



(51) International Patent Classification:

C12N 15/82 (2006.01) *A01H 4/00* (2006.01)
C12N 15/90 (2006.01) *C07K 14/415* (2006.01)
C12Q 1/6895 (2018.01) *C12N 15/64* (2006.01)

(21) International Application Number:

PCT/US2024/035993

(22) International Filing Date:

28 June 2024 (28.06.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/511,323 30 June 2023 (30.06.2023) US

(71) Applicant: **ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA** [US/US];
The University Of Arizona, Tech Launch Arizona, 1600 E. Idea Lane, Suite 110, Tucson, AZ 85713 (US).

(72) Inventors: **SCHMIDT, Monica**; The University Of Arizona, 1657 E Helen St., Tucson, AZ 85721 (US). **HER-**

MAN, Eliot, M.; The University Of Arizona, 1657 E Helen St., Tucson, AZ 85721 (US).

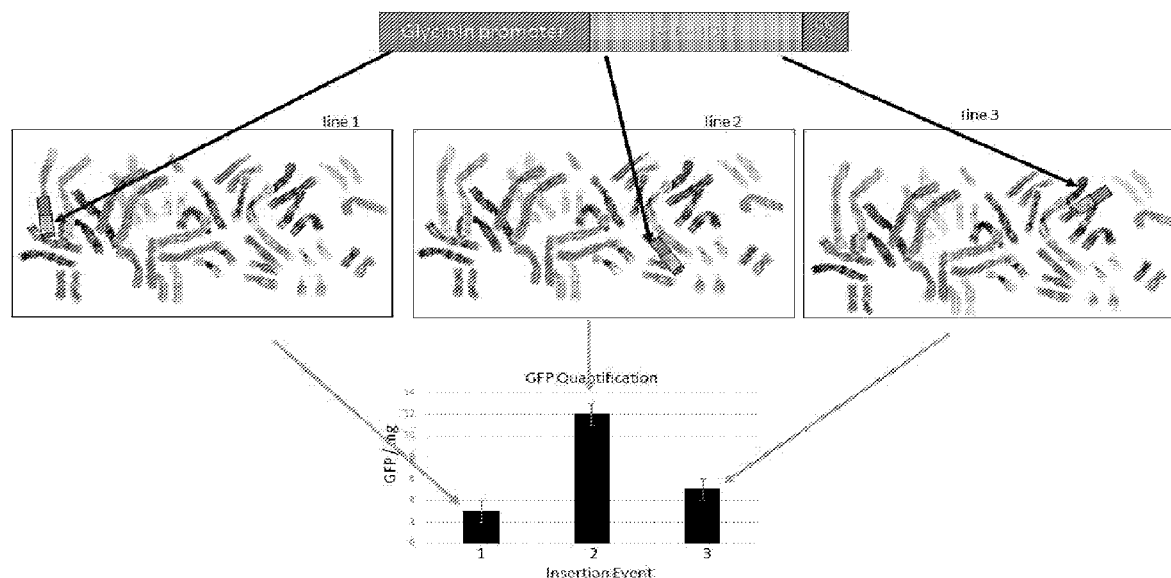
(74) Agent: **NGUYEN, Quan**; Nguyen Tarbet IP Law, 4199 Campus Drive, Suite 550, Irvine, CA 92612 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ,

(54) Title: METHODS FOR GENERATING HIGH PROTEIN SOYBEAN MUTANTS

FIG. 1



(57) Abstract: Soybean, one of the world's most significant protein and oil crops, has been enhanced through genetic engineering and mutagenesis to introduce or enhance agronomic traits. Both transgenic modification and mutagenesis have shown considerable variation in phenotypes observed in crop species. The effects of these methods on soybean seed composition and gene expression were determined by subjecting stable transgenic soybean lines to mutagenesis, generating 100 mutagenized lines with altered GFP expression, which were subsequently analyzed for protein and oil content.

RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS FOR GENERATING HIGH PROTEIN SOYBEAN MUTANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of U.S. Provisional Application No. 63/511,323 filed June 30, 2023, the specification of which is incorporated herein in their entirety by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under Grant No. 2014-33522-22531 awarded by USDA/NIFA. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention features transgenic legume plants, such as soybeans, with altered seed compositions, including high protein or oil content. Additionally, the present invention encompasses methods for generating these transgenic legume seeds with altered seed compositions.

BACKGROUND OF THE INVENTION

[0004] Soybean (*Glycine max* (L.) Merr.) is a global commodity as a source of edible oil and protein meal that make up 18% and 38% of its seeds, respectively. The success of soybean cultivation, and agriculture in general, lies in the variation that exists or is introduced in plant genomes, with the desired genomic variants being successively selected through generations. Mutagenesis and transformation are currently the two primary mechanisms to introduce genetic variation into plant genomes. Although these two techniques have been used for decades, there remains uncertainty about the unintended consequences.

[0005] Selecting crop mutants with desired phenotypes has occurred since civilization's dawn. The construction of a genetic linkage map provides helpful information to breed important agronomic traits such as resistance to cyst nematodes, seed weight, oil content, and protein content. To enable the introduction of additional genetic diversity, plant breeders/geneticists use induced mutagenesis via exposing plants to chemicals or radiation to increase the rate of mutations. The chemical mutagen ethyl methanesulfonate (EMS) is often used to produce mutant populations of crops, including maize, barley, wheat, and soybean. For soybeans, a 40 mM EMS solution treatment of seeds results in a mutant rate between 1/140 kb to 1/550 kb. As a mechanism, EMS induces a single nucleotide polymorphism, specifically a G/C to A/T transition, and in soybeans has been shown to result in a mutation population distribution of 33-45% missense (resulting in a different amino acid in a protein), 51-58% silent (no phenotype at amino acid level) and 4-8% truncation (stopped the translation of the protein early) mutations. Chemically induced mutants have aided in identifying soybean genes functional in nodulation,

pathogen susceptibility, decreased raffinose, and altered oil composition. However, inherent to this genetic variation technique is that chemical mutagens introduce genetic alterations in multiple genome locations simultaneously. As a consequence, without a simple genome mapping procedure, it is difficult to determine which of the many mutations results in an observed phenotype. A parallel concern is that a selected mutant line with a desired trait may also carry other less desirable collateral mutations.

[0006] The ability to integrate foreign genes into plants is now a routine procedure that enables functional genomics and biotechnology. However, little is currently known about how introduced DNA is taken up or inserted into a plant nucleus and its genome to be stably incorporated. For any transgenic event, there are a large number of potential genetic factors that may be considered in the evaluation of the transgenic safety/risk as well as the effectiveness of altering the output trait. Factors to consider in an engineered transgenic plant include the number and position(s) of the inserted gene and how these variables affect the output trait. From the perspective of safety and regulation, how/if gene insertion can introduce unintended collateral effects remains a largely unknown variable that is difficult to quantitate from the perspective of risk. Collateral unintended consequences are a prominent unknown factor that is among the most often cited concerns by critics of biotechnology. Gene expression can differ between independent insertion events, that is, different transgenic lines when the same inserted cassette is engineered into plants, a result referred to as the 'position effect.' The position effect is presumed to result from the differential integration of the inserted cassette into the plant genome. With respect to the capacity or susceptibility of areas of the genome where transgenes are inserted, some are quiescent (i.e., heterochromatin), while other genome domains contain structural or regulatory elements such as transcriptional enhancers or inhibitors that may affect the newly integrated gene's expression. Position effect is a problematic variable in transgenomics, so numerous techniques have been developed to minimize or circumvent this variability, namely site-specific integration techniques like Cre/lox, Tale effectors, and Zinc finger endonucleases, and insulating elements such as Matrix Attachment Regions. Although all of these mechanisms have added to the reproducibility of transgene expression, there is still much uncertainty and variability between transgenic lines.

[0007] An additional significant variable of transgene expression is copy number. A general trend in producing transgenics is that the more copies of the transgene within the plant genome, the lesser the summed expression of the constituent incorporated transgenes. This phenomenon is referred to as homology-dependent gene silencing (HDGS), and it has been observed to arise when the insertion consists of tandem and inverted repeats. Curiously, this effect is the exact opposite of the consequences of polyploidy, whether in seeds or fruit flies,

where endoduplication results in enhanced expression and accumulation of certain intrinsic gene products, including seed storage proteins. This likely reflects an underlying mechanism that the semi-random insertion of multiple copies of single genes in the genome is not functionally equivalent to endoduplication of the entire genome and how it confers enhanced output traits. The nature of the mechanism(s) of the multiple gene insertion fratricides remains largely uncharacterized and may have simple mechanisms such as inefficient competition for nuclear regulatory proteins or may have complex genome mechanisms that result from parallel and independent regulation of multiple copies of a gene in varying, and perhaps inappropriate, genome locations. The observation of multiple gene fratricide is not ubiquitous, and there is a body of observational data that also reports no relationship between copy number and gene expression, while in contrast, other investigations have reported a strong correlation between multiple copies and diminished trait output. For each inserted gene, there are additional regulatory elements that not only are positively and negatively controlled by the diverse circumstances of the plant's metabolism but may also interact by cooperative feedback and control mechanisms where the regulation of one gene exerts some control over many others. Together, the interaction of regulatory controls and their hierarchy(ies) result in a systems, or game type, model where many decision tree elements ultimately control the overt manifestation of the final output trait. From an experimental perspective, the challenge is to design *in situ* probes that can dissect out the variations of traits and how this interacts with basic genetic elements such as gene position, copy number, and the effects of distal mutation and/or variations in other interactive genes.

BRIEF SUMMARY OF THE INVENTION

[0008] It is an objective of the present invention to provide compositions and methods that allow for high protein soybean mutants, as specified in the independent claims. Embodiments of the invention are given in the dependent claims. Embodiments of the present invention can be freely combined with each other if they are not mutually exclusive.

[0009] The present invention utilizes a mutation platform comprising a stable transgenic soybean line expressing an inserted seed-specific detectable marker, e.g., comprising a detectable reporter operatively linked to a seed-specific protein promoter. Without wishing to limit the present invention to any theory or mechanism, it is believed that the inserted seed-specific detectable marker allows for quick visual screening of alterations in the inserted detectable reporter (e.g., green fluorescent protein, GFP), which functions as a proxy for soybean seed protein, driven by the GLY1 promoter (one of the major seed storage proteins, glycinin). Seeds from this GFP transgenic line are subsequently mutagenized by fast neutron mutagenesis (FMS), which has the distinct advantage of producing large deletions (mostly 1-4

kb) that are easier to locate compared to single point mutations produced by chemical mutagenesis. Thousands of seeds may be quickly and effectively screened for altered production of the GFP marker protein, with changes in GFP indicating that the seed protein content of the mutagenized seed is affected.

[0010] In some embodiments, the present invention features, a transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation (e.g., via mutagenesis, transformation, or a combination thereof) to a transgenic legume seed comprising a seed-specific detectable marker. In some embodiments, the transgenic legume seed composition comprises an increased amount of protein compared to a legume seed without genetic variation. In other embodiments, the transgenic legume seed composition comprises an increased amount of oil compared to a legume seed without genetic variation. In further embodiments, the transgenic legume seed composition comprises an increase in carbohydrate content compared to the legume seed without genetic variation. In some embodiments, the legume is selected from a group consisting of alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans.

[0011] In other embodiments, the present invention may feature a transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation (e.g., via mutagenesis, transformation, or a combination thereof) to a transgenic soybean seed comprising a seed-specific detectable marker. In some embodiments, the transgenic soybean seed composition comprises an increased amount of protein compared to a soybean seed without genetic variation. In other embodiments, the transgenic soybean seed composition comprises an increased amount of oil compared to a soybean seed without genetic variation. In further embodiments, the transgenic soybean seed composition comprises an increase in carbohydrate content compared to the soybean seed without genetic variation.

[0012] In certain embodiments, the present invention may feature a transgenic legume (e.g., a soybean) comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising at least two seed-specific detectable markers. In such embodiments, the transgenic legume seed composition may comprise an increased amount of protein and carbohydrate content compared to a legume seed without genetic variation. In other embodiments, the transgenic legume seed composition may comprise an increased amount of oil and carbohydrate content compared to a legume seed without genetic variation.

[0013] The seed-specific detectable marker as used herein may comprise a detectable reporter operatively linked to a seed-specific promoter. In some embodiments, the seed-specific

detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter (e.g., a seed-specific protein storage promoter). In other embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter. In further embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific carbohydrate promoter. In certain embodiments, the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker. Alternatively, or in addition to, in some embodiments, the seed-specific detectable marker further comprises an ER-retention signal (e.g., a HDEL sequence) operatively linked to the 3' end of the detectable marker.

[0014] One of the unique and inventive technical features of the present invention is the implementation of a novel seed-specific detectable marker in a transgenic model, which is then mutagenized to effectively identify alterations in the expression of the mutant GFP marker gene9 (e.g., the detectable reporter). Without wishing to limit the invention to any theory or mechanism, it is believed that this technical feature advantageously enables the identification of genes crucial for storage protein seed production and regulation. None of the presently known prior references or work has the unique inventive technical feature of the present invention.

[0015] The present invention combines established methods of inducing genetic variation, such as mutagenesis and transformation, to develop soybean seeds with higher protein content. The integration of transgenic and mutagenesis techniques is a powerful approach. A particularly unique aspect is the use of a seed-specific marker transgene (e.g., a seed-specific detectable marker). Without being bound to any specific theory or mechanism, it is believed that employing a reporter gene (e.g., seed-specific detectable marker) made from GFP, driven by the strongest seed protein promoter, enabled efficient and rapid screening of mutant transgenic seeds with altered protein (and inversely, oil) content due to the mutagenesis process. Any genomic change in the soybean seed affecting this promoter will result in varying levels of GFP expression. Given the thousands of unknown factors that control seed protein, mutating this population, for instance, with neutrons or chemicals, creates a screenable population where increased GFP serves as a proxy for increased protein content.

[0016] In some embodiments, the methods described herein may comprise manual or robotic selection. Moreover, the methods described herein are adaptable to other traits, such as looking for higher oil or carbohydrates in other seed plants.

[0017] Any feature or combination of features described herein are included within the scope of the present invention provided that the features included in any such combination are not

mutually inconsistent as will be apparent from the context, this specification, and the knowledge of one of ordinary skills in the art. Additional advantages and aspects of the present invention are apparent in the following detailed description and claims.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING(S)

[0018] The features and advantages of the present invention will become apparent from a consideration of the following detailed description presented in connection with the accompanying drawings in which:

[0019] FIG. 1 shows an outline of a transgenomics experiment to assay the variability in the genome of soybean seed composition and transgene expression using an ER-targeted GFP transgene.

[0020] FIG. 2 shows a mutagenesis approach to assay the variability in soybean seed composition

[0021] FIG. 3 shows an overview of transgenomics and mutagenesis effect on soybean seed composition and transgene expression variability.

[0022] FIG. 4 shows GFP as a marker protein to show the suppression of freed-up reserves when beta conglycinin is suppressed, and GFP expressed under glycinin regulatory elements; CS, beta conglycinin suppressed; GFP, green fluorescent protein, circle notes proteins identified by both western blot and Mass Spectrometry (MS) analysis to be GFP. Note GFP visual increase under UV light in GFPx CS seed.

[0023] FIG. 5A and 5B show GFP-HDEL soybean seed proteome separated by 2D gel electrophoresis. FIG. 5A shows the first dimension is a non-linear 3-10 pH range isoelectric point gel, and the second phase is an 8-16% polyacrylamide gel to separate by molecular weight. Circles are the GFP proteins as determined by western blot analysis (shown in FIG. 5B) of a replicate gel using a monoclonal anti-GFP antibody.

DETAILED DESCRIPTION OF THE INVENTION

[0024] For purposes of summarizing the disclosure, certain aspects, advantages, and novel features of the disclosure are described herein. It is to be understood that not necessarily all such advantages may be achieved in accordance with any particular embodiments of the disclosure. Thus, the disclosure may be embodied or carried out in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other advantages as may be taught or suggested herein.

[0025] Additionally, although embodiments of the disclosure have been described in detail, certain variations and modifications will be apparent to those skilled in the art, including embodiments that do not provide all the features and benefits described herein. It will be

understood by those skilled in the art that the present disclosure extends beyond the specifically disclosed embodiments to other alternative or additional embodiments and/or uses and obvious modifications and equivalents thereof. Moreover, while a number of variations have been shown and described in varying detail, other modifications, which are within the scope of the present disclosure, will be readily apparent to those of skill in the art based upon this disclosure. It is also contemplated that various combinations or sub-combinations of the specific features and aspects of the embodiments may be made and still fall within the scope of the present disclosure. Accordingly, it should be understood that various features and aspects of the disclosed embodiments can be combined with or substituted for one another in order to form varying modes of the present disclosure. Thus, it is intended that the scope of the present disclosure herein disclosed should not be limited by the particular disclosed embodiments described herein.

[0026] As used herein, the singular forms “a,” “an,” and “the” are intended to include the plural forms as well unless the context clearly indicates otherwise. Furthermore, to the extent that the terms “including,” “includes,” “having,” “has,” “with,” or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising.”

[0027] Suitable methods and materials for the practice and/or testing of embodiments of the disclosure are described below. Such methods and materials are illustrative only and are not intended to be limiting. Other methods and materials similar or equivalent to those described herein can be used. For example, conventional techniques well known in the art to which the disclosure pertains are described in various general and more specific references, including, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, 1989; Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3d ed., Cold Spring Harbor Press, 2001; Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates, 1992 (and Supplements to 2000); Ausubel et al., *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 4th ed., Wiley & Sons, 1999; Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1990; and Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1999, the disclosures of which are incorporated in their entirety by reference herein.

[0028] All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. Although methods and materials similar or equivalent to those described herein can be used to practice or test the disclosed technology, suitable methods and materials are described below. The materials, methods, and examples are

illustrative only and not intended to be limiting.

[0029] The term "vector or construct" may refer to any nucleic acid that acts as a carrier for other (e.g., foreign) nucleic acid sequences that are not native to the vector. When introduced into an appropriate host cell, a vector may replicate itself (and, thereby, the foreign nucleic acid sequence) or express at least a portion of the foreign nucleic acid sequence. In one context, a vector is a linear or circular nucleic acid into which a nucleic acid sequence of interest is introduced (for example, cloned) for replication (e.g., production) and/or manipulation using standard recombinant nucleic acid techniques (e.g., restriction digestion). A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements known in the art. Typical vectors include, for example, plasmids, cosmids, phage, phagemids, artificial chromosomes (e.g., BAC, PAC, HAC, YAC), and hybrids that incorporate features of more than one of these types of vectors. Typically, a vector includes one or more unique restriction sites (and, in some cases, a multi-cloning site) to facilitate the insertion of a target nucleic acid sequence.

[0030] The term "transgenic plant" refers to any plant whose DNA has been modified using genetic engineering techniques, well known in the art, to introduce a new trait not naturally occurring in the species. This includes plants whose genomes have been altered by the stable integration of recombinant DNA, as well as plants regenerated from originally transformed plant cells and progeny transgenic plants from later generations or crosses of a transformed plant.

[0031] As used herein, "recombinant DNA" refers to DNA that has been genetically engineered and constructed outside of a cell. This includes DNA containing naturally occurring DNA, cDNA, or synthetic DNA.

[0032] As used herein, "operably linked" refers to the association of two or more DNA fragments in a DNA construct (e.g., an expression cassette), such that the function of one fragment (e.g., protein-encoding DNA) is regulated by the other fragment (e.g., a promoter)

[0033] As used herein, "expressed" refers to the production process in which a protein is synthesized in a plant cell. This occurs when its corresponding DNA is transcribed into mRNA, which is then translated into the protein.

[0034] Referring now to FIGs. 1-5B, the present invention features methods of producing soybean mutants comprising a soybean seed with a high protein and/or oil content as well as the soybean mutants generated.

[0035] The present invention may feature a transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation (e.g., via mutagenesis, transformation, or a combination thereof) to a transgenic legume seed comprising a seed-specific detectable marker. In some embodiments, the transgenic legume seed composition comprises an increased amount of protein compared to a legume seed without genetic variation. In other embodiments, the transgenic legume seed composition comprises an increased amount of oil compared to a legume seed without genetic variation. In further embodiments, the transgenic legume seed composition comprises an increase in carbohydrate content compared to the legume seed without genetic variation. Legumes may include but are not limited to alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans. However, the present invention is not restricted to legumes and can encompass any crop with its associated high-value trait, provided the appropriate transgene/promoter combination is utilized.

[0036] In some embodiments, the present invention may feature a transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation (e.g., via mutagenesis, transformation, or a combination thereof) to a transgenic soybean seed comprising a seed-specific detectable marker. In some embodiments, the transgenic soybean seed composition comprises an increased amount of protein compared to a soybean seed without genetic variation. In other embodiments, the transgenic soybean seed composition comprises an increased amount of oil compared to a soybean seed without genetic variation. In further embodiments, the transgenic soybean seed composition comprises an increase in carbohydrate content compared to the soybean seed without genetic variation.

[0037] The seed-specific detectable marker as used herein may comprise a detectable reporter operatively linked to a seed-specific promoter. In some embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter (e.g., a seed-specific protein storage promoter; e.g., the detectable reporter is driven by the seed-specific protein promoter). In other embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter (e.g., e.g., the detectable reporter is driven by the seed-specific oil promoter). In further embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific carbohydrate promoter (e.g., the detectable reporter is driven by the seed-specific carbohydrate promoter). In certain embodiments, the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker. Alternatively, or in addition to, in some embodiments, the seed-specific detectable marker further comprises an ER-retention signal (e.g., a HDEL sequence) operatively linked to the 3' end of the

detectable marker. In some embodiments, the ER-retention signal comprises the sequence HDEL.

[0038] The present invention may also feature a transgenic legume with an altered seed composition produced by introducing genetic variation to a transgenic legume seed comprising at least two seed-specific detectable markers. In some embodiments, the transgenic legume seed composition comprises an increased amount of protein and carbohydrate content compared to a legume seed without genetic variation. In other embodiments, the transgenic legume seed composition comprises an increased amount of oil and carbohydrate content compared to a legume seed without genetic variation. In the aforementioned embodiment of the present invention, the transgenic legume seed may comprise a first seed-specific detectable marker and a second seed-specific detectable marker. In some embodiments, the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific protein promoter, and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter. In other embodiments, the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific oil promoter, and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter. Legumes may include but are not limited to alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans. However, the present invention is not restricted to legumes and can encompass any crop with its associated high-value traits, provided the appropriate transgene/promoter combination is utilized.

[0039] In some embodiments, the present invention may also feature a transgenic soybean with an altered seed composition produced by introducing genetic variation to a transgenic soybean seed comprising at least two seed-specific detectable markers. In some embodiments, the transgenic soybean seed composition comprises an increased amount of protein and carbohydrate content compared to a soybean seed without genetic variation. In other embodiments, the transgenic soybean seed composition comprises an increased amount of oil and carbohydrate content compared to a soybean seed without genetic variation. In the aforementioned embodiment of the present invention, the transgenic soybean seed may comprise a first seed-specific detectable marker and a second seed-specific detectable marker. In some embodiments, the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific protein promoter, and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter. In other embodiments, the first seed-specific detectable marker

comprises a first detectable reporter operatively linked to a seed-specific oil promoter, and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter.

[0040] The first seed-specific detectable marker or the second seed-specific detectable marker, as described above, may further comprise an ER-signal sequence operatively linked to the 5' end of the detectable marker. In certain embodiments, the first seed-specific detectable marker or the second seed-specific detectable marker, as described above, may further comprise an ER-retention signal operatively linked to the 3' end of the detectable marker. In some embodiments, the ER-retention signal comprises the sequence HDEL.

[0041] In some embodiments, the detectable reporter comprises a fluorescent reporter. Non-limiting examples of fluorescent reporters include yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), or green fluorescent proteins (GFP). In some embodiments, the detectable reporter comprises a light-emitting luciferase protein. The present invention is not limited to the aforementioned detectable reports and may also include any marker that can be detected and quantified through an appropriate assay.

[0042] In some embodiments, genetic variation may be introduced via mutagenesis, transformation, or a combination thereof of both methods. Mutagenesis methods may include insertional techniques (such as transposons or T-DNA), chemical methods, as well as particle (neutron) and photon (X-ray) irradiation. For example, genetic variation may be introduced by fast neutron mutagenesis or by chemical mutagenesis. Without restricting the present invention to any specific theory or mechanism, it is believed that fast neutron mutagenesis, with its average deletion size of 1-4 kb, may facilitate easier downstream detection compared to the single base-pair mutations typically induced by chemical mutagenesis.

[0043] The present invention may further feature a method to determine genes regulating legume seed composition. In some embodiments, the method comprises transforming a legume seed with a seed-specific detectable marker and subsequently introducing genetic variation to the transgenic legume seed (e.g., via mutagenesis). The method may further comprise measuring a signal from the seed-specific detectable marker (i.e., the detectable reporter) to determine the legume seed composition. In some embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter. For example, the seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific protein promoter (e.g., a seed-specific protein storage promoter). Or, alternatively, the seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific oil promoter. In certain embodiments, the seed-specific

detectable marker may comprise a detectable reporter operatively linked to a seed-specific carbohydrate promoter.

[0044] In other embodiments, the present invention may further feature a method to determine genes regulating legume seed composition. In some embodiments, the method comprises transforming a legume seed with two seed-specific detectable markers (e.g., a first seed-specific detectable marker and a second seed-specific detectable marker), and subsequently introducing genetic variation to the transgenic legume seed (e.g., via mutagenesis). The method may further comprise measuring a signal from the seed-specific detectable markers (i.e., the detectable reporters) to determine the legume seed composition. In some embodiments, the first seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific protein promoter (e.g., a seed-specific protein storage promoter), and the second seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific carbohydrate promoter. In other embodiments, the first seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific oil promoter, and the second seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific carbohydrate promoter.

[0045] In certain embodiments, the method comprises introducing genetic variation to a transgenic legume seed (e.g., via mutagenesis) comprising a seed-specific detectable marker (e.g., comprising a detectable reporter operatively linked to a seed-specific promoter) and measuring a signal from the detectable marker (i.e., the detectable reporter) to determine the legume seed composition. In other embodiments, the method comprises introducing genetic variation to a transgenic legume seed (e.g., via mutagenesis) comprising at least two seed-specific detectable markers (e.g., a first seed-specific detectable marker and a second seed-specific detectable marker) and measuring a signal from the first detectable marker (i.e., the first detectable reporter) and the second detectable marker (i.e., the second detectable reporter) to determine the legume seed composition.

[0046] In some embodiments, an increase in a signal from the detectable reporter indicates the transgenic legume seed composition comprises an increased amount of protein and a decrease in a signal from the detectable reporter indicates the transgenic legume seed composition comprises a decreased amount of protein. In other embodiments, an increase in a signal from the detectable reporter indicates the transgenic legume seed composition comprises an increased amount of oil and a decrease in a signal from the detectable reporter indicates the transgenic legume seed composition comprises a decreased amount of oil. In further embodiment, an increase in a signal from the detectable reporter indicates the transgenic legume seed composition comprises an increased amount of carbohydrate content and a

decrease in a signal from the detectable reporter indicates the transgenic legume seed composition comprises a decreased amount of carbohydrate content.

[0047] In some embodiments, the present invention may further feature a method to determine genes regulating soybean seed composition. In some embodiments, the method comprises transforming a soybean seed with a seed-specific detectable marker and subsequently introducing genetic variation to the transgenic soybean seed (e.g., via mutagenesis). The method may further comprise measuring a signal from the seed-specific detectable marker (i.e., the detectable reporter) to determine the soybean seed composition. In some embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter. For example, the seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific protein promoter (e.g., a seed-specific protein storage promoter). Or, alternatively, the seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific oil promoter. In certain embodiments, the seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific carbohydrate promoter.

[0048] In other embodiments, the method comprises transforming a soybean seed with two seed-specific detectable markers (e.g., a first seed-specific detectable marker and a second seed-specific detectable marker), and subsequently introducing genetic variation to the transgenic soybean seed (e.g., via mutagenesis). The method may further comprise measuring a signal from the seed-specific detectable markers (i.e., the detectable reporters) to determine the soybean seed composition. In some embodiments, the first seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific protein promoter (e.g., a seed-specific protein storage promoter), and the second seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific carbohydrate promoter. In other embodiments, the first seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific oil promoter, and the second seed-specific detectable marker may comprise a detectable reporter operatively linked to a seed-specific carbohydrate promoter.

[0049] In certain embodiments, the method comprises introducing genetic variation to a transgenic soybean seed (e.g., via mutagenesis) comprising a seed-specific detectable marker (e.g., comprising a detectable reporter operatively linked to a seed-specific promoter) and measuring a signal from the detectable marker (i.e., the detectable reporter) to determine the soybean seed composition. In other embodiments, the method comprises introducing genetic variation to a transgenic soybean seed (e.g., via mutagenesis) comprising at least two seed-specific detectable markers (e.g., a first seed-specific detectable marker and a second

seed-specific detectable marker) and measuring a signal from the first detectable marker (i.e., the first detectable reporter) and the second detectable marker (i.e., the second detectable reporter) to determine the soybean seed composition.

[0050] In some embodiments, an increase in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises an increased amount of protein and a decrease in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises a decreased amount of protein. In other embodiments, an increase in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises an increased amount of oil and a decrease in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises a decreased amount of oil. In further embodiments, an increase in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises an increased amount of carbohydrate content and a decrease in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises a decreased amount of carbohydrate content.

[0051] Without wishing to limit the present invention to any theory or mechanism, it is believed that the combination of using a seed-specific detectable marker (e.g., comprising a detectable reporter, e.g., GFP) as a proxy for soybean seed content and then mutagenizing the seeds uniquely enables a quick and efficient screening method of mutant seeds for altered protein/oil content.

[0052] In certain embodiments, the present invention features a method to determine genes regulating legume seed composition. The method may comprise introducing genetic variation to a transgenic legume seed via mutagenesis. In some embodiments, the transgenic legume seed comprises a seed-specific detectable marker comprising a detectable reporter (e.g., GFP) driven by a major storage protein promoter GLY1.

[0053] In some embodiments, the detectable reporter (e.g., GFP reporter) is driven by a seed protein promoter, and an increase in the GFP signal indicates an increase in the amount of protein in the legume seed composition, and a decrease in the GFP signal indicates a decrease in the amount of protein in the legume seed composition.

[0054] In other embodiments, the detectable reporter (e.g., GFP reporter) is driven by a seed oil promoter, and an increase in the GFP signal indicates an increase in the amount of oil in the legume seed composition, and a decrease in the GFP signal indicates a decrease in the amount of oil in the legume seed composition.

[0055] Non-limiting examples of promoters that may be utilized include those from the soybean

storage protein families glycinin and beta-conglycinin, which serve as markers for protein regulation. Additionally, considering that oil content typically shows an inverse relationship with protein content, promoters associated with oil production, such as Oleosin, BCCP (biotin carboxyl carrier protein), and beta-ketoacyl-acyl carrier protein (a small gene family), may also be utilized.

[0056] The present invention features methods and compositions for increasing seed compositions (e.g., proteins, oil, carbohydrates, or a combination thereof) in legumes. Legumes may include but are not limited to alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans.

[0057] The present invention may further feature an expression cassette comprising a seed-specific detectable marker comprising a detectable reporter operatively linked to a seed-specific promoter. In some embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter. In other embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter. In further embodiments, the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific carbohydrate promoter.

[0058] **EXAMPLE 1**

[0059] The following is a non-limiting example of the present invention. It is to be understood that said example is not intended to limit the present invention in any way. Equivalents or substitutes are within the scope of the present invention.

[0060] Worldwide, soybeans are the most important source of edible vegetable oil and high-quality protein. Approximately 25% of the world's edible oils and 75% of the world's protein meal are from soybeans. Oil and protein are then understandably the major economic products from soybean seed. Protein meal largely goes to animal feed. By 2050, estimates are 9 billion people on the planet, and a need to increase animal feed by almost 250%. Oil has several uses, including cooking, cosmetics, industrial, and biofuels. As the source of 79% of edible oils, soybeans play a significant role in the American diet. One in every four deaths in the US is due to heart disease. Coronary heart disease is the most common type of heart disease, and it is the leading cause of death for Americans. Coronary heart disease costs the US nearly \$109 billion/year, considering the health care costs, medications, and loss of productivity. Foods manufactured with a healthier soybean oil will contribute to the reduction of the incidence of coronary heart disease in the US. Soybean improvements are sought by the stakeholder industry comprising the "Better Bean" initiative that includes conventional breeding as well as transgenic and mutational approaches. As a crop that can be manipulated, soybean has intrinsic

characteristics of a lack of related species in its primary production sites in North and South America coupled with an enclosed self-fertilizing crop there is essentially no pollen flow even to adjacent plants. These characteristics make soybeans a highly desirable target for improvement and deployment. The lack of relatives and pollen flow limit some of the common risk concerns associated with many other crops. Collateral effects and potential variation of output trait remain as a variable for both transgenic and mutational approaches to crop improvement. The research outlined in this proposal will directly evaluate these remaining variables.

[0061] All crop improvement efforts hinge on genetic variation existing with plant genomes. This genetic variation is ultimately derived from one of two sources: mutations or transgenics. The present invention investigates to what extent unintended genetic variations occur during these two processes in the two most important seed traits in soybeans: protein and oil.

[0062] A seed-storage protein proxy marker, GFP (HDEL), regulated by the glycinin promoter, was generated as an experimental model to examine the effects of transgene insertion and copy number and the consequences of mutation on a protein-storage output trait. To construct the marker trait, the open reading frame (ORF) of GFP was modified to include an N-terminal domain endoplasmic reticulum (ER)-signal sequence and a C-terminal HDEL sequence to both direct the protein to be synthesized by the ER and sequestered in the ER lumen. The HDEL sequence retards the exit of the protein from the ER, leveraging the intrinsic mechanism of retarding the progression of ER lumen proteins from the ER by its retrieval by the cis-Golgi ER-retention receptor. Retarding proteins within the ER lumen often promote their accretion and, as a consequence, the formation of ER-bodies that are analogous to the protein bodies formed by many types of seeds, especially monocots. GFP-HDEL as a glycinin mimic is similarly synthesized during mid-late seed maturation; it accumulates as a normalized endpoint quantity of protein, showing the transgene is incorporated into the seed's genetic program as another glycinin allele. The GFP-HDEL accretes to form ER-bodies that accumulate in the seed, so at the endpoint of the dry seed, the total GFP content of the seed represents the integrated accumulation throughout the development of the transgene GFP-glycinin-mimic allele without any complicating factor of protein turnover. That the GFP is synthesized and accumulated as if it was a glycinin family member enables its use as a marker to test how a seed protein trait is impacted by its encoding gene's position in the genome and by the effects of induced mutation at other places in the genome distal to the GFP gene's location.

[0063] For soybeans, economic value is accrued from their protein and oil content. Thus, there is increasing effort to use both induced mutation collections and transgenic traits to produce value-added traits needed to make future crops for feeding and expanding animal production. Soybeans are one of the key global commodities of the animal feed industry and the crop that

provides a primary input of reduced nitrogen. To employ the dual strategies of mutation and transgenics effectively, how the important output traits, protein in soybean, for instance, varies within a single potential breeding line was evaluated in response to distal mutations and in response to transgene placement within the genome.

[0064] This model GFP-HDEL biotechnology trait is easily standardized, observed, and quantified, and, as a seed trait, it is focused on one of the primary agricultural biotechnology goals: improving food, feed, and fuel. By using the GFP accumulation as a model marker, seed traits statistically valid set of lines were compared for protein accumulation as a proxy for general seed trait improvements, and the insertion site of each of the transgenes either as a single copy or for multiple copy transgenics were mapped with the reference genome. A library of mutants was created by mutating single-copy GFP accumulating lines with EMS.

[0065] To evaluate the potential variations of a soybean transgene-produced protein output trait, the effects of position, copy number, and distal mutation were tested.

[0066] To assess the position effect on the transgene-encoded protein output trait, a minimum of 100 distinct lines of glycinin-promoter/terminator-regulated GFP-HDEL were produced. These lines encompass the variations produced by the biolistic transformation approach that is employed by academic and industry projects. Each line produced will be regenerated to homozygosity, and the overt output trait will be evaluated by quantification of the GFP. Within each line, the individual seed GFP accumulation as a standard output can be measured by fluorometry with small error bars averaging with other seeds within a given population produced by an individual plant. For all 100 lines produced, the average GFP glycinin-proxy accumulation was assessed and compared to another 100 line population produced via mutagenesis, and an output trait distribution curve was constructed encompassing the variations resulting from the insertion site. The copy number and the insertion site for each transgenic line was determined with the insertion site mapped against the reference soybean genome. The resulting data will evaluate how copy number and position yield variation of the seed protein/oil output traits. Transgenic experiment is depicted in FIG 1.

[0067] To examine the effects of distal mutations on the transgene's protein/oil output traits, seeds from a single stable transgenic soybean line producing GFP were mutated by EMS, and a population of mutated soybeans was selected. These mutant soybeans selected were into individual lines, and each line was assessed for changes in seed GFP accumulation in comparison to the parental reference line. One hundred mutant lines displaying altered seed GFP expression were used as the seed-enhanced mutagenized population and used to analyze the variation of the seed protein/oil output traits. The mutagenic experiment is depicted in FIG. 2.

[0068] The resulting datasets provide a systematic assessment of the range that proteins can be produced in soybean seeds and how the background genetics affects the output trait (FIG. 3).

[0069] **EXAMPLE 2**

[0070] The following is a non-limiting example of the present invention. It is to be understood that said example is not intended to limit the present invention in any way. Equivalents or substitutes are within the scope of the present invention.

[0071] As soybean seeds are nearly 40% protein, they are an ideal platform to produce introduced proteins, including enhanced food/feed and industrial protein biologics. Yet if proteins are merely introduced by seed-expression in seeds, the yield is typically low (<1%) and even then only if the protein is not degraded in late maturation and seed desiccation. Heterologous protein production can be enhanced by producing 8% of the soybean seed proteome as a GFP-HDEL proxy of glycinin by exploiting the plasticity of the seed proteome where one storage protein, glycinin, can replace a shortage of the other major storage protein conglycinin (FIG. 4).

[0072] Two strategies were developed and exploited to enhance heterologous protein yield that when combined resulted in the enhanced accumulation of foreign protein in soybean seeds. By stabilizing the protein by both shuttling and retaining the protein into the endoplasmic reticulum (ER) with the carboxy-terminal HDEL retention sequence and, in parallel, suppressing conglycinin one of soybean's storage proteins to free up nutrient flux to support enhanced heterologous protein yield (FIG. 4). The detail of this engineering strategy was determining that when the beta-conglycinin storage protein in soybean seed was suppressed, either by RNAi or co-suppression technology, the glycinin storage protein was overexpressed as a compensation mechanism that quantitatively replaced the conglycinin shortfall. Of the glycinin produced in compensation for the conglycinin shortage, a fraction did not progress from the ER to the vacuole but instead accreted into ER-bodies, essentially an analogue to the cereal ER-derived protein bodies. Of particular note, the sum of the abundance of glycinin in the vacuole plus the abundance of proglycinin retained in ER-bodies compensated for the conglycinin shortage, resulting in a seed with the standard protein content but with a proteome dominated by glycinin as the sole major storage protein. A glycinin allele mimic was engineered by constructing an expression cassette of an enhanced green fluorescent protein targeted to the ER by a 5' ER signal tag and a 3' HDEL retention tag controlled by glycinin regulatory elements. When this construct was introgressed into the conglycinin silenced soybean the additional glycinin allele in the form of a GFP-HDEL mimic participated in the glycinin compensation of the conglycinin shortfall, resulting in an eight-fold increase of GFP accumulation compared to the same construct's expression in the standard nontransgenic Jack. The result shows the glycinin mimic

GFP allele was able to participate in the proteome rebalancing of shortage of beta-conglycinin.

[0073] The seed-specific ER-targeted GFP trait in the cv Jack background produces about 1% (w/w) of the total heterologous protein in ER-derived protein bodies that has been grown for numerous generations, and this output trait has proven stable. ER retention sequences will retain and stabilize post-translationally unstable proteins in seeds. The engineering strategy of retaining proteins in the ER has become a common technique for stabilizing foreign proteins expressed in transgenic seeds. Due to the ease of visualization of this trait by GFP fluorescence and as a glycinin storage protein proxy, this expression cassette is now used as a visual marker, instead of antibiotic resistance markers, in plant species, such as Camelina, that can be transformed through a floral dip.

[0074] Since GFP is easily visually detected through a fluorescent dissecting scope, large numbers of seeds can readily be efficiently screened. In addition, there are commercially available molecular means to detect, cytochemically localize, and quantitate GFP, specifically a GFP monoclonal antibody and GFP quantitative standards to measure GFP by a fluorometer. All of these molecular means to detect and quantitate GFP makes it an ideal marker to perform large-scale analysis of transgene insertion sights on introduced gene expression cassettes.

[0075] A primary approach is the characterization of the seed proteome, particularly to quantify GFP used as a seed storage protein proxy marker to assess the variations in output trait. The GFP storage protein proxy system enabled a rapid, reproducible screening and quantitation system. The assay system illustrated in FIG. 5A and 5B shows standard 2D gels of a GFP transgenic seed.

[0076] *Mutagenesis*: The effects of mutagenesis on the production of unintended phenotypes using the transgene-encoded expression of GFP as a seed storage protein proxy were determined. Mutagenesis can be a powerful tool to assess gene function. There are various means to produce mutagenized populations including insertional (transposons or T-DNA), chemical as well as particle (neutron), and photon (X-ray) irradiation. Insertional mutants have the distinct advantage of knowing the inserted gene sequence so downstream, the chromosomal location could be determined, allowing for position effects to be analyzed. Such insertional mutant libraries have proven to be powerful tools, and considerable effort has been expended to develop these collections as community toolkits. The limitation of insertional mutants is it results in only loss-of-function phenotypes as the insertion event disrupts open reading frames or critical regulatory sequences. An additional limitation for insertional mutagenesis is since soybean's endogenous transposons are quiescent, the insertion event for a T-DNA would have to take place by a transformation event. This would add a substantial

additional transgenic variable to the mutant effect. To limit these variables, this mutagenized population will be created by the chemical mutagen ethyl methansulphonate (EMS). Chemical mutagens are capable of producing both loss- and gain-of-functions phenotypes due to their ability to cause single point mutations.

[0077] The same GFP-HDEL cassette was used, and a single already characterized GFP stably transformed soybean line was subjected to EMS to generate a mutagenized population. The GFP mutagenized population will ultimately be assessed for the expression of the stable GFP transgene and alterations in seed protein and oil content and composition. A key aspect of the production of a mutagenized population is the balance of obtaining actual mutations in the genome while simultaneously maintaining viability/fertility. Typically a series of EMS concentrations are used to determine this optimum concentration. The GFP-HDEL seeds of an already characterized stable transgenic soybean line will be used as the seed stock. Batches of 1,000 fresh GFP-HDEL seeds will be treated with either 25, 37.5, 50 mM EMS for 9 hrs. EMS will be neutralized by adding 10% sodium thiosulfate solution. After thorough washing, M1 seeds will be planted in a greenhouse with 16 hr photoperiod. To estimate EMS effects, the germination rate of the M1 seeds will be quantified for each treatment. Mutation frequency in plants can be determined by the degree of embryo lethality assayed in the pods of M1 plants. Upon visual inspection of defective embryos in pods, the treatment producing defective embryos on a frequency comparable to that of the normal rate (40:1 normal/defective) will be used to produce a large-scale mutagenized population. To produce M2 lines, seeds collected from a single soybean pod from a mature M1 plant will be grown, and only one plant will be allowed to reach maturity. In this manner, each M1 plant will only give rise to a single M2 line so that each individual in the M2 population descended from a different mutagenized parental line. As the seed stock of the mutagenized population contains a seed-specific GFP expression cassette, seeds from M1 plants will be visually screened to selectively choose seeds with a visual phenotype. That is, seeds displaying an alteration in the inserted seed GFP expression will be selected to move into the M2 generation. M1 lines will continue to be screened by fluorescent stereo microscopy and visually compared to the seed stock parental line for the degree of GFP expression. If visual inspection isn't sensitive enough, a small amount of tissue will be chipped off the side of the seeds opposite the embryo, total protein extracted, and GFP quantification performed using a fluorometer and by immunoassay (ELISA) that permits rapid screening of a large number of samples using commercially available GFP as a standard. Both over and under expressing lines will be selected to produce a total of 100 mutagenized M2 lines. By using a seed-specific inserted trait as a marker, lines that contain mutations that affect seed phenotypes will be selected. Through this approach, 100 M2 lines will be enriched for seed mutations.

[0078] The mutagenized population of the single GFP event removes the position effect as a variable in the resulting seed phenotype. GFP will be quantitated from seed lysates by fluorometry. Total protein of the seeds will be estimated by Bradford assays and both protein and oil content will be determined by nondestructive NMR. As with the transgenomic lines, the top 20 lines that show an alteration in either/both protein or oil as determined by deviation from the wild-type levels will be selected for comprehensive proteomic and triglyceride analysis. The proteomic analysis will initially consist of 2D gel analysis. If any differential spots are observed, then LC/MS analysis of the spots of interest will be used to determine the identification of the proteins. If a large number of differences are observed, the 2D gel analysis will be supplemented with additional MuDPIT analysis. The mass spectroscopy will yield a comprehensive catalog of proteins that comprise that line's seed proteome. Similarly, mutant lines that exhibit an altered oil content by NMR assay will be further analyzed by GC/MS and lipidomic profiling. If no lines exhibit an altered NMR determined oil content, then mutant lines can be screened by GC/MS analysis to determine oil composition changes. The top 20 lines shown to have either altered overall oil content or oil composition will be subjected to extensive lipidomics analysis.

[0079] Each mutant line will have the level of GFP determined by fluorometry and seed composition changes determined by proteomic analysis (Bradford, NMR, 2D gels and/or MuDPIT) and oil analysis (NMR, GC/MS and/or lipidomic analysis).

[0080] As used herein, the term "about" refers to plus or minus 10% of the referenced number.

[0081] Although there has been shown and described the preferred embodiment of the present invention, it will be readily apparent to those skilled in the art that modifications may be made thereto which do not exceed the scope of the appended claims. Therefore, the scope of the invention is only to be limited by the following claims. In some embodiments, the figures presented in this patent application are drawn to scale, including the angles, ratios of dimensions, etc. In some embodiments, the figures are representative only and the claims are not limited by the dimensions of the figures. In some embodiments, descriptions of the inventions described herein using the phrase "comprising" includes embodiments that could be described as "consisting essentially of" or "consisting of", and as such the written description requirement for claiming one or more embodiments of the present invention using the phrase "consisting essentially of" or "consisting of" is met.

WHAT IS CLAIMED IS:

1. A transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising a seed-specific detectable marker, wherein the transgenic legume seed composition comprises an increased amount of protein compared to a legume seed without genetic variation.
2. The transgenic legume of claim 1, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter.
3. The transgenic legume of claim 1 or claim 2, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter.
4. A transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising a seed-specific detectable marker, wherein the transgenic legume seed composition comprises an increased amount of oil compared to a legume seed without genetic variation.
5. The transgenic legume of claim 4, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter.
6. The transgenic legume of claim 4 or claim 5, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter.
7. The transgenic legume of any one of claims 1-6, wherein the transgenic legume seed composition further comprises an increase in carbohydrate content compared to the legume seed without genetic variation.
8. The transgenic legume of any one of claims 1-7, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
9. The transgenic legume of any one of claims 1-8, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
10. The transgenic legume of any one of claims 1-9, wherein the detectable reporter comprises a fluorescent reporter.
11. The transgenic legume of claim 10, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), and green fluorescent proteins (GFP).
12. The transgenic legume of any one of claims 1-11, wherein the genetic variation is introduced via mutagenesis, transformation, or a combination thereof.

13. The transgenic legume of any one of claims 1-12, wherein the legume is selected from a group consisting of alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans.
14. A transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising a seed-specific detectable marker, wherein the transgenic soybean seed composition comprises an increased amount of protein compared to a soybean seed without genetic variation.
15. The transgenic soybean of claim 14, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter.
16. The transgenic soybean of claim 14 or claim 15, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter.
17. A transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising a seed-specific detectable marker, wherein the transgenic soybean seed composition comprises an increased amount of oil compared to a soybean seed without genetic variation.
18. The transgenic soybean of claim 17, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter.
19. The transgenic soybean of claim 17 or claim 18, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter.
20. The transgenic soybean of any one of claims 14-19, wherein the transgenic soybean seed composition further comprises an increase in carbohydrate content compared to the soybean seed without genetic variation.
21. The transgenic soybean of any one of claims 14-20, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
22. The transgenic soybean of any one of claims 14-21, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
23. The transgenic soybean of any one of claims 14-22, wherein the detectable reporter comprises a fluorescent reporter.
24. The transgenic soybean of claim 23, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP),

and green fluorescent proteins (GFP).

25. The transgenic soybean of any one of claims 14-24, wherein the genetic variation is introduced via mutagenesis, transformation, or a combination thereof.
26. A transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising at least two seed-specific detectable markers, wherein the transgenic legume seed composition comprises an increased amount of protein and carbohydrate content compared to a legume seed without genetic variation.
27. The transgenic legume of claim 26, wherein transgenic legume seed comprises a first seed-specific detectable marker and a second seed-specific detectable marker.
28. The transgenic legume of claim 27, wherein the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific protein promoter and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter.
29. A transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising at least two seed-specific detectable markers, wherein the transgenic legume seed composition comprises an increased amount of oil and carbohydrate content compared to a legume seed without genetic variation.
30. The transgenic legume of claim 29, wherein transgenic legume seed comprises a first seed-specific detectable marker and a second seed-specific detectable marker.
31. The transgenic legume of claim 30, wherein the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific oil promoter and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter.
32. The transgenic legume of any one of claims 26-31, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
33. The transgenic legume of any one of claims 26-32, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
34. The transgenic legume of any one of claims 26-33, wherein the detectable reporter comprises a fluorescent reporter.
35. The transgenic legume of claim 34, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP),

- and green fluorescent proteins (GFP).
36. The transgenic legume of any one of claims 26-35, wherein the genetic variation is introduced via mutagenesis, transformation, or a combination thereof.
37. The transgenic legume of any one of claims 26-36, wherein the legume is selected from a group consisting of alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans.
38. A transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising at least two seed-specific detectable markers, wherein the transgenic soybean seed composition comprises an increased amount of protein and carbohydrate content compared to a soybean seed without genetic variation.
39. The transgenic soybean of claim 38, wherein transgenic soybean seed comprises a first seed-specific detectable marker and a second seed-specific detectable marker.
40. The transgenic soybean of claim 39, wherein the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific protein promoter and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter.
41. A transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising at least two seed-specific detectable markers, wherein the transgenic soybean seed composition comprises an increased amount of oil and carbohydrate content compared to a soybean seed without genetic variation.
42. The transgenic soybean of claim 41, wherein transgenic soybean seed comprises a first seed-specific detectable marker and a second seed-specific detectable marker.
43. The transgenic soybean of claim 42, wherein the first seed-specific detectable marker comprises a first detectable reporter operatively linked to a seed-specific oil promoter and the second seed-specific detectable marker comprises a second detectable reporter operatively linked to a seed-specific carbohydrate promoter.
44. The transgenic soybean of any one of claims 38-43, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
45. The transgenic soybean of any one of claims 38-44, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
46. The transgenic soybean of any one of claims 38-45, wherein the detectable reporter

- comprises a fluorescent reporter.
47. The transgenic soybean of claim 46, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), and green fluorescent proteins (GFP).
48. The transgenic soybean of any one of claims 38-47, wherein the genetic variation is introduced via mutagenesis, transformation, or a combination thereof.
49. A method to determine genes regulating legume seed composition, the method comprising:
- a) introducing genetic variation to a transgenic legume seed; wherein the transgenic legume seed comprises a seed-specific detectable marker, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter; and
 - b) measuring the detectable reporter to determine the legume seed composition.
50. The method of claim 49, wherein the genetic variation is introduced via mutagenesis, transformation, or a combination thereof.
51. A method to determine genes regulating legume seed composition, the method comprising:
- a) introducing genetic variation to a transgenic legume seed via mutagenesis; wherein the transgenic legume seed comprises a seed-specific detectable marker, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter; and
 - b) measuring the detectable reporter to determine the legume seed composition.
52. The method of any one of claims 49-51, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter.
53. The method of claim 52, wherein an increase in a signal from the detectable reporter indicates the transgenic legume seed composition comprises an increased amount of protein, and a decrease in a signal from the detectable reporter indicates the transgenic legume seed composition comprises a decreased amount of protein.
54. The method of any one of claims 49-51, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter.
55. The method of claim 54, wherein an increase in a signal from the detectable reporter indicates the transgenic legume seed composition comprises an increased amount of oil, and a decrease in a signal from the detectable reporter indicates the transgenic legume seed composition comprises a decreased amount of oil.
56. The method of any one of claims 49-51, wherein the seed-specific detectable marker

- comprises a detectable reporter operatively linked to a seed-specific carbohydrate promoter.
57. The method of claim 56, wherein an increase in a signal from the detectable reporter indicates the transgenic legume seed composition comprises an increased amount of carbohydrate content, and a decrease in a signal from the detectable reporter indicates the transgenic legume seed composition comprises a decreased amount of carbohydrate content.
58. The method of any one of claims 49-57, wherein the detectable reporter comprises a fluorescent reporter.
59. The method of claim 58, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), and green fluorescent proteins (GFP).
60. The method of any one of claims 49-59, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
61. The method of any one of claims 49-60, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
62. The method of any one of claims 49-61, wherein the legume is selected from a group consisting of alfalfa, clover, mesquite, tamarind, carob, peas, beans, peanuts, or other legume nuts, lentils, and soybeans.
63. A method to determine genes regulating soybean seed composition, the method comprising:
- a) introducing genetic variation to a transgenic soybean seed; wherein the transgenic soybean seed comprises a seed-specific detectable marker, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific promoter; and
 - b) measuring the detectable reporter to determine the soybean seed composition.
64. The method of claim 63, wherein the genetic variation is introduced via mutagenesis, transformation, or a combination thereof.
65. A method to determine genes regulating soybean seed composition, the method comprising:
- a) introducing genetic variation to a transgenic soybean seed via mutagenesis; wherein the transgenic soybean seed comprises a seed-specific detectable marker, wherein the seed-specific detectable marker comprises a detectable

- reporter operatively linked to a seed-specific promoter; and
- b) measuring the detectable reporter to determine the soybean seed composition.
66. The method of any one of claims 63-65, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter.
67. The method of claim 66, wherein an increase in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises an increased amount of protein, and a decrease in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises a decreased amount of protein.
68. The method of any one of claims 63-65, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter.
69. The method of claim 68, wherein an increase in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises an increased amount of oil, and a decrease in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises a decreased amount of oil.
70. The method of any one of claims 63-65, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific carbohydrate promoter.
71. The method of claim 70, wherein an increase in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises an increased amount of carbohydrate content, and a decrease in a signal from the detectable reporter indicates the transgenic soybean seed composition comprises a decreased amount of carbohydrate content.
72. The method of any one of claims 63-71, wherein the detectable reporter comprises a fluorescent reporter.
73. The method of claim 72, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), and green fluorescent proteins (GFP).
74. The method of any one of claims 63-73, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
75. The method of any one of claims 63-74, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
76. An expression cassette comprising a seed-specific detectable marker comprising a detectable reporter operatively linked to a seed-specific promoter.

77. The expression cassette of claim 76, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific protein promoter.
78. The expression cassette of claim 76, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific oil promoter.
79. The expression cassette of claim 76, wherein the seed-specific detectable marker comprises a detectable reporter operatively linked to a seed-specific carbohydrate promoter.
80. The expression cassette of any one of claims 76-79, wherein the detectable reporter comprises a fluorescent reporter.
81. The expression cassette of claim 80, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), and green fluorescent proteins (GFP).
82. The expression cassette of any one of claims 76-81, wherein the seed-specific detectable marker further comprises an ER-signal sequence operatively linked to the 5' end of the detectable marker.
83. The expression cassette of any one of claims 76-82, wherein the seed-specific detectable marker further comprises an ER-retention signal operatively linked to the 3' end of the detectable marker; wherein the ER-retention signal comprises the sequence HDEL.
84. An expression cassette comprising a seed-specific detectable marker comprising:
- a) a detectable reporter;
 - b) an ER-signal sequence operatively linked to the 5' end of the detectable marker;
 - and
 - c) an ER-retention signal operatively linked to the 3' end;
- all operatively linked to a seed-specific promoter.
85. An expression cassette comprising a seed-specific detectable marker consisting essentially of:
- a) a detectable reporter;
 - b) an ER-signal sequence operatively linked to the 5' end of the detectable marker;
 - and
 - c) an ER-retention signal operatively linked to the 3' end;
- all operatively linked to a seed-specific promoter.
86. The expression cassette of claim 84 or claim 85, wherein the seed-specific promoter is a seed-specific protein promoter.
87. The expression cassette of claim 84 or claim 85, wherein the seed-specific promoter is a seed-specific oil promoter.

88. The expression cassette of claim 84 or claim 85, wherein the seed-specific promoter is a seed-specific carbohydrate promoter.
89. The expression cassette of any one of claims 84-88, wherein the detectable reporter comprises a fluorescent reporter.
90. The expression cassette of claim 89, wherein the fluorescent reporter is selected from a group consisting of yellow fluorescent proteins (YFP), red fluorescent proteins (RFP), and green fluorescent proteins (GFP).

FIG. 1

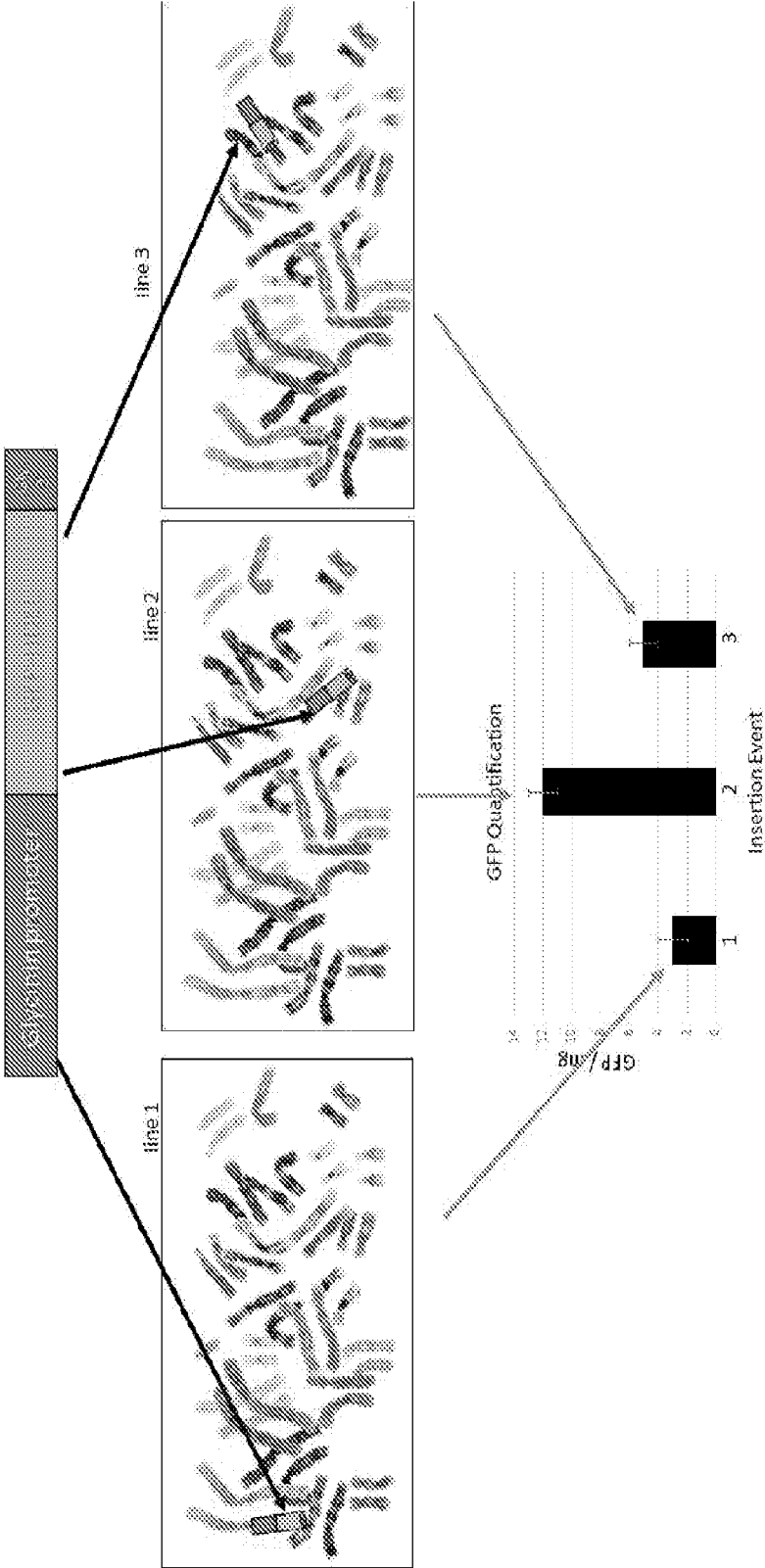


FIG. 2

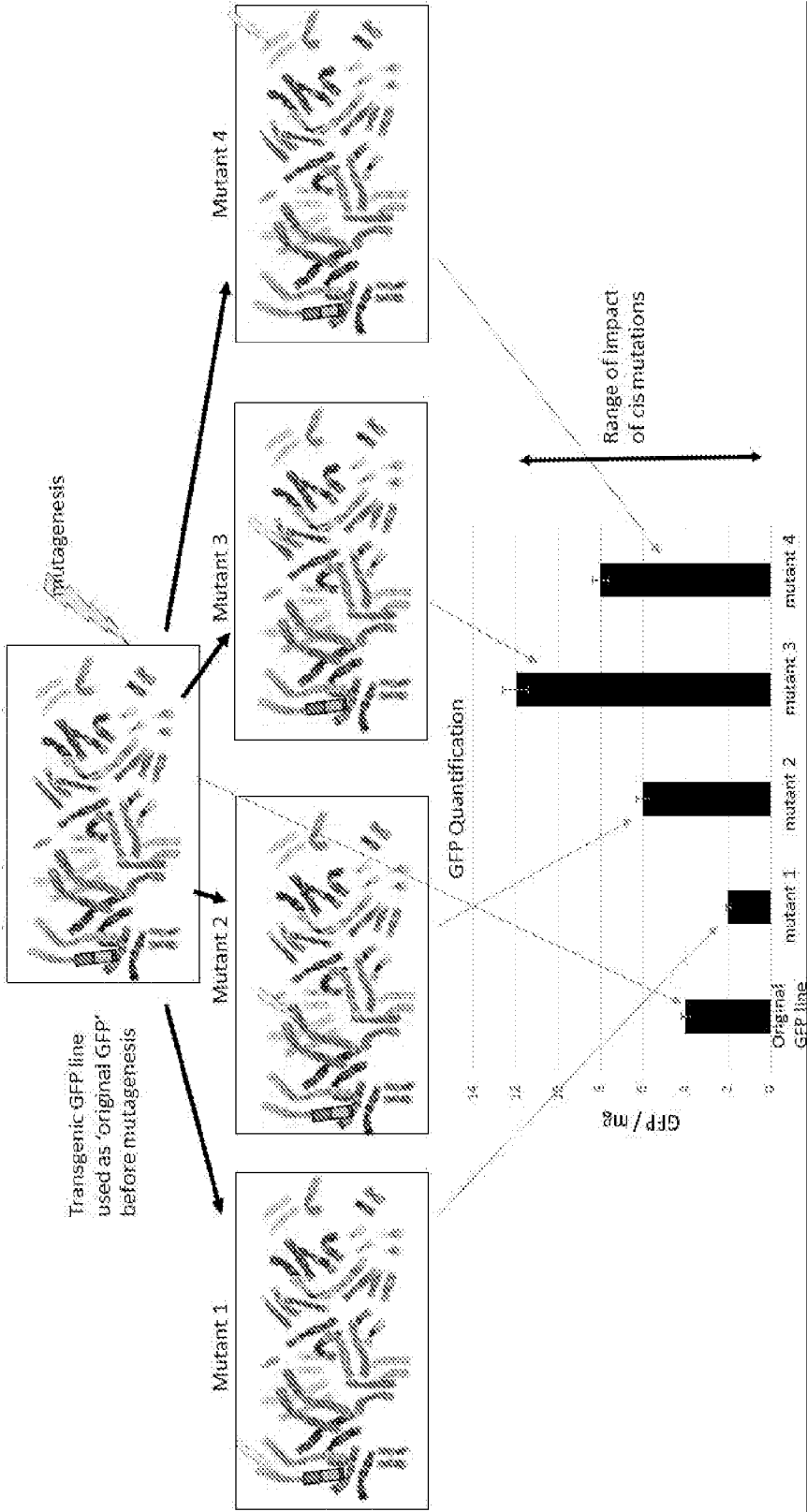


FIG. 3

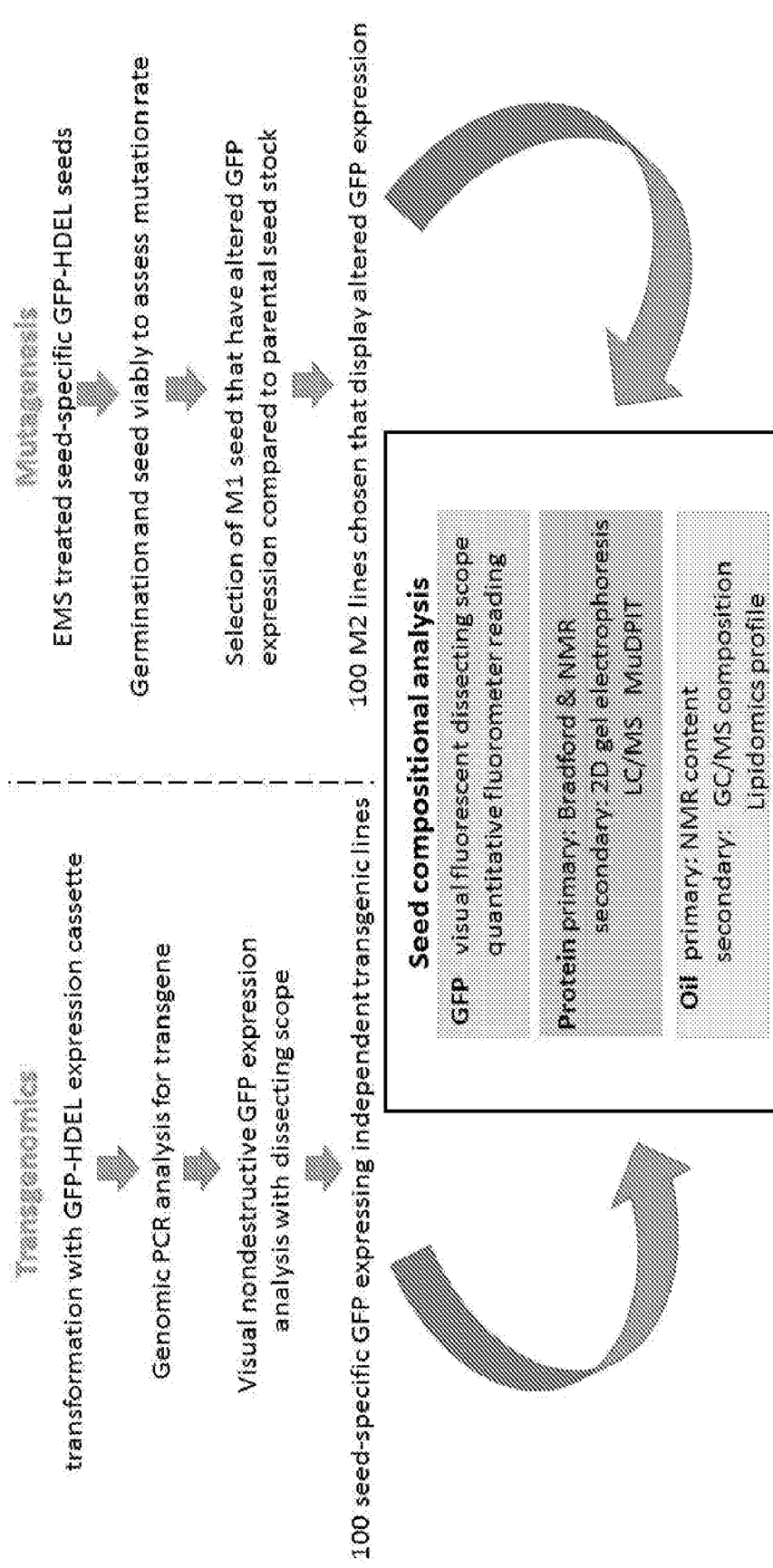


FIG. 4

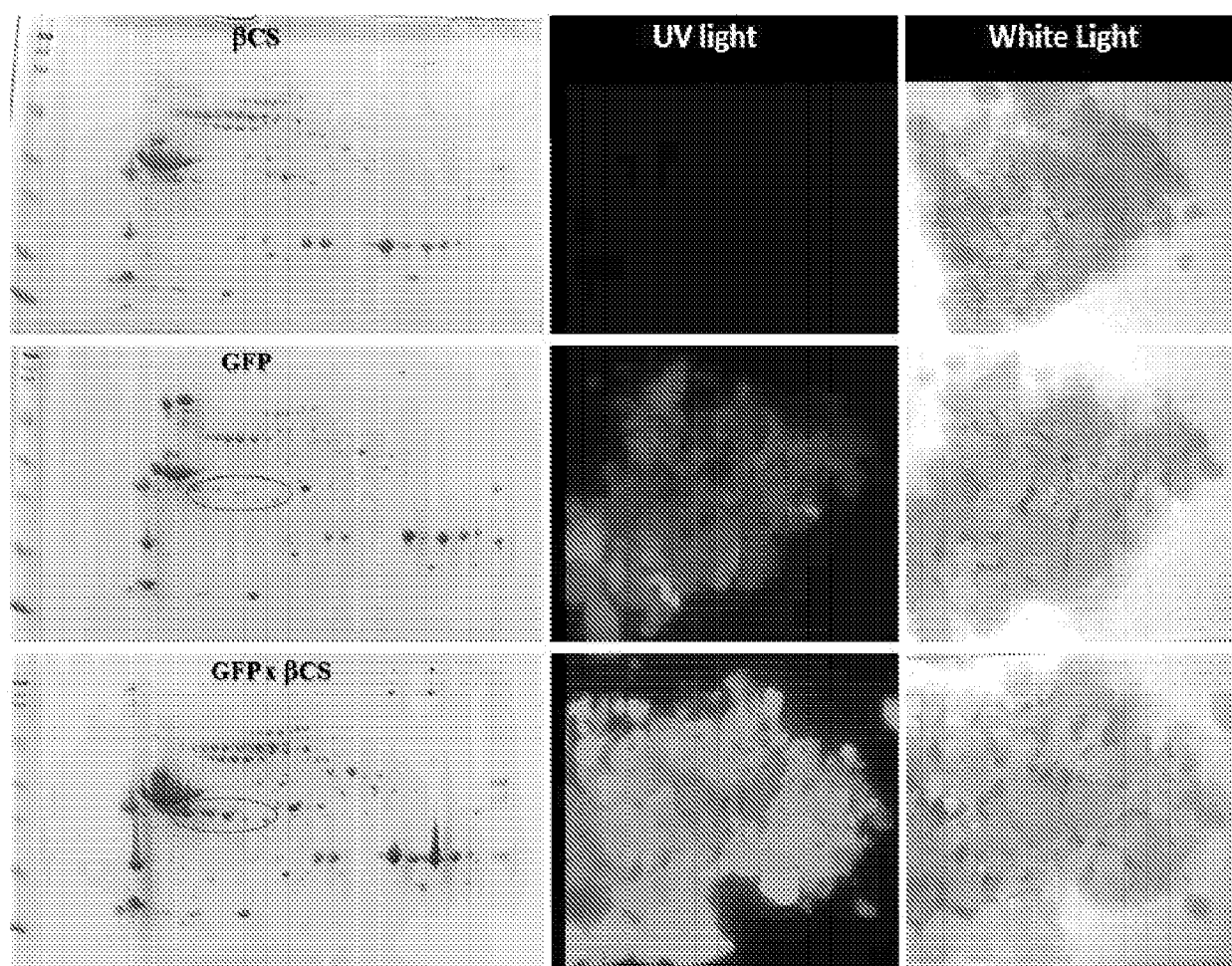


FIG. 5A

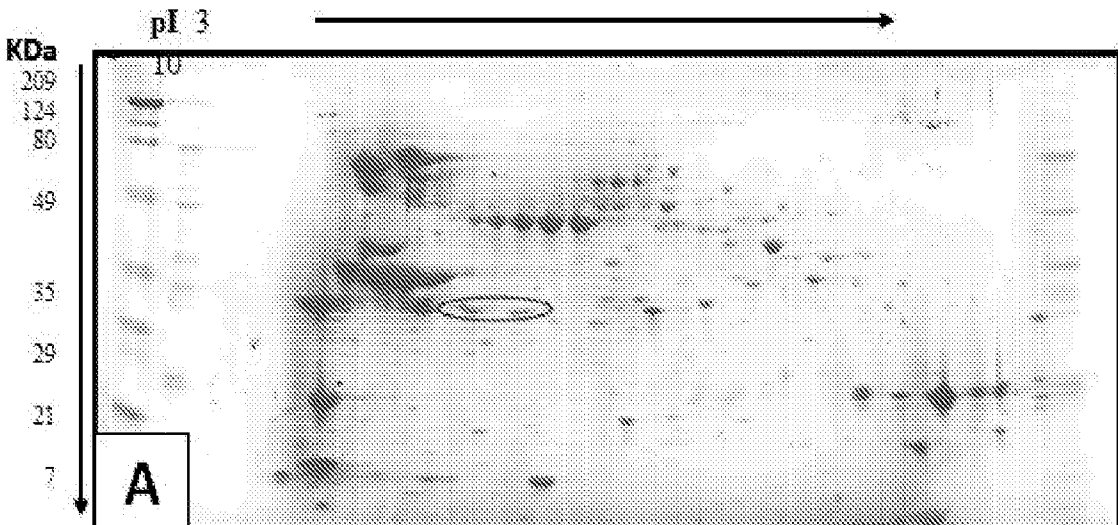
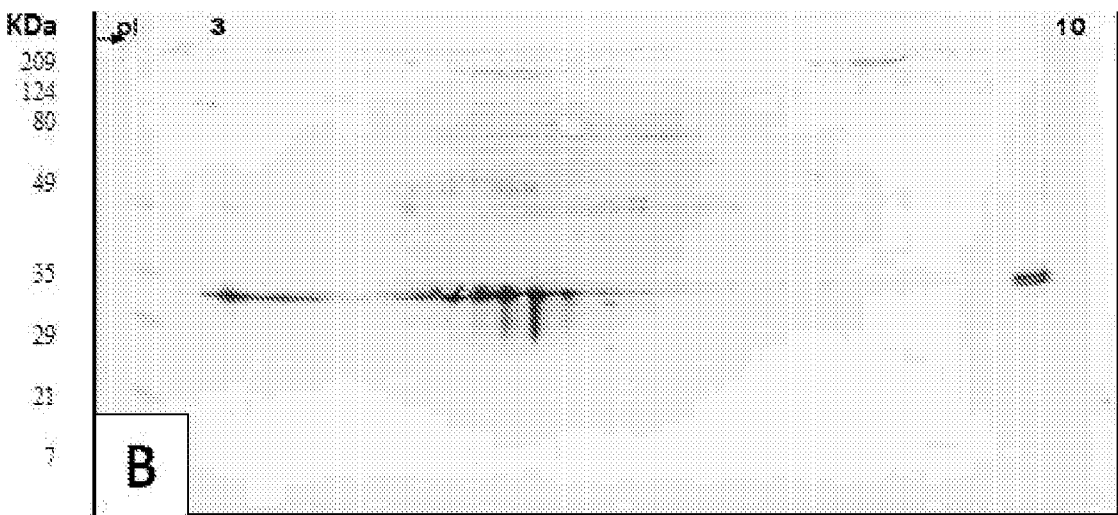


FIG. 5B



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/035993

A. CLASSIFICATION OF SUBJECT MATTERIPC: *C12N 15/82* (2024.01); *C12N 15/90* (2024.01); *C12Q 1/6895* (2024.01); *A01H 4/00* (2024.01); *C07K 14/415* (2024.01); *C12N 15/64* (2024.01)CPC: *C12N 15/8247*; *C12N 15/8261*; *C12N 15/64*; *C12N 15/8234*; *C12N 15/821*; *C07K 14/415*; *C12Q 1/6895*; *A01H 4/008*; *Y02A 40/146*

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2005/0246796 A1 (CIGAN et al.) 03 November 2005 (03.11.2005) entire document	1-3, 14-16, 26-28, 31, 32, 38-40, 43, 44
Y	US 2009/0083883 A1 (ARIAS et al.) 26 March 2009 (26.03.2009) entire document	1-6, 14-19, 26-32, 38-44
Y	US 2017/0037419 A1 (THE ARIZONA BOARD OF REGENTS ON BEHALF OF THE UNIVERSITY OF ARIZONA) 09 February 2017 (09.02.2017) entire document	4-6, 17-19, 29-31, 41-43
A	US 2003/0041342 A1 (SUN et al.) 27 February 2003 (27.02.2003) entire document	1-6, 14-19, 26-32, 38-44



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

19 August 2024 (19.08.2024)

Date of mailing of the international search report

12 November 2024 (12.11.2024)

Name and mailing address of the ISA/US

COMMISSIONER FOR PATENTS
MAIL STOP PCT, ATTN: ISA/US
P.O. Box 1450
Alexandria, VA 22313-1450 UNITED STATES OF AMERICA

Authorized officer

TAINA MATOS

Facsimile No. 571-273-8300

Telephone No. 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/035993

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SCHMIDT et al., Characterization and functional biology of the soybean aleurone layer, BMC Plant Biology, Vol. 18:354, 13 December 2018, Pgs. 1-9, [retrieved on 19.08.2024], Retrieved from the internet: <URL:https://doi.org/10.1186/s12870-018-1579-8>. entire document	32, 44

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: **7-13, 20-25, 33-37, 45-48, 58-62, 72-75, 82, 83, 89, 90**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I: claims 1-6, 14-19, 26-32, and 38-44 are drawn to transgenic legumes and transgenic soybeans.

Group II: claims 49-57 and 63-71 are drawn to methods to determine genes regulating legume seed composition and methods to determine genes regulating soybean seed composition.

Group III: claims 76-81 and 84-88 are drawn to expression cassettes.

The inventions listed in Groups I-III do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The special technical features of Group I, transgenic legumes and transgenic soybeans, are not present in Groups II and III; the special technical features of Group II, methods to determine genes regulating legume seed composition and methods to determine genes regulating soybean seed composition, are not present in Groups I and III; and the special technical features of Group III, expression cassettes, are not present in Groups I and II.

Additionally, even if Groups I-III were considered to share the technical features of a transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising a seed-specific detectable marker, wherein the transgenic legume seed composition comprises an increased amount of protein compared to a legume seed without genetic variation; a transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising a seed-specific detectable marker, wherein the transgenic legume seed composition comprises an increased amount of oil compared to a legume seed without genetic variation; a transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising a seed-specific detectable marker, wherein the transgenic soybean seed composition comprises an increased amount of protein compared to a soybean seed without genetic variation; a transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising a seed-specific detectable marker, wherein the transgenic soybean seed composition comprises an increased amount of oil compared to a soybean seed without genetic variation; a transgenic legume comprising an

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising at least two seed-specific detectable markers, wherein the transgenic legume seed composition comprises an increased amount of protein and carbohydrate content compared to a legume seed without genetic variation; a transgenic legume comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic legume seed comprising at least two seed-specific detectable markers, wherein the transgenic legume seed composition comprises an increased amount of oil and carbohydrate content compared to a legume seed without genetic variation; a transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising at least two seed-specific detectable markers, wherein the transgenic soybean seed composition comprises an increased amount of protein and carbohydrate content compared to a soybean seed without genetic variation; a transgenic soybean comprising an altered seed composition produced by a method comprising introducing genetic variation to a transgenic soybean seed comprising at least two seed-specific detectable markers, wherein the transgenic soybean seed composition comprises an increased amount of oil and carbohydrate content compared to a soybean seed without genetic variation., these shared technical features do not represent a contribution over the prior art as disclosed by US 20170037419 to The Arizona Board of Regents on Behalf of The University of Arizona (hereinafter, "Arizona"), US 2009/0083883 to Arias et al. (hereinafter, "Arias"), and US 2003/0041342 to Sun et al. (hereinafter, "Sun").

Arizona teaches a transgenic legume (a method for the production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; the plant is... soybean, Para. [0015]; New plant genotypes are described herein that mimic the genotypes of plants grown in nitrogen-limited environments that manifest the phenotype of higher C:N ratios. Altering perceived source input, yields a range of high C:N ratio phenotypes that, for instance, can be leveraged to create, for example, soybean having higher oil content, Para. [0052]; a nucleic acid (preferably DNA) sequence encoding an Asnase gene may be used for the production of a plant having an altered allocation of fixed carbon, wherein the ratio of C:N in the plant or a particular organ within a plant is higher than in a control plant ... introduction of a nucleic acid sequence encoding an Asnase into a plant that would benefit from reprogramming of its fixed carbon allocation. One suitable plant capable of providing a nucleic acid sequence encoding an Asnase having high specificity is soybean, Para. [0059]) comprising an altered seed composition (An Asnase gene expressed in soybean seeds was thus selected as an experimental model for using Asnase activity to modulate the perceived nitrogen status in target plants, Para. [0056]) produced by a method comprising introducing genetic variation to a legume seed comprising a seed-specific detectable marker (A suitable nucleic acid sequence encoding Asnase (e.g., SEQ ID NO: 1) may be transferred to a suitable recipient plant, Para. [0060]; Expression constructs can include at least one marker gene, operably linked to a regulatory element (such as a promoter) that allows transformed cells containing the marker to be either recovered by negative selection, Para. [0070]; Generally after transformation, plant cells or cell groupings are selected for the presence of the gene of interest. Wherein a marker gene was co-transferred with the gene of interest, plants may be selected by observing or detecting the marker gene, Para. [0076]; The promoter may also be a tissue specific promoter, which results in increased expression of the nucleic acid in a particular tissue or organ... tissue specific promoters useful in the methods described herein include... the glycinin seed specific promoter, Para. [0090], See instant disclosure, Para. [0013], which states a seed-specific detectable marker, comprises a detectable reporter operatively linked to a seed-specific promoter; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]), wherein the legume seed composition comprises an increased amount of oil compared to a legume seed without genetic variation (the plants or parts thereof comprise... at least 1% increase in oil content relative to a wild-type control plant grown in parallel, Para. [0016]; production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]); a transgenic soybean (a method for the production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; the plant is... soybean, Para. [0015]; New plant genotypes are described herein that mimic the genotypes of plants grown in nitrogen-limited environments that manifest the phenotype of higher C:N ratios. Altering perceived source input, yields a range of high C:N ratio phenotypes that, for instance, can be leveraged to create, for example, soybean having higher oil content, Para. [0052]; a nucleic acid (preferably DNA) sequence encoding an Asnase gene may be used for the production of a plant having an altered allocation of fixed carbon, wherein the ratio of C:N in the plant or a particular organ within a plant is higher than in a control plant.... introduction of a nucleic acid sequence encoding an Asnase into a plant

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

that would benefit from reprogramming of its fixed carbon allocation. One suitable plant capable of providing a nucleic acid sequence encoding an Asnase having high specificity is soybean, Para. [0059]) comprising an altered seed composition (An Asnase gene expressed in soybean seeds was thus selected as an experimental model for using Asnase activity to modulate the perceived nitrogen status in target plants, Para. [0056]) produced by a method comprising introducing genetic variation to a soybean seed comprising a seed-specific detectable marker (A suitable nucleic acid sequence encoding Asnase (e.g., SEQ ID NO: 1) may be transferred to a suitable recipient plant, Para. [0060]; Expression constructs can include at least one marker gene, operably linked to a regulatory element (such as a promoter) that allows transformed cells containing the marker to be either recovered by negative selection, Para. [0070]; Generally after transformation, plant cells or cell groupings are selected for the presence of the gene of interest. Wherein a marker gene was co-transferred with the gene of interest, plants may be selected by observing or detecting the marker gene, Para. [0076]; The promoter may also be a tissue specific promoter, which results in increased expression of the nucleic acid in a particular tissue or organ... tissue specific promoters useful in the methods described herein include... the glycinin seed specific promoter, Para. [0090], See instant disclosure, Para. [0013], which states a seed-specific detectable marker, comprises a detectable reporter operatively linked to a seed-specific promoter; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]), wherein the soybean seed composition comprises an increased amount of oil compared to a soybean seed without genetic variation (the plants or parts thereof comprise... at least 1% increase in oil content relative to a wild-type control plant grown in parallel, Para. [0016]; production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]); a transgenic legume (a method for the production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; the plant is... soybean, Para. [0015]; New plant genotypes are described herein that mimic the genotypes of plants grown in nitrogen-limited environments that manifest the phenotype of higher C:N ratios. Altering perceived source input, yields a range of high C:N ratio phenotypes that, for instance, can be leveraged to create, for example, soybean having higher oil content, Para. [0052]; a nucleic acid (preferably DNA) sequence encoding an Asnase gene may be used for the production of a plant having an altered allocation of fixed carbon, wherein the ratio of C:N in the plant or a particular organ within a plant is higher than in a control plant ... introduction of a nucleic acid sequence encoding an Asnase into a plant that would benefit from reprogramming of its fixed carbon allocation. One suitable plant capable of providing a nucleic acid sequence encoding an Asnase having high specificity is soybean, Para. [0059]) comprising an altered seed composition (An Asnase gene expressed in soybean seeds was thus selected as an experimental model for using Asnase activity to modulate the perceived nitrogen status in target plants, Para. [0056]) produced by a method comprising introducing genetic variation to a legume seed comprising at least two seed-specific detectable markers (A suitable nucleic acid sequence encoding Asnase (e.g., SEQ ID NO: 1) may be transferred to a suitable recipient plant, Para. [0060]; Expression constructs can include at least one marker gene, operably linked to a regulatory element (such as a promoter) that allows transformed cells containing the marker to be either recovered by negative selection, Para. [0070]; Generally after transformation, plant cells or cell groupings are selected for the presence of the gene of interest. Wherein a marker gene was co-transferred with the gene of interest, plants may be selected by observing or detecting the marker gene, Para. [0076]; The promoter may also be a tissue specific promoter, which results in increased expression of the nucleic acid in a particular tissue or organ... tissue specific promoters useful in the methods described herein include... the glycinin seed specific promoter, Para. [0090], See instant disclosure, Para. [0013], which states a seed-specific detectable marker, comprises a detectable reporter operatively linked to a seed-specific promoter; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]), wherein the legume seed composition comprises an increased amount of oil and carbohydrate content compared to a legume seed without genetic variation (the plants or parts thereof comprise 20-50% greater cellulose content per unit mass greater than a wild-type control plant grown in parallel... a 2-3 fold increase in sugar content relative to a wild-type control plant grown in parallel; and] at least 1% increase in oil content relative to a wild-type control plant grown in parallel, Para. [0016]; production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]); a transgenic soybean (a method for the production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; the plant is... soybean, Para. [0015]; New plant

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

genotypes are described herein that mimic the genotypes of plants grown in nitrogen-limited environments that manifest the phenotype of higher C:N ratios. Altering perceived source input, yields a range of high C:N ratio phenotypes that, for instance, can be leveraged to create, for example, soybean having higher oil content, Para. [0052]; a nucleic acid (preferably DNA) sequence encoding an Asnase gene may be used for the production of a plant having an altered allocation of fixed carbon, wherein the ratio of C:N in the plant or a particular organ within a plant is higher than in a control plant ... introduction of a nucleic acid sequence encoding an Asnase into a plant that would benefit from reprogramming of its fixed carbon allocation. One suitable plant capable of providing a nucleic acid sequence encoding an Asnase having high specificity is soybean, Para. [0059]) comprising an altered seed composition (An Asnase gene expressed in soybean seeds was thus selected as an experimental model for using Asnase activity to modulate the perceived nitrogen status in target plants, Para. [0056]) produced by a method comprising introducing genetic variation to a soybean seed comprising at least two seed-specific detectable markers (A suitable nucleic acid sequence encoding Asnase (e.g., SEQ ID NO: 1) may be transferred to a suitable recipient plant, Para. [0060]; Expression constructs can include at least one marker gene, operably linked to a regulatory element (such as a promoter) that allows transformed cells containing the marker to be either recovered by negative selection, Para. [0070]; Generally after transformation, plant cells or cell groupings are selected for the presence of the gene of interest. Wherein a marker gene was co-transferred with the gene of interest, plants may be selected by observing or detecting the marker gene, Para. [0076]; The promoter may also be a tissue specific promoter, which results in increased expression of the nucleic acid in a particular tissue or organ... tissue specific promoters useful in the methods described herein include... the glycinin seed specific promoter, Para. [0090], See instant disclosure, Para. [0013], which states a seed-specific detectable marker, comprises a detectable reporter operatively linked to a seed-specific promoter; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]), wherein the soybean seed composition comprises an increased amount of oil and carbohydrate content compared to a soybean seed without genetic variation (the plants or parts thereof comprise 20-50% greater cellulose content per unit mass greater than a wild-type control plant grown in parallel... a 2-3 fold increase in sugar content relative to a wild-type control plant grown in parallel; and at least 1% increase in oil content relative to a wild-type control plant grown in parallel, Para. [0016]; production of a transgenic plant having increased sugar, starch, cellulose, and/or oil content relative to a control plant, Para. [0021]; Ectopic Asnase has been successfully introduced and expressed in soybean in a tissue-specific manner utilizing the nucleic acid sequences described herein and the glycinin seed-specific promoter, Para. [0172]).

Arias teaches introducing genetic variation to a transgenic legume seed (methods for transformation of soybean (*Glycine max*) based on a D-alanine and/or D-serine selection, Para. [0001]; introducing into a soybean cell or tissue a DNA construct comprising at least one first expression construct comprising a promoter active in said soybean plant and operably linked thereto a nucleic acid sequence encoding an enzyme capable to metabolize D-alanine and/or D-serine, Para. [0008]; incubating said soybean cell or tissue of step a) on a selection medium comprising D-alanine and/or D-serine and/or a derivative thereof, Para. [0009]; The methods and compositions of the invention allow for subsequent transformation ... It is therefore possible to transform existing transgenic plants comprising another selection marker with the constructs of the invention or to subsequently transform the plants obtained by the method of the invention (and comprising the expression constructs for the D-serine and/or D-alanine metabolizing enzyme) with another marker. In consequence, another embodiment of the invention relates to a method for subsequent transformation of at least two DNA constructs into a soybean plant, Para. [0389]).

Sun teaches an increased amount of protein content compared to a legume seed without genetic variation (recombinant expression cassettes, host cells, transgenic plants, Abstract; Expression of the nucleic acids encoding the proteins of the present invention can be increased... relative to a non-transformed control plant, Para. [0040]; such heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed, Para. [0099]; an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell, Para. [0123]; concentration of the polypeptides is increased... by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned expression cassette, Para. [0124]; the polypeptides of the present invention are modulated in... soybean, Para. [0126]; Preferred plants include... soybean, Para. [0159]); and an increased amount of protein content compared to a soybean seed without genetic variation (recombinant expression cassettes, host cells, transgenic plants, Abstract; Expression of the nucleic acids encoding the proteins of the present invention can be increased... relative to a non-transformed control plant, Para. [0040]; such heterologous DNA encodes RNA

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

and proteins that are not normally produced by the cell into which it is expressed, Para. [0099]; an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell, Para. [0123]; concentration of the polypeptides is increased... by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned expression cassette, Para. [0124]; the polypeptides of the present invention are modulated in... soybean, Para. [0126]; Preferred plants include... soybean, Para. [0159]).

The inventions listed in Groups I-III therefore lack unity under Rule 13 because they do not share a same or corresponding special technical feature.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **1-6, 14-19, 26-32, 38-44**

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.