R130K) or semi-determinate (Glyma.18g273600 intron SNP). As the functional *E1* donor, two soybean varieties were utilized. Jake is a high yielding soybean cultivar with determinate stem termination type

and MG V, which was released by the University of Missouri (Shannon et al., 2007). Ellis is a high yielding MG IV determinate soybean cultivar developed by the University of Tennessee (Pantalone et al., 2017). As the donor of target stem termination types, three soybean lines were utilized. L91-8060 and L91-8052 are the classical tall determinate isolines which were generated from crosses between a soybean cultivar Clark and either soybean cultivars, Soysota or Peking, respectively (Thompson et al., 1997). LG90-2550 is a MG III semi-determinate soybean line cooperatively developed and released in 1997 by the USDA-ARS and the Illinois Agricultural Experiment Station (Thompson et al., 1999). The four populations made for this study were as follows (Supplementary Table 2 in Appendix): Cross 1 was KB17-16: L91-8052, the tall determinate isoline carrying *dt1R130K* allele, was crossed to Jake, the high yielding MG V soybean cultivar having the functional *E1* allele. Cross 2 was KB17-17: L91-8060, the other tall determinate isoline having *dt1R62S* allele, was crossed to Jake. Cross 3 was KB17-7: LG90-2550, the semi-determinate soybean line carrying *Dt2* allele, was crossed to Jake. Cross 4 was KB17-8: LG90-2550 was crossed to Ellis, the high yielding, determinate MG IV soybean cultivar having the functional *E1* allele.

The crosses for the experimental soybean populations were made in the 2017 soybean growing season at the South Farm Research Center near Columbia, MO. The F_1 seeds of the KB17-16 and KB17-17 were self-pollinated to produce F_2 seeds in January 2018 in Upala, Costa Rica. The KB17-16 (378) and KB17-17 (338) F_2 seeds were returned to Missouri and planted in the field in May 2018 and harvested as single plant threshes of $F_{2:3}$ seeds in November 2018; A mixture of three seeds from each line was used for DNA extraction using the DNeasy Plant Mini Kit (Qiagen, Inc., Valencia, CA)

and genotyping assays for R130K or R162S *Dt1* alleles, functional *E1* alleles identified eight lines for KB17-16 and six lines for KB17-17 with the targeted genotype. Three $F_{2:3}$ seeds from each selected line were sent to the winter nursery in Costa Rica to advance a single F_3 plant per line, with $F_{3:4}$ seeds returned to Missouri in 2019. Also, in the winter nursery in Costa Rica, the F_1 seeds of KB17-7 and KB17-8 were self-pollinated to produce F_2 seeds in January 2018; the F_2 plants were sampled with leaf presses on Whatman FTA cards and processed as DNA templates for genotyping assays for functional *Dt1* alleles, as well as *E1*. Plants selected with the targeted genotypes were then single plant threshed, and the $F_{2:3}$ seeds were returned for field planting in Columbia, Missouri in May 2018, confirmed by genotype to have *Dt2* alleles, and harvested as $F_{3:4}$ seeds in October 2018. Seeds from selected lines were then advanced a single generation in the winter nursery in Costa Rica and returned to Missouri as $F_{4:5}$ seeds in April 2019.

Allele-specific molecular marker assays

A total of five separate SimpleProbe assays were used to evaluate alleles at the *E1* maturity gene and the two genes regulating stem termination type, *Dt1* and *Dt2*, of the tested soybean lines. The SimpleProbe assays of *E1* alleles were conducted as described Langewisch et al. (2017). For the *Dt1* gene, three separate SimpleProbe assays were developed to distinguish each of the missense mutations at the *Dt1* locus targeted for the population developments, *dt1R62S, dt1R130K*, and *dt1R166W*. The SimpleProbes used in each of the assays consisted of 5'-Fluorescein-SPC-GGACCTCATATCACCACCCTCAAT-phosphate-3' for *dt1R62S*, 5'-Fluorescein-SPC-TGGAGTAACACACTGTCTACGCTT-phosphate-3' for *dt1R130K*, 5'-Fluorescein-

SPC- TGCACAGAGGGAAACGGCT-phosphate-3' for *dt1R166W* allele. Each of the mutations are indicated by bold font with underline. Two different sets of PCR primers were designed to amplify *Dt1* exon 1 and exon 4 regions where the *dt1R62S* allele, and *dt1R130K* and *dt1R166W* alleles are located, respectively. For the *dt1R62S* allele, genotyping reactions were performed with a 5:2 (forward to reverse) asymmetric mix of primers (Dt1upf1: 5'-CACACTCGATCTACCT-3' at 0.5 μM final concentration, and Dt1in1r: 5'-ACATACCGTGTGACCATG-3' at 0.2 μM final concentration). And for *dt1R130K* and *dt1R166W* alleles, genotyping reactions were performed with a 2:5 (forward to reverse) asymmetric mix of primers (Dt1in31f: 5'-CATGAGAGAGATCACTGAC-3' at 0.2 μM final concentration, Dt1endr1: 5'-GCAAAACCAGCAGCTACTT-3' at 0.5 μM final concentration). In terms of the *Dt2* gene, a SimpleProbe assay was developed for the associated nucleotide polymorphism described above associated with overexpression of the *Dt2* gene Glyma.18g273600 (Pos: 55642486 on chromosome 18). The SimpleProbes consisted of 5'-Fluorescein-SPC-GTGCAGACTACCA**C**GCATGC -phosphate-3'. A set of PCR primers was designed to amplify the surrounding region of the SNP. Genotyping reactions were performed with a 5:2 (forward to reverse) asymmetric mix of primers (Dt2fa: 5'-CACAGGTTCGTAGTTATAG-3' at 0.5 μM final concentration, and Dt2reva: 5'-CATAGGATACTAACCAACG-3' at 0.2 μM final concentration). The SimpleProbes were designed using Roche Applied Science LightCycler Probe Design software 2.0 (version 1.0, February 2004) and the probes were ordered from Flourescentric, Inc. (Park City, UT).

Reactions for each of the genotyping assays were carried out in 20 µl total volume containing 5-50ng DNA template, primers, 0.2 µM final concentration of SimpleProbe, buffer (40 mM Tricine- KOH [pH 8.0] 16 mM KCl, 3.5 mM $MgCl_2$, 3.75 $\mu g \cdot ml^{-1}$ BSA,), 5% DMSO, 200 µM dNTPs, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Genotyping reactions were performed using a Lightcycler 480 II real-time PCR instrument (Roche Life Sciences, Indianapolis, IN), using the following PCR parameters: 95 °C for 5 min followed by 40 cycles of 95 °C for 20 s, 60 °C for 20 s, 72 °C for 20 s. The melting curves of each of the genotyping assay were as follows: *dt1R62S* assay – from 54 °C to 72 °C (reference alleles produced a peak at 66 °C, mutant alleles produced a peak at 58.5 °C, and heterozygous samples produced both peaks); *dt1R130K* assay – from 50 °C to 80 °C (reference alleles produced a peak at 64.5 °C, mutant alleles produced a peak at 57 °C, and heterozygous samples produced both peaks); *dt1R166W* assay – from 50 °C to 70 °C (reference alleles produced a peak at 57 °C, mutant alleles produced a peak at 63 °C, and heterozygous samples produced both peaks); *Dt2* assay – from 54 °C to 72 °C (reference alleles produced a peak at 59.5 °C, mutant alleles produced a peak at 67.8 °C, and heterozygous samples produced both peaks).

Growth conditions

For MG VI–VII environment in the US, a set of 21 soybean lines including 16 experimental and 5 control soybean lines under 7 genotype categories were planted in 2019 at University of Georgia Iron Horse Plant Science Farm, Watkinsville, GA (19GA; Table 2.2). The field trial in GA consisted of a completely randomized design without replication. The soybeans were planted on 6 June 2019, at the seeding rate of 33 seeds/m.

Individual plots consisted of two rows 6 ft long spaced 30 in apart. No artificial irrigation was applied during the whole growth period.

In terms of the MG IV–V environment in the US, a total of 20 soybean lines including 14 experimental and 6 control soybean lines were planted in 2020 at East Tennessee Agriculture and Education Center, Knoxville, TN (20TN; Table 2.2). The soybean lines were planted on 26 May 2020 in a randomized complete block design with three replications. Individual plots were 2-row plots with the row length of 4.88 m and spaced 76.2 cm apart. The seeding rate was 200 seeds per row. During the growing season, irrigation has been applied depending on the field condition. A total 20 plots with poor germination were not included from the data analysis.

For the MG III–IV environment in the US, the 14 soybean lines with the *e1-as* background with different stem termination genotypes were planted in 2020 at the South Farm Research Center near Columbia, MO (20MO; Table 2.2). The soybean lines were planted on 1 June 2020 in 7 feet rows with 2 feet alleys and row spacing of 30 inches. The seeding rate was 50 seeds per row. No artificial irrigation was applied.

Phenotype measurements

Several agronomic traits were evaluated as parameters to assess potential agronomic merits depending on the plant architecture genes in each of the three production environments. For the field experiment in 2019 in Georgia (19GA), four morphological characteristics associated with yield and lodging of soybean plants, plant height, number of nodes, stem diameter, and lodging, were measured at maturity. Plant height, number of nodes, and stem diameter were measured from five randomly selected individual plants from each plot, and lodging was scored per plot. Plant height was

evaluated as the distance from the soil surface at the base of the plants to the stem apex in centimeters. The number of nodes was counted from the main stem of each plant. In terms of stem diameter, the diameters at the first, middle, and last internodes of each plant were measured and then the mean of the three diameters was calculated to consider overall thickness of each stem. Lodging was subjectively scored per plot on a scale of 1 (all plants erect) to 5 (almost all plants prostrate).

For the Tennessee (20TN) production environment, a total of eight agronomic traits including the four morphological parameters assessed in 19GA plus four other parameters, days to maturity (DTM), number of pods, raceme length, and number of branches, were evaluated at maturity. Lodging and DTM were evaluated per plot, while the other morphological characteristics were measured on 10 individual plants from each plot. The plant height, the number of nodes, stem diameter, and lodging were measured in the same manner as in 19GA. For DTM, it was scored as the number of days from planting to the date when the soybean plants in a plot reached R8 stage, on which 95% of the pods had turned brown. The number of pods was counted at the apical stem of each plant. Length of terminal raceme was recorded in centimeters and number of branches was counted from the main stem of each plant.

In the field trial at Missouri in 2020 (20MO), a total of nine agronomic traits were recorded, including the eight parameters measured in 19TN and an additional trait, days to flowering (DTF). Lodging, DTM, and DTF were evaluated per plot, and the other characteristics were evaluated from five individual plants per plot. For the DTF, it was scored as the number of days from planting to the date when more than three plants had

opened flowers within each plot. The other eight agronomic traits were measured in the same manner as in 19GA and 20TN.

Data analysis

For the phenotype data collected in MG VI-VII environment in the US, the differences in agronomic traits depending on genotype combinations were analyzed using descriptive statistics, since the experimental design used was completely randomized design without replication. In terms of the phenotype data collected in MG IV-V and MG III-IV environments in the US, the statistical differences in agronomic traits depending on genotype combinations were tested using PROC GLM in SAS version 9.4 (SAS Institute Inc., Cary, NC, 2013). Means were separated by Fisher's least significant difference (LSD) procedure at $P = 0.05$ probability level.

RESULTS

The classical *dt1-t* allele responsible for tall determinate stem termination type is caused by two of the identified missense alleles of *dt1*

Since the gene controlling the tall determinate stem type was found to be allelic to *dt1* (Thompson et al., 1997), we first examined the DNA sequence of the coding region of *Dt1* (Glyma.19g194300; 45183357 – 45185175 Wm82.a2.v1) in the classical tall determinate soybean cultivars, Peking and Soysota, and isolines with those cultivars as donors, L91-8060 and L91-8052, respectively (Table 2.1; Supplementary Table 1 in Appendix). Glyma.19g194300 is an 1819 bp gene with four exons that encodes a 173 amino acid protein, alternatively termed *GmTFL1* (http://www.phytozome.net/) (Tian et al., 2010). In the coding region of *Dt1*, six missense *dt1* alleles –*dt1R62S*, *dt1L67Q*, *dt1P113L*, *dt1R130K*, *dt1H141R*, and *dt1R166W*– have been identified in the search for the recessive variants (Tian at al., 2010; Liu et al., 2010; Yue et al., 2021). In the Williams 82 Assembly 2 Genomic Sequence (Wm82.a2), the *Dt1* gene Glyma.19g194300 is oriented on the opposite strand. For the tall determinate cultivars Peking and Soysota, we utilized resequencing data from Zhou et al. (2015) and analyzed nucleotide polymorphisms at the *Dt1* locus. For the two isolines, a primer set specific for *Dt1* gene and PCR amplification was developed, and the sequence of the gene was confirmed by PCR analysis and Sanger sequencing of PCR fragments.

The classical tall determinate soybean line Peking and its isoline, L91-8060, contained a guanine to thymine causative SNP on chromosome 19 at position 45,184,804 (Wm82.a2.v1), g186t in coding sequence. It creates a missense mutation resulting in a change from R62S in the protein amino acid sequence, which correspond to the identified

dt1R62S allele (Tian et al., 2010). The other tall determinate soybean line Soysota and its isoline, L91-8052, contained a guanine to adenine causative SNP on chromosome 19 at position 45,183,808 (Wm82.a2.v1), g389a in coding sequence. The nonsynonymous SNP corresponded to the identified *dt1R130K* allele, which results in a change from R130K in the protein amino acid sequence (Tian et al., 2010). These results suggest that the tall determinate stem type resulted from two identified missense alleles of *dt1*.

High disparity in distinguishing stem termination types in soybean accessions with tall determinate alleles

In an attempt to address how stem growth habits of soybean accessions with the tall determinate alleles have been phenotypically scored, as well as to broaden our understanding of the relationships between the known genotypes in stem growth habits and the actual phenotype displayed, we analyzed publicly available genome resequence data, along with phenotypic measurement data of the evaluated soybean lines. A total of 528 soybean accessions which have both phenotype and resequencing data was examined. The phenotyping results of the stem termination type descriptor were based on observations on the abruptness of the stem tip for three categories: D (determinate; stem abruptly terminating), I (indeterminate; stem tapering gradually toward tip), and S (Semi-determinate; intermediate between determinate and indeterminate). The whole-genome resequence data of the evaluated accessions were obtained from several different datasets (Valliyodan et al., 2016; Zhou et al., 2015). The genotype of each accession was designated depending on the allele combinations at the *Dt1* and *Dt2* loci (Table 2.1). In

the case of *Dt2*, since no causative mutation has been assigned, the genotype at the locus was assigned based on a nucleotide polymorphism of a highly associated SNP.

The rates of concordance between the observed stem termination type phenotypes downloaded from the GRIN database and the expected phenotypes conferred from the genetic allele combination of stem growth habit genes for the 531 *G. max* accessions are shown in Figure 2.1. When soybean accessions have the genetically indeterminate *Dt1 dt2* genotype, observers assigned the indeterminate stem termination type with a rate of 92.8% (Figure 2.1A). This high accuracy might be due to the distinct morphological characteristics that indeterminate genotypes have. A high rate of discrepancy was observed for the genetically semi-determinate *Dt1 Dt2* genotype, where only about 35% of accessions were scored with a semi-determinate phenotype (Figure 2.1B). This high discrepancy rate might result from the intermediate morphological characteristics of semi-determinate stem type which is conferred from the genotype *Dt1 Dt2*. Despite the distinct phenotypic characteristics of the determinate stem type, not quite 70% of the genetically determinate accessions were scored as determinate when evaluating soybeans with any of the missense *dt1* alleles (Figure 2.1C). The group of soybean accessions with missense *dt1* alleles were further subdivided into four genotype categories depending on the presence of one of the four missense mutations (Figure 2.1D–G). Interestingly, the concordance between different *dt1* alleles and observed phenotypes was specific to the allele status. When *G. max* accessions have the *dt1R166W* allele, 84% of these were considered determinate types (Figure 2.1D). In the case of *dt1P113L*, about 64% of the accessions were scored determinate (Figure 2.1E). Notably, when either the *dt1R130K* or *dt1R62S* alleles were present, the soybean accessions were more frequently evaluated as

indeterminate or semi-determinate rather than the determinate type, with only a minority 34.8% or 48.6%, respectively scored as determinate. This disparity, which was also observed for the genetically semi-determinate *Dt1 Dt2* genotype, suggests that soybeans having the *dt1R130K* or *dt1R62S* allele are likely to have intermediate stem characteristics, similar to those with the semi-determinate genotype allele combination.

Effects of plant architecture genes and alleles on yield and lodging components in a MG VI–VII environment in the US

In order to evaluate potential agronomic merits of the tall determinate alleles and other plant architecture genes in MG VI–VII environment in the US, a total of 21 experimental and control soybean lines having each of seven different genotype combinations were produced in a field experiment at Athens, GA (33.72°N, -83.30°W) for evaluation of their morphological characteristics. Allele combinations of each of the tested soybean lines are listed in Table 2.2. The genotype categories for the soybean lines were assigned based on the allele status of the two genes for stem termination type, *Dt1* and *Dt2*, and one gene for maturity, *E1*. Among the seven genotype categories, *dt1R166W* with *E1* is the most predominantly found from commercial soybean varieties grown in the southern US environments. Four agronomic traits, plant height, number of main stem nodes, internode diameter, and lodging, were measured as parameters for evaluating agronomic performance of each genotype category under the production environment.

Plant height and number of nodes are directly associated with yield potential of soybean. In the functional *E1* background, the soybean line with *dt1R166W* alleles had

the shortest plant heights with an average plant height of 64.8 cm compared to other soybean lines under different genotype categories (Figure 2A). Even the control soybean line with the early flowering *e1-as* gene and indeterminate alleles of *Dt1* was taller (79.1 cm) than the soybeans with *dt1R166W* and late flowering *E1* alleles (Figure 2A). Notably, in the functional *E1* background, the soybean lines having either *dt1R130K* or *dt1R62S* missense mutations at the *dt1* locus were about 1.7- or 1.5-times, respectively, taller than those with *dt1R166W* alleles, confirming the earlier characterization of tall determinate genetic types in the *e1-as* background (Thompson et al., 1997). Interestingly, the soybean lines with *dt1R130K* alleles had statistically equivalent plant heights compared to lines in the *Dt1 Dt2* category, which showed intermediate plant heights typically expected from the semi-determinate stem termination type (Ping et al., 2014). The ranking of plant heights in the *E1* background had the *Dt1* line as the tallest, *dt1R166W* lines as the shortest, and the other three categories as intermediate in height with no significant difference between *dt1R130K* and *Dt1 Dt2* category lines, but *dt1R62S* lines being significantly shorter (Figure 2A).

In the case of the number of nodes, soybean lines with taller plants tended to have more nodes, and the pattern was strikingly similar between plant height and number of nodes (Figure 2). In the functional *E1* background, soybean lines with *dt1R166W* alleles had the fewest nodes with an average number of nodes of 14.3 compared to those with other stem termination genotypes (Figure 2B). Even early flowering soybean lines having *Dt1* or *Dt1 Dt2* genotypes in the *e1-as* background had more or statistically equivalent number of nodes than those with *dt1R166W* alleles. In the functional *E1* maturity gene background, genotype *Dt1 dt2* had the highest number of nodes with an average number

of nodes of 23, and soybean lines with *dt1R130K* or *Dt2* stem termination genotypes had the second most nodes with an average number of nodes of 20.7 or 20, respectively. Soybeans with *dt1R62S* alleles also had significantly higher number of nodes than those with *dt1R166W* alleles. Soybeans with the *dt1R62S* alleles also had about 5 more nodes on average than those with the *dt1R166W* allele. The ranking of number of nodes in the *E1* background had the *Dt1* line with the most, *dt1R166W* lines with the fewest, and the other three categories as intermediate in height with no significant difference between *dt1R130K* and *Dt1 Dt2* category lines, but *dt1R62S* lines having significantly fewer nodes (Figure 2B).

The soybean plant architecture parameters, internode diameter and lodging score, are closely associated with lodging and yield loss. For considering thickness of the whole parts of a main stem in each soybean plant, the mean of the diameters measured at the first, middle, and last internodes were used as a scale of the stem diameter in this study. There were significant differences in the lodging components depending on plant architecture genotype categories (Table 2.3). In the functional *E1* background, compared to the typical stem termination genotype in this environment, *dt1R166W*, most soybean lines which exhibited taller plant heights and more nodes had a stem thickness that was not significantly different to the *dt1R166W* genotypes with an average scale of stem diameter of 6.4 mm. Only the soybeans with semi-determinate *Dt2* alleles were statistically less thick than those with *dt1R166W* alleles. In the functional *E1* maturity gene background, soybean lines with either functional *Dt1* or one of three missense mutations at the *dt1* locus showed an average stem diameter ranging from 6.1 to 6.8 mm; the variations depending on the stem termination allele at the *dt1* locus were not

significant for the stem diameter. There were significant differences in lodging scores for soybean lines depending on genotype category (Table 2.3). The lodging scores of the soybeans with either of the two tall determinate alleles were significantly higher than the average lodging score of 1.0 for those with *dt1R166W* alleles. There was substantial variation for lodging scores for the lines with tall determinate alleles and to a lesser degree also for semi-determinate lines (Table 2.3).

Effects of plant architecture genes and alleles on yield and lodging components in MG IV-V environment in the US

The potential agronomic merits of the plant architecture genes were further evaluated in 2020 at Knoxville TN (35.96°N, -83.86°W), equivalent to MG IV-V in the US. A total of 20 experimental and control soybean lines under seven different genotype categories were produced to be assessed for their morphological characteristics under the production environment (Table 2.2). Like the previous field trial in GA, the genotype categories of each soybean line were classified based on the allele status at the two genes for stem termination type, *Dt1* and *Dt2*, and one maturity gene, *E1* (Table 2.2). The seven genotype categories tested in TN were slightly different from those tested in the GA, since two control lines with either Dt1E1 or Dt2e1 genotypes categories were not included; the control line under the Dt1e1 category in the GA environment (Clark), was changed to S13-10592, and two additional control lines in either the R130Ke1 or R62Se1 category were included in this field experiment. For evaluating the potential agronomic merits of each of the genotype categories in the TN production environment, eight agronomic traits, plant height, number of nodes in a main stem, stem diameter (SD),

lodging score, days to maturity (DTM), number of branches in a main stem (Bran), number of pods at the stem tip (Pod), and raceme length (RL), were measured as parameters.

The planting date in TN was later than in GA, and the plant heights of experimental and control soybean lines grown in TN were overall shorter (ranging from 25.4 cm to 86.4 cm) than those grown in GA (Figure 2.3A). As observed in the previous field trial in GA, the soybean lines with *dt1R166W* allele had the shortest plant heights with an average plant height of 45.9 cm among all the tested genotype categories in the functional *E1* maturity gene background (Figure 2.3A). The plant heights of soybean lines with the functional *E1* and the two tall determinate alleles, either *dt1R130K* or *dt1R62S*, were significantly increased with average plant heights of 61.6 cm and 52.3 cm, respectively, compared to those with the *dt1R166W* alleles in this production environment, which aligns with the observations in the field experiment in GA. Unlike the plant heights of soybean lines under the *Dt2 E1* genotype category grown in GA which showed statistically equivalent plant heights with those under the R130K *E1* category, the *Dt2 E1* lines in the TN production environment had significantly shorter plant heights with an average plant height of 49.0 cm compared the R130K *E1* lines. The ranking of plant heights in the *E1* background had the *dt1R130K* line as the tallest, *dt1R166W* lines as the shortest although statistically equivalent to the *Dt2* lines, and the *dt1R62S* lines as intermediate in height between *dt1R130K* and *dt1R166W* lines.

The ranking pattern of the number of nodes in this production environment differed from the TN environment plant heights (Figure 2.3). In the functional *E1* background, the number of nodes of soybean lines with either *dt1R130K* or *dt1R62S*

alleles were significantly increased than those with *dt1R166W* allele in this production environment, which further confirms the earlier characterization of tall determinate genetic types in the *e1-as* background (Thompson et al., 1997). The ranking of number of nodes in the *E1* background had the *Dt1 Dt2* and *dt1R130K* lines with the most (18.5 ea and 18.4 ea, respectively) with no significant difference between the two stem termination types, *dt1R166W* lines with the fewest (14.3 ea), and *dt1R130K* had an intermediate number of nodes (15.8 ea) (Figure 2B).

Similar to the observations in the GA production environment, significant differences in lodging components, stem diameter and lodging score were seen depending on the tested genotype categories (Table 2.3). In the functional *E1* background, soybean lines with the tall determinate alleles, either *dt1R130K* or *dt1R62S*, showed the thickest stems (6.9 mm or 6.8 mm, respectively) with no significant differences. The other five genotype categories had statistically equivalent stem diameters, ranging from 4.7 mm to 5.2 mm, regardless of the allele status at the *E1* maturity gene. In terms of the lodging score, soybean lines in the functional *E1* background with different stem termination types were statistically different, but the actual mean value of the lodging score by genotype category was phenotypically indistinguishable (Table 2.3). The soybean lines with *e1-as* allele showed mean lodging scores ranging from 1.5 mm to 1.7 mm, overall better in lodging score than those with *E1* alleles.

Effects of plant architecture genes and alleles on yield and lodging components in MG III–IV environment in the US

The differences in important agronomic traits related to yield potential of soybean depending on plant architecture genes were also evaluated in 2020 in Columbia, MO (38.91°N, -92.29°W), which is a MG III–IV environment in the US. A total of fourteen soybean lines including eight Clark isolines and six control lines were produced in the MO field experiment (Table 2.2). Each of the fourteen soybean lines were grouped into five genotype categories depending on the allele status of the two genes regulating stem termination type, *Dt1* and *Dt2*. In terms of the *E1* maturity gene, all the tested soybean lines in the 20MO had the *e1-as* allele, which is the maturity gene used in this production environment. Among the five tested genotype categories, Dt1e1 is the most predominantly found in the Midwest environment in the US (Bernard, 1972; Tian et al., 2010; Liu et al., 2015). A total of nine agronomic traits, plant height, number of nodes, stem diameter, lodging score, days to flowering (DTF), days to maturity (DTM), number of pods at stem tip, raceme length, and number of branches, were measured as parameters for evaluating agronomic performance depending on the allele combination in the MG III–IV environment.

In terms of plant height, the soybean lines with the functional *Dt1* allele had the tallest plants with an average plant height of 95.8 cm among the five different stem termination types tested in this environment; the plant height of soybean lines having a recessive *dt1R166W* allele had the shortest plant height with an average plant height of 36.7 cm (Figure 4A). Notably in presence of either of the two tall determinate alleles, *dt1R130K* or *dt1R62S*, the plant heights were increased about 1.7- or 1.9- times,

respectively, compared to those with the *dt1R166W* allele. In the case of the soybean lines with *dt1R62S* allele having a mean plant height of 68.7 cm, they had statistically equivalent plant heights to the soybean lines having *Dt2* with a mean plant height of 71.9 cm. In summary, the ranking of plant height in the *e1-as* background had the *Dt1* lines as the tallest, *dt1R166W* lines as the shortest and the three categories as intermediate in height with a significant difference in *dt1R130K* compared to *dt1R62S* and *Dt2* (Figure 4A). The 20MO environment results with *dt1R130K* and *dt1R62S* alleles in the *e1-as* background were consistent with the original characterization of tall determinate genetic types (Thompson et al., 1997).

For the number of nodes of the tested soybean lines in the 20MO experiment, there was a tendency that the taller plants tended to have more nodes, and the pattern of ranking was remarkably similar with that of plant height (Figure 4). In the *e1-as* background, soybean lines with the *Dt1* allele had the most nodes with an average number of nodes of 21.6, while those with *dt1R166W* had the fewest nodes with an average number of nodes of 11.8 (Figure 4B). Consistent with the earlier observation by Thompson et al. (1997), the soybean lines with either *dt1R130K* or *dt1R62S* alleles had about 5 more nodes compared to those with *dt1R166W* alleles. In terms of soybean lines with *Dt2* alleles, these had about 7 more nodes compared to those with *dt1R166W* alleles. Therefore, the ranking of number of nodes in the *e1-as* background had the *Dt1* lines with the most, *dt1R166W* lines with the fewest, and the other three categories as intermediate in number of nodes, while significant differences were observed between *Dt2* and the tall determinate category lines.

There were significant differences in the other agronomic traits measured in the 20MO production environment (Table 2.4). For stem diameter, the soybean lines having the missense alleles, *dt1R166W*, *dt1R130K* and *dt1R62S*, had the thickest stems without significant differences between the alleles (Table 2.4). The stem diameter of the soybean lines having recessive *dt1* alleles were about 0.8 to 1.6 mm thicker compared to those having either the *Dt1* or *Dt2* alleles, respectively. The mean lodging scores of all the tested genotype categories were at or below 1.3, which reflects very low lodging for lines in any of the categories (Table 2.4). As for the flowering time, the *Dt2* lines were significantly earlier flowering compared to other stem termination types with an average DTF of 42.1 days (Table 2.4). In comparison of DTF among soybean lines with either functional *Dt1* or one of the three recessive *dt1* alleles, statistically significant differences were observed, although the actual differences were less than 4 days at the most. In terms of maturity, soybean lines with *Dt1* allele matured significantly later compared to other genotype categories with an average DTM of 4.7 days later than the next closest genotype category (Table 2.4). The soybean lines with *Dt2* alleles were the earliest to mature, with the difference in DTM between *Dt2* and *Dt1* lines of about 8 days. The DTM of soybean lines with recessive *dt1* alleles was intermediate between those of soybean lines with either *Dt1* or *Dt2.* In the comparison in DTM among the soybean lines with one of the three missense mutations at the *dt1* locus, the differences were less than 2 days at the most. The average number of branches of all the tested genotype categories was less than 2, and there were no significant differences in number of branches depending on the genotype categories (Table 2.4). In terms of morphological characteristics at the stem tip, the *Dt1* lines had the shortest raceme with the fewest

number of pods at the apical stems (Table 2.4). The soybean lines with one of the three recessive alleles at the *dt1* locus had significantly longer racemes than either the *Dt1* or *Dt2* lines. There were at least about 10-fold more pods at the apical stem for the three recessive *dt1* lines and the *Dt2* lines compared to the *Dt1* lines, although there were minor significant differences in number of pods at the apical stem among the *dt1* and *Dt2* genotype categories. The soybean lines with *Dt2* alleles had intermediate raceme lengths.

DISCUSSION

Soybean stem termination type is an important agronomic trait which not only is associated with domestication and adaptation of soybean, but also affects other important agronomic traits, such as plant height, node number, flowering and maturity and so forth. Three types of stem termination type, indeterminate, determinate, and semi-determinate, have been generally accepted, although an additional stem type, called tall determinate, was introduced about 25 years ago. Only two out of three generally accepted stem termination types, either indeterminate or determinate, have been typically utilized as genetic sources for the development of soybean cultivars targeting northern and southern production environments (Cooper, 1985). Semi-determinate, the other generally accepted stem termination with intermediate stem characteristics between indeterminate and determinate, has been recognized for its potential to improve productivity in high-yielding and lodging-prone environments but is still not widely used in various breeding programs (Cooper, 1985; Ping et al., 2014). Although there are potential agronomic benefits using the other alternative stem termination type, tall determinate, in soybean variety development, it has not been broadly introduced in breeding programs due to the lack of knowledge about the trait (Thompson et al., 1997). Here, we presented the molecular basis of the alternative genetic sources to modify plant architecture of soybean and evaluated their responses in multiple production environments to provide insight for further uses in breeding programs, with the goal of genetic improvement. We found that the tall determinate stem termination type is caused by two of the previously identified missense alleles of the determinate gene *dt1*, *dt1R62S* and *dt1R130K*. The results from field experiments in three different latitudinal environments in the US revealed that

soybeans with tall determinate stem type have taller plant height and greater number of nodes while having similar stem thickness and similar pod density at the stem tip compared to the typical determinate type in both Midwest and Southern environments. We speculate that taller plants with additional nodes and therefore the potential for more pods combined with lodging resistance are expected to result in improved yield, especially under high yield production environments. To incorporate more genetic diversity in soybean variety development, it is worthwhile to utilize the tall determinate soybeans into breeding programs for soybeans with better environmental adaptability.

Through sequence analysis at the *Dt1* locus, we found that the two tall determinate alleles were rarely present (35 lines for *dt1R62S* and 23 lines with *dt1R130K* were found out of 528 soybean accessions evaluated) in soybean germplasm pools (Supplementary Table 1 in Appendix). The semi-determinate allele (*Dt2*) was also rarely found among the analyzed soybean accessions (only 23 lines out of 528 accessions have the *Dt2* allele). In particular, none of the tall determinate and semi-determinate alleles were found among 16 North American Ancestor lines evaluated (Tian et al., 2010; Zhou et al., 2015). The rare occurrences of these alleles in soybean genetic pools suggests the reason for limited use of these alleles in modern soybean breeding programs.

There are difficulties for accurate phenotyping in stem termination type due to the effects of other factors, such as environmental, management, and genetic (Li et al., 2020). In this study, we presented an objective measure of the disconnect between the stem termination phenotype for soybean accessions with the actual stem termination genotype. The tendency of higher rates of disparity was also observed from the analysis with a panel of 1120 Chinese soybean varieties (Liu et al., 2015). The higher rates of disparity

between the expected phenotype resulting from each genotype in stem termination type and how observers evaluated the phenotypes suggests the necessities of more detailed understanding of the morphological characteristics each stem termination type has under various production environments.

Our results failed to reveal obvious improvements in lodging scores. However, given the fact that generation of the experimental lines having *dt1R130K* or *dt1R62S* allele was F3:4 (Supplementary Table 2 in Appendix) and the high variation in lodging scores observed from the lines, there the possibility remains that additional evaluation in diverse environments will enable line selection with improved lodging scores based on stem termination alleles. Therefore, considering the morphological characteristics observed, the soybeans with *dt1R130K* allele are expected to have greater yield potential in the southern environments compared to the typical *dt1R166W* allele due to its effect on stem characteristics with taller plants, higher number of nodes, and stronger stems. Considering the intermediate stem characteristics of the *dt1R130K* allele which were also observed in semi-determinate stem types, the soybean lines with the tall determinate alleles are expected to have potential to improve yield in particular production environments, such as high planting density.

For a prolonged period, yield has been a major concern in soybean breeding programs. And there have been diverse approaches to improve the yield potential, since the yield is a complex system of traits affected by diverse genetic and environmental factors. A previous study on the relationship between flowering time and yield suggested that the yield improvement can be achieved by developing full-season soybeans with longer reproductive periods (Cooper, 2003). Choosing soybeans with higher plant

heights, a greater number of nodes, strong stems, and longer growth periods was suggested as an ideal strategy for the genetic improvement by optimizing plant architecture (Li et al., 2020; Liu et al., 2020). Previous studies quantified the genetic changes to yield and yield stability which occurred across eight decades from 1928 to 2008 in soybean breeding in Northern and Southern environments in the US (Boehm Jr et al., 2019; Rincker et al., 2014). The authors found that the yield selections over the time have resulted in shorter plants with better lodging resistance (Boehm Jr et al., 2019; Rincker et al., 2014). Information about the stem termination types of the released soybean varieties is missing in the articles. However, considering the absence of tall determinate alleles in US soybean ancestor lines as well as the released soybean varieties were targeted either to northern or southern parts of the US where the majority of soybean varieties grown are indeterminate or determinate, respectively, it is highly probable that the tall determinate alleles have not been utilized and thus have not been examined. Therefore, it is still worthwhile to utilize the alternative stem termination type, tall determinate, into breeding programs since it has potential to improve yield by generating taller plants with lodging resistance. Also, additional studies need to be conducted preferably with near isogenic lines in various production environments to reveal the impact of the tall determinate stem termination type on yield of soybean.

TABLES

Table 2.1. Description of alleles and genes controlling stem termination type and maturity used in this study

Gene	Gene Identifier	Allele	Polymorphism	Position[†]	Protein Change	Reference
Dt1 (*GmTFL1*)	Glyma.19g194300[‡]	*dt1R62S*	SNP: C/A	45184804	R62S	Tian et al. (2010)
		dt1P113L	SNP: G/A	45183859	P113L	Tian et al. (2010)
		dt1R130K	SNP: C/T	45183808	R130K	Tian et al. (2010)
		dt1R166W	SNP: T/A	45183701	R166W	Tian et al. (2010)
Dt2	Glyma.18g273600	*dt2*	SNP: G/A[§]	55642486		
E1	Glyma.06g207800	*e1-as*	SNP: C/G	20207322	T15R	
E2	Glyma.10g221500					Watanabe et al. (2011)
E3	Glyma.19g224200					Watanabe et al. (2009)

[†] The positions are based on the Williams 82 soybean reference genome sequence (Wm82.a2.v1)
[‡] Since the glyma.19g194300 is on the reverse strand in the genomic sequence, SNPs at the locus are reverse complement on chromosome 19
[§]The recessive *dt2* was designated based on allele status at the position of an associated SNP marker, ss715632223, (Pos: 55642486 on chromosome 18)

Table 2.2. Allele combination of experimental and control lines used for field experiments in three different latitudinal environments

Genotype Category	Gene					Soybean lines produced in each production experiment			
	Dt1	*Dt2*	*E1*	*E2*	*E3*	Name	19GA	20TN	20MO
Dt1E1	*Dt1*	*dt2*	*E1*	*E2*	*E3*	G19-197	Incl.[†]		
R166WE1	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Jake	Incl.	Incl.	
						Ellis		Incl.	
						S11-20242C		Incl.	
						G19-192	Incl.		
R130KE1	*dt1R130K*	*dt2*	*E1*	*E2*	*E3*	KB17-16 #1	Incl.	Incl.	
						KB17-16 #2	Incl.	Incl.	
						KB17-16 #3	Incl.		
						KB17-16 #4	Incl.		
						KB17-16 #5	Incl.		
						KB17-16 #6	Incl.		
						KB17-16 #7	Incl.		
						KB17-16 #8	Incl.		
R62SE1	*dt1R62S*	*dt2*	*E1*	*E2*	*E3*	KB17-17 #1	Incl.	Incl.	
						KB17-17 #2	Incl.	Incl.	
						KB17-17 #3	Incl.	Incl.	
						KB17-17 #4	Incl.	Incl.	
						KB17-17 #5	Incl.		
						KB17-17 #6	Incl.		
Dt2E1	*Dt1*	*Dt2*	*E1*	*E2*	*E3*	KB17-7 #1	Incl.		
						KB17-7 #2		Incl.	
						KB17-7 #3		Incl.	
						KB17-7 #4		Incl.	
						KB17-7 #5		Incl.	
						KB17-8 #1	Incl.		
						KB17-8 #2		Incl.	
						KB17-8 #3		Incl.	
						KB17-8 #4		Incl.	
						KB17-8 #5		Incl.	
Dt1e1	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Clark	Incl.		
						Williams 82			Incl.
						Jack			Incl.
						LG04-6000			Incl.
						S07-5049			Incl.
						S13-10592		Incl.	

R166We1	*dt1R166W*	*dt2*	*e1-as*	*E2*	*E3*	L63-3016			Incl.
						L63-3297			Incl.
						L65-792			Incl.
						L72-1737			Incl.
R130Ke1	*dt1R130K*	*dt2*	*e1-as*	*E2*	*E3*	L91-8052		Incl.	Incl.
R62Se1	*dt1R62S*	*dt2*	*e1-as*	*E2*	*E3*	L91-8060		Incl.	Incl.
Dt2e1	*Dt1*	*Dt2*	*e1-as*	*E2*	*E3*	LG90-2550	Incl.		Incl.
						L62-1251			Incl.
						L73-811			Incl.
						KB18-14-1375			Incl
Total number of evaluated soybeans in each of field experiments							21	20	14

† Each of the soybean lines followed by 'Incl.' means the line was included in each field experiment marked on the top of the column

Table 2.3. Mean values of internode diameter and lodging score for soybeans having different genotype combinations in field experiments at two different southern environments in 2019 and 2020 (19GA and 20TN)

Genotype Category	Gene			19GA			20TN		
	Dt1	*Dt2*	*E1*	n[†]	SD[‡]	Lodging[§]	n[†]	SD[‡]	Lodging[§]
Dt1E1	*Dt1*	*dt2*	*E1*	1	6.2 ± 0.6 ab	1.0 ± 0.0 d			
R166WE1	*dt1R166W*	*dt2*	*E1*	2	6.4 ± 1.0 a	1.0 ± 0.0 d	3	5.2 ± 1.3 b	2.2 ± 0.6 c
R130KE1	*dt1R130K*	*dt2*	*E1*	8	6.8 ± 1.2 a	2.8 ± 1.0 b	2	6.9 ± 1.7 a	2.4 ± 0.3 ab
R62SE1	*dt1R62S*	*dt2*	*E1*	6	6.1 ± 0.8 ab	1.6 ± 0.9 c	4	6.8 ± 1.6 a	2.4 ± 0.3 b
Dt2E1	*Dt1*	*Dt2*	*E1*	2	5.5 ± 1.5 bc	3.0 ± 0.5 a	8	5.2 ± 1.3 b	2.2 ± 0.3 c
Dt1e1	*Dt1*	*dt2*	*e1-as*	1	5.2 ± 0.8 c	1.0 ± 0.0 d	1	4.9 ± 1.0 b	2.5 ± 0.0 a
R166We1	*dt1R166W*	*dt2*	*e1-as*						
R130Ke1	*dt1R130K*	*dt2*	*e1-as*				1	4.7 ± 1.0 b	1.7 ± 0.2 d
R62Se1	*dt1R62S*	*dt2*	*e1-as*				1	4.7 ± 0.9 b	1.5 ± 0.0 e
Dt2e1	*Dt1*	*Dt2*	*e1-as*	1	3.1 ± 0.9 d	1.0 ± 0.0 d			
			LSD (0.05)		0.8	0.0		0.6	0.1

[†] Number of lines within each of genotype categories
[‡] SD, mean of stem diameters (mm) measured at first, middle, and last internode at maturity
[§] Lodging was scored at maturity on a scale of one (all plants erect) to five (almost all plants prostrate)

Table 2.4. Mean values of agronomic traits of soybeans having different genotype combinations measured at Columbia, MO in 2020 (20MO)

Genotype Category	Gene			n[†]	Agronomic traits						
	Dt1	*Dt2*	*E1*		SD[‡]	Lodging[§]	DTF[¶]	DTM[#]	Branch[$]	Pod[≈]	RL[≠]
Dt1e1	*Dt1*	*dt2*	*e1-as*	4	4.9 ± 1.1 b	1.3 ± 0.2 a	45.3 ± 2.2 b	122.7 ± 2.6 a	1.9 ± 2.0 a	0.8 ± 0.5 c	0.3 ± 0.4 c
R166We1	*dt1R166W*	*dt2*	*e1-as*	4	6.5 ± 1.5 a	1.1 ± 0.3 bc	44.7 ± 1.9 b	116.8 ± 4.4 bc	1.7 ± 1.5 ab	8.9 ± 4.0 ab	4.3 ± 2.2 a
R130Ke1	*dt1R130K*	*dt2*	*e1-as*	1	6.5 ± 1.3 a	1.0 ± 0.0 c	43.7 ± 0.5 c	116.0 ± 2.5 cd	1.3 ± 1.0 ab	10.3 ± 2.6 a	4.6 ± 1.7 a
R62Se1	*dt1R62S*	*dt2*	*e1-as*	1	6.1 ± 1.4 a	1.2 ± 0.2 ab	47.0 ± 0.0 a	118.0 ± 2.9 b	1.1 ± 1.4 b	8.1 ± 3.2 b	4.4 ± 2.0 a
Dt2e1	*Dt1*	*Dt2*	*e1-as*	4	5.3 ± 1.0 b	1.0 ± 0.1 bc	42.1 ± 2.9 d	114.8 ± 3.0 d	1.5 ± 1.2 ab	7.8 ± 2.9 b	2.2 ± 0.7 b
			LSD (0.05)		0.7	0.1	0.7	1.4	0.8	1.4	0.7

[†] Number of lines within each of genotype categories
[‡] SD, mean of stem diameters (mm) measured at first, middle, and last internode at maturity
[§] Lodging was scored at maturity on a scale of one (all plants erect) to five (almost all plants prostrate)
[¶] DTF, Days to Flowering
[#] DTM, Days to Maturity
[$] Number of branches in the main stem
[≈] Pod, represents nunber of pods at apical stem
[≠] RL, Raceme Length in cm

FIGURES

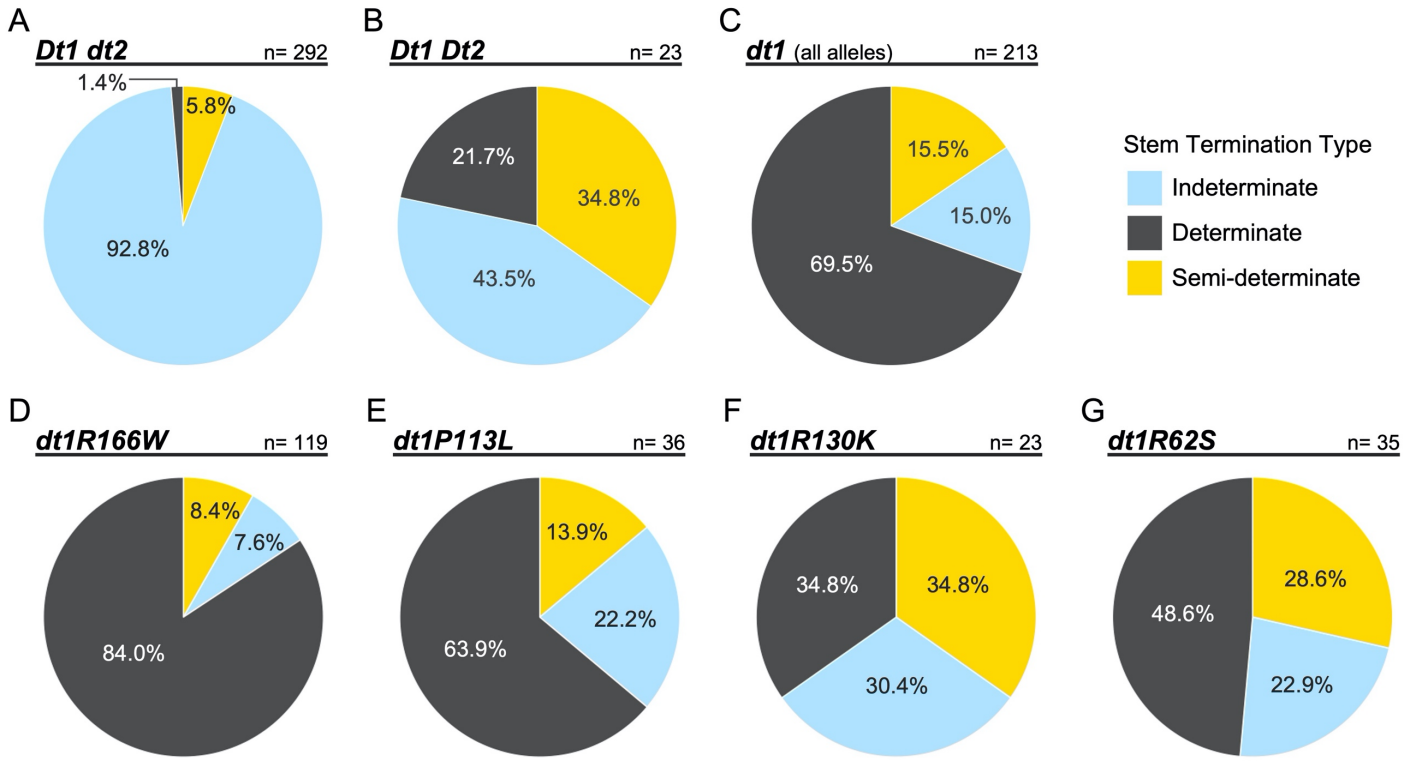

Figure 2.1. Pie charts showing the rates of concordance between publicly available phenotype and genotype for stem termination types of 528 *G. max* accessions. The phenotype data were downloaded from GRIN database. Each of the evaluated phenotypes was highlighted by colors – blue: indeterminate, yellow: semi-determinate, grey: determinate. Genotype categories about which each of the pie charts illustrates were written in bold and italic characters above of lines drawn over each chart. Percentages written in pie charts represent rates on how often people evaluate the stem growth habits of soybean accessions within each genotype category. Digits following by n= represented numbers of the *G. max* accessions within each genotype category. **(A–C)** illustrated how stem termination types of soybean accessions with each genotype combination for *Dt1* and *Dt2* genes have been phenotypically scored **(D–G)** showed variations in scored phenotypes depending on the presence of each of the four missense mutations at the *dt1* locus.

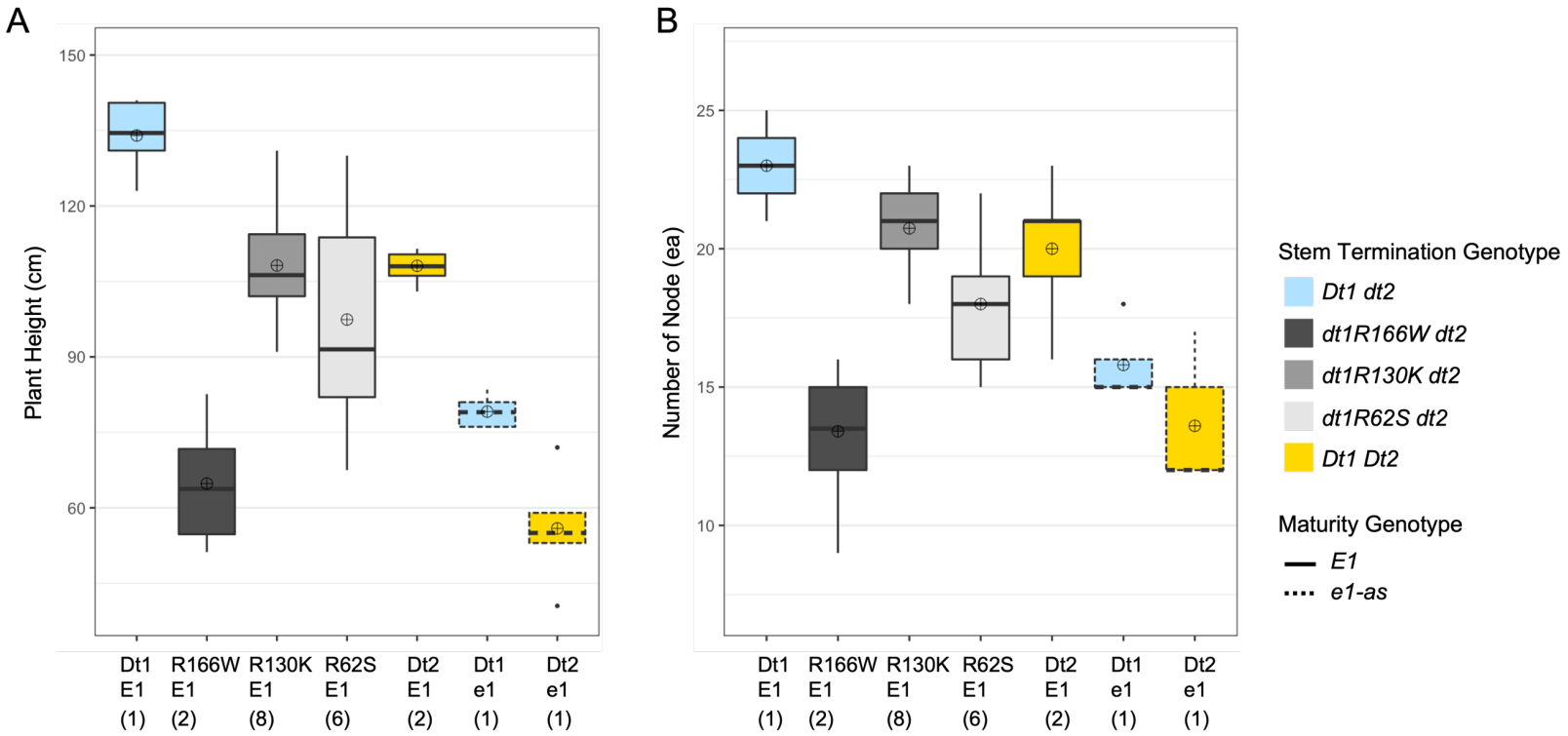

Figure 2.2. Boxplots of plant height and number of nodes depending on genotype categories examined in 2019 at Athens, GA (19GA). Total 21 soybean lines with each allele combination for genes related to stem termination type and maturity were evaluated. The colors of each box were designated based on the allele status of the two genes conferring stem termination type, *Dt1* and *Dt2*. The outlines of each box were designated based on the allele status of the *E1* maturity gene. Sixteen soybean lines under three genotype categories, R166WE1, R62SE1, and Dt2E1, were experimental lines from KB17-16 ($F_{3:4}$), KB17-17 ($F_{3:4}$), and either KB17-7 or KB17-8 RIL populations ($F_{4:5}$), respectively. The other five soybean lines were control varieties. Each genotype category has an unequal number of soybean lines, and the number of lines were written in parentheses below each genotype category. The plant height and number of nodes were measured from five individual plants per line. ⊕ in each box represents mean values of each genotype category. **(A)** displays the distribution of final plant height at maturity in cm. **(B)** shows the distribution of number of nodes counted from a main stem at maturity.

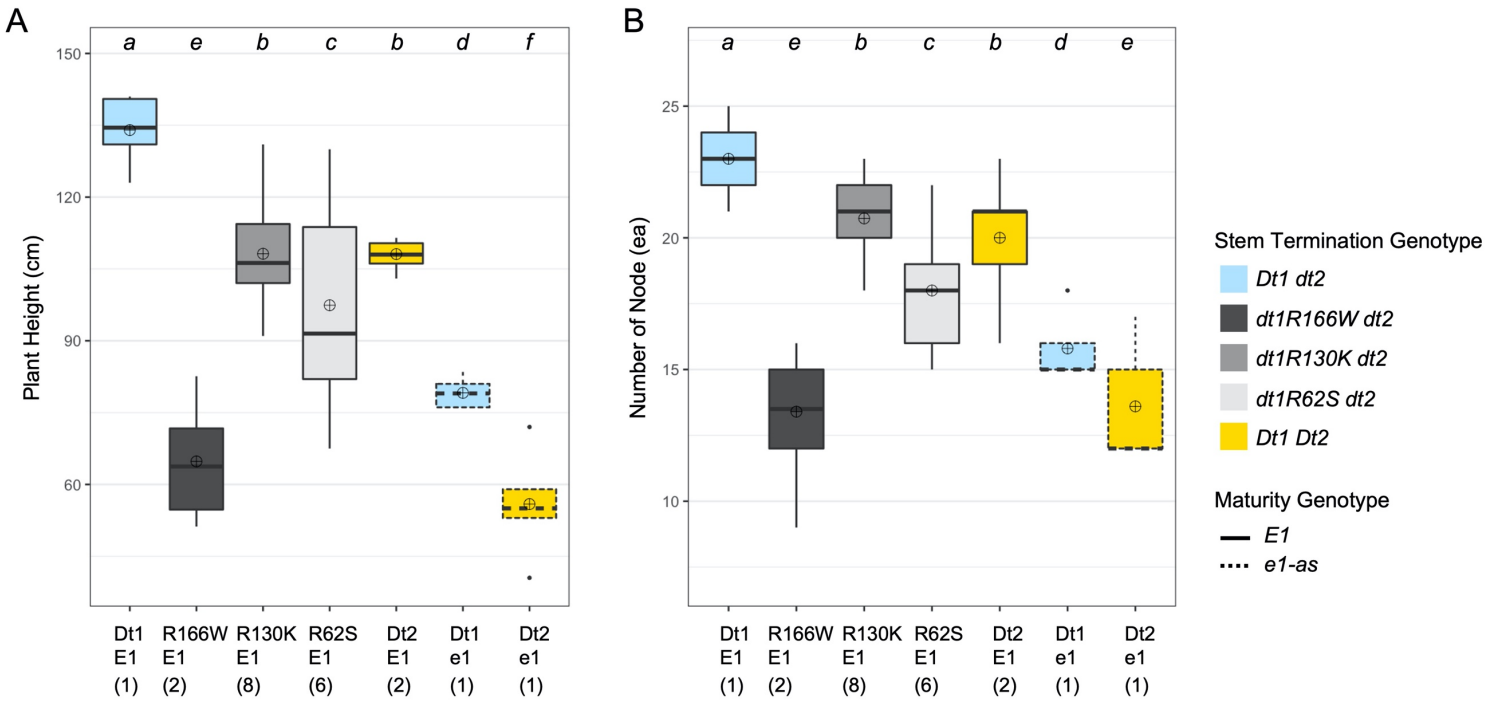

Figure 2.3. Boxplots of plant height and number of nodes depending on genotype categories examined in 2020 at Knoxville, TN (20TN). A total 20 soybean lines with each allele combination for genes related to stem termination type and maturity were evaluated. The colors of each box were designated based on the allele status of the two genes conferring stem termination type, *Dt1* and *Dt2*. The outlines of each box were designated based on the allele status of the *E1* maturity gene. Fourteen soybean lines under three genotype categories, R166WE1, R62SE1, and Dt2E1, were experimental lines from KB17-16($F_{3:5}$), KB17-17($F_{3:5}$), and either KB17-7 or KB17-8 RIL populations ($F_{4:6}$), respectively. The other six soybean lines were control varieties. Each genotype category has an unequal number of soybean lines, and the number of lines were written in parentheses below each genotype category. The plant height and number of nodes were measured from ten individual plants per plot. ⊕ in each box represents mean values of each genotype category. The genotype categories marked with a common lower-case letter indicate that the means are not significantly different (p=0.05) according to Fisher's Least Significant Difference (LSD) procedure. **(A)** displays the distribution of final plant height at maturity in cm. **(B)** shows the distribution of number of nodes counted from a main stem at maturity.

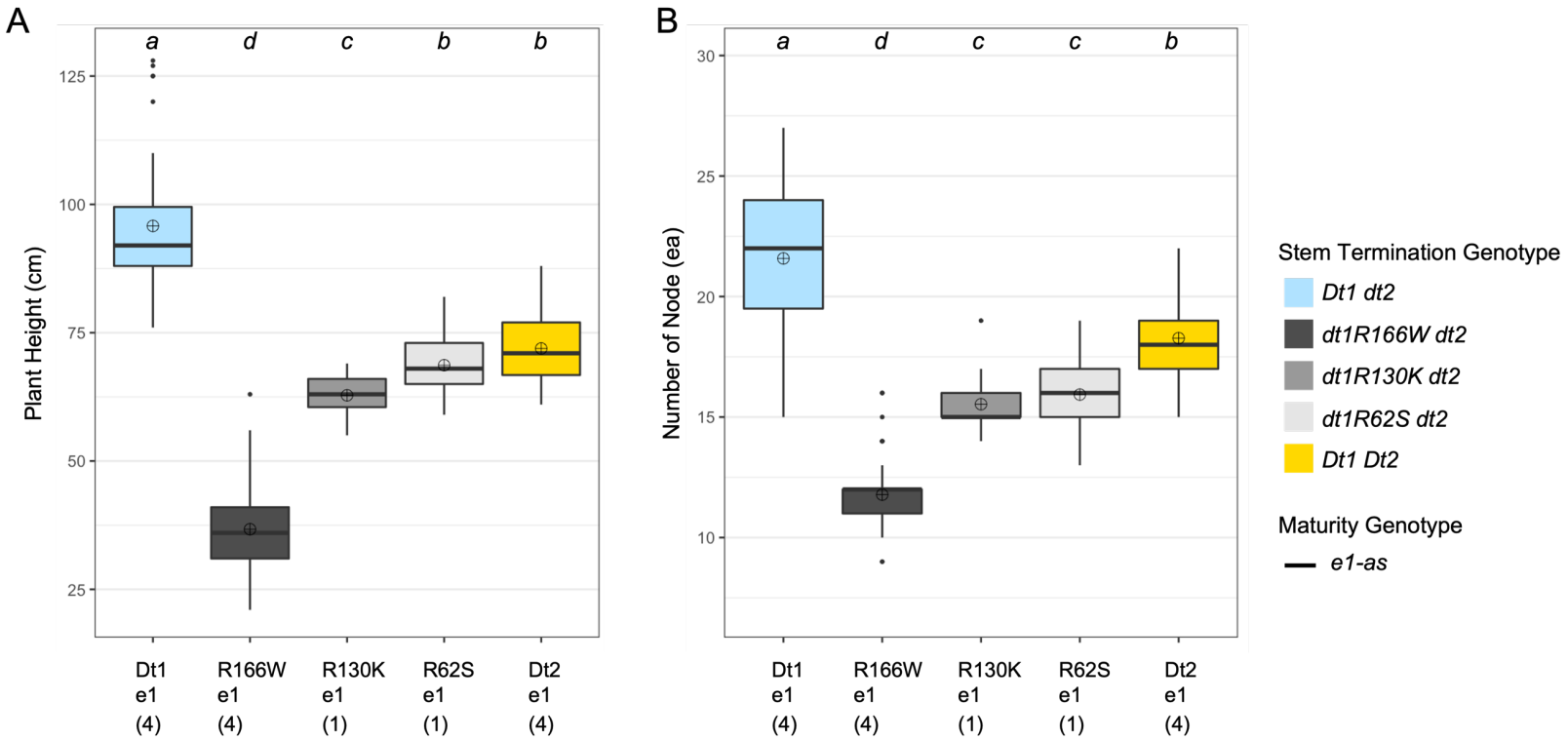

Figure 2.4. Boxplots of plant height and number of nodes depending on genotype categories examined in 2020 at Columbia, MO (20MO). Total 14 soybean lines with each allele combination for genes related to stem termination type and maturity were evaluated. The colors of each box were designated based on the allele status of the two genes conferring stem termination type, *Dt1* and *Dt2*. All the experimental and control soybean lines had *e1-as* maturity allele. Each genotype category has an unequal number of soybean lines, and the number of lines were written in parentheses below each genotype category. The plant height and number of nodes were measured from 15 individual plants per plot. ⊕ in each box represents mean values of each genotype category. The genotype categories marked with a common lower-case letter indicate that the means are not significantly different (p=0.05) according to Fisher's Least Significant Difference (LSD) procedure. **(A)** displays the distribution of final plant height at maturity in cm. **(B)** shows the distribution of number of nodes counted from a main stem at maturity.

REFERENCES

Bernard, R.L., Two genes affecting stem termination in soybeans. Crop Science, 1972. 12(2): p.235-239.

Boehm Jr, J.D., Abdel-Haleem, H., Schapaugh Jr, W.T., Rainey, K., Pantalone, V.R., Shannon, G., Klein, J., Carter Jr, T.E., Cardinal, A.J., Shipe, E.R. and Gillen, A.M., Genetic improvement of us soybean in maturity groups V, VI, and VII. Crop Science, 2019. 59(5), p.1838-1852.

Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A. and Kong, F., Molecular mechanisms of flowering under long days and stem growth habit in soybean. Journal of Experimental Botany, 2017. 68(8): p.1873-1884.

Cooper, R.L., Breeding semidwarf soybeans. Plant breeding reviews, 1985. 3: p.289-311.

Cooper, R.L., A delayed flowering barrier to higher soybean yields.Field Crops Research, 2003. 82(1), p.27-35.

Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database. [Rome] :FAO, 1997.

Hartung, R.C., Specht, J.E. and Williams, J.H., Modification of soybean plant architecture by genes for stem growth habit and maturity. Crop Science, 1981. 21(1): p.51-56.

Heatherly, L.G. and Smith, J.R., Effect of soybean stem growth habit on height and node number after beginning bloom in the midsouthern USA. Crop science, 2004. 44(5): p.1855-1858.

Kilgore-Norquest, L. and Sneller, C.H., Effect of stem termination on soybean traits in southern US production systems. Crop science, 2000. 40(1): p.83-90.

Langewisch, T., Lenis, J., Jiang, G.-L., Wang, D., Pantalone, V., and Bilyeu, K., The development and use of a molecular model for soybean maturity groups. BMC Plant Biology, 2017. 17(1): p. 91.

Li, M., Liu, Y., Wang, C., Yang, X., Li, D., Zhang, X., Xu, C., Zhang, Y., Li, W. and Zhao, L., Identification of traits contributing to high and stable yields in different soybean varieties across three Chinese latitudes. Frontiers in plant science, 2020. 10: p.1642.

Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., Nagamatsu, A., Arai, M., Yamada, T., Kitamura, K. and Masuta, C., The soybean stem growth habit gene *Dt1* is an ortholog of *Arabidopsis TERMINAL FLOWER1*. Plant Physiology, 2010. 153(1): p.198-210.

Liu, G., Zhao, L., Averitt, B.J., Liu, Y., Zhang, B., Chang, R., Ma, Y., Luan, X., Guan, R. and Qiu, L., Geographical distribution of *GmTfl1* alleles in Chinese soybean varieties. The Crop Journal, 2015. 3(5): p.371-378.

Liu, S., Zhang, M., Feng, F. and Tian, Z., Toward a "green revolution" for soybean. Molecular plant, 2020. 13(5): p.688-697.

Pantalone, V., Smallwood, C. and Fallen, B., Development of 'Ellis' soybean with high soymeal protein, resistance to stem canker, southern root knot nematode, and frogeye leaf spot. Journal of Plant Registrations, 2017. 11(3): p.250-255.

Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., Sui, Y., Lin, F., Liu, X., Tang, Z. and Nguyen, H., *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. The Plant Cell, 2014. 26(7): p.2831-2842.

Rincker, K., Nelson, R., Specht, J., Sleper, D., Cary, T., Cianzio, S.R., Casteel, S., Conley, S., Chen, P., Davis, V. and Fox, C., Genetic improvement of US soybean in maturity groups II, III, and IV. Crop Science, 2014. 54: p.1419-1432

Shannon, J.G., Wrather, J.A., Sleper, D.A., Robbins, R.T., Nguyen, H.T. and Anand, S.C., Registration of 'Jake' soybean. Journal of plant registrations, 2007. 1(1): p.29-30.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B., Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PloS one, 2013. 8(1): p.e54985.

Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J.E., Nelson, R.L., McClean, P.E., Qiu, L. and Ma, J., Artificial selection for determinate growth habit in soybean. Proceedings of the National Academy of Sciences, 2010. 107(19): p.8563-8568.

Thompson, J.A., Bernard, R.L. and Nelson, R.L., A third allele at the soybean *dt1* locus. Crop science, 1997. 37(3): p.757-762.

Thompson, J.A., Amdor, P.J. and Nelson, R.L., Registration of LG90-2550 and LG91-7350R soybean germplasm. Crop science, 1999. 39(1): p.302-303.

Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., Chen, C., Li, Y., Joshi, T., Song, L. and Vuong, T.D., Landscape of genomic diversity and trait discovery in soybean. Scientific reports, 2016. 6(1): p.1-10.

Yue, L., Li, X., Fang, C., Chen, L., Yang, H., Yang, J., Chen, Z., Nan, H., Chen, L., Zhang, Y. and Li, H., FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. Journal of Integrative Plant Biology, 2021. 63(6): p.1004-1020.

Zeng, S., Škrabišová, M., Lyu, Z., Chan, Y.O., Dietz, N., Bilyeu, K. and Joshi, T., Application of SNPViz v2. 0 using next-generation sequencing data sets in the discovery of potential causative mutations in candidate genes associated with phenotypes. International Journal of Data Mining and Bioinformatics, 2021. 25(1-2): p.65-85.

Zhang, D., Wang, X., Li, S., Wang, C., Gosney, M.J., Mickelbart, M.V. and Ma, J., A post-domestication mutation, *Dt2*, triggers systemic modification of divergent and convergent pathways modulating multiple agronomic traits in soybean. Molecular plant, 2019. 12(10): p.1366-1382.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y. and Fang, C., Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nature biotechnology, 2015. 33(4): p.408-414.

CHAPTER 3

Redesigning Soybean with Improved Oil and Meal Traits

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] is an important crop worldwide, and its overall value comes from vegetable oil and high protein meal primarily used for animal feeding. To increase the value of the oilseed, the high oleic and low linolenic acid oil trait (HOLL; >70% oleic and <3% linolenic acid) is targeted to maximize the functionality of soybean oil while capturing all health properties. For soybean seed meal, research is directed toward increasing its metabolizable energy (ME) by altering the carbohydrate profile with increased sucrose and decreased anti-nutritional factors; raffinose and stachyose, the raffinose family of oligosaccharides (RFOs) negatively affect ME while sucrose positively impacts ME in the diets of monogastric animals Previous research identified four variant alleles of major fatty acid desaturase (*FAD*) genes necessary for the HOLL trait in soybean oil, and two alleles of key raffinose synthase (*RS*) genes responsible for reduced or nearly eliminated levels of RFOs in soybean meal. The development of soybean varieties with the desired allele combinations for improved oil and meal quality are expected to provide a compositional value bundle for soybean. This research aims to evaluate the interactions of the variant alleles on modified fatty acid profiles in the oil and carbohydrate profiles in experimental soybean germplasm containing different allele combinations. The results from eight environments indicated that the four alleles of the *FAD* genes increased oleic acid content over 80% and reduced linolenic acid contents to less than 3% of total fatty acid profile of the seed oil regardless of combinations with variant alleles of the *RS* genes. Also, whichever genotype combinations of *FAD* genes, experimental soybean germplasm with the two variant alleles of the *RS* genes increased seed sucrose content significantly and reduced stachyose content to less than 1% of the

total carbohydrate composition. The results will determine the feasibility of soybean variety development with this unique combination of oil and meal traits.

INTRODUCTION

Soybean is an important crop worldwide which is widely used as an oilseed as well as a protein source for animal feeding. For its commercial uses, the soybean seeds first undergo an array of processing procedures including cleaning, crushing, dehulling, flaking, and pre-press or solvent extraction, for the oil extraction. After the removal of the oil, the remaining flakes are used to produce soybean meal mostly for animal feeds or are further processed to produce diverse soybean protein products. In 2020, United States soybean growers planted 83.1 million acres of soybean and harvested 4.14 billion bushels of soybean, which were valued at $30.5 billion dollars (SoyStats, 2021). Given the fact that one bushel of soybean weights 60 pounds and produces about 12 pounds of oil and 47 pounds of protein-rich meal, technically about 2.08 billion bushels of soybean were consumed in the production of soybean oil and approximately 2.15 billion bushels of soybean were consumed in the production of soybean meal in the US, which eventually turned into 24.9 billion pounds of soybean oil and 99 billion pounds of soybean meal, respectively (SoyStats, 2021). Considering the fact that soybean is the largest oilseed crop in the world representing 59% of the total oilseed production, and 55% of the vegetable oil consumed in the US as well as 70% of the protein meal utilized worldwide were from soybean, it is obvious that overall value of soybean comes from the vegetative oil and high protein meal for animal feeds.

While soybean is an excellent source of vegetable oil and protein, the soybean seed composition can be further improved to enhance its functionality. For improving functionality of soybean as a source for vegetable oil, soybean varieties with superior oil fatty acid profiles have been developed through traditional plant breeding efforts (Pham

et al. 2012). The oil extracted from commodity soybeans has typically undergone a chemical hydrogenation process to reduce the amount of polyunsaturated fatty acids (PUFA), linoleic acid and linolenic acid oil components; as a consequence, the more beneficial and oxidatively stable oleic acid oil component was increased. The hydrogenation process contributes to the improved oxidative stability of soybean oil, but it also created an unwanted by-product in the oil, 10–40 % *trans* fats, which have negative effects on health and thus have been regulated in foods (Hu et al. 1997; FDA 2003; FDA 2015). Genes controlling oleic acid and PUFA content in soybean seed oil have been characterized: two oleate desaturase genes *FAD2-1A* (Glyma.10g278000) and *FAD2-1B* (Glyma.20g111000) and the three linoleate desaturase genes *FAD3A* (Glyma.14g194300), *FAD3B* (Glyma.02g227200) and *FAD3C* (Glyma.18g062000) (Bilyeu et al. 2003; Bilyeu et al. 2005; Bilyeu et al. 2006; Heppard et al. 1996; Pham et al. 2011; Pham et al. 2012; Schlueter et al. 2007; Schmutz et al. 2010). Our earlier research reported four alleles responsible for the most dramatic increase in oleic acid and decrease in linolenic acids, null alleles of *FAD2-1A* and *FAD3C* as well as missense alleles of *FAD2-1B* and *FAD3C* (Bilyeu et al. 2018a). It has been confirmed that soybean lines having the combination of these four alleles can successfully produce the high oleic and low linolenic acid (HOLL) seed oil phenotype with over 80% oleic acid, 3–7% linoleic acid, and less than 3% linolenic acid (Hagely et al. 2021; Pham et al. 2011; Pham et al. 2012).

For improving the functionality of soybean seeds as a source for animal feeds, research has been directed toward increasing metabolizable energy of soybean meal by modifying soluble carbohydrate composition in soybean seeds, which is closely

associated with the digestibility of the soybean meal (Dierking and Bilyeu 2008; Dierking and Bilyeu 2009; Hagely et al. 2013; Jo et al. 2018; Jo et al. 2019; Hagely et al. 2020). Soybean meal represents about 70% of world protein meal consumption and majority of it is consumed by poultry and swine (SoyStats, 2021). The dry weight of a typical soybean mainly consists of about 20% oil, 40% protein, and 15% soluble carbohydrate (Openshaw and Hadley 1978). Soybean meal, the by-product of the extraction of soybean oil, contains the protein fraction which make it a valuable source for animal feeding; also in the soybean meal is the soluble carbohydrate fraction which primarily consists of sucrose, raffinose, and stachyose (Hsu et al. 1973). Sucrose is fully digested in the animal gut and thus utilized as a net positive for metabolizable energy; in contrast, raffinose and stachyose, the raffinose family of oligosaccharides (RFO), which are derived from sucrose, cannot be digested by monogastric animals -including humans as well as poultry and swine- due to the absence of the relevant enzymes required for its digestion. Alternative processing procedures, such as ethanol extraction, have been explored to remove RFO in soybean meal, but the procedures result in the reduction of sucrose at the same time with the reduction of RFO, which lowered the overall quality of soybean meal (Coon et al. 1990). Conventional soybean breeding has led to the development of soybean varieties having traits of reduced amounts of seed RFO content, termed as the Low RFO and Ultra-Low (UL) RFO traits, in soybean seeds (Hagely et al. 2020). From the earlier genetic studies, genes reducing RFO content in soybean seed have been characterized: two raffinose synthase genes, *RS2* (Glyma.06g179200) and *RS3* (Glyma.05g003900) that, when functional, produce raffinose and *myo*-inositol from sucrose and galactinol. An allele with a three base pair nucleotide deletion normally

encoding a highly conserved tryptophan at position 331 from the start codon of the *RS2* gene, referred to as *rs2W331-*, was found to be associated with the Low RFO trait causing less than 2% seed RFO content (Hagely et al. 2013; Hagely et al. 2020; Jo et al. 2018; Jo et al. 2019). Further reductions of up to less than 1% RFO content were observed when the *rs2W331-* alleles were combined with either of two independent alleles of *RS3,* referred to as *rs3snp6* and *rs3G75E* (Hagely et al. 2020). Therefore, it has been confirmed that soybean lines having the mutations in *RS2* and *RS3* genes can successfully reduce the RFO content up to less than 1% of the total carbohydrate composition; along with this reduction in negative components, the net content of the positive component sucrose was significantly increased (Hagely et al. 2020; Jo et al. 2018; Jo et al. 2019).

Considering soybean value is derived from the oil and meal, a combination of the oil value trait, HOLL, with one of the meal value traits, Low or UL RFO, would be a good strategy to improve overall values of the soybean crop. Combining the desired oil and meal value traits through molecular-driven breeding is expected to improve the overall quality of soybean by creating higher metabolizable energy soybean meal and improved stability soybean oil without health issues. However, knowledge about potential interactions between the improved fatty acid profile and reduced RFO contents in soybean seeds is still limited. Here, we utilized a molecular marker-assisted breeding strategy which successfully combined the six alleles responsible for the desired soybean oil and meal value traits into a matrix of oil and meal traits culminating with the HOLL plus Low or UL RFO combinations, and reported the analysis of seed composition for the novel soybean germplasm lines produced in field studies.

MATERIALS AND METHODS

Plant materials and population developments

In order to evaluate possible interactions of the variant alleles for the seed modified fatty acids and carbohydrate profiles, experimental and control soybean lines with different genotype combinations were produced in a total of eight different locations during 2019 and 2020 field seasons. Each allele had one original donor source (Table 2.1). The high oleic acid trait is conditioned by the combination of mutant alleles in the *FAD2-1A* and *FAD2-1B* genes (Pham et al. 2010; Pham et al. 2011; Bilyeu et al. 2015a; Bilyeu et al. 2015b; Bilyeu et al. 2018b; Bilyeu et al. 2019; Bilyeu et al. 2020). Altered seed carbohydrate profiles are conditioned by one or two variant raffinose synthase genes (Bilyeu et al. 2008; Hagely et al. 2020; Schillinger et al. 2013; Schillinger et al. 2018). Ongoing soybean germplasm development utilized a molecular breeding approach that consisted of soybean crossing at the South Farm Research Center near Columbia, Missouri during the annual field season (May-October) typically followed by two generations of advancement and genotype selection in a winter nursery; additional genotype or phenotyping assays were used to combine the desired alleles. The F_1 seeds were sent to a winter nursery near Upala, Costa Rica and advanced one cycle to produce F_2 seeds. In the second off-season generation, the F_2 plants were sampled with Whatman® FTA® cards (Whatman, Clifton, NJ, USA) for genotyping with molecular marker assays for the desired alleles, and selected $F_{2:3}$ seeds from single plant threshes were returned to Missouri to be planted and used as parents in subsequent Missouri field seasons. Genotyping assays were used again to confirm the status of targeted alleles or identify selections that were still segregating from some of the genes. These schemes

were followed to generate the experimental soybean lines (Figure 3.1). In some cases of earlier germplasm development, a chipped portion of F_2 or F_3 seeds was used for fatty acid analysis and selection of the remnant seed with the desired fatty acid profile as described previously (Pham et al. 2010).

Allele-specific molecular marker assays

Seven separate allele-specific molecular marker assays developed from an array of previous studies were utilized to distinguish soybean lines with each targeted mutation. for the six genes: *FAD2-1A*, *FAD2-1B*, *FAD3A*, *FAD3C*, *RS2* and *RS3*. The molecular marker assay of the *FAD2-1A* indel allele was conducted as described by Pham et al. (2011). For the *FAD2-1B* P137R allele, the assay was carried out as described in Pham et al. (2012). SimpleProbe assays for *FAD3A* splice site (G810A) and *FAD3C* G128E were conducted as described by Bilyeu et al. (2011). In terms of the *rs2 W331-*, *rs3 snp6* alleles, and *rs3 G75E* alleles, SimpleProbe assays were conducted as described by Hagely et al. 2020.

Growth conditions

A total of 23 soybean experimental and control lines were planted at three different locations over 2 years in 2019 and 2020. The three locations were as follows: South Farm Research Center in Columbia, MO; Greenley Research Center near Novelty, MO; and the Purdue University Agronomy Center for Research and Education near West Lafayette, IN. For the production environment of South Farm Research Center in Columbia, MO, the subsets of soybean lines were planted in two different planting dates

in both years. Therefore, considering each location-year as a single environment, there were a total of eight different production environments. For each environment, ten seeds of each soybean line were planted, and the experimental design was a randomized complete block design with three replications. The soybeans were planted per plot by hand into 91 cm plots with 30 cm spacing. Three single plants within each plot were harvested together for seed samples to further analyze the seed fatty acid and carbohydrate components.

Analysis of seed carbohydrate compositions

A subset of total soluble carbohydrates was determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing a Dionex ICS-5000 with Electrochemical Detector (Thermo Scientific Dionex, Waltham, MA). A total of 15 seeds of each soybean line were lyophilized to dryness in Speed Vac prior to powdering. A 12.5 mg portion of ground sample was extracted with 1 mL of 50 % ethanol at 70 °C for 30 min including intermittent shaking three times in a 2 mL microcentrifuge tube. The samples were then centrifuged 15 min at 16,000g so that the supernatant was passed through a 0.2-μm filter. Around 600 mL of the supernatant was taken and stored at 4 °C before further experiments. Following filtration, A 50-mL aliquot of each sample was dried under speed vacuum and resuspended in 250 μL deionized water. The resuspended samples were arrayed in a 96-well plate and automatically applied to the column with an injection volume of 10 mL. For the separation of the soluble carbohydrates, a Dionex Carbo Pac PA 10 analytical column (250 mm × 4 mm, 10 μm) connected to a Carbo Pac PA 10 guard column (50 nm × 4 nm)

arrangement was used. The mobile phase was 90 mM NaOH (blanketed with helium) with a flow rate of 1.5 mL min^{-1}. A gold electrode was used in the electrochemical cell of the detector, and the settings were (time in s/V): 0/0.1; 0.2/0.1; 0.4/0.1; 0.41/− 2.0; 0.42/− 2.0; 0.43/0.6; 0.44/− 0.1; 0.5/− 0.1. Runtime was a total of 48 min, with the first 18 min for sample separation followed by a 15-min washing step with 200 mM NaOH, and a 15-min re-equilibration step with 90 mM NaOH. Peak areas were integrated for galactinol, sucrose, raffinose, and stachyose. Carbohydrates were quantified based on standard curves generated for each carbohydrate. The content of galactinol, sucrose, raffinose, and stachyose was reported as the percent of dry seed weight, which can be converted to g kg^{-1} by multiplying the percent of dry seed result by ten.

Analysis of seed fatty acid compositions

The fatty acid profiles for seeds from each soybean line were determined using the established method of gas chromatography of total fatty acid methyl esters of extracted oil (Beuselinck et al. 2006; Bilyeu et al., 2005). The individual fatty acid components were reported as the relative percentages of palmitic, stearic, oleic, linoleic, and linolenic acids in the extracted oil. For the seeds produced from eight different environments for two years, five whole crushed individual seeds were used as composite samples to determine the fatty acid profiles.

Data analysis

Since each location-year was considered as a single environment, data of each fatty acid and carbohydrate components collected from a total of eight different

environments were analyzed. To assess statistical differences in fatty acid and carbohydrate components depending on the genotype combination, analysis of variance was conducted using PROC GLM in SAS version 9.4 (SAS Institute Inc., Cary, NC, 2013). The mean values of fatty acid and carbohydrate contents per genotype categories presented were generated by calculating average values of each component from the soybean lines within the specific genotype combination. The composite samples of three seeds per plot were utilized to analyze fatty acid profiles using GC, and those of separate fifteen seeds per plot were used to analyze soluble carbohydrate profiles using HPLC.

RESULTS

Molecular breeding strategy to combine the HOLL seed oil trait with the Low or UL RFO meal trait

For developing soybeans with improved oil and meal value traits, soybean lines having variant alleles conferring the high oleic acid and low linolenic acid (HOLL) seed oil phenotype and Low or Ultra-Low RFO (UL RFO) seed meal phenotypes were selected from our experimental germplasm development collection. Each of the variant alleles was originally from a unique donor soybean accession, and four genes were required for the HOLL trait plus two additional genes that controlled the carbohydrate meal trait (Table 3.1). A system of molecular marker-based breeding was used in which each targeted gene was coded by a single letter for tracking the alleles (Table 3.1; Figure 3.1). For the seed oil trait, we selected experimental germplasm lines having four variant alleles (abcd in Figure 3.1) of the previously reported *FAD* genes, which when combined, increased oleic acid over about 80% and reduced linolenic acid to less than 3% of the seed oil (Hagely et al. 2021). For the seed meal traits Low RFO or UL RFO carbohydrate profile, we have previously reported that variant alleles of two *RS* genes are responsible for the reductions in seed RFO content along with significant increases in seed sucrose composition (Hagely, et al., 2020). We targeted the variant alleles with the most dramatic improvement in carbohydrate profiles from our germplasm collection when combined: the single mutant of *rs2W331-* for the Low RFO trait (e); or the double mutants of either *rs2W331-* combined with *rs3snp6* (Dierking and Bilyeu 2008; Schillinger et al. 2013; Schillinger et al. 2018) or *rs2W331-* combined with *rs3G75E* (Hagely et al. 2020) for the UL RFO trait (either ef or ef', respectively, in Figure 3.1). We selected lines with as

much genetic diversity as possible for the study, although many of the lines were closely related or had an experimental line as a parent (Figure 3.1). The soybean germplasm lines and their genotype combinations were selected to target either normal oil or the HOLL oil trait in combination with either normal, Low RFO, or UL RFO meal traits (Table 3.2). The final desired genotype was utilized for the six genes (*FAD2-1A*, *FAD2-1B*, *FAD3A*, *FAD3C*, *RS2*, and *RS3*) so that lines were categorized by genotype into reciprocal categories of either normal or modified oil composition combined with normal or two different modified meal composition categories; this set of soybean germplasm also contained in the *e1-as E2 E3* maturity gene background, which is appropriate for the maturity group III-IV environments (Langewisch et al. 2017).

Soybean seed oil and meal traits can be combined by genotype selection that results in the targeted phenotypes without interference

Soybean germplasm lines categorized into eight groups having normal or variant oil and meal genotype combinations (Table 3.2) of the four fatty acid desaturase genes (coded A, B, C, or D) and the two raffinose synthase genes (coded E and F) were field tested in eight environments and analyzed for seed oil fatty acid components and seed meal carbohydrate profiles. All of the soybean lines with the four mutant alleles (abcd) of *FAD* genes had over 83% oleic acid and 2.5% or less linolenic acid in the seed oil, regardless of the allele combinations for the raffinose synthase genes (Table 3.3). There were no significant differences in seed oleic acid content among the genotype categories having the mutant alleles of the *FAD* genes (abcd). For linolenic acid content, there were large significant differences between the normal oil genotype categories and the mutant

oil genotype categories, and small significant differences within the normal or mutant oil categories. One soybean line with the incomplete mutant *FAD* oil genotype (abcD ef) was included in this experiment to capture additional germplasm with the double mutant raffinose synthase genotype, and a small significant difference was observed for increased linolenic acid content for the genotype category that included this line (abcDef) compared to the categories with the complete mutant *FAD* oil genotype (abcd). Differences in the other fatty acids in the oil (palmitic, stearic, and linoleic acid) followed the expected results based on the oil genotype such that there were large significant differences between the normal oil genotype categories (ABCD) and the mutant oil genotype categories (abcd), and small significant differences within the normal or mutant oil categories. The status of the raffinose synthase genotype did not interfere with the fatty acid profile in the seed oil of the soybean germplasm with mutant *FAD* genotypes, and those lines all produced the HOLL oil profile.

The seed carbohydrate profiles were also examined for the same set of field-produced seeds of the soybean germplasm lines used in this study. Galactinol, sucrose, raffinose, and stachyose contents are reported as percentages of the seed dry weight (Table 3.4). Similar to the analysis of seed oil profiles, the seed carbohydrate components were quantified and compared to investigate if there was interference from the oil trait. Consistent with prior research for carbohydrate profile meal value traits (Hagely et al. 2020; Jo et al. 2018; Jo et al. 2019), soybean germplasm with normal oil genotypes and the three mutant raffinose synthase categories (eF, ef, and ef') produced significantly increased sucrose and significantly reduced RFO content compared to the normal soybean genotype lines (Table 3.4). Comparing the two double mutant raffinose synthase

genotype categories with normal oil (ABCD ef and ABCD ef') there were no significant differences in the very low contents of seed raffinose or stachyose, but there were small significant differences in seed sucrose and galactinol content, which is different from earlier results, although soybean seed sucrose content has been found to be variable (Hagely et al. 2020; Jo et al. 2018; Jo et al. 2019).

The direct comparison of soybean germplasm with the same raffinose synthase genotype but contrasting oil genotypes (ABCD or abcd) revealed there was no significant difference in raffinose between any of the pairwise comparisons, except for a minor significant difference (0.1%) for the normal raffinose synthase categories (ABCD EF and abcd EF) (Table 3.4). For stachyose content, there were no significant differences between the lowest stachyose double mutant line comparisons (ABCD ef and abcd ef; ABCD ef' and abcd ef'), but there were small significant differences between the pairwise comparisons with either normal raffinose synthase categories or one mutant raffinose synthase (ABCD EF and abcd EF; ABCD eF and abdc eF). For total RFO content, there were no significant differences between the pairwise comparisons for lines with either of the double mutant raffinose synthase categories (ef or ef'); there was significantly increased RFO content for HOLL oil lines with one mutant raffinose synthase gene compared to the normal oil lines (abcd eF versus ABCD eF). The situation for RFO content was reversed for the comparison between normal raffinose synthase genotypes (ABCD EF was higher than abcd EF for RFO content). For sucrose comparisons between contrasting oil genoytpes, there were no significant differences for the normal (EF) or two mutant raffinose synthase genotype categories (ef or ef'), but normal oil single raffinose synthase genotype lines (ABDC eF) contained significantly

higher seed sucrose than the HOLL oil genotype lines (abcd eF). Seed galactinol content was not significantly different between contrasting oil genotypes with normal raffinose synthase genotypes or those with one mutant raffinose synthase gene (EF or eF), but the HOLL lines had small but significant increases in galactinol content for the lines with double mutant raffinose synthase genotypes compared to their normal oil genotype accessions (ef or ef'). The status of the oil genotype did not interfere with the soybean seed carbohydrate profile in the soybean germplasm with the most dramatic phenotype controlled by the double mutant raffinose synthase genotypes, and those lines all produced the UL RFO meal profile; for soybean germplasm lines with only one raffinose synthase mutant gene, there were small but significant differences in the expression of the Low RFO meal profile based on the oil genotype.

A Better soybean with enhanced quality seed oil and meal composition phenotypes

Soybean varieties with the HOLL seed oil trait have enhanced value from more oxidatively stable oil compared to normal soybean varieties. Likewise, soybean varieties with increased seed sucrose and decreased RFO have enhanced value from higher metabolizable energy in the meal. We developed soybean germplasm lines with contrasting seed oil traits (normal or HOLL) combined with seed carbohydrate meal traits (normal, Low RFO, and UL RFO carbohydrate profile). Since the extreme oil and meal phenotypes were expressed in seeds without interference, this compositional bundle of HOLL with UL RFO represents a better soybean with a unique combination of oil and meal composition characteristics (Table 3.5). Compared with normal soybean, seeds from the "Better bean" soybean type (the abcd ef and abcd ef' soybean germplasm lines) meet

the target for improved oil (>75% oleic acid and < 3% linolenic acid) and have an improved carbohydrate profile with increased sucrose and almost no RFO (Table 3.5).

DISCUSSION

The results of this study demonstrated that soybean with both enhanced oil and meal traits can be successfully produced with molecular marker assisted selection for the combination of the four fatty acid desaturase alleles responsible for the HOLL seed oil trait plus either one or two alleles responsible for the Low or UL RFO seed meal traits. No substantial interference was observed between the desired oil and meal phenotypes. The modified seed composition traits were combined together to enhance the overall functionality of soybean seeds and thus are expected to increase value generated from the soybean crop. The molecular breeding strategy presented, and the "Better bean", non-GMO soybean germplasm lines developed in this study, can contribute to the creation of healthier, more oxidatively stable soybean oil and higher metabolizable energy soybean meal.

From the direct comparison of seed fatty acid profiles between the normal versus the mutant oil genotype categories, all the soybean lines with the four mutant alleles (abcd) of *FAD* genes successfully increased oleic acid content over 83% and reduced linolenic acid to 2.5% or less in the seed oil, regardless of the allele combinations for the raffinose synthase genes. However, there were small significant differences within the normal or mutant oil categories for all fatty acids except the oleic acid content. For linolenic acid in the oil, one of the target fatty acids desired to be modified, the significant differences observed were mainly due to the inclusion of a germplasm line with the incomplete mutant *FAD* oil genotype (abcD ef). Therefore, in the situation of a complete combination of the mutant alleles of the four *FAD* genes, significant differences in linolenic acid are not expected to be detected.

For the carbohydrate profiles, in presence of the single mutant allele of the *RS2* gene (either ABCD eF or abcd eF), sucrose content was significantly increased over 7% and RFO contents were significantly decreased to less than 2 % compared to the lines having wildtype allele of *RS* genes (either ABCD EF or abcd EF); even much greater increases in the levels of seed sucrose (over 7.4%) as well as a decrease in RFO content (to less than 0.5%) were observed when the soybean lines had either of the double mutant raffinose synthase categories (ef or ef'). These significant increases in sucrose content and reduction in RFO contents in presence of either of a single or a double mutation for the *RS* genes were consistent with the phenotypic observation from an array of our previous studies (Dierking and Bilyeu 2008, 2009; Jo et al. 2018; Jo et al. 2019; Hagely et al. 2020). Our previous study showed that the impact of the *rs3snp6* and r*s3G75E* in the reduction of seed RFO contents was found to be similar, and there were no significant differences in galactinol and sucrose contents generated from soybeans having each of the mutant alleles Hagely et al. (2020). The current research reconfirmed earlier findings, the similar impact of the *rs3G75E* allele (f') compared to the *rs3snp6* (f) in significantly increasing sucrose contents and reducing RFO contents, using the soybean lines having different genetic background grown difference production environments: but inconsistent results were observed in galactinol and sucrose contents measured from the current study which showed small significant differences between the two double mutant raffinose synthase genotype categories with normal oil (ABCD ef and ABCD ef').

In addition to the effects of modifier genes, environmental conditions, especially temperature during the pod-filling phase, are also known to influence the fatty acids and carbohydrate profiles in soybean seeds. It has been observed that cooler temperature

during pod fill correlated with the increase in sucrose and decrease in RFO and oleic acid accumulation in soybean seeds (Bilyeu and Wiebold, 2016; Kumar et al., 2010; Lee et al. 2009). Results from our earlier study about fatty acid composition in two different latitudinal environments also showed instability of linolenic acid content across environments (Pham et al 2012). Our earlier research examined the carbohydrate profiles of soybean lines from different genotype classes with different targeted maturity groups (MGs) produced the targeted latitudinal environment (Jo et al. 2019). The results showed a highly significant increase in seed sucrose content, especially from the soybeans produced in earlier maturity group environments having comparatively cooler temperatures. Since the soybean germplasm lines developed in this study were only targeted to the maturity group III-IV environments and the current study primarily aimed to evaluate the feasibility of the development of soybeans having both of the oil and meal value traits across the environment, the potential differences in the seed composition traits depending on the different production environments were not specifically analyzed. For its broader production across the US as well as for the appropriate pricing for these new value-added soybeans, the environmental stability of the seed composition phenotypes and the yield potential of the “Better bean” need to be further confirmed in additional experiments with more diverse production environments across different years.

TABLES

Table 3.1. Description of alleles of genes used to modify fatty acid and carbohydrate components of soybean seed

Wm82.a2.v1	Gene	Allele	Type	Genotype Code	Source	Reference
Glyma.10g278000	*FAD2-1A*	indel	Null-frameshift	**a**	PI 603452	Pham et al. (2011)
Glyma.20g111000	*FAD2-1B*	P137R	Missense	**b**	PI 283327	Pham et al. (2010)
Glyma.14g194300	*FAD3A*	Splice site (G810A)	Null-splice	**c**	CX1512-44	Bilyeu et al. (2005)
Glyma.18g062000	*FAD3C*	G128E	Missense	**d**	CX1512-44	Bilyeu et al. (2005);
Glyma.06g179200	*RS2*	*rs2W331-*	In-frame deletion	**e**	PI 200508	Kerr and Sebastian (2000); Hitz et al. (2002) Dierking and Bilyeu (2008)
Glyma.05g003900	*RS3*	*rs3snp6*	SNP in non-coding region	**f**	Patriot/SGUL	Jo et al. (2018)
Glyma.05g003900	*RS3*	*rs3G75E*	Missense	**f'**	W82 rs3	Thapa et al. (2019); Hagely et al. (2020)

Table 3.2. Nomenclature and allele combination of experimental and control soybean lines used to evaluate interference among the variant alleles responsible for the seed modified fatty acids and carbohydrate contents

Name	Oil genotype	Meal Genotype	Predicted phenotype	
			Oil trait	Meal trait
Williams 82	ABCD	EF	Normal	Normal
LG04-6000	ABCD	EF	Normal	Normal
Jack	ABCD	EF	Normal	Normal
Jack rs2 RS3	ABCD	eF	Normal	Low RFO
KB12-31 A1	ABCD	eF	Normal	Low RFO
KH145	ABCD	eF	Normal	Low RFO
Jack rs2 rs3	ABCD	ef	Normal	UL RFO
KB12-31 B1	ABCD	ef	Normal	UL RFO
KB17-41 A1	ABCD	ef	Normal	UL RFO
KB17-41 B1	ABCD	ef'	Normal	UL RFO
KB17-41 B2	ABCD	ef'	Normal	UL RFO
KH144	ABCD	ef'	Normal	UL RFO
KB13-15 A1	abcd	EF	HOLL	Normal
KB15-6 A1	abcd	EF	HOLL	Normal
KB15-16 A1	abcd	EF	HOLL	Normal
KB15-5 A1	abcd	eF	HOLL	Low RFO
KB16-24 A1[†]	abcd	eF	HOLL	Low RFO
KB17-35 A1	abcd	eF	HOLL	Low RFO
KB17-38 A1	abcd	eF	HOLL	Low RFO
KB14-22 A1	abcD	ef	HOL[i]	UL RFO
KB16-24 B1	abcd	ef	HOLL	UL RFO
KB17-40 A1	abcd	ef'	HOLL	UL RFO
KB18-35 A1	abcd	ef'	HOLL	UL RFO
KB18-35 A2[‡]	abcd	ef'	HOLL	UL RFO

[†] KB16-24 A1 was only grown in 2019 field season
[‡] KB18-35 A2 was only grown in 2020 field season

Table 3.3. Genotype and seed fatty acid profiles for control and experimental soybean lines with each of allele combinations for modified fatty acids and carbohydrate profiles from experiments in 2019 and 2020

Genotype		n[§]	Fatty acid; percent of total fatty acid content					Resulting phenotype for Oil trait
Oil	Meal		Palmitic (16:0)	Stearic (18:0)	Oleic (18:1)	Linoleic (18:2)	Linolenic (18:3)	
ABCD	EF	3	10.5 c[†]	4.0 c	24.0 c	54.5 ab	7.0 b	Normal
ABCD	eF	3	10.0 d	3.7 d	24.6 bc	54.3 b	7.5 a	Normal
ABCD	ef	3	11.2 a	4.4 a	25.5 b	52.1 c	6.9 b	Normal
ABCD	ef'	3	10.8 b	4.1 b	22.2 d	55.3 a	7.5 a	Normal
abcd	EF	3	6.8 g	3.3 e	84.1 a	3.6 d	2.2 d	HOLL
abcd	eF	4	7.0 f	3.1 f	83.9 a	3.8 d	2.3 d	HOLL
abcd[‡]	ef	3	7.0 f	3.4 e	83.8 a	3.3 d	2.5 c	HOLL
abcd	ef'	3	7.3 e	3.2 f	83.4 a	3.8 d	2.3 cd	HOLL

[†] Mean value was obtained by averaging means of three replications which were averaged from fatty acid values of three composited seeds per plot; means followed by the same letter were not significantly different from each other at 95% confidence with Fisher's least significant difference (LSD) procedure

[‡] A soybean experimental line having a functional FAD3C with the three variant alleles for the other target FAD genes (abcD), KB14-22 A1, was included in this genotype category

[§] Number of soybean lines within each genotype category used for the field experiment in 2019 and 2020

Table 3.4. Genotype and seed carbohydrate profiles for control and experimental soybean lines with each of allele combinations for modified fatty acids and carbohydrate profiles from experiments in 2019 and 2020

Genotype		n[§]	Carbohydrate; percent of total carbohydrate content					Resulting phenotype for Meal trait
Oil	Meal		Galactinol	Sucrose	Raffinose	Stachyose	RFO	
ABCD	EF	3	0.1 e†	5.6 e	0.7 a	4.4 a	5.2 a	Normal
abcd	EF	3	0.1 e	5.4 e	0.7 b	4.2 b	4.9 b	Normal
ABCD	eF	3	0.5 d	7.2 c	0.1 c	1.3 d	1.4 d	Low RFO
abcd	eF	4	0.5 d	7 d	0.1 c	1.6 c	1.7 c	Low RFO
ABCD	ef	3	0.5 c	7.8 a	0.0 d	0.3 ef	0.4 e	UL RFO
abcd‡	ef	3	0.6 b	7.8 a	0.0 d	0.3 e	0.4 e	UL RFO
ABCD	ef'	3	0.6 b	7.5 b	0.0 d	0.2 f	0.3 e	UL RFO
abcd	ef'	3	0.8 a	7.4 bc	0.0 d	0.3 ef	0.3 e	UL RFO

† Mean value was obtained by averaging means of three replications which were averaged from fatty acid values of three composited seeds per plot; means followed by the same letter were not significantly different from each other at 95% confidence with Fisher's least significant difference (LSD) procedure

‡ A soybean experimental line having a functional FAD3C with the three variant alleles for the other target FAD genes (abcD), KB14-22 A1, was included in this genotype category

§ Number of soybean lines within each genotype category used for the field experiment in 2019 and 2020

Table 3.5. Summary comparison of seed composition components between the normal soybean type and a Better bean type of soybean germplasm with targeted variant alleles for the HOLL oil trait and the UL RFO meal trait

Type	Genotype	n[†]	Resulting phenotype		Fatty acid (% of oil)		Carbohydrate (% seed)	
			Oil	Meal	Oleic	Linolenic	Sucrose	RFO
Soybean	ABCDEF	61	Normal	Normal	24.0	7.0	5.6	5.2
Better bean	abcdef/f'	74	HOLL	UL RFO	83.5	2.3	7.4	0.3

[†] Number of soybean samples produced from the eight different environments across 2 years (2019 and 2020)

FIGURES

A

FAD2-1A Glyma.10g278000	
A	a
	A50014633-
FAD2-1A	Null
WT	indel

FAD2-1b Glyma.20g111000	
B	b
	C35318088G
FAD2-1B	Missense
WT	P137R

FAD3A Glyma.14g194300	
C	c
	G45937922A
FAD3A	Null
WT	splice

FAD3C Glyma.18g062000	
D	d
	G5647138A
FAD3C	Missense
WT	G128E

RS2 Glyma.06g179200	
E	e
	CCA15222105---
RS2	in-frame deletion
WT	indel

RS3 Glyma.05g003900		
F	f	f'
	C308412T	G307744A
RS3	non-coding	missense
WT	SNP	G75E

B

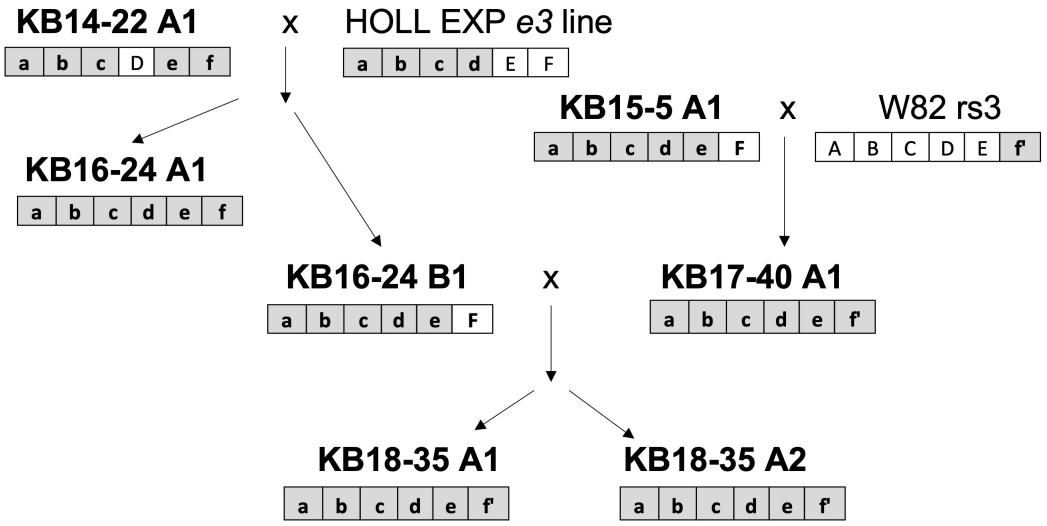

Figure 3.1. Molecular marker-based breeding scheme to combine the HOLL seed oil trait plus Low/UL RFO traits. **(A)** Gene names, identification in Wm82.a2.v1, and allele information for six targeted genes. For simplicity, each targeted gene is coded by an uppercase (functional allele) or lowercase letter (alternate allele). The details of the alleles are provided, and bold plus gray shading represents the variant alleles targeted in the current study. **(B)** Breeding scheme with experimental line names and allele code for the six targeted genes.

REFERENCES

Anai, T., Yamada, T., Hideshima, R., Kinoshita, T., Rahman, S.M. and Takagi, Y., Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by *GmFAD2-1a*. Breeding Science, 2008. 58(4): p.447-452.

Beuselinck, P.R., Sleper, D.A. and Bilyeu, K.D., An assessment of phenotype selection for linolenic acid using genetic markers. Crop science, 2006. 46(2): p.747-750.

Bilyeu, K.D., Palavalli, L.A.V.A.N.Y.A., Sleper, D.A. and Beuselinck, P.R., Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. Crop Science, 2003. 43(5): p.1833-1838.

Bilyeu, K.D., L. Palavalli, D.A. Sleper, and P.R. Beuselinck., Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Science, 2005. 45(5): p.1830–1836.

Bilyeu, K., Palavalli, L., Sleper, D.A., and Beuselinck, P., Molecular genetic resources for development of 1% linolenic acid soybeans. Crop Science, 2006. 46(5): p. 1913-1918.

Bilyeu, K., Gillman, J.D. and LeRoy, A.R., Novel *FAD3* mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Science, 2011. 51(1): p.259-264.

Bilyeu, K., Shannon, J.G., Lee, J.D., and Pham, A.T., U.S. Patent Number 9,035,129. Method to develop high oleic acid soybeans using conventional soybean breeding techniques. May 2015a.

Bilyeu, K., Shannon, J.G., Lee, J.-D., and Pham, A.T., U.S. Patent Number 9,198,365. Method to develop high oleic acid soybeans using conventional soybean breeding techniques. December 2015b.

Bilyeu, K.D. and Wiebold, W.J., Environmental stability of seed carbohydrate profiles in soybeans containing different alleles of the raffinose synthase 2 (*RS2*) gene. Journal of agricultural and food chemistry, 2016. 64(5): p.1071-1078.

Bilyeu, K., Škrabišová, M., Allen, D., Rajcan, I., Palmquist, D.E., Gillen, A., Mian, R. and Jo, H., The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. Journal of the American Oil Chemists' Society, 2018a. 95(1): p.39-49.

Bilyeu, K.D., Shannon, J.G., Lee, J.D., and Pham, A.T., U.S. Patent Number 10,087,454. Method to develop high oleic acid soybeans using conventional soybean breeding techniques. October 2018b.

Bilyeu, K., Shannon, G., Lee, J.D., and Pham, A.T., U.S. Patent Number 10,329,576. Method to develop high oleic acid soybeans using conventional soybean breeding techniques. June 2019.

Bilyeu, K., Shannon, G., Lee, J.D., and Pham, A.T., U.S. Patent Number 10,774,337. Method to develop high oleic acid soybeans using conventional soybean breeding techniques. September 2020.

Combs, R. and Bilyeu, K., Novel alleles of *fad2-1a* induce high levels of oleic acid in soybean oil. Molecular Breeding, 2019. 39(6): p. 79.

Coon, C.N., Leske, K.L., Akavanichan, O. and Cheng, T.K., Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. Poultry Science, 1990. 69(5): p.787-793.

Dierking, E.C. and Bilyeu, K.D., Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome, 2008. 1(2): p. 135-145.

Dierking, E.C. and Bilyeu, K.D., New sources of soybean seed meal and oil composition traits identified through TILLING. BMC plant biology, 2009. 9(1): p.1-11.

FDA (2003) Food labeling: trans fatty acids in nutrition labeling, nutrient content claims, and health claims. Federal Register Rule Number 68 FR 41433 Session Number 21 CFR Part 101, 68:41433–41506

FDA (2015) Final determination regarding partially hydrogenated oils. Federal Register Rule Number 80 FR 34650, 80:34650– 34670

Hagely, K.B., Palmquist, D. and Bilyeu, K.D., Classification of distinct seed carbohydrate profiles in soybean. Journal of agricultural and food chemistry, 2013. 61(5): p.1105-1111.

Hagely, K.B., Jo, H., Kim, J.-H., Hudson, K.A., and Bilyeu, K., Molecular-assisted breeding for improved carbohydrate profiles in soybean seed. Theoretical and Applied Genetics, 2020. 133(4): p. 1189-1200.

Hagely, K., Konda, A.R., Kim, J.-H., Cahoon, E.B., and Bilyeu, K., Molecular-assisted breeding for soybean with high oleic/low linolenic acid and elevated vitamin e in the seed oil. Molecular Breeding, 2021. 41(1): p. 3.

Heppard, E.P., Kinney, A.J., Stecca, K.L. and Miao, G.H., Developmental and growth temperature regulation of two different microsomal [omega]-6 desaturase genes in soybeans. Plant Physiology, 1996.110(1): p.311-319.

Hsu, S.H., Hadley, H.H. and Hymowitz, T., Changes in carbohydrate contents of germinating soybean seeds. Crop science, 1973. 13(4): pp.407-410.

Hu, F.B., Stampfer, M.J., Manson, J.E., Rimm, E., Colditz, G.A., Rosner, B.A., Hennekens, C.H. and Willett, W.C., Dietary fat intake and the risk of coronary heart disease in women. New England journal of medicine, 1997. 337(21): p.1491-1499.

Jo, H., Lee, J.-D., and Bilyeu, K.D., Environmental stability of carbohydrate profiles in different soybean genotypes. Crop Science, 2018. 58(2): p. 773-782.

Jo, H., Lorenz, A.J., Rainey, K.M., Shannon, J.G., Chen, P., and Bilyeu, K.D., Environmental stability study of soybeans with modified carbohydrate profiles in maturity groups 0 to v. Crop Science, 2019. 59(4): p. 1531-1543.

Kerr, P.S., and Sebastian S.A., US Patent 6,147,193. Soybean products with improved carbohydrate composition and soybean plants. November 2000.

Kumar, V., Rani, A., Goyal, L., Dixit, A.K., Manjaya, J.G., Dev, J. and Swamy, M., Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. Journal of agricultural and food chemistry, 2010. 58(8): p.5081-5085

Langewisch, T., Lenis, J., Jiang, G.-L., Wang, D., Pantalone, V., and Bilyeu, K., The development and use of a molecular model for soybean maturity groups. BMC Plant Biology, 2017. 17(1): p. 91.

Lee, J.D., Woolard, M., Sleper, D.A., Smith, J.R., Pantalone, V.R., Nyinyi, C.N., Cardinal, A. and Shannon, J.G., Environmental effects on oleic acid in soybean seed oil of plant introductions with elevated oleic concentration. Crop science, 2009. 49(5): p.1762-1768.

Openshaw, S.J., and Hadley, H.H., Maternal effects on sugar content in soybean seeds. Crop Science, 1978. 18(4): p.581-584.

Pham, A.T., Lee, J.D., Shannon, J.G., and Bilyeu, K., Mutant alleles of *fad2-1a* and *fad2-1b* combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology, 2010. 10(1): p. 195.

Pham, A.T., Lee, J.D., Shannon, J., and Bilyeu, K., A novel *fad2-1a* allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. TAG Theoretical and Applied Genetics, 2011. 123(5): p. 793-802.

Pham, A.T., Shannon, J.G., and Bilyeu, K., Combinations of mutant *FAD2* and *FAD3* genes to produce high oleic acid and low linolenic acid soybean oil. Theoretical and Applied Genetics, 2012. 125(3), 503-515.

SoyStats (2021) http://www.soystats.com.

Schillinger, J.A., Dierking, E.C., and Bilyeu, K.D., U.S. Patent Number 8,471,107. Soybeans having high germination rates and ultra-low raffinose and stachyose content. June 2013.

Schillinger, J.A., Dierking, E.C. and Bilyeu, K.D., U.S. Patent 10,081,814. Soybeans having high germination rates and ultra-low raffinose and stachyose content. September 2018.

Schlueter, J.A., Vasylenko-Sanders, I.F., Deshpande, S., Yi, J., Siegfried, M., Roe, B.A., Schlueter, S.D., Scheffler, B.E. and Shoemaker, R.C., The *FAD2* gene family of soybean: insights into the structural and functional divergence of a paleopolyploid genome. Crop Science, 2007. 47: p.S-14.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.-C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A., Genome sequence of the palaeopolyploid soybean. Nature, 2010. 463(7278): p.178-183.

Haun, W., Coffman, A., Clasen, B.M., Demorest, Z.L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A., Cedrone, F. and Mathis, L., Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant biotechnology journal, 2014. 12(7): p.934-940.

Hoshino, T., Takagi, Y. and 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. Breeding science, 2010. 60(4): p.419-425.

Sweeney, D.W., Carrero-Colón, M. and Hudson, K.A., Characterization of new allelic combinations for high-oleic soybeans. Crop Science, 2017. 57(2): p.611-616.

Thapa, R., Carrero-Colón, M., Rainey, K.M. and Hudson, K., Tilling by sequencing: a successful approach to identify rare alleles in soybean populations. Genes, 2019. 10(12): p.1003.

APPENDIX

Supplementary Tables for Chapter 2

Supplementary Table 1. Summary of soybean accessions used in the study to evaluate rates of concordance between the observed stem termination type phenotypes

Accession	Name	Taxonomy[§]	Improvement status[§]	Classification[§]	Origin[§]	*Dt1*[†]	*Dt2*[†]	*E1*[‡]	*E2*[‡]	*E3*[‡]	Observed stem termination type[§]	MG[§]	Source of resequenced data
PI 166105	Bhart	*G. max*	Landrace	Other	Uttar Pradesh, India	*Dt1*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	VII	USB481
PI 458505	Da Bai mei	*G. max*	Landrace	Other	Liaoning Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	II	USB481
PI 632650	DT 22	*G. max*	Landrace	Other	Vietnam	*Dt1*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	V	USB481
PI 567489 A	Er da li huang dou	*G. max*	Landrace	Other	Hebei Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	IV	USB481
PI 592523	Glacier	*G. max*	Elite	Other	Minnesota, United States	*Dt1*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	00	USB481
PI 407965	KAERI 504-4	*G. max*	Elite	Other	Jeollanam-do, Korea, South	*Dt1*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	V	Valliyodan et al. (2006)
PI 437265 D	(Dobruzanca D)	*G. max*	Landrace	Other	Moldova	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	0	USB481
PI 361066 B	(F. 56-17)	*G. max*	Landrace	Other	Romania	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	I	USB481
PI 468408 B	(Qi Huang No. 1)	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	USB481
PI 602502 B	(Xiong yue xiao huang dou)	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 438347	35S.277	*G. max*	Landrace	Other	Australia	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VII	USB481
PI 153231	B-63	*G. max*	Landrace	Other	Unknown	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	USB481
PI 567418 A	Bai hei dou	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	II	USB481
PI 567428	Bai ji yao	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 179935	Bhart	*G. max*	Landrace	Other	Himachal Pradesh, India	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VII	USB481
PI 361070	Faur	*G. max*	Landrace	Other	Romania	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	0	USB481
PI 603162	GL 2631 /96	*G. max*	Landrace	Other	Korea, North	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 417581	H-060072	*G. max*	Landrace	Other	United States	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	V	USB481
PI 603494	Hai dou zi	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 567173	Hei he 51	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	00	USB481
PI 603526	Hei you dou	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 567439	Hong jia huang dou	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	V	USB481
PI 603555	Hua da hei dou	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 567548	Hua li hu zi	*G. max*	Landrace	Other	Shandong Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 603389	Huang ke	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	II	USB481
PI 424195 A	ISZ-3	*G. max*	Landrace	Other	Hungary	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	0	USB481
PI 587804	Jing 789	*G. max*	Landrace	Other	Hubei Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 398965	KLS 628-1	*G. max*	Landrace	Other	Jeollanam-do, Korea, South	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 437160	Krasnodarscaja 13	*G. max*	Landrace	Other	Krasnodar, Russian Federation	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	I	USB481
PI 567361	Lu fang huang dou	*G. max*	Landrace	Other	Ningxia Huizi Zizhiqu, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	USB481
PI 567343	Ma huang dou	*G. max*	Landrace	Other	Gansu Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	V	USB481

PI 548364	Macoupin	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 603549	Mei dou	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	USB481
PI 123440	No. 2	*G. max*	Landrace	Other	Myanmar	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VI	USB481
PI 548479	Otootan	*G. max*	Landrace	Other	Taiwan	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VIII	USB481
PI 567746	Pei xian da bai jiao	*G. max*	Landrace	Other	Jiangsu Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 417242	Pekin dai seitou	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	II	USB481
PI 495017 C	(Beijing da qing don)	*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 404198 B	(Sun huan do)	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 567519	Bai hua chi	*G. max*	Elite	Other	Shandong Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 548317	Columbia	*G. max*	Landrace	Other	Hebei Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 567357	Du jia qiao huang dou	*G. max*	Landrace	Other	Ningxia Huizi Zizhiqu, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 567719	Fu yang (43)	*G. max*	Elite	Other	Anhui Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 567305	Hei dou zi	*G. max*	Landrace	Other	Gansu Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 567387	Huang huai dou	*G. max*	Landrace	Other	Shaanxi Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 561271	Pei xian da quing dou	*G. max*	Elite	Other	Zhejiang Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	V	Valliyodan et al. (2006)
PI 84987A	(Oni Hadaka)	*G. max*	Landrace	Other		*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	Zhou et al. (2015)
PI 567293	Ben di huang dou	*G. max*	Landrace	Other	Gansu Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 567395	Lai wa dou	*G. max*	Landrace	Other	Shaanxi Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 548593	Maple Arrow	*G. max*	Elite	NA Cultivar	Ontario, Canada	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	00	Zhou et al. (2015)
PI 548643	Maple Glen	*G. max*	Elite	NA Cultivar	Ontario, Canada	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	00	Zhou et al. (2015)
PI 548391	Mukden	*G. max*	Landrace	NA Ancestor	Liaoning Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 567364	Ping luo huang da dou	*G. max*	Landrace	Other	Ningxia Huizi Zizhiqu, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	Zhou et al. (2015)
PI 438496 C	(Peking)	*G. max*	Landrace	Other	United States	*Dt1*	*dt2*	*E1*	*E2*	*e3*	Indeterminate	IV	USB481
PI 567726	Fu yang (50)	*G. max*	Landrace	Other	Anhui Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*e3*	Indeterminate	IV	USB481
PI 438323	Grignon 53-F-3	*G. max*	Landrace	Other	France	*Dt1*	*dt2*	*E1*	*E2*	*e3*	Indeterminate	I	USB481
PI 209333	No. 3	*G. max*	Landrace	Other	Hokkaidô, Japan	*Dt1*	*dt2*	*E1*	*E2*	*e3*	Indeterminate	VI	USB481
PI 602993	Pi xian ruan tiao zhi	*G. max*	Landrace	Other	Jiangsu Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*e3*	Indeterminate	IV	USB481

PI 157421	Ebony	*G. max*	Landrace	Other	Korea, South	*Dt1*	*dt2*	*E1*	*E2*	*e3*	Indeterminate	III	Zhou et al. (2015)
PI 424038 B	74053	*G. max*	Landrace	Other	Kyonggi, Korea, South	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	V	USB481
PI 567435 B	(Hei hei dou)	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	USB481
PI 603495 B	(Hong mi lan dou zi)	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	V	USB481
PI 438496 B	(Peking)	*G. max*	Landrace	Other	United States	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	USB481
PI 567410 B	(Yang huang dou)	*G. max*	Landrace	Other	Shaanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	VII	USB481
PI 567415 A	Bai da huang dou	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 567416	Bai dou	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 567426	Bai huang dou	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 391577	Cha ye sheng tou	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	II	USB481
PI 548316	Cloud	*G. max*	Landrace	Other	Zhejiang Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	USB481
PI 548452	Dixie	*G. max*	Landrace	Other	Phyeongyang, Korea, North	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	V	USB481
PI 417500	Escura A	*G. max*	Landrace	Other	Brazil	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	VIII	USB481
PI 561371	Fen dou 15	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 574477	Fen dou 31	*G. max*	Landrace	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 437662	Gun-tszu-lin 658	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	II	USB481
PI 578495	Jin dou No. 4	*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 548359	Kingwa	*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 507017	Madara ooha tsuru mame	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	VII	USB481
PI 209332	No. 4	*G. max*	Landrace	Other	Hokkaidô, Japan	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 567782	OAC Dorado	*G. max*	Elite	Other	Ontario, Canada	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	I	USB481
PI 603492	Qi hei dou	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 324924	Rhosa	*G. max*	Landrace	Other	South Africa	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	V	USB481
PI 567516 C	(Ba yue zha)	*G. max*	Landrace	Other	Shandong Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 567336 B	(Lao hei dou)	*G. max*	Landrace	Other	Gansu Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 475783 B	(Tsing 2)	*G. max*	Elite	Other	Shanxi Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 612611	Browngilgun	*G. max*	Landrace	Other	Korea, North	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 467312	Cha-mo-shi-dou	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	II	Valliyodan et al. (2006)
PI 437655	Er-huan-jan	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 548349	Ilsoy	*G. max*	Landrace	Other	Phyeongyang, Korea, North	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 548298	A.K. (Harrow)	*G. max*	Landrace	NA Ancestor	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Zhou et al. (2015)

FC 33243	Anderson	*G. max*	Elite	NA Ancestor	Unknown	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 437654	Er-hej-jan	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Zhou et al. (2015)
PI 548348	Illini	*G. max*	Landrace	NA Ancestor	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Zhou et al. (2015)
PI 578457 A	May den	*G. max*	Landrace	Other	Vietnam	*Dt1*	*dt2*	*E1*	*N*	*E3*	Indeterminate	VIII	Zhou et al. (2015)
PI 424078	74077	*G. max*	Landrace	Other	Gangwon-do, Korea, South	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	III	USB481
PI 372403 B	(Caloria)	*G. max*	Landrace	Other	Austria	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	00	USB481
PI 578309	Bhatmash	*G. max*	Landrace	Other	Nepal	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	VI	USB481
PI 592960	Dong nong 38	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	I	USB481
PI 514671	Feng shou No. 7	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	0	USB481
PI 548325	Flambeau	*G. max*	Elite	Other	Russian Federation	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	00	USB481
PI 548336	Habaro	*G. max*	Landrace	Other	Habarovskij kraj, Russian Federation	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	I	USB481
PI 548571	Harlon	*G. max*	Landrace	Other	Ontario, Canada	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	I	USB481
PI 548582	McCall	*G. max*	Elite	Other	Minnesota, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	00	USB481
PI 424298	KAS 300-10	*G. max*	Landrace	Other	Chungcheongnam-do, Korea, South	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 424608 A	KAS 681-21	*G. max*	Landrace	Other	Gyeongsangbuk-do, Korea, South	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 508083	Dassel	*G. max*	Elite	NA Cultivar	Minnesota, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	0	Zhou et al. (2015)
PI 513382	Glenwood	*G. max*	Elite	NA Cultivar	Minnesota, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	0	Zhou et al. (2015)
PI 548379	Mandarin (Ottawa)	*G. max*	Landrace	NA Ancestor	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Indeterminate	0	Zhou et al. (2015)
PI 416751	A-B(D)	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	USB481
PI 417529	A38	*G. max*	Landrace	Other	Germany	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	0	USB481
PI 548521	BSR 201	*G. max*	Elite	Other	Iowa, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	USB481
PI 297505	Czi ti No. 5	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	USB481
PI 578412	Gong jiao 6308-1	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	USB481
PI 567171	Hei he No. 1	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	00	USB481
PI 407701	Hei long No. 3	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	USB481
PI 548561	Hodgson	*G. max*	Elite	Other	Minnesota, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	USB481

PI 398633	KAS 390-17-2	*G. max*	Landrace	Other	Chungcheongbuk-do, Korea, South	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	V	USB481
PI 548360	Korean	*G. max*	Landrace	Other	Korea, North	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	USB481
PI 567558	Liu shi ri jin huang da dou	*G. max*	Landrace	Other	Shandong Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	III	USB481
PI 497967	PLSO 96	*G. max*	Landrace	Other	Jammu and Kashmir, India	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	VII	USB481
PI 437169 B	(VNIISC-4)	*G. max*	Elite	Other	Krasnodar, Russian Federation	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Valliyodan et al. (2006)
PI 548511	Beeson 80	*G. max*	Elite	NA Cultivar	Indiana, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Valliyodan et al. (2006)
PI 518751	NS-20	*G. max*	Elite	Other	Former Serbia and Montenegro	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Valliyodan et al. (2006)
PI 533655	Burlison	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 548512	Century	*G. max*	Elite	NA Cultivar	Indiana, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 542403	Dawson	*G. max*	Elite	NA Cultivar	Minnesota, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	0	Zhou et al. (2015)
PI 548573	Harosoy	*G. max*	Elite	NA Cultivar	Ontario, Canada	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 547680	L62-17	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 547686	L62-956	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 547690	L63-1212	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	II	Zhou et al. (2015)
PI 253658 B	No. 9	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	Zhou et al. (2015)
PI 518750	NS-16	*G. max*	Elite	Other	Former Serbia and Montenegro	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	Zhou et al. (2015)
PI 548638	OAC Libra	*G. max*	Elite	NA Cultivar	Ontario, Canada	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	0	Zhou et al. (2015)
PI 548644	OAC Musca	*G. max*	Elite	NA Cultivar	Ontario, Canada	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	0	Zhou et al. (2015)
PI 591435	OT94-41	*G. max*	Elite	NA Cultivar	Ontario, Canada	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	I	Zhou et al. (2015)
PI 561389 B	(Okura Natto)	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*	*e1-as*	*E2*	*e3*	Indeterminate	0	USB481
PI 361080	Kormovaia 15	*G. max*	Landrace	Other	Russian Federation	*Dt1*	*dt2*	*e1-as*	*E2*	*e3*	Indeterminate	II	USB481
PI 266806 C	No. 4	*G. max*	Landrace	Other	Hebei Sheng, China	*Dt1*	*dt2*	*e1-as*	*E2*	*e3*	Indeterminate	II	USB481
PI 548520	Preston	*G. max*	Elite	Other	Iowa, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*e3*	Indeterminate	II	USB481
PI 548311	Capital	*G. max*	Elite	NA Ancestor	Ontario, Canada	*Dt1*	*dt2*	*e1-as*	*E2*	*e3*	Indeterminate	0	Zhou et al. (2015)

PI 548540	Corsoy	*G. max*	Elite	NA Cultivar	Iowa, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*e3*	Indeterminate	II	Zhou et al. (2015)
PI 578375 B	(Aan tu dang di hei dou)	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	I	USB481
PI 639550 E	(KSHI 713)	*G. max*	Landrace	Other	Moldova	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	II	USB481
PI 605765 B	(Ninh minh)	*G. max*	Landrace	Other	Tuyên Quang, Vietnam	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	II	USB481
PI 556511	A3127	*G. max*	Elite	Other	United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	USB481
PI 548313	Chestnut	*G. max*	Landrace	Other	Habarovskij kraj, Russian Federation	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	USB481
PI 378663	Habarovskaja II	*G. max*	Landrace	Other	Russian Federation	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	I	USB481
PI 561318 A	Hui nan bai hua xiao hei dou	*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	I	USB481
PI 603442	Ke qi xiao hei dou	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	USB481
PI 548383	Mansoy	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	USB481
PI 253661 B	No. 12	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	USB481
PI 548400	Patoka	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	USB481
PI 552538	Dunbar	*G. max*	Elite	NA Cultivar	Nebraska, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 542044	Kunitz	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 639740	LD00-3309	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 593258	Macon	*G. max*	Elite	Other	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 597387	Pana	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 547460	L64-1083	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 547562	L72-2157	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 547862	L83-570	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 591511	L89-1581	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Zhou et al. (2015)
PI 591539	L91-8558	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 591495	L93-2740	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 548362	Lincoln	*G. max*	Elite	NA Ancestor	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	III	Zhou et al. (2015)
PI 515961	Pennyrile	*G. max*	Elite	NA Cultivar	Kentucky, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	Zhou et al. (2015)

PI 548603	Perry	*G. max*	Elite	NA Ancestor	Indiana, United States	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 639559 B	(VYTKA 2)	*G. max*	Landrace	Other	Ukraine	*Dt1*	*dt2*	*e1-as*	*N*	*E3*	Indeterminate	II	USB481
PI 547716	L62-667	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*	*e1-as*	*N*	*N*	Indeterminate	II	Zhou et al. (2015)
PI 378658	Dnepropetrovsk 12	*G. max*	Landrace	Other	Dnipropetrovsk, Ukraine	*Dt1*	*dt2*	*N*	*e2*	*E3*	Indeterminate	0	USB481
PI 548572	Harly	*G. max*	Elite	Other	Ontario, Canada	*Dt1*	*dt2*	*N*	*e2*	*E3*	Indeterminate	I	USB481
PI 504288	S	*G. max*	Landrace	Other	Iwate, Japan	*Dt1*	*dt2*				Indeterminate	V	USB481
PI 438335	SAO 196-C	*G. max*	Landrace	Other	Algeria	*Dt1*	*dt2*				Indeterminate	III	USB481
PI 548411	Seneca	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 479735	Silihuang	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*dt2*				Indeterminate	III	USB481
PI 548619	Sparks	*G. max*	Elite	Other	Kansas, United States	*Dt1*	*dt2*				Indeterminate	IV	USB481
PI 180501	Strain No. 18	*G. max*	Landrace	Other	Germany	*Dt1*	*dt2*				Indeterminate	0	USB481
PI 593953	Sui nong No. 10	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	I	USB481
PI 548193	T201	*G. max*	Landrace	Other	Iowa, United States	*Dt1*	*dt2*				Indeterminate	IV	USB481
PI 548200	T211H	*G. max*	Landrace	Other	Illinois, United States	*Dt1*	*dt2*				Indeterminate	IV	USB481
PI 587588 A	Tai xing niu mao huang yi	*G. max*	Landrace	Other	Jiangsu Sheng, China	*Dt1*	*dt2*				Indeterminate	IV	USB481
PI 548490	Tanner	*G. max*	Landrace	Other	Taiwan	*Dt1*	*dt2*				Indeterminate	VII	USB481
PI 632418	Tara	*G. max*	Elite	Other	Maryland, United States	*Dt1*	*dt2*				Indeterminate	V	USB481
PI 417381	Tenpoku shirome	*G. max*	Landrace	Other	Hokkaidô, Japan	*Dt1*	*dt2*				Indeterminate	0	USB481
PI 578503	Tie jia si li huang	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	I	USB481
PI 518668	TN 4-86	*G. max*	Elite	Other	Tennessee, United States	*Dt1*	*dt2*				Indeterminate	IV	USB481
PI 437165 A	Toncostebelnaja 27	*G. max*	Landrace	Other	Krasnodar, Russian Federation	*Dt1*	*dt2*				Indeterminate	I	USB481
PI 507467	Tousan kei F 764	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*				Indeterminate	IV	USB481
PI 507471	Tousan kei na 16	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*				Indeterminate	III	USB481
PI 594307	Tsurusengoku	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*				Indeterminate	VIII	USB481
PI 437376 A	Ussurijscaja 308	*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*				Indeterminate	I	USB481
PI 437991 B	VIR 1657	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	0	USB481
PI 438019 B	VIR 1883	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 639528 B	VIR 233	*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 437110 A	VIR 244	*G. max*	Landrace	Other	Russian Federation	*Dt1*	*dt2*				Indeterminate	III	USB481
PI 437112 A	VIR 249	*G. max*	Landrace	Other	Russian Federation	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 438083	VIR 2506	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 437788 A	VIR 3018	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 639543	VIR 3715	*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*				Indeterminate	II	USB481
PI 437500 A	VIR 3810	*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*				Indeterminate	I	USB481

PI 437505	VIR 3853	*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*	Indeterminate	II	USB481
PI 438230 A	VIR 4521	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	I	USB481
PI 438239 B	VIR 4536	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	I	USB481
PI 639570	VIR 7010	*G. max*	Landrace	Other	Philippines	*Dt1*	*dt2*	Indeterminate	V	USB481
PI 438500	Virginia	*G. max*	Landrace	Other	United States	*Dt1*	*dt2*	Indeterminate	III	USB481
PI 567238	W6 6210	*G. max*	Landrace	Other	Yunnan Sheng, China	*Dt1*	*dt2*	Indeterminate	IX	USB481
PI 548524	Weber	*G. max*	Elite	Other	Iowa, United States	*Dt1*	*dt2*	Indeterminate	I	USB481
PI 548427	Wilson	*G. max*	Landrace	Other	Liaoning Sheng, China	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 445824 A	Wolfsthaler	*G. max*	Landrace	Other	Germany	*Dt1*	*dt2*	Indeterminate	000	USB481
PI 548633	Wye	*G. max*	Elite	Other	Maryland, United States	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 603399	Xiao bai qi	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	II	USB481
PI 567407	Xiao dou	*G. max*	Landrace	Other	Shaanxi Sheng, China	*Dt1*	*dt2*	Indeterminate	V	USB481
PI 567408	Xiao jin huang	*G. max*	Landrace	Other	Shaanxi Sheng, China	*Dt1*	*dt2*	Indeterminate	V	USB481
PI 495020	Xu dou 2	*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 603290	Zao shu 18	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	I	USB481
PI 592937	ZDD 18846	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 592940	ZDD 18849	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 603556	ZDD08563	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	III	USB481
PI 603559	ZDD08590	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 467347	Zi-hua-cuo-zi	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*dt2*	Indeterminate	II	USB481
PI 612754	ZY 645	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	Indeterminate	I	USB481
PI 549017	ZYD 3938	*G. max*	Landrace	Other	Ningxia Huizi Zizhiqu, China	*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 549018	ZYD 3939	*G. max*	Landrace	Other	Ningxia Huizi Zizhiqu, China	*Dt1*	*dt2*	Indeterminate	V	USB481
FC 029333		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
FC 031697		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 054591		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 054614		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 058955		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 062203		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	V	USB481
PI 070080		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 071465		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	V	USB481
PI 081041		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 081785		*G. max*	Landrace	Other	Hokkaido, Japan	*Dt1*	*dt2*	Indeterminate	III	USB481
PI 083881		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 084637		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	II	USB481
PI 084656		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 084973		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 086904		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	VI	USB481
PI 087620		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 088788		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481

PI 089775		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	VI	USB481
PI 090763		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 091160		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 092651		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 291294		*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	Indeterminate	I	USB481
PI 468908		*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*dt2*	Indeterminate	000	USB481
PI 475820		*G. max*	Landrace	Other	Xinjiang Uygur Zizhiqu, China	*Dt1*	*dt2*	Indeterminate	II	USB481
PI 291309 D		*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	Indeterminate	II	USB481
PI 291310 C		*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	Indeterminate	II	USB481
PI 342619 A		*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*	Indeterminate	0	USB481
PI 054615 -1		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 068732 -1		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	USB481
PI 091159 -4		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	IV	USB481
PI 548415	Sooty	*G. max*	Landrace	Other	Zhejiang Sheng, China	*Dt1*	*dt2*	Indeterminate	IV	Valliyodan et al. (2006)
PI 438258	VIR 4714	*G. max*	Elite	Other	China	*Dt1*	*dt2*	Indeterminate	II	Valliyodan et al. (2006)
PI 567230	WJK-PRC-23	*G. max*	Landrace	Other	Shaanxi Sheng, China	*Dt1*	*dt2*	Indeterminate	V	Valliyodan et al. (2006)
PI 567354	You huang dou	*G. max*	Landrace	Other	Gansu Sheng, China	*Dt1*	*dt2*	Indeterminate	IV	Valliyodan et al. (2006)
FC 031721		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	VI	Valliyodan et al. (2006)
PI 086006		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	Valliyodan et al. (2006)
PI 087617		*G. max*	Elite	Other		*Dt1*	*dt2*	Indeterminate	III	Valliyodan et al. (2006)
PI 407729		*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	Indeterminate	IV	Valliyodan et al. (2006)
PI 468915		*G. max*	Landrace	Other	Liaoning Sheng, China	*Dt1*	*dt2*	Indeterminate	II	Valliyodan et al. (2006)
PI 549031		*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*dt2*	Indeterminate	III	Valliyodan et al. (2006)
PI 603176 A		*G. max*	Elite	Other	Korea, North	*Dt1*	*dt2*	Indeterminate	IV	Valliyodan et al. (2006)
PI 087631 -1		*G. max*	Landrace	Other		*Dt1*	*dt2*	Indeterminate	III	Valliyodan et al. (2006)

PI 548488	S-100	*G. max*	Landrace	NA Ancestor	Heilongjiang Sheng, China	*Dt1*	*dt2*				Indeterminate	V	Zhou et al. (2015)
PI 548631	Williams	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*				Indeterminate	III	Zhou et al. (2015)
PI 603318	Xiao zhu yao	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	I	Zhou et al. (2015)
PI 548634	Zane	*G. max*	Elite	NA Cultivar	Ohio, United States	*Dt1*	*dt2*				Indeterminate	III	Zhou et al. (2015)
PI 603424 A	ZDD007871	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	0	Zhou et al. (2015)
PI 603420	ZDD01501	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Indeterminate	II	Zhou et al. (2015)
PI 89138	Zontanorukon	*G. max*	Landrace	Other	Hamkyeongpukto, Korea, North	*Dt1*	*dt2*				Indeterminate	II	Zhou et al. (2015)
PI 339734		*G. max*	Landrace	Other	Gangwon-do, Korea, South	*Dt1*	*dt2*				Indeterminate	IV	Zhou et al. (2015)
PI 603675	Huai yin gua dou jia	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	III	USB481
PI 594599	Chang de chun hei dou	*G. max*	Elite	Other	Hunan Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	IV	Valliyodan et al. (2006)
PI 437321	Dunganscaja 462	*G. max*	Landrace	Other	Primorye, Russian Federation	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	III	Zhou et al. (2015)
PI 88479	Kungchuling Improved No. 77	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	II	Zhou et al. (2015)
PI 594615	Liu yue zao	*G. max*	Landrace	Other	Guizhou Sheng, China	*Dt1*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	IV	Zhou et al. (2015)
PI 407708 A	Feng shou No. 10	*G. max*	Landrace	Other	Heilongjiang Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Semi-determinate	0	USB481
PI 548406	Richland	*G. max*	Landrace	NA Ancestor	Jilin Sheng, China	*Dt1*	*dt2*	*e1-as*	*e2*	*e3*	Semi-determinate	II	Zhou et al. (2015)
PI 297520	Iregi Universal	*G. max*	Landrace	Other	Hungary	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Semi-determinate	0	USB481
PI 603426 G	(Ben di yuan huang dou)	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*e1-as*	*E2*	*E3*	Semi-determinate	II	USB481
PI 437776	VIR 1302	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Semi-determinate	III	USB481
PI 603345	ZDD00403	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Semi-determinate	II	USB481
PI 094159 -3		*G. max*	Landrace	Other		*Dt1*	*dt2*				Semi-determinate	IV	USB481
PI 458515	Tie Zhugan	*G. max*	Landrace	Other	Shandong Sheng, China	*Dt1*	*dt2*				Semi-determinate	IV	Valliyodan et al. (2006)

PI 80822	Shiheigai Shirobana	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Semi-determinate	III	Zhou et al. (2015)
PI 548182	T157	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*dt2*				Semi-determinate	III	Zhou et al. (2015)
PI 437944	VIR 569	*G. max*	Landrace	Other	China	*Dt1*	*dt2*				Semi-determinate	II	Zhou et al. (2015)
PI 587848	Wu chang hei dong dou	*G. max*	Landrace	Other	Hubei Sheng, China	*Dt1*	*dt2*				Semi-determinate	V	Zhou et al. (2015)
PI 578499 A	Lu yue bai	*G. max*	Landrace	Other	China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Determinate	II	USB481
PI 587552	Nan jing da ping ding huang yi No. 1	*G. max*	Landrace	Other	Jiangsu Sheng, China	*Dt1*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	Zhou et al. (2015)
PI 506942	Koushurei 235	*G. max*	Landrace	Other	Japan	*Dt1*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	II	USB481
PI 248515	White Hilum Iwata Variety No. 2	*G. max*	Elite	Other	Japan	*Dt1*	*dt2*				Determinate	IV	Valliyodan et al. (2006)
PI 567488 A	Di liu huang dou No. 2	*G. max*	Landrace	Other	Hebei Sheng, China	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 437838	DV-254	*G. max*	Landrace	Other	Russian Federation	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Indeterminate	II	USB481
PI 567225	Kisinevskaja 90	*G. max*	Landrace	Other	Moldova	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Indeterminate	0	USB481
PI 171428	Large Yellow Soybean	*G. max*	Landrace	Other	Beijing Shi, China	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 437127 A	Imeretinscaja	*G. max*	Landrace	Other	Georgia	*Dt1*	*Dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 417091	Kuro mame	*G. max*	Landrace	Other	Japan	*Dt1*	*Dt2*	*E1*	*E2*	*E3*	Indeterminate	II	Valliyodan et al. (2006)
PI 437863 A	DV-2841	*G. max*	Elite	Other	China	*Dt1*	*Dt2*	*E1*	*N*	*E3*	Indeterminate	II	Valliyodan et al. (2006)
PI 438112 B	VIR 2623	*G. max*	Landrace	Other	China	*Dt1*	*Dt2*				Indeterminate	III	USB481
PI 549041 A	ZYD 2709	*G. max*	Landrace	Other	Liaoning Sheng, China	*Dt1*	*Dt2*				Indeterminate	III	USB481
PI 054608 -1		*G. max*	Landrace	Other		*Dt1*	*Dt2*				Indeterminate	II	USB481
PI 407716	Jin nung No. 3	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*Dt2*	*E1*	*e2*	*e3*	Semi-determinate	I	Zhou et al. (2015)
PI 391583	Jilin No. 10	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Semi-determinate	II	USB481
PI 547409	L62-1251	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*Dt2*	*e1-as*	*E2*	*E3*	Semi-determinate	IV	Zhou et al. (2015)
PI 547459	L64-1081	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*Dt2*	*e1-as*	*E2*	*E3*	Semi-determinate	IV	Zhou et al. (2015)
PI 548169	T117	*G. max*	Landrace	Other	Illinois, United States	*Dt1*	*Dt2*				Semi-determinate	IV	USB481
PI 464923	Tie Fen 16	*G. max*	Landrace	Other	Liaoning Sheng, China	*Dt1*	*Dt2*				Semi-determinate	I	USB481
PI 548190	T176	*G. max*	Elite	NA Cultivar	Illinois, United States	*Dt1*	*Dt2*				Semi-determinate	II	Zhou et al. (2015)
PI 467343	Yan-nong No. 2	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*Dt2*				Semi-determinate	I	Zhou et al. (2015)
PI 458510	Ji Ti No. 1	*G. max*	Landrace	Other	Liaoning Sheng, China	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Determinate	III	USB481
PI 464896	Jou Nong No. 5	*G. max*	Landrace	Other	Jilin Sheng, China	*Dt1*	*Dt2*	*E1*	*e2*	*E3*	Determinate	I	USB481

PI 603357	Du Lu Dou	*G. max*	Landrace	Other	China	*Dt1*	*Dt2*	*e1-as*	*E2*	*e3*	Determinate	I	Zhou et al. (2015)
PI 476352 B	(Colnon)	*G. max*	Landrace	Other	Kyrgyzstan	*Dt1*	*Dt2*	*e1-as*	*E2*	*E3*	Determinate	II	USB481
PI 088313		*G. max*	Landrace	Other		*Dt1*	*Dt2*				Determinate	II	USB481
PI 417345 B	(Shou outou)	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 578493	Huang bao zhu	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	II	USB481
PI 548474	Nanda	*G. max*	Landrace	Other	Hwanghaipukto, Korea, North	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VIII	USB481
PI 159925	Glycine H	*G. max*	Landrace	Other	Lima, Peru	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	VIII	USB481
PI 548198	T209	*G. max*	Landrace	Other	Illinois, United States	*dt1R166W*	*dt2*				Indeterminate	III	USB481
PI 587811 A	ZDD005777	*G. max*	Landrace	Other	Hubei Sheng, China	*dt1R166W*	*dt2*				Indeterminate	VIII	USB481
PI 549040	ZYD 2704	*G. max*	Landrace	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*				Indeterminate	IV	USB481
PI 091100 -3		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Indeterminate	III	USB481
PI 079691 -4		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Indeterminate	III	Valliyodan et al. (2006)
PI 567532	Dai ye xiao huang dou	*G. max*	Landrace	Other	Shandong Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Semi-determinate	IV	USB481
PI 416890	Gokuwase natsu daizu	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Semi-determinate	0	Zhou et al. (2015)
PI 548382	Manitoba Brown	*G. max*	Landrace	NA Ancestor	Unknown	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Semi-determinate	00	Zhou et al. (2015)
PI 506933	Kouiku 1	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	IV	USB481
PI 628812	MG/BR-46 (Conquista)	*G. max*	Landrace	Other	Brazil	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	V	USB481
PI 154189	No. 57	*G. max*	Landrace	Other	Netherlands	*dt1R166W*	*dt2*	*E1*	*E2*	*e3*	Semi-determinate	0	USB481
PI 628913	BR-30	*G. max*	Landrace	Other	Brazil	*dt1R166W*	*dt2*	*e1-as*	*e2*	*E3*	Semi-determinate	VI	USB481
PI 567262 A	Similar to: Gu tian type	*G. max*	Landrace	Other	Fujian Sheng, China	*dt1R166W*	*dt2*				Semi-determinate	II	USB481
PI 423926	Tousan 72	*G. max*	Landrace	Other	Nagano, Japan	*dt1R166W*	*dt2*				Semi-determinate	IV	USB481
PI 317336	Shinsei	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R166W*	*dt2*				Semi-determinate	0	Zhou et al. (2015)
PI 507293 B	(Shoukin ou)	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	III	USB481
PI 416838	Choutan	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	USB481
PI 549028	Feng da li	*G. max*	Landrace	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	USB481
PI 603397	Hei qi huang da dou	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	IV	USB481
PI 548356	Kanro	*G. max*	Landrace	Other	Phyeongyang, Korea, North	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	II	USB481
PI 438471	Fiskeby III	*G. max*	Elite	Other	Östergötlands län, Sweden	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	00	Valliyodan et al. (2006)

PI 603154	GL 2622 /96	*G. max*	Elite	Other	Korea, North	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	Valliyodan et al. (2006)
PI 603170	GL 2683 /96	*G. max*	Elite	Other	Korea, North	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	IV	Valliyodan et al. (2006)
PI 603175	GL 2688 /96	*G. max*	Elite	Other	Korea, North	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	IV	Valliyodan et al. (2006)
PI 200471	Hanayome Ibaragi No. 1	*G. max*	Elite	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	III	Valliyodan et al. (2006)
PI 398593	KAS 390-4	*G. max*	Elite	Other	Chungcheongbuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	Valliyodan et al. (2006)
PI 398595	KAS 390-5	*G. max*	Elite	Other	Chungcheongbuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	Valliyodan et al. (2006)
PI 398610	KAS 390-8	*G. max*	Elite	Other	Chungcheongbuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	Valliyodan et al. (2006)
PI 398614	KAS 390-9	*G. max*	Elite	Other	Chungcheongbuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	V	Valliyodan et al. (2006)
PI 200508	Natsu Daizu	*G. max*	Elite	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	I	Valliyodan et al. (2006)
PI 229343	Nonaka No. 1	*G. max*	Elite	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	IV	Valliyodan et al. (2006)
PI 407788 A	ORD 8113	*G. max*	Elite	Other	Kyonggi, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	IV	Valliyodan et al. (2006)
PI 398296	KAS 173-3	*G. max*	Landrace	Other	Kyonggi, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	II	Zhou et al. (2015)
PI 407849	KAS 510-1	*G. max*	Landrace	Other	Jeollabuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	III	Zhou et al. (2015)
PI 424391	KAS 521-15	*G. max*	Landrace	Other	Jeollabuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	VI	Zhou et al. (2015)
PI 317334 A	Kitamishiro	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	I	Zhou et al. (2015)
PI 591541	L74-102	*G. max*	Elite	NA Cultivar	Illinois, United States	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	II	Zhou et al. (2015)
PI 84987	Oni Hadaka	*G. max*	Landrace	Other	Saitama, Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	III	Zhou et al. (2015)
PI 591431	OT94-49	*G. max*	Elite	NA Cultivar	Ontario, Canada	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	0	Zhou et al. (2015)
PI 603336	Qing pi si li huang	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*e3*	Determinate	II	Zhou et al. (2015)
PI 437685 D	(Phun-zhun)	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	USB481
PI 430595	58-161	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	IV	USB481
PI 567788	Bienville	*G. max*	Elite	Other	Louisiana, United States	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VIII	USB481
PI 464912	Dan Dou 1	*G. max*	Landrace	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	IV	USB481
PI 490766	Dawudou	*G. max*	Landrace	Other	Hebei Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	USB481
PI 171451	Kosamame	*G. max*	Landrace	Other	Kanagawa, Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VII	USB481

PI 549021 A	Na hei dou	*G. max*	Landrace	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	USB481
PI 507088	Nattou Kotsubu	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VI	USB481
PI 417215	Ooita Aki Daizu 2	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VIII	USB481
PI 471938	197	*G. max*	Elite	Other	Nepal	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	Valliyodan et al. (2006)
PI 464920 B	(Jin Dou 33)	*G. max*	Elite	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	Valliyodan et al. (2006)
PI 594012	Heuksatangdu	*G. max*	Elite	Other	Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	Valliyodan et al. (2006)
PI 416937	Houjaku Kuwazu	*G. max*	Elite	Other	Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VI	Valliyodan et al. (2006)
PI 518664	Hutcheson	*G. max*	Elite	NA Cultivar	Virginia, United States	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	Valliyodan et al. (2006)
PI 548657	Jackson	*G. max*	Elite	NA Ancestor	North Carolina, United States	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VII	Valliyodan et al. (2006)
PI 408105 A	KAS 633-19	*G. max*	Elite	Other	Gyeongsangbuk-do, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	IV	Valliyodan et al. (2006)
PI 417015	Kawanagare (Iwate)	*G. max*	Elite	Other	Iwate, Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	Valliyodan et al. (2006)
PI 548342	Higan	*G. max*	Landrace	Other	Tôkyô, Japan	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	IV	Zhou et al. (2015)
PI 548985	Kershaw	*G. max*	Elite	NA Cultivar	South Carolina, United States	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VI	Zhou et al. (2015)
PI 399043	KLS 903	*G. max*	Landrace	Other	Jeju-teukbyeoljachido, Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	Zhou et al. (2015)
PI 196166	No. 2296	*G. max*	Landrace	Other	Korea, South	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	Zhou et al. (2015)
PI 548477	Ogden	*G. max*	Elite	NA Ancestor	Tennessee, United States	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VI	Zhou et al. (2015)
PI 603384	Ping ding xiang	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	Zhou et al. (2015)
PI 548485	Roanoke	*G. max*	Landrace	NA Ancestor	Jiangsu Sheng, China	*dt1R166W*	*dt2*	*E1*	*e2*	*E3*	Determinate	VII	Zhou et al. (2015)
PI 506862	Karikei 86	*G. max*	Elite	Other	Japan	*dt1R166W*	*dt2*	*E1*	*E2*	*e3*	Determinate	IV	USB481
PI 209334	No. 9	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R166W*	*dt2*	*E1*	*E2*	*e3*	Determinate	III	USB481
PI 548456	Haberlandt	*G. max*	Landrace	NA Ancestor	Phyeongyang, Korea, North	*dt1R166W*	*dt2*	*E1*	*E2*	*e3*	Determinate	VI	Zhou et al. (2015)
PI 561701	G88-20092	*G. max*	Landrace	Other	Georgia, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	IV	USB481
PI 548978	Gail	*G. max*	Elite	Other	Texas, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VI	USB481

PI 594922	Graham	*G. max*	Elite	Other	North Carolina, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	V	USB481
PI 542972	H7190	*G. max*	Landrace	Other	United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	USB481
PI 561387	Kosuzu	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	V	USB481
PI 559932	Manokin	*G. max*	Elite	Other	Maryland, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	IV	USB481
PI 548473	Monetta	*G. max*	Landrace	Other	Jiangsu Sheng, China	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	USB481
PI 594512 A	Bian zi jiang se dou	*G. max*	Elite	Other	Sichuan Sheng, China	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	Valliyodan et al. (2006)
PI 548667	Essex	*G. max*	Elite	NA Cultivar	Virginia, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	V	Valliyodan et al. (2006)
PI 647086	N8001	*G. max*	Elite	NA Cultivar	North Carolina, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VIII	Valliyodan et al. (2006)
PI 553047	Gordon	*G. max*	Elite	NA Cultivar	Georgia, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	Zhou et al. (2015)
PI 533602	Lloyd	*G. max*	Elite	NA Cultivar	Arkansas, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	VI	Zhou et al. (2015)
PI 548604	Pershing	*G. max*	Elite	NA Cultivar	Missouri, United States	*dt1R166W*	*dt2*	*E1*	*E2*	*E3*	Determinate	IV	Zhou et al. (2015)
PI 591432	OT94-51	*G. max*	Elite	NA Cultivar	Ontario, Canada	*dt1R166W*	*dt2*	*E1*	*N*	*e3*	Determinate	0	Zhou et al. (2015)
PI 546044	OT89-06	*G. max*	Elite	NA Cultivar	Ontario, Canada	*dt1R166W*	*dt2*	*e1-as*	*e2*	*e3*	Determinate	0	Zhou et al. (2015)
PI 591433	OT94-37	*G. max*	Elite	NA Cultivar	Ontario, Canada	*dt1R166W*	*dt2*	*e1-as*	*e2*	*e3*	Determinate	0	Zhou et al. (2015)
PI 86024	Daidzuhinshu satei	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R166W*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	III	Zhou et al. (2015)
PI 547779	L72D-4110	*G. max*	Elite	NA Cultivar	Illinois, United States	*dt1R166W*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	II	Zhou et al. (2015)
PI 360957	Karafuto No. 1	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R166W*	*dt2*	*e1-as*	*e2*	*N*	Determinate	00	USB481
PI 232992	Kono-Kuradaizu	*G. max*	Landrace	Other	Saga, Japan	*dt1R166W*	*dt2*	*e1-as*	*E2*	*e3*	Determinate	III	USB481
PI 540552	Hoyt	*G. max*	Elite	NA Cultivar	Ohio, United States	*dt1R166W*	*dt2*	*e1-as*	*E2*	*e3*	Determinate	II	Zhou et al. (2015)
PI 507180	Rikuu 21	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*	*e1-as*	*E2*	*E3*	Determinate	IV	USB481
PI 548565	Gnome	*G. max*	Elite	NA Cultivar	Ohio, United States	*dt1R166W*	*dt2*	*e1-as*	*E2*	*E3*	Determinate	II	Zhou et al. (2015)
PI 547488	L67-3207	*G. max*	Elite	NA Cultivar	Illinois, United States	*dt1R166W*	*dt2*	*e1-as*	*E2*	*E3*	Determinate	IV	Zhou et al. (2015)
PI 549026	Gao li huang	*G. max*	Landrace	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*	*N*	*N*	*E3*	Determinate	V	USB481
PI 548178	T145	*G. max*	Landrace	Other	Illinois, United States	*dt1R166W*	*dt2*				Determinate	III	USB481
PI 548256	T279	*G. max*	Landrace	Other	Mississippi, United States	*dt1R166W*	*dt2*				Determinate	VII	USB481
PI 436684	Tie-feng 8	*G. max*	Landrace	Other	Liaoning Sheng, China	*dt1R166W*	*dt2*				Determinate	III	USB481
PI 598358	TN 5-95	*G. max*	Elite	Other	Tennessee, United States	*dt1R166W*	*dt2*				Determinate	V	USB481
PI 507458	Tousan kei BL 521	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*				Determinate	IV	USB481
PI 507480	Tousan kei YL 24	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*				Determinate	IV	USB481

PI 417479	Yougetsu	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*				Determinate	IV	USB481
PI 083942		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Determinate	V	USB481
PI 088468		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Determinate	II	USB481
PI 095860		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Determinate	VI	USB481
PI 090479 P		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Determinate	IV	USB481
PI 196175	Yu tae	*G. max*	Elite	Other	Korea, South	*dt1R166W*	*dt2*				Determinate	V	Valliyodan et al. (2006)
PI 84631	S-56	*G. max*	Landrace	Other	Kyonggi, Korea, South	*dt1R166W*	*dt2*				Determinate	III	Zhou et al. (2015)
PI 243541	Shakujo	*G. max*	Landrace	Other	Akita, Japan	*dt1R166W*	*dt2*				Determinate	IV	Zhou et al. (2015)
PI 423954	Shirome	*G. max*	Landrace	Other	Kumamoto, Japan	*dt1R166W*	*dt2*				Determinate	0	Zhou et al. (2015)
PI 536635	Sprite	*G. max*	Elite	NA Cultivar	Ohio, United States	*dt1R166W*	*dt2*				Determinate	III	Zhou et al. (2015)
PI 507355	Tokei 423	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R166W*	*dt2*				Determinate	I	Zhou et al. (2015)
PI 417398	Touhou torotou	*G. max*	Landrace	Other	China	*dt1R166W*	*dt2*				Determinate	III	Zhou et al. (2015)
PI 594301	Toyomusume	*G. max*	Landrace	Other	Japan	*dt1R166W*	*dt2*				Determinate	I	Zhou et al. (2015)
PI 508266	Young	*G. max*	Elite	NA Cultivar	North Carolina, United States	*dt1R166W*	*dt2*				Determinate	VI	Zhou et al. (2015)
PI 594579	Zhong he tian cheng dou	*G. max*	Landrace	Other	Hunan Sheng, China	*dt1R166W*	*dt2*				Determinate	V	Zhou et al. (2015)
PI 080837		*G. max*	Landrace	Other		*dt1R166W*	*dt2*				Determinate	IV	Zhou et al. (2015)
PI 407801		*G. max*	Landrace	Other	Kyonggi, Korea, South	*dt1R166W*	*dt2*				Determinate	VI	Zhou et al. (2015)
PI 567611	Ba yue zha	*G. max*	Elite	Other	Henan Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Valliyodan et al. (2006)
PI 587666	Er dao zao	*G. max*	Landrace	Other	Anhui Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VI	Zhou et al. (2015)
PI 416971	Kaifuu gyuumou ou 1	*G. max*	Landrace	Other	Japan	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	Zhou et al. (2015)
PI 404187	Suj nii hun mao ju	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Indeterminate	II	USB481
PI 567231	WJK-PRC-46	*G. max*	Landrace	Other	Sichuan Sheng, China	*dt1P113L*	*dt2*				Indeterminate	VIII	USB481
PI 567675	Yu cheng xiao tie jiao huang	*G. max*	Landrace	Other	Henan Sheng, China	*dt1P113L*	*dt2*				Indeterminate	IV	USB481
PI 567685	Zhong mou tie jiao er cao	*G. max*	Landrace	Other	Henan Sheng, China	*dt1P113L*	*dt2*				Indeterminate	IV	USB481
PI 605869 A	Sample 140	*G. max*	Elite	Other	Lào Cai, Vietnam	*dt1P113L*	*dt2*				Indeterminate	V	Valliyodan et al. (2006)
PI 567780 B	(Tong shan zheng ji dou)	*G. max*	Landrace	Other	Jiangsu Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	IV	USB481

PI 567690	Fu yang (7)	*G. max*	Elite	Other	Anhui Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	III	Valliyodan et al. (2006)
PI 538386 A	1886	*G. max*	Landrace	Other	Hebei Sheng, China	*dt1P113L*	*dt2*	*E1*	*E2*	*E3*	Semi-determinate	III	USB481
PI 437695 A	S-185	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Semi-determinate	I	USB481
PI 594456 A	Xiao jin huang	*G. max*	Landrace	Other	Sichuan Sheng, China	*dt1P113L*	*dt2*				Semi-determinate	III	USB481
PI 594777	Liu yue huang	*G. max*	Landrace	Other	Yunnan Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*e3*	Determinate	IV	Zhou et al. (2015)
PI 587712 B	(E dou No. 1)	*G. max*	Landrace	Other	Hubei Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 548447	Cherokee	*G. max*	Landrace	Other	Zhejiang Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Determinate	VIII	USB481
PI 518727	Ju huang	*G. max*	Landrace	Other	Guangdong Sheng, China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Determinate	VI	USB481
PI 603596	Bai hua gu tian dou	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Determinate	III	Zhou et al. (2015)
PI 567189 A	Ekhabac	*G. max*	Landrace	Other	Vietnam	*dt1P113L*	*dt2*	*E1*	*e2*	*E3*	Determinate	IV	Zhou et al. (2015)
PI 603463	Dong jie No. 1	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	II	USB481
PI 603458 A	Shui dou	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Determinate	IV	USB481
PI 379618	TC 1	*G. max*	Landrace	Other	Taiwan	*dt1P113L*	*dt2*				Determinate	V	USB481
PI 578504	Xiang dou No. 3	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Determinate	II	USB481
PI 592954	ZDD 11242	*G. max*	Elite	Other	China	*dt1P113L*	*dt2*				Determinate	II	USB481
PI 603488	ZDD19294	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Determinate	III	USB481
PI 597464	Zhe chun No. 3	*G. max*	Landrace	Other	Zhejiang Sheng, China	*dt1P113L*	*dt2*				Determinate	II	USB481
PI 342434		*G. max*	Landrace	Other	Iwate, Japan	*dt1P113L*	*dt2*				Determinate	V	USB481
PI 587752	Xian ning dong huang dou jia	*G. max*	Landrace	Other	Hubei Sheng, China	*dt1P113L*	*dt2*				Determinate	V	Zhou et al. (2015)
PI 594629	Xiao hua lian	*G. max*	Landrace	Other	Guizhou Sheng, China	*dt1P113L*	*dt2*				Determinate	VI	Zhou et al. (2015)
PI 588053 A	Xiao li huang	*G. max*	Landrace	Other	Guangdong Sheng, China	*dt1P113L*	*dt2*				Determinate	V	Zhou et al. (2015)
PI 603516	Xiao ma yi dan	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Determinate	VI	Zhou et al. (2015)
PI 603756	ZDD05996	*G. max*	Landrace	Other	China	*dt1P113L*	*dt2*				Determinate	II	Zhou et al. (2015)
PI 548656	Lee	*G. max*	Elite	Other	Mississippi, United States	*dt1P113L & dt1L67Q*	*dt2*	*E1*	*E2*	*E3*	Determinate	VI	USB481
PI 548658	Lee 74	*G. max*	Elite	Other	Arkansas, United States	*dt1P113L & dt1L67Q*	*dt2*	*E1*	*E2*	*E3*	Determinate	VI	USB481
PI 165675	Nanking 332	*G. max*	Landrace	Other	Jiangsu Sheng, China	*dt1P113L & dt1L67Q*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	USB481
PI 548445	CNS	*G. max*	Landrace	NA Ancestor	Jiangsu Sheng, China	*dt1P113L & dt1L67Q*	*dt2*	*E1*	*E2*	*E3*	Determinate	VII	Zhou et al. (2015)

PI 189873	Miko Saumon	*G. max*	Landrace	Other	France	*dt1R130K*	*dt2*	*E1*	*e2*	*e3*	Indeterminate	0	USB481
PI 378680 E	(VNIIMK 9186)	*G. max*	Landrace	Other	Russian Federation	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	I	USB481
PI 606374	Cao bang 8	*G. max*	Landrace	Other	Vietnam	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	IV	USB481
PI 361093	Novosadska Br. 1	*G. max*	Landrace	Other	Serbia	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	I	USB481
PI 567307	Hei huang dou	*G. max*	Landrace	Other	Gansu Sheng, China	*dt1R130K*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	IV	USB481
PI 437485	VIR 1048	*G. max*	Landrace	Other	Primorye, Russian Federation	*dt1R130K*	*dt2*				Indeterminate	II	USB481
PI 404182	Sin i tu li rau	*G. max*	Landrace	Other	China	*dt1R130K*	*dt2*				Indeterminate	III	Zhou et al. (2015)
PI 567525	Cao qing huang dou	*G. max*	Landrace	Other	Shandong Sheng, China	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	II	USB481
PI 437653	Er-da-li	*G. max*	Landrace	Other	China	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	II	Zhou et al. (2015)
PI 594451	Liu yue bao	*G. max*	Landrace	Other	Sichuan Sheng, China	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	III	Zhou et al. (2015)
PI 437814 A	Anda	*G. max*	Landrace	Other	China	*dt1R130K*	*dt2*	*e1-as*	*E2*	*E3*	Semi-determinate	II	USB481
PI 438309	VIR 3017	*G. max*	Landrace	Other	China	*dt1R130K*	*dt2*				Semi-determinate	I	USB481
PI 437793	VIR 3024	*G. max*	Landrace	Other	China	*dt1R130K*	*dt2*				Semi-determinate	II	USB481
PI 567651	Shang cai er cao ping ding shi	*G. max*	Elite	Other	Henan Sheng, China	*dt1R130K*	*dt2*				Semi-determinate	IV	Valliyodan et al. (2006)
PI 153262	Roumanie	*G. max*	Landrace	Other	Belgium	*dt1R130K*	*dt2*				Semi-determinate	0	Zhou et al. (2015)
PI 361087	Medias 23	*G. max*	Landrace	Other	Romania	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Determinate	I	USB481
PI 567298	Chan yao dou	*G. max*	Landrace	Other	Gansu Sheng, China	*dt1R130K*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	Zhou et al. (2015)
PI 603698 J	(Dan yang shui bai dou)	*G. max*	Landrace	Other	China	*dt1R130K*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	0	USB481
PI 437240	CSchi 1069	*G. max*	Landrace	Other	Moldova	*dt1R130K*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	0	USB481
PI 567226	Harkovskaja Zernoukosnaja	*G. max*	Landrace	Other	Russian Federation	*dt1R130K*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	00	USB481
PI 372418	Novosadska Br. 4	*G. max*	Landrace	Other	Serbia	*dt1R130K*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	I	USB481
PI 438336	Sao 208	*G. max*	Landrace	Other	Algeria	*dt1R130K*	*dt2*				Determinate	0	USB481
PI 548417	Soysota	*G. max*	Landrace	Other	Italy	*dt1R130K*	*dt2*				Determinate	I	Zhou et al. (2015)
PI 603722	Nan chong ba yue huang	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	VIII	USB481
PI 567576	Ping ding huang	*G. max*	Landrace	Other	Shandong Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Indeterminate	III	USB481
PI 404166	Krasnoarmejskaja	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*	*E1*	*E2*	*E3*	Indeterminate	III	Valliyodan et al. (2006)
PI 165563	Bhart	*G. max*	Landrace	Other	Uttar Pradesh, India	*dt1R62S*	*dt2*	*e1-as*	*e2*	*E3*	Indeterminate	VII	USB481
PI 548171	T134	*G. max*	Elite	Other	Illinois, United States	*dt1R62S*	*dt2*				Indeterminate	III	USB481
PI 567352 A	Yang yan qing dou	*G. max*	Landrace	Other	Gansu Sheng, China	*dt1R62S*	*dt2*				Indeterminate	IV	USB481
PI 567353	Yang yan ren dou	*G. max*	Landrace	Other	Gansu Sheng, China	*dt1R62S*	*dt2*				Indeterminate	IV	USB481

PI 089772		*G. max*	Landrace	Other		*dt1R62S*	*dt2*				Indeterminate	IV	USB481
PI 567258	NC 9173	*G. max*	Landrace	Other	Jiangxi Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*e3*	Semi-determinate	II	Zhou et al. (2015)
PI 532463 B	(He bei No. 1)	*G. max*	Landrace	Other	Hebei Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	III	USB481
PI 567698 A	Fu yang (17)	*G. max*	Landrace	Other	Anhui Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	IV	USB481
PI 567731	Fu yang (56)	*G. max*	Elite	Other	Anhui Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Semi-determinate	III	Valliyodan et al. (2006)
PI 437690	Pin-din-guan	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*	*E1*	*E2*	*E3*	Semi-determinate	III	Valliyodan et al. (2006)
PI 437679	Nan-cou	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*	*E1*	*E2*	*E3*	Semi-determinate	IV	Zhou et al. (2015)
PI 567503	Niu mao huang	*G. max*	Landrace	Other	Hebei Sheng, China	*dt1R62S*	*dt2*	*E1*	*E2*	*E3*	Semi-determinate	IV	Zhou et al. (2015)
PI 548162	T48	*G. max*	Landrace	Other	Illinois, United States	*dt1R62S*	*dt2*				Semi-determinate	IV	USB481
PI 437725	Te-zu-gan	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*				Semi-determinate	IV	Valliyodan et al. (2006)
PI 438498	Sable	*G. max*	Landrace	Other	United States	*dt1R62S*	*dt2*				Semi-determinate	IV	Zhou et al. (2015)
PI 603497	Hua dou	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*	*E1*	*e2*	*e3*	Determinate	III	USB481
PI 103088	Ming Chuan	*G. max*	Landrace	Other	Henan Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*e3*	Determinate	III	USB481
PI 407742	16	*G. max*	Landrace	Other	Shaanxi Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 587588 B	Tai xing niu mao huang yi	*G. max*	Landrace	Other	Jiangsu Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 567383	Da ke huang dou	*G. max*	Landrace	Other	Shaanxi Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 597476	Deogyukong	*G. max*	Elite	Other	Korea, South	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 548696	Dortchsoy 67	*G. max*	Elite	Other	Arkansas, United States	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 567346	Niu mao huang dou	*G. max*	Landrace	Other	Gansu Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	V	USB481
PI 602991	Niu jiao qi da hei dou	*G. max*	Landrace	Other	Shandong Sheng, China	*dt1R62S*	*dt2*	*E1*	*e2*	*E3*	Determinate	IV	Zhou et al. (2015)
PI 548402	Peking	*G. max*	Landrace	Other	Beijing Shi, China	*dt1R62S*	*dt2*	*E1*	*E2*	*E3*	Determinate	IV	USB481
PI 597478 B	(Paldalkong)	*G. max*	Landrace	Other	Korea, South	*dt1R62S*	*dt2*	*e1-as*	*e2*	*E3*	Determinate	III	USB481
PI 594170 B	(Geden shirazu)	*G. max*	Landrace	Other	Akita, Japan	*dt1R62S*	*dt2*	*e1-as*	*E2*	*e3*	Determinate	I	USB481
PI 594880	Song zi dou	*G. max*	Landrace	Other	Yunnan Sheng, China	*dt1R62S*	*dt2*				Determinate	V	USB481
PI 567604 A	Xin huang dou	*G. max*	Landrace	Other	Shandong Sheng, China	*dt1R62S*	*dt2*				Determinate	IV	USB481
PI 592952	ZDD 10095	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*				Determinate	III	USB481
PI 612730	Zhong huong No. 10	*G. max*	Landrace	Other	China	*dt1R62S*	*dt2*				Determinate	II	USB481
PI 507354	Tokei 421	*G. max*	Landrace	Other	Hokkaidô, Japan	*dt1R62S*	*dt2*				Determinate	I	Valliyodan et al. (2006)

† The genotype at stem termination type loci (*Dt1* and *Dt2*) were assigned based on the allele state at each of the genes from publicly available whole genome resequenced datasets
‡ The allele status at the maturity gene loci (*E1*, *E2* and *E3*) were obtained from a previous work of Langewisch et al. (2014)
§ The detail information of soybean accessions used in the analysis were downloaded from the USDA Germplasm Resources Information Network (GRIN)

Supplementary Table 2. Summary of recombinant inbred line (RIL) population development with plans per year for the developing experimental lines of the four RIL populations used in this study

	Population			
	KB17-16	**KB17-17**	**KB17-7**	**KB17-8**
Pedigree	L91-8052† x Jake‡	L91-8060† x Jake‡	LG90-2550† x Jake‡	KG90-2550† x Ellis‡
Target genotype	R130K E1	R62S E1	Dt2 E1	Dt2 E1
Population development history				
In summer 2017	Crosses made at the South Farm Research Center, Columbia, MO	Crosses made at the South Farm Research Center, Columbia, MO	Crosses made at the South Farm Research Center, Columbia, MO	Crosses made at the South Farm Research Center, Columbia, MO
In winter 2017	Planted F_1 seeds in Costa Rica	Planted F_1 seeds in Costa Rica	Planted F_1 seeds in Costa Rica	Planted F_1 seeds in Costa Rica
	Allele at *Dt1* locus was confirmed from DNA extracted from each of F1 plants	Allele at *Dt1* locus was confirmed from DNA extracted from each of F_1 plants	Allele at *Dt2* locus was confirmed from DNA extracted from each of F_1 plants	Allele at *Dt2* locus was confirmed from DNA extracted from each of F_1 plants
In spring 2018	Bulk harvested F_2 seeds in Costa Rica	Bulk harvested F_2 seeds in Costa Rica	Bulk harvested F_2 seeds in Costa Rica	Bulk harvested F_2 seeds in Costa Rica
			Planted F_2 seeds again in Costa Rica	Planted F_2 seeds again in Costa Rica
			Allele at *Dt2* locus was confirmed from DNA extracted from each of F_2 plants	Allele at *Dt2* locus was confirmed from DNA extracted from each of F_2 plants
			Harvested $F_{2:3}$ seeds from the three selected F_2 plants by single seed thresh	Harvested $F_{2:3}$ seeds from the four selected F_2 plants by single seed thresh
In summer 2018	Planted F_2 seeds at the South Farm Research Center, Columbia, MO	Planted F_2 seeds at the South Farm Research Center, Columbia, MO	Planted $F_{2:3}$ seeds the South Farm Research Center, Columbia, MO	Planted $F_{2:3}$ seeds at the South Farm Research Center, Columbia, MO
			Allele at *E1* locus was confirmed from DNAs extracted from $F_{2:3}$ plants	Allele at *E1* locus was confirmed from DNAs extracted from $F_{2:3}$ plants
In winter 2018	Harvested $F_{2:3}$ seeds by single plant thresh	Harvested $F_{2:3}$ seeds by single plant thresh	Harvested $F_{2:4}$ seeds from the selected by pod pick from 10 plants per plot (total 30 pods were harvested)	Harvested $F_{2:4}$ seeds from the selected by pod pick from 10 plants per plot (total 40 pods were harvested)

	Alleles at *dt1* and *E1* loci were confirmed from DNAs extracted from three $F_{2:3}$ seeds	Alleles at *dt1* and *E1* loci were confirmed from DNAs extracted from three $F_{2:3}$ seeds[§]		
	Planted $F_{2:3}$ seeds in Costa Rica	Planted $F_{2:3}$ seeds in Costa Rica	Planted $F_{2:4}$ seeds in Costa Rica	Planted $F_{2:4}$ seeds in Costa Rica
In spring 2019	Harvested $F_{3:4}$ seeds by single plant thresh	Harvested $F_{3:4}$ seeds by single plant thresh	Harvested $F_{4:5}$ seeds by single seed thresh	Harvested $F_{4:5}$ seeds by single seed thresh
		Homozygosity of *dt1R62S* allele was confirmed from DNAs extracted from three $F_{3:4}$ seeds		
In summer 2019	Planted $F_{3:4}$ seeds in progeny rows for field experiment 19GA	Planted $F_{3:4}$ seeds in progeny rows for field experiment 19GA	Planted $F_{4:5}$ seeds in progeny rows for field experiment 19GA	Planted $F_{4:5}$ seeds in progeny rows for field experiment 19GA
	Morphological characteristics were measured from five randomly selected plants per experimental line	Morphological characteristics were measured from five randomly selected plants per experimental line	Morphological characteristics were measured from five randomly selected plants per experimental line	Morphological characteristics were measured from five randomly selected plants per experimental line
In winter 2019	Bulk harvested $F_{3:5}$ seeds	Bulk harvested $F_{3:5}$ seeds	Bulk harvested $F_{4:6}$ seeds	Bulk harvested $F_{4:6}$ seeds
In summer 2020	Planted $F_{3:5}$ seeds for the field experiment 20TN	Planted $F_{3:5}$ seeds for the field experiment 20TN	Planted $F_{4:6}$ seeds for the field experiment 20TN	Planted $F_{4:6}$ seeds for the field experiment 20TN
	Morphological characteristics were measured from ten randomly selected plants per plot	Morphological characteristics were measured from ten randomly selected plants per plot	Morphological characteristics were measured from ten randomly selected plants per plot	Morphological characteristics were measured from ten randomly selected plants per plot

[†] Donor parents of each of target alleles on stem termination type for each population

[‡] Donor parents of the functional *E1* allele for each population

[§] Genotypes having heterozygous *dt1R62S* with homozygous *E1* were selected, since there were no plants having homozygous *dt1R62S* and *E1*

Supplementary Table 3. Detail information of field experiments for evaluating morphological characteristics depending on genotype combination in three different latitudinal environments in 2019 and 2020

	Field Experiments			
	19GA	**20TN**	**19MO†**	**20MO**
Year	2019	2020	2019	2020
Location	University of Georgia Iron Horse Plant Science Farm, Watkinsville, Georgia	East Tennessee Agriculture and Education Center, Knoxville, Tennessee	South Farm Research Center, Columbia, Missouri	South Farm Research Center, Columbia, Missouri
Latitude	33.72˚N, -83.30˚W	35.96˚N, -83.86˚W	38.91˚N, -92.29˚W	38.91˚N, -92.29˚W
Experimental design	Completely randomized design	Randomized complete block design	Randomized complete block design	Randomized complete block design
Number of lines	21	20	2	14
Number of genotypes	7	7	1	5
Boarder rows	No	No	No	Yes
Number of rows per plot	2	2	2	3
Number of replications	none	3	3	3
Planting density	48 seed per row	200 seeds per row	50 seeds per row	50 seeds per row
Row length	6 ft (1.83 m)	16 ft (4.88 m)	7 ft (2.13 m)	7 ft (2.13 m)
Row spacing	30 in (76.2 cm)	30 in (76.2 cm)	30 in (76.2 cm)	30 in (76.2 cm)
Planting date	5th June, 2019	27th May, 2020	30th May, 2019	2nd June, 2020
Measurment parameters (abbr.; unit)				
Number of plants per plot	5	10	5	5
Plant height	Yes	Yes	Yes	Yes
Number of nodes	Yes	Yes	Yes	Yes
Stem diameter (SD; mm)	Yes	Yes	Yes	Yes
Lodging	Yes	Yes		Yes
Days to maturity (DTM)		Yes		Yes
Days to flowering (DTF)				Yes
Number of pods at stem tip (Pod; ea)		Yes	Yes	Yes
Raceme length (RL; cm)		Yes	Yes	Yes
Number of branches (Bran; ea)		Yes		Yes

† 19MO was a preliminary field trial of 20MO which was aimed to compare the differences between Clark and Williams 82

Supplementary Table 4. Mean values of other morphological characteristics for soybeans with different genotype combinations in field experiments at two different southern environments in 2019 and 2020 (19GA and 20TN)

Genotype	Gene			19GA			20TN				
	Dt1	*Dt2*	*E1*	n[†]	Pod[‡]	RL[§]	n[†]	Pod[‡]	RL[§]	DTM[¶]	Bran[#]
Dt1E1	*Dt1*	*dt2*	*E1*	1	1.2 ± 0.8 e	0.0 ± 0.0 c					
R166WE1	*dt1R166W*	*dt2*	*E1*	2	8.6 ± 2.7 a	3.1 ± 3.8 a	3	7.7 ± 3.1 ab	2.9 ± 1.5 a	140.1 ± 3.1 a	2.7 ± 1.2 a
R130KE1	*dt1R130K*	*dt2*	*E1*	8	6.4 ± 2.7 b	2.7 ± 1.0 a	2	6.9 ± 3.9 bc	1.8 ± 1.2 b	137.7 ± 1.3 b	2.8 ± 1.6 a
R62SE1	*dt1R62S*	*dt2*	*E1*	6	5.1 ± 3.2 bc	2.7 ± 1.3 a	4	8.7 ± 3.4 a	3.0 ± 1.4 a	140.1 ± 3.7 a	3.1 ± 1.4 a
Dt2E1	*Dt1*	*Dt2*	*E1*	2	4.1 ± 2.2 cd	2.4 ± 0.9 ab	8	6.2 ± 2.5 c	1.0 ± 0.6 c	138.5 ± 3.1 b	3.1 ± 1.5 a
Dt1e1	*Dt1*	*dt2*	*e1-as*	1	1.2 ± 0.8 e	2.7 ± 1.8 a	1	3.1 ± 1.7 e	0.2 ± 0.4 d	138.5 ± 0.5 b	3.1 ± 1.5 a
R166We1	*dt1R166W*	*dt2*	*e1-as*								
R130Ke1	*dt1R130K*	*dt2*	*e1-as*				1	4.6 ± 2.1 d	1.6 ± 0.7 b	131.3 ± 1.3 d	1.8 ± 1.8 b
R62Se1	*dt1R62S*	*dt2*	*e1-as*				1	5.7 ± 2.2 cd	1.7 ± 0.7 b	132.5 ± 0.5 c	1.3 ± 1.2 b
Dt2e1	*Dt1*	*Dt2*	*e1-as*	1	2.4 ± 1.5 de	1.4 ± 0.7 b					
			LSD (0.05)		2.0	1.2		1.4	0.5	0.9	0.7

[†] Number of lines within each of genotype categories
[‡] Number of pods at apical stem
[§] Lodging was scored at maturity on a scale of one (all plants erect) to five (almost all plants prostrate)
[#] Number of branches in the main stem

Supplementary Table 5. Mean values of agronomic traits measured from soybean cultivars, Clark and Williams 82, grown at Columbia, MO in 2019 (19MO)

Genotype category	Gene			n[†]	Agronomic traits					
	Dt1	*Dt2*	*E1*		Plant height	Node[‡]	SD[§]	DTF[¶]	DTM[#]	Pod[$]
Clark	*Dt1*	*dt2*	*e1-as*	14	90.1 ± 3.6	22.1 ± 1.9	6.0 ± 1.1	43.1 ± 1.5	120.4 ± 0.5	1.0 ± 0.0
Williams 82	*Dt1*	*dt2*	*e1-as*	15	89.7 ± 8.9	22.1 ± 2.5	5.3 ± 0.9	44.7 ± 3.4	122.0 ± 1.5	0.8 ± 0.4
				LSD (0.05)	0.88	0.99	0.06	0.12	0.00	0.08

[†] Number of plant samples per each line
[‡] Number of nodes in a main stem at maturity
[§] SD, mean of stem diameters (mm) measured at first, middle, and last internode at maturity
[¶] DTF, Days to Flowering
[#] DTM, Days to Maturity
[$] Number of pods at apical stem

VITA

Jeonghwa Kim was born on January 11, 1991 and raised in Daegu, South Korea. Jeonghwa graduated from Youngsong Girls' High School in 2009. She earned her bachelor's degree in applied bioscience from Kyungpook National University in Daegu, South Korea in 2014. During her undergraduate degree, she worked as an undergraduate research assistant for six months in 2012 in Dr. Felix Fritschi's lab at University of Missouri. In 2016, she earned her master's degree in applied biosciences with an emphasis of soybean breeding and genetics under supervision of Dr. Jeong-Dong Lee at Kyungpook National University. After completing her master's degree, she briefly worked as a research associate at The Agricultural Genome Center, National Institute of Agriculture Science, Rural Development Administration, in Jeonju, South Korea. In 2017, she joined Dr. Kristin Bilyeu's lab in the Division of Plant Sciences at University of Missouri to pursue a PhD with an emphasis of soybean genetics and genomics. Jeonghwa has accepted a post-doctoral position at North Dakota State University and will continue research in crop genomics.