TRANSGENE STABILITY, TRANSPOSON ACTIVATION AND POTENTIAL FOR GENE TAGGING IN PEANUT (*Arachis hypogaea* L).

by

MANJU CHANDRAN

(Under the Direction of Peggy Ozias-Akins)

ABSTRACT

Crop improvement in peanut (*Arachis hypogaea* L.) is aimed at improving yield, product quality, reduction in seed allergen content and resistance to biotic and abiotic stresses. Application of RNA interference (RNAi) was successful in developing a transgenic line B11.1.1/11 with complete reduction of Ara h 2 and reduced levels of Ara h 6, two major seed allergens. The transgene was stable across three transgenic generations tested as Ara h 2 signal was absent and Ara h 6 signal was reduced in quantitative western blot assay. Stability of transgene silencing to altered soil sulfur, specifically on partially silenced Ara h 6, was tested by growing the transgenic line under three soil sulfur levels. Transgenic lines showed no Ara h 2 expression and significantly reduced Ara h 6 expression under different sulfur levels, providing further evidence for stability of the transgene and the effectiveness of RNAi even under nutritional conditions optimal for synthesis of sulfur-rich proteins. High soil sulfur concentrations significantly increased the levels of Ara h 2 and Ara h 6 in non-transgenic lines compared to low sulfur. Ara h 3 levels were significantly reduced under low sulfur in both transgenic and non-transgenic lines and Ara h 1 levels were unaffected by sulfur.

Activation of the peanut transposon, *AhMITE1*, was studied under two stress situations, namely ethylmethane sulfonate (EMS) mutagenesis and prolonged tissue culture. Transposon display provided evidence that the peanut Miniature Inverted-repeat Transposable Element (MITE) was activated under both stresses and variation was observed with regard to the different stress situations and genotypes studied ('Tifrunner' and 'Florunner'). Peanut MITE sequence capture, Sanger sequencing, and primer design to MITE and flanking DNA sequences enabled the development of 17 markers polymorphic across 29 peanut genotypes. Sequences flanking MITEs showed similarity to protein coding regions via BLAST searches. High throughput Illumina sequencing with five genotypes was carried out to recover additional flanking sequence for marker development and gene identification.

Stability of protein silencing across generations and under different nutritional conditions showed that the polyploid peanut genome is amenable to stable transgene expression. Genome instability associated with MITE transposons may occur under conditions of stress and may provide a source of positive and negative variation for gene functional analysis and breeding.

INDEX WORDS: Peanut transgene stability, sulfur and transgene stability, transposon activation and stress, transposon-based sequence capture.

TRANSGENE STABILITY, TRANSPOSON ACTIVATION AND POTENTIAL FOR GENE TAGGING IN PEANUT (*Arachis hypogaea* L).

by

MANJU CHANDRAN

B.Sc. (Agri.), Kerala Agricultural University, 1993

M.Sc. (Hort.), Kerala Agricultural University, 1996

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

TRANSGENE STABILITY, TRANSPOSON ACTIVATION AND POTENTIAL FOR GENE TAGGING IN PEANUT (*Arachis hypogaea* L).

by

MANJU CHANDRAN

Major Professor: Peggy Ozias-Akins

Committee: Joann Conner
Cecilia McGregor
John Beasley
Arthur Grider

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

DEDICATION

I dedicate my dissertation to my family, friends and all those individuals who have believed and encouraged me throughout my life.

ACKNOWLEDGEMENTS

At the very outset I thank Dr. Peggy Ozias-Akins, my major Professor & chair of committee for her guidance, support and encouragement. Through her approach and way of dealing she taught me how to be a true professional without losing personal touch. I sincerely appreciate her patience, especially with my 'writing skills', and all her contribution of time for review, ideas and funding which make my Ph.D. life insightful and an amazing learning experience .

I would like to thank my committee members Dr. Joann Conner, for her guidance in molecular techniques, which helped me many times when I was challenged, Dr. John Beasley for his expertise in field studies, Dr. Arthur Grider for his guidance in protein work and Dr. Cecilia McGregor for her thought-provoking suggestions.

I would like to thank Brian Abernathy for helping with the Bioinformatic works related to my research.

I have immensely enjoyed working with all the professionals (present and past) in Dr. Ozias-Akins' lab for the last four years of my research in Tifton. I thank Dr. Ye Chu who had helped me with many aspects of this research. I would like to express my gratitude to Ms. Evelyn Morgan and Mr. Greg Thomas for the technical support.

Many thanks to Joe, Anjan, Jackie, Jane, Congling, Bindu, Murug, Rattan, Walid, Guna, Brant, Yufang, Madhu, Lekshmi, Mrs. & Mr. Rajeev and members of The Tift Indian Group for their positive energy and constant encouragement. I also extend my appreciation to staff members of the Department of Horticulture and NESPAL for the friendship and help, especially Mary Jane and Evelyn Folds.

Running short of words to express my gratitude for my loving, supportive, encouraging, and patient daughters, Nayana and Nirupama, as well as my husband who have stood with me, at all times.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Peanut (*Arachis hypogaea* L.), widely grown in the semi-arid tropics, is a rich and economical source of protein and oil. The oil, protein and carbohydrate contents in peanut seeds are 40-60%, 20-40% and 10-20%, respectively (Pandey *et al*., 2012). Cultivated peanut is a recent allotetraploid with an AABB genome (2n=4x=40) derived from the wild progenitors *A.duranensis* (AA) and *A.ipaensis* (BB) (Favero *et al*., 2006; Kochert *et al*., 1996).

Crop improvement in cultivated peanut is directed towards resistance to abiotic stress such as drought, increased yield, resistance to biotic stresses such as pests and diseases, high oleic acid content in seeds and reduction of seed protein allergens. Yield and quality improvement coupled with biotic and abiotic stress resistance is needed to ensure sustained viability of the crop in the US and global productivity. Crop improvement employs conventional breeding, marker-assisted breeding and genetic engineering tools. Genes for resistance to biotic and abiotic stresses are found in the wild relatives of peanut but their introgression into cultivated types is limited by fertility barriers, differences in ploidy levels and linkage drag (Bertioli *et al*., 2011; Holbrook, 2010). Genetic engineering is an alternative tool to overcome the reproductive barriers and linkage drag in traditional breeding, and plant regeneration and transgene insertion techniques have already been standardized for peanut (Ozias-Akins, 2007; Ozias-Akins and Gill, 2001; Sharma and Anjaiah, 2000).

Broad application of molecular technology in breeding depends on the identification of genes or the biosynthetic pathway affecting a particular trait or the development of polymorphic markers

associated with traits of interest in cultivated peanut and its wild relatives. Cultivated peanut shows little variation at the molecular level as evidenced from restriction fragment length polymorphism (RFLP) and random amplification of polymorphic DNA (RAPD) analysis, although phenotypic variation exists (Halward *et al*., 1992; Kochert *et al*., 1991). More recently, amplified fragment length polymorphism (AFLP) markers (He and Prakash, 1997), simple sequence repeat (SSR) markers (Nagy *et al*., 2012; Pandey *et al*., 2012) and transposon based markers (Shirasawa *et al*., 2012b) were identified which were polymorphic among cultivated genotypes. The high density genetic maps constructed so far cover 2166.4 cM with 1114 loci mapped using SSR and transposon markers in tetraploid peanut (Shirasawa *et al*., 2012b) and 1081.3 cM with 1724 loci mapped using SNP and SSR markers in *A.duranensis* (Nagy *et al*., 2012). More markers have to be identified in wild diploids and tetraploid cultivated peanut for developing ultra-high density genetic maps for use in accelerated breeding by marker assisted selection, to identify traits linked to markers and for gene identification. Consensus maps that integrate information from cultivated and wild *Arachis* (Shirasawa *et al.* 2013) will be valuable for molecular breeding.

Peanut is an allotetraploid with a genome size of 2800 Mbp, which is ~ 20 times larger than the model species *Arabidopsis*. The complexity of the genome makes the assembly of the genome sequence computationally demanding. As a complexity reduction strategy, targeted sequencing can be conducted to identify gene rich areas. A peanut DNA transposon from the Miniature Inverted Repeat Transposable element class (MITE-*AhMITE1*) showed an insertion frequency of 11.5% in genic regions and exhibited insertional polymorphism among cultivars (Shirasawa *et al*., 2012a; Shirasawa *et al*., 2012b). Targeted sequencing of the MITE rich regions will generate data on genes and markers which will accelerate breeding and can be integrated with the whole

genome sequence data. MITEs are also found to transpose under various stresses such as chemical mutation and tissue culture. Such transposition events generally occur to genic regions, and could be utilized in a transposon tagging strategy (Gowda *et al*., 2011; Nakazaki *et al*., 2003).

For peanut crop improvement by genetic engineering, RNA interference (RNAi) technology was used to develop peanut lines in which allergen genes are silenced. Major allergen proteins in seeds are Ara h 1, Ara h 2, Ara h 3 and Ara h 6. Ara h 1 and Ara h 2 are recognized by serum IgE from 95% of peanut allergic individuals and are considered as major allergens (Burks *et al*., 1997; Stanley *et al*., 1997). Current research provided evidence that both Ara h 2 and Ara h 6 are potent allergens (Bernard *et al*., 2007; Kulis *et al*., 2012). A peanut line with complete silencing of *Ara h 2* and *Ara h 6* genes was developed by RNAi technology (Chu *et al*., 2008). Arah 2 and Ara h 6 are 2S albumin seed storage proteins belonging to the prolamin super family with conserved cysteine residues (Becker *et al*., 2011; Burks *et al*., 1992; Kleber-Janke *et al*., 1999). Ara h 2 comprises ~5.9-9.3% of total protein in the peanut seed (Koppelman *et al*., 2001). Attempts to regulate expression of such a protein via transgene silencing can be influenced by plant endogenous factors such as methylation of the promoters, site of integration of the transgene, chromatin modifications and exogenous factors like temperature, light intensity and soil nutrition (Assaad *et al*., 1993; Fagard and Vaucheret, 2000; tenLohuis *et al*., 1995). Testing the stability of transgene silencing across generations and varied environments is needed before any transgenic crop is used for further research and crop improvement.

Genomic flux can be induced by mutation, transposon activation, polyploidy, insertion or deletion of sequences, including transgenes. Polyploidy can buffer mutation, fix heterozygosity, and create allelic diversity and genes with novel functions. At the same time polyploidy can lead

to gene redundancy, meiotic irregularities, imbalanced gene expression, transposon activation, gene deletion, gene silencing and cytosine methylation (Comai, 2005; Hegarty and Hiscock, 2008). Evidence for the activation of *AhMITE 1*, a peanut transposon, by polyploidy-induced genomic shock is limited, although Southern blotting showed few *AhMITE 1*copies in the diploid species, *A.duranensis* (Shirasawa et al., 2012a). The current study was undertaken to test the transposition of peanut MITE, *AhMITE1*, by tissue culture and ethylmethane sulfonate treatment, with the objectives to compare transposition under varied stresses and to determine whether transposition preferentially targets gene regions. Targeted sequencing of the MITE rich regions was also done for marker development in the gene space. This study will generate information for development of a saturated genetic map and could assist in genome sequence assembly. The stability of a transgenic line, silenced for sulfur-rich Ara h 2 and Ara h 6 proteins was tested across three generations and under various levels of sulfur nutrition. Transgene expression stability would support the applicability of genetic engineering as a tool for peanut crop improvement.

References

Assaad F.F., Tucker K.L., Signer E.R. (1993) Epigenetic repeat-induced gene silencing *(*RIGS*)* in *Arabidopsis.* Plant Mol. Biol. 22:1067-1085.

Becker W.M., Petersen A., Jappe U. (2011) Peanut allergens: new consolidated findings on structure, characteristics and allergom. Allergologie 34:398-411.

Bernard H., Mondoulet L., Drumare M.F., Paty E., Scheinmann P., Thai R., Wal J.M. (2007) Identification of a new natural Ara h 6 isoform and of its proteolytic product as major allergens in peanut. J. Agric. Food Chem. 55:9663-9669.

Bertioli D.J., Seijo G., Freitas F.O., Valls J.F.M., Leal-Bertioli S.C.M., Moretzsohn M.C. (2011) An overview of peanut and its wild relatives. Plant genetic resources-characterization and utilization 9:134-149.

Burks A.W., Williams L.W., Connaughton C., Cockrell G., Obrien T.J., Helm R.M. (1992) Identification and characterization of a 2nd major peanut allergen, Ara h 2, with use of the sera of patients with atopic-dermatitis and positive peanut challenge. J. Allergy Clin. Immunol 90:962-969.

Burks A.W., Shin D., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A. (1997) Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. Eur. J. Biochem. 245:334-339.

Chu Y., Faustinelli P., Laura Ramos M., Hajduch M., Stevenson S., Thelen J.J., Maleki S.J., Cheng H., Ozias-Akins P. (2008) Reduction of IgE Binding and nonpromotion of *Aspergillus flavus* fungal growth by simultaneously silencing Ara h 2 and Ara h 6 in peanut. J. Agric. Food Chem. 56:11225-11233.

Comai L. (2005) The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6:836-846.

Fagard M., Vaucheret H. (2000) (Trans)gene silencing in plants: How many mechanisms? Ann. Rev. Plant Physiol. Plant Mol. Biol. 51:167-194.

Favero A.P., Simpson C.E., Valls J.F.M., Vello N.A. (2006) Study of the evolution of cultivated peanut through crossability studies among *Arachis ipaensis, A duranensis, and A hypogaea*. Crop Sci. 46:1546-1552.

Gowda M.V.C., Bhat R.S., Sujay V., Kusuma P., Varshakumari, Bhat S., Varshney R.K. (2011) Characterization of *AhMITE1* transposition and its association with the mutational and

evolutionary origin of botanical types in peanut (*Arachis spp*.). Plant Syst. Evol. 291:153-158

Halward T., Stalker T., Larue E., Kochert G. (1992) Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant Mol. Biol.18:315-325.

He G.H., Prakash C.S. (1997) Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). Euphytica 97:143-149

Hegarty M.J., Hiscock S.J. (2008) Genomic clues to the evolutionary success of polyploid plants. Curr. Biol. 18:R435-R444.

Holbrook C.C., Stalker H.T. (2010) Peanut Breeding and Genetic Resources. Plant Breeding Reviews 22:297-356.

Kleber-Janke T., Crameri R., Appenzeller U., Schlaak M., Becker W.M. (1999) Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. Int.Arch. Allergy Immunol. 119:265-274

Kochert G., Halward T., Branch W.D., Simpson C.E. (1991) RFLP variability in peanut *(Arachis hypogaea* L.*)* cultivars and wild-species. Theor. Appl. Genet. 81:565-570.

Kochert G., Stalker H.T., Gimenes M., Galgaro L., Lopes C.R., Moore K. (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am.J. Bot. 83:1282-1291.

Koppelman S.J., Vlooswijk R.A.A., Knippels L.M.J., Hessing M., Knol E.F., van Reijsen F.C., Bruijnzeel-Koomen C. (2001) Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties Runner, Spanish, Virginia, and Valencia, bred in different parts of the world. Allergy 56:132-137

Kulis M., Chen X., Lew J., Wang Q., Patel O.P., Zhuang Y., Murray K.S., Duncan M.W., Porterfield H.S., Burks A.W., Dreskin S.C. (2012) The 2S albumin allergens of *Arachis hypogaea*, Ara h 2 and Ara h 6, are the major elicitors of anaphylaxis and can effectively desensitize peanut-allergic mice. Clin. Exp. Allergy 42:326-336.

Nagy E.D., Guo Y.F., Tang S.X., Bowers J.E., Okashah R.A., Taylor C.A., Zhang D., Khanal S., Heesacker A.F., Khalilian N., Farmer A.D., Carrasquilla-Garcia N., Penmetsa R.V., Cook D., Stalker H.T., Nielsen N., Ozias-Akins P., Knapp S.J. *(*2012*)* A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13:469

Nakazaki T., Okumoto Y., Horibata A., Yamahira S., Teraishi M., Nishida H., Inoue H., Tanisaka T. (2003) Mobilization of a transposon in the rice genome. Nature 421:170-172.

Ozias-Akins P. (2007) Peanut. In: Biotechnology in Agriculture and Forestry 61 – Transgenic Crops VI. E.C. Pua and M.R. Davey, eds. Springer, Heidelberg, pp 81-105.

Ozias-Akins P., Gill R. (2001) Progress in the development of tissue culture and transformation methods applicable to the production of transgenic peanut. Peanut Sci. 28:123-131.

Pandey M.K., Monyo E., Ozias-Akins P., Liang X., Guimaraes P., Nigam S.N., Upadhyaya H.D., Janila P., Zhang X., Guo B., Cook D.R., Bertioli D.J., Michelmore R., Varshney R.K. (2012) Advances in *Arachis* genomics for peanut improvement. Biotechnol. Adv.30:639-651.

Sharma K.K., Anjaiah V. (2000) An efficient method for the production of transgenic plants of peanut (*Arachis hypogaea* L.) through *Agrobacterium tumefaciens*-mediated genetic transformation. Plant Sci. 159:7-19.

Shirasawa K., Hirakawa H., Tabata S., Hasegawa M., Kiyoshima H., Suzuki S., Sasamoto S., Watanabe A., Fujishiro T., Isobe S. (2012a) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor. Appl. Genet.124:1429-1438.

Shirasawa K., Koilkonda P., Aoki K., Hirakawa H., Tabata S., Watanabe M., Hasegawa M., Kiyoshima H., Suzuki S., Kuwata C., Naito Y., Kuboyama T., Nakaya A., Sasamoto S., Watanabe A., Kato M., Kawashima K., Kishida Y., Kohara M., Kurabayashi A., Takahashi C., Tsuruoka H., Wada T., Isobe S. (2012b) *In silico* polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol. 12:80

Shirasawa K., Bertioli D. J., Varshney R. K., Moretzsohn M.C., Leal-Bertioli S.C. M., Thudi M., Pandey M.K., Rami J.F., Fonc`eka D., Gowda M.V., Qin H., Guo B., Hong Y., Liang X., Hirakawa H., Tabata S., Isobe S. (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. DNA Res. 1-12

Stanley J.S., King N., Burks A.W., Huang S.K., Sampson H., Cockrell G., Helm R.M., West C.M., Bannon G.A. (1997) Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. Arch. Biochem. Biophys. 342:244-253.

tenLohuis M., Muller A., Heidmann I., Niedenhof I., Meyer P. (1995) A repetitive DNA fragment carrying a hot spot for *de novo* DNA methylation enhances expression variegation in tobacco and petunia. Plant J. 8:919-932.

CHAPTER 2

LITERATURE REVIEW

Introduction

Advances in the fields of genetics, genomics and molecular biology, have led to the development of tools such as transgenics and gene tagging that can be utilized for identifying genes, deducing gene functions, understanding metabolic pathways and translating this information for crop improvement. RNA interference, a reverse genetics technique, has been employed in peanut to develop transgenic lines with reduced allergens. The advancement of a transgenic line depends on the stability of transgene inheritance and expression, which is determined by endogenous and environmental factors. Endogenous factors include the transgene copy number, site of integration, methylation factors and transposons. Genome sequencing projects have advanced our knowledge about transposable elements, their distribution, functions, epigenetic regulation and their role in genome evolution. Transposons are also evolving as tools in molecular biology for targeted sequencing and marker development. Stability studies in transgenics, the utility of transposons in marker development and targeted sequencing in plants are reviewed here.

Distribution and Classification of Transposable Elements

Transposable elements (TE) are mobile genetic elements, widely distributed in bacteria, fungi, yeast, plants and animals (Wicker *et al.*, 2007). They move and spread within the genome, which can lead to genetic variation and impact genome evolution (Casacuberta and Santiago, 2003). Evidence from sequencing projects shows that transposable elements (TEs) constitute a large proportion of the genetic material in eukaryotes (Feschotte *et al.*, 2002). They account for

approximately 50-80% of the genetic material in most members of the grass family (Meyers *et al.*, 2001), 35% in rice (Takata *et al.*, 2007), 85% in maize (Schnable *et al.*, 2009), 80% in wheat (Cantu *et al.*, 2010) and 59% in soybean (Schmutz *et al*., 2010).

Transposable elements were first discovered by Barbara McClintock while studying pigmentation of maize kernels. She identified breaks in maize chromosome 9, which could be associated with presence or absence of kernel color by genetic analysis. The element at the site of chromosome breakage was termed 'Dissociation (Ds)' and was found to be mobile in the presence of another element, 'Activator (Ac)' (McClintock, 1951). The Ac/Ds transposons, discovered by Barbara McClintock and recently thoroughly characterized at the sequence level in the maize genome (Du *et al*., 2011), are Class II transposons/ DNA transposons, which move by a cut and paste mechanism. Class II transposons are often present in low copy numbers in genomes (Casacuberta and Santiago, 2003). Class I transposons, or retrotransposons, transpose through an RNA intermediate and can reach high copy numbers within the genome (Wicker *et al*. 2007).

Class I, or retrotransposons, are grouped into long terminal repeat (LTR) and non-LTR retrotransposons. LTR-retrotransposons have long terminal repeats in direct orientation. Autonomous LTR- retrotransposons have *gag* and *pol* genes. *Gag* encodes the packing proteins and *pol* encodes protease, reverse transcriptase, integrase and RNAse H proteins needed to complete the retrotranposon life cycle. Retrotransposon mobilization is dependent on reverse transcriptase. Non-autonomous LTR-retrotransposons lack the coding sequences, fully or partially. The two groups of non-LTR retrotransposons are LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements). LINEs are autonomous and have two open reading frames (ORFs), ORF1 and ORF2. ORF1 has RNA binding activity and nucleic

acid chaperone activity. ORF2 encodes the endonuclease and reverse transcriptase. SINEs are non-autonomous and possess a RNA *pol III* promoter internally. Transposition of SINEs is dependent on LINEs, which show 3' sequence similarity (Feschotte *et al.*, 2002; Han, 2010).

Class II or DNA transposons have short terminal inverted repeats (TIR), target-site duplications (TSD) and they encode the transposase enzyme for transposition (Feschotte and Pritham, 2007). A transposase recognizes and binds to the TIR, resulting in the excision of the TE and integration to other genomic regions (Jiang *et al.*, 2004). Internal deletions of transposase sequences from the autonomous DNA transposons can give rise to non-autonomous DNA transposons. Non-autonomous DNA transposons can be trans-activated by other autonomous transposons which show similarity in TIRs.

Mutant analysis and computer assisted searches of repeated sequences led to the discovery of a new class of mobile elements namely Miniature Inverted Repeat Transposable Elements (MITEs). *Tourist*, the first reported MITE was identified in studies with the *wx*-B2 mutant of maize (Bureau and Wessler, 1992). Database searches showed that the *Tourist* MITE is associated with about 30 genes in the grass family. This was followed by the identification of *Stowaway* MITEs in sorghum, by computer-assisted database searches, a family found in other monocots as well as dicots (Bureau and Wessler, 1994).

MITEs resemble non-autonomous DNA transposons, having a small size (<600 bp) and a TIR but no transposase. Copy number of most DNA transposons per haploid genome is <100 whereas that of MITEs exceeds a thousand (Jiang and Wessler, 2001). The high copy number of MITEs and target site duplication (TSD-TA or TAA) are characteristic of MITEs. MITEs, like non-autonomous DNA transposons, originate as internal deletion derivatives of autonomous DNA

transposons, some of which attain high copy numbers for reasons still to be elucidated (Feschotte *et al.,* 2002).

Mutant analysis and database searches have identified a broad distribution of MITEs across organisms such as humans (Morgan, 1995), *C. elegans* (Oosumi *et al.*, 1996), mosquitoes (Tu, 1997), maize (Walker *et al.*, 1997), fish (Izsvak *et al.*, 1999), *Medicago* (Charrier *et al.*, 1999), rice (Jiang *et al.,* 2003; Kikuchi *et al.,* 2003; Nakazaki *et al.*, 2003), and peanut (Patel *et al.,* 2004).

Classification of MITEs

MITEs do not code for a transposase. Classification of MITEs is based on similarity of features to existing DNA transposons (which code for transposase). "Bottom up" and "top down" are the two approaches employed to classify MITEs. In the "bottom up" approach the MITE sequences are used as queries to identify autonomous DNA transposons with similar TSD and TIR (Feschotte and Mouches, 2000; Smit and Riggs, 1996). In the "top down" approach a DNA transposon is used as query to identify MITEs with similar TSD and TIR (Zhang *et al.,* 2001). Accordingly MITEs are classified into super families in relation to the super families of DNA transposons. First similarity is established between the length and sequence of TSDs. When similarity exists in the sequence and length of TSD between the MITE and DNA transposon, sequences in the TIR are subsequently considered for grouping MITEs into DNA transposon super families.

MITEs have been classified into the following transposase super families:-

1) Tc1/*mariner*
2) *PIF/Harbinger*
3) TTAA/*piggyBac*
4) *hAT*
5) *Mutator*-like

Tc1/*mariner* Super Family

The MITEs classified as *Stowaway*, show a TSD of TA dinucleotide (Bureau and Wessler, 1994), a feature shared by the Tc1/*mariner* family of DNA transposons. The "bottom up" approach was employed initially in *C. elegans* (Oosumi *et al.*, 1996) to establish the relationship between the MITEs discovered and the Tc1/*mariner* super family of transposases. Computational analysis led to the discovery of many MITE families in *C. elegans*, which could be associated to the Tc1/*mariner* transposons earlier identified in this species. The same strategy followed to assign the human MITE MER2 to the Tc1/*mariner* super family. MER2 belongs to the *Tiggers* transposon, which is similar to the pogo subgroup of Tc1/*mariner* transposons (Smit and Riggs, 1996). The *pony* MITEs in *Aedes aegypti* show a TSD of TA and similarity in TIR to the *C. elegans* Tc2 transposon, belonging to the Tc1/*mariner* super family (Tu, 2000). In plants the first report of a MITE related to the Tc1/*mariner* super family was that of *Emigrant* in *Arabidopsis thaliana*. It originated from *Lemi1* which codes for a pogo-like transposase belonging to Tc1/*mariner* group (Feschotte and Mouches, 2000). The multiple *Stowaway* MITEs reported in plants show similarity to the coding elements identified in rice and soybean, the ORF's of which resemble the animal

mariner transposons, a sub-class of Tc1/*mariner* transposons (Jarvik and Lark, 1998; Turcotte *et al*., 2001).

***PIF/Harbinger* Super Family**

Mining the *Arabidopsis* genome led to the discovery of the *Harbinger* element that possessed an ORF, encoding a transposase related to a bacterial insertion sequence (Kapitonov and Jurka, 1999). *PIF* is an autonomous DNA element identified in a mutant of the anthocyanin regulatory gene R of maize (Walker *et al.*, 1997). Maize *PIF* elements show a TSD of TTA and a 14-bp TIR. Of the various *PIF* elements the autonomous element *PIFa* was proved to mobilize the MITE *mPIF* in maize as evidenced from the sequence similarity at TIRs and sub terminal regions and a 3 bp TSD (Zhang *et al.*, 2001). Similarity exists between *PIFa* and *Harbinger* in the coding sequences. The *PIF/Harbinger* super family represented in maize, rice, *Arabidopsis*, *C. elegans* and fungi was used in the top-down approach to identify the related MITEs of the *Tourist* family.

Analysis of MITEs from different genomes suggests that most of them belong to the Tc1/*mariner* or PIF/*Harbinger* super family of transposases.

TTAA/*piggyback* Super Family

The MITEs included in this class show a TSD of TTAA and the sequence TTAAAGGRR in their TIR (Unsal and Morgan, 1995). The TTAA TSD of several MITEs suggests a link to the *piggyBac* autonomous DNA transposons in *Trichoplusiani* (Fraser *et al*., 1996).

***hAT (*hobo-Activator-Tam) Super Family**

The transposases of the *hAT* super family of transposons are characterized by approximately 50 amino acids at the C terminus. They have short TIRs and show a TSD of 8 bp

(Rubin *et al.*, 2001). Many MITEs from *Oryza sativa* are classified under the *hAT* super family of transposases.

***Mutator* like Elements (MULEs)**

MULEs are characterized by a 9-bp TSD with 40-50% identity in TIRs. They are found in rice and *Arabidopsis*. The Bigfoot MITEs of *Medicago* (Charrier *et al.*, 1999*)* and the peanut MITE share the 9-bp TSD as seen in MULEs (Patel *et al.*, 2004).

Activation of Transposons

Movement of transposons within genomes can create mutations. Many of the transposons were initially identified by analyzing mutants. The major causes for activation of transposons, and the subsequent mutations that arise, are various forms of stress such as mutagen treatment, hybridization shock, tissue culture, or pathogen infection (Casacuberta and Santiago, 2003).

Effect of tissue culture

The first active retrotransposon in plants *Tnt1*, was characterized due to its transposition into a nitrate reductase gene in tobacco (Grandbastien *et al.*, 1989). Transposition was discovered after a nitrate reductase deficient mutant was isolated from cell cultures derived from tobacco mesophyll protoplasts. Studies by Hirochika (1993) had reported that tissue culture and production of transgenics activated transposition of tobacco retrotransposons. Tobacco retrotransposons, *Tto1-Tto3,* were identified by PCR of cDNA with primers specific to the reverse transcriptase domain. Tissue culture resulted in a 10-fold increase of *Tto1*, whereas *Tto2* and *Tnt 1* were increased only 2-fold, the variation possibly due to differences in transcription levels. Production of transgenics and mutants also increased the copy number of *Tto1*. Similar results were obtained in rice tissue culture, where plants regenerated from tissue culture of the *O. japonica* variety Nipponbare showed increased copy number of the retrotransposon *Tos 17*. The

copy number of the retrotransposons *Tos 10* and *Tos 17* were increased in the culture of the Oc cell line derived from C5924, an *indica* cultivar of rice (Hirochika *et al.,* 1996).

Activity of the *Tos17* retrotransposon in rice, under tissue culture is dependent on the genotype and the site of insertion. Tissue culture of the genetically similar introgression lines RZ2 and RZ35 and the parental variety Matsumae showed different responses with respect to *Tos17* transposition. No transposition was seen in Matsumae and RZ2 under tissue culture. The introgression line RZ35, containing DNA of wild rice, showed *Tos17* transposition possibly due to introgression of wild rice DNA. This difference in response between genotypes and among introgression lines was attributed to the difference in chromatin structure due to wild DNA introgression and methylation status within the transposon (Liu *et al.,* 2004).

The rice *gh2-1* mutant, with reddish brown pigment in the internodes, hull and basal leaf sheath, identified in tissue culture of *japonica* cultivar Nipponbare, is due to the insertion of DNA transposon *nDiaZ0* into the *gh2* gene. The Nipponare genome has 16 elements with sequence similarity to *nDiaZ0* and *nDiaZ0* originated from *nDiaZ9* in chromosome 5. Of the 16 *nDiaZ* sequences, only *nDiaZ9* was found to be active under anther and scutellum tissue culture, with more transposition in scutellum culture (29.3%) than anther tissue culture (8.6%) (Huang *et al.*, 2009).

Effect of mutagens

The first MITE family was identified in the study with the maize *wx*-B2 mutant, a spontaneous null mutant (Bureau and Wessler, 1992). MITES identified in the various genera showed no coding capacity and hence were incapable of catalyzing their transposition (Bureau and Wessler, 1992; Jiang *et al.,* 2003). The first active MITE reported was the rice *mPing* by three groups independently and *Pong* was identified as the autonomous element catalyzing its transposition

(Jiang *et al.*, 2003; Kikuchi *et al.*, 2003; Nakazaki *et al.*, 2003). They were found to be active in callus derived from anther tissue culture (Kikuchi *et al.*, 2003) and in tissue cultured *indica* lines (Jiang *et al.*, 2003). The slender-glume mutant of rice was determined to be due to *mPing* insertion into the slender glume locus, upon gamma irradiation. Excision of *mPing* from the mutant allele restored the wild-type phenotype (Nakazaki *et al.*, 2003).

In peanut the first active MITE was identified by the genetic analysis of a mutant showing the high oleic acid trait. The mutation resulted from the transposition of a MITE into the *ahFAD2* gene following chemical mutation by ethylmethane sulfonate and diethyl sulfate (Patel *et al.*, 2004). Studies with the various peanut cultivars demonstrated that the peanut MITE (*AhMITE1*) can be inserted or deleted from certain chromosomal locations, spontaneously in nature and by gamma irradiation, ethylmethane sulfonate and tissue culture (Gowda *et al.*, 2011; Shirasawa *et al.*, 2012a). These insertions or deletions were found to alter plant phenotypes such as branching, seed dormancy, pod size and late leaf spot resistance.

Effect of pathogens

The first retrotransposon in plants, *Bs1*, was identified as an insertional mutant in the maize *Adh* gene, following barley stripe mosaic virus infection (Johns *et al.*, 1985). The tobacco retrotransposon *Tnt 1* is transcriptionally activated by biotic and abiotic factors such as pathogens (Tobacco Mosaic Virus), biotic elicitors (*Trichoderma viridae* fungal extracts, cryptogenin, *Erwinia chrysanthemi* culture supernatant), wounding, $CuCl_2$ and tissue culture. The transcriptional activation of *Tnt1* in response to plant defense is associated with the *cis* elements in the U3 region of the LTR, as shown in tobacco transformed with a U3-GUS construct under control of the CaMV 35S promoter (Grandbastien *et al.*, 1997). The sub-families of *Tnt1* are activated under different stress conditions. *Tnt1A* is activated by methyl jasmonate

and cryptogen, *Tnt1B* by cell culture and salicylic acid, and *Tnt 1C* by 2,4-D (Beguiristain *et al.*, 2001). This variability in activation is correlated with the sequence variability of the U3 region of the LTR.

Effect of hybridization shock

Interspecific hybridization activated the *mPing* transposon and the corresponding transposase *Pong* in rice Recombinant Inbred Lines (RIL) derived from crosses between wild rice *Zizania latifolia* and the rice cultivar Matsumae (Shan *et al.*, 2005). Wild rice does not contain *mPing* as tested by Southern blotting or *mPing* specific PCR. The RILs showed similar numbers of *mPing* bands as Matsumae but with different banding patterns. Thus the difference in banding pattern in RILs was attributed to the "genomic shock" brought about by introgression of *Zizania* DNA.

Association of MITEs with Gene Rich Areas of Genomes

After the discovery of MITEs, computer assisted searches showed that they are predominantly seen in the non-coding regions near genes. *Tourist* MITEs are associated with the genes of maize, barley, sorghum and rice (Bureau and Wessler, 1992; Bureau and Wessler, 1994). *Stowaway* MITEs are associated with the genes of both monocots and dicots (Bureau and Wessler, 1994). Computational analysis of 105 rice genomic gene sequences identified nine small inverted repeat element families, including earlier reported *Tourist, Stowaway,* p-SINE*1* and six new families (*Gaijin, Castaway, Ditto, Wanderer, Explorer,* Amy/LTP). Members of *Tourist, Stowaway* and the new families were observed to be associated with genes. Small inverted repeat members of the new families were commonly found in the 5' flanking sequences, followed by introns and 3'UTR (Bureau *et al.*, 1996).

Analysis of sequencing data has given more evidence of association of MITEs with the gene rich regions of several species. Studies on rice chromosome 4 showed that MITEs are usually

associated with euchromatic regions. They were absent in exons but found in introns (Feng *et al.*, 2002; Zhang *et al.*, 2000). In rice cell lines derived from the *indica* cultivar C5924, Transposon Display (TD) showed that 32/35 new insertions of MITE *mPing* were near single copy sequences suggesting the association of MITEs with gene rich areas (Jiang *et al.*, 2003; Kikuchi *et al.*, 2003; Nakazaki *et al.*, 2003). Similar results were obtained in database searches of the rice cultivar Nipponbare (Oki *et al.*, 2008).

Insertion preference of maize MITE *Heartbreaker* into gene rich areas was confirmed as 17/24 *Heartbreaker* flanking sequence probes showed less than 10 hybridization bands in maize genomic DNA blots (Zhang *et al.*, 2000).

Association of TEs with genes was studied in 26 long wheat genomic sequences by *in silico* methods. Among the various TEs studied 43.1% of MITEs were found to be associated with genes, even though they represented only 17.3% of the total TEs studied (Sabot *et al.*, 2005).

In a peanut genomic library enriched with *AhMITE1*, 11.5% of the flanking sequences obtained showed identity to known genes in BLASTx analysis (Shirasawa *et al.*, 2012a), but this analysis would only have identified MITEs in or very near coding sequence.

MITEs as Molecular Markers

The insertion preference of MITEs to euchromatic regions, their polymorphism between and within species, sequence similarity among members within a family, and high copy number make them an excellent molecular marker type.

The *Heartbreaker* (*Hbr*) family of MITEs in maize shows insertion polymorphism between maize and teosinte lines. To identify *Hbr* specific markers, transposon display was carried out on 100 RILs derived from maize inbreds B73 X Mo17, with 8 primer pairs corresponding to *Hbr* and *Mse*I/*Bfa*I adapter sequences. Of the 418 fragments amplified, 60.3% were polymorphic, and

adding these polymorphic markers to the maize RFLP map increased the genetic distance by 150 cM (Casa *et al.*, 2000).

In barley, Inter-MITE Polymorphism (IMP) was discovered using primers corresponding to the TIRs of *Stowaway* and *Barfly* families of MITES (Chang *et al.*, 2001). Regions between MITEs are amplified to detect IMP. *Stowaway* specific primers (TEM-1) and *Barfly* specific primers (TEM-10) produced 62.5% and 21.1% polymorphic bands. Eighty-eight polymorphic loci were mapped to the 7 linkage groups in barley in a doubled haploid segregating population. This technique eliminates the restriction and adapter ligation steps followed in TD and primers could be extended to other cereal crops too.

Transposon display in 93 doubled haploid lines derived from *Hordeum vulgare ssp.vulgare* and *H.v.ssp.spontaneum*, identified 240 segregating markers with 16 primer combinations, with one primer in each pair specific to MITE sequences in *Hordeum* (Takahashi *et al.*, 2006). A total of 369 MITE-specific polymorphic markers were developed with 11 primer combinations in sequence specific amplification polymorphism (SSAP) reactions with 38 wild and 11 cultivated varieties of barley (Lyons *et al.*, 2008).

In peanut, polymerase chain reaction with primers corresponding to flanking sequences identified insertional polymorphism of *AhMITE1* among three Virginia and one Spanish type cultivars. Of the 411 primer pairs showing bands, 41.1% were polymorphic among the four cultivars tested (Shirasawa *et al.*, 2012a). *In silico* polymorphism analysis identified *AhMITE1* markers which showed 56.4% polymorphism between peanut cultivars Satonoka and Kintoki as compared with 28.8% for genomic SSR markers. A high density peanut genetic linkage map

was constructed in the F_2 generation of 'Satonoka' X 'Kintoki', with *AhMITE1* markers and SSR markers developed from various studies, covering 2166.4 cM and 1114 loci (Shirasawa *et al.*, 2012b).

Target Enrichment and Next Generation Sequencing to Develop Molecular Markers and Identify Genes

In next generation sequencing (NGS) technologies DNA is sheared, ligated to adapters and sequenced as synthesis occurs. The entire genome is read randomly across thousands of templates and hence the process is called massively parallel sequencing. These sequencing methods produce short reads of 50 -500 bp and hence adequate coverage of the templates is needed for accurate assembly. Coverage refers to the number of overlapping reads for a particular region of the DNA.

Next Generation Sequencing Platforms

Roche GS-FLX 454 Genome Sequencing

The single stranded DNA molecules with previously ligated biotinylated adapter sequences bind to streptavidin-coated beads. The beads and PCR reagents are emulsified in water-in-oil mixture within micro reactors. PCR amplification occurs in the micro reactors for amplifying the fragments for sequencing. The beads carrying the PCR amplified fragments are added to picotiter plates in which NGS is done. Roche 454 uses the pyrosequencing technology where the pyrophosphate released by the addition of dNTP undergoes a series of reactions and produces light using the luciferase enzyme. Light emitted during each nucleotide addition is recorded. While read lengths started in the 200-400 nt range, they are now approaching 1 kb (Mardis, 2008; Varshney *et al*., 2009; Zhang *et al*., 2011).

Illumina/Solexa Genome Analyzer

Currently this is the most commonly used system. The adapter-ligated DNA is attached to a flow cell and each strand is amplified by bridge amplification for cluster generation. The sequence of DNA is determined through sequencing by synthesis approach. All four fluorescently labeled nucleotides are added for incorporation into the oligo-primed clusters. The nucleotides are synthesized using reversible terminator chemistry and each incorporation is an independent event. Each nucleotide addition is imaged for base calling. Illumina HiSeq sequencing produces 50, 100, or 150 nt single- or paired-end reads (Mardis, 2008; Varshney *et al*., 2009; Zhang *et al*., 2011), and the MiSeq now pushes read length to 2 x 250 nt.

ABI SOLiD

Single stranded DNA is attached to magnetic beads, PCR amplified and deposited onto flow cells. This technology uses DNA ligase-mediated sequencing, where the primer corresponding to an adapter is attached, followed by the addition of an 8 bp oligonucleotide, which is fluorescently labeled at the fourth and fifth nucleotides. The 8 bp oligonucleotide attaches to the 3' end after the primer. The ends are sealed by DNA ligase, followed by fluorescence detection at the fourth and fifth bp. After fluorescence detection the octamer is cleaved after the fifth base pair and prepares for a second round of sequencing (Mardis, 2008; Varshney *et al*., 2009; Zhang *et al*., 2011).

While next generation sequencing technologies are applicable to polyploid species, sequence assembly and comparative genome analysis of a polyploid species is computationally challenging and not cost-effective. Complexity reduction is possible by targeted sequencing. This method reduces the complexity of data by generating sequences from the regions of interest

and increasing the depth of sequencing. Also multiplexing with bar-coded genotypes reduces the cost of sequencing (Grover *et al*., 2012; Mamanova *et al*., 2010).

Targeted sequencing can be done to mRNA, exons, genes, organelle genes and to repetitive sequences. Sequence enrichment methods can be PCR-, hybridization-, or molecular inversion probe-based. PCR-based methods require a high degree of specificity at the target regions, but are simple and reproducible. In the hybridization-based capture method, an oligonucleotide probe complementary to the target sequence is used. In array-based hybridization, the probes are attached to a solid support and the target DNA is applied to the solid support for complementary binding to the probes, followed by washing to eliminate the unbound DNA fragments. In solution-based hybridization, biotin-labeled probes in a liquid phase bind to the target DNA and are separated by streptavidin beads. Molecular inversion probes are useful for SNP detection. The probes containing sequences complementary to the target sequence bind to the target DNA in a circular conformation with a single gap, which is filled by DNA polymerase and ligase. The circularized probes are cleaved by exonuclease, and then amplified by adapter specific primers (Cronn *et al*., 2012; Mamanova *et al*., 2010; Mertes *et al*., 2011).

Targeted sequencing was successfully employed to determine variants in the human genome (Altshuler *et al*., 2010). Sequence capture and re-sequencing of a 2.2 Mb maize chromosome interval and 43 genes dispersed in the maize genome identified approximately 2500 SNPs in the target regions of maize inbred Mo17 compared with the reference B73 genome (Fu *et al*., 2010). In the soybean reference cultivar, Williams 82, intracultivar variability was identified by sequence capture to an oligonucleotide microarray followed by Illumina sequencing (Haun *et al*., 2011).

In maize photosynthetic mutant lines, sequence capture and Illumina sequencing of the high copy *Mutator* (*Mu*) transposon-based sequences identified four genes linked to chloroplast biogenesis (Williams-Carrier *et al*., 2010).

As MITEs are commonly found to be associated with genes, they may play a role in gene regulation and expression. Furthermore, the insertion site polymorphism of MITEs makes them excellent markers. Targeted sequencing of the MITE-rich regions of a genome also is expected to provide information about the gene rich regions. Insertional polymorphisms of the target sequence can be identified by sequencing multiple genotypes.

Transposons as Insertional Mutagens and Gene Tagging

Mutations are employed to study gene function or for developing novel genotypes. Ionizing radiation or chemical mutagens are efficient mutagens as they cause DNA damage. An alternative approach is insertional mutagenesis, wherein a T-DNA or transposon is introduced into the genome. Insertions cause gene modification depending on the site of insertion such as in regulatory or coding sequences. The inserted sequences serve as tags, as their sequence is known, and the genes disrupted are identified with respect to the tags (Tadege *et al*., 2005).

The most widely studied transposons as insertional mutagens are the rice *Tos17*, tobacco *Tnt1* and maize *Ac/Ds* systems (Ayliffe and Pryor, 2009). The rice transposon *Tos17* is widely activated under tissue culture, and analysis of sequences flanking transposed *Tos17* in regenerated plants showed amino acid similarity to genes involved in thiamine biosynthesis in *Escherichia coli*, $NADP^+$-dependent malic enzyme, phytochrome A in rice and S-receptor kinase-like in *Brassica*, indicating the preferential insertion near genes (Hirochika *et al*., 1996). BLASTx analysis of 123 flanking sequences derived by TAIL-PCR and Inverse PCR revealed

that one third of the *Tos17* insertions were near genes. Further confirmation for this was obtained when probes corresponding to the flanking sequences detected only one or two bands in genomic Southern blotting experiments (Yamazaki *et al.*, 2001). Preference of *Tos17* for genic regions was also observed during analysis of >42,000 new insertion sites of *Tos17* in rice (Miyao *et al*., 2003). During *in vitro* transformation and regeneration of *Medicago truncatula*, the tobacco transposon *Tnt1* was found to be transposed to coding regions in 12 out of 38 sequences analyzed (d'Erfurth *et al.*, 2003). In a large-scale insertional mutant collection of *M. truncatula*, developed by *Tnt1* activation in the transgenic line Tnk88-7-7 by somatic embryogenesis, analysis of flanking sequence tagged sites identified *Tnt1* insertions in coding exons (34.1%), introns plus UTRs (23.2%) and intergenic regions (42.6%) (Tadege *et al*., 2008). In the heterologous transformation system with the tobacco transposon *Tnt1* in lettuce, analysis of 25 *Tnt1* flanking sequences showed that the *Tnt1* insertions were near gene-rich areas (Mazier *et al.*, 2007). A dwarf phenotypic mutant identified in the study was due to disruption of the gibberellin 3 beta-hydroxylase gene by *Tnt1* insertion. Similarly in *Arabidopsis thaliana* analysis of sequences flanking the *Tnt1* transposon upon transformation and *in vitro* culture showed that 73% of insertions were near coding regions (Courtial *et al.*, 2001). In soybean the rice MITE, *mPing*, is more suitable for activation tagging than the maize *Ac/Ds* or rice *Tos17* transposons. Activation of *mPing* in soybean is developmentally regulated and shows an insertion preference within 2.5 kb of genes (Hancock *et al*., 2011).

Apart from the heterologous transformation systems, endogenous transposons can be activated by a variety of factors such as tissue culture, chemical mutagens or ionizing radiation, pathogens, hybridization shock, etc. Large-scale sequencing data reveal the presence of transposons in genomes (Sabot *et al*., 2005; Schnable *et al*., 2009). Activation of endogenous or heterologous

transposons can produce mutants, which can be utilized for gene functional analysis (Nakazaki *et al*., 2003; Patel *et al*., 2004).

Silencing of Transgenes in Plants

Transgene silencing is epigenetic and can be transcriptional or post-transcriptional. Transcriptional gene silencing (TGS) leads to repression of mRNA synthesis by blocking transcription. In post-transcriptional gene silencing (PTGS) degradation of the transcribed mRNA occurs.

Transcriptional gene silencing (TGS) can occur due to integration of multiple copies of the transgene (Assaad *et al*., 1993), or when transgenes get integrated near transcriptionally inactive heterochromatic regions (Prols and Meyer, 1992) or near methylated repeated elements (tenLohuis *et al*., 1995), or by modification of chromatin structure at transgene–genomic junctions (Fagard and Vaucheret, 2000; van Blokland *et al*., 1997). The role of microRNA (miRNA), endogenous antisense small interfering RNA (siRNA) and tiny noncoding RNA (tncRNA) in TGS is currently being investigated as they have been found to create epigenetic modifications in the genome by cytosine methylation and histone modification (Matzke *et al.*, 2004). TGS is associated with hypermethylation of cytosines along with di-methylation of histone H3 at lysine in position nine (H3K9me2), resulting in a repressed chromatin state (Law & Jacobsen, 2010).

The phenomenon of post-transcriptional gene silencing of an endogenous gene was first observed in transgenic *Petunia* carrying the coding region of the chalcone synthase gene for over-expression of anthocyanin pigment (Napoli *et al*., 1990). Instead of increasing petal pigmentation, it was eliminated resulting in white flowers. This phenomenon was termed co-suppression as some of the transgenic *Petunia* lost the activity of both the endogenous gene and

the introduced gene, though run-on transcription tests showed no reduction in mRNA. The phenomenon of co-suppression observed in *Petunia* was also observed in other crops (Ingelbrecht *et al*., 1994), fungi (Cogoni *et al*., 1996) and worms (Fire *et al*., 1991). The introduction of antisense RNA was also effective in suppressing gene expression in *Caenorhabditis elegans* (Fire *et al*., 1991). The role of double-stranded RNA in PTGS, where sequence similarity to transgene mRNA lead to complete silencing of the transgene, was demonstrated by studies in *C. elegans* (Fire *et al*., 1998). Co-suppression in plants or RNA interference (RNAi) in animals (Rocheleau *et al*., 1997) or quelling in fungi are mechanistically similar, where translation of the mRNA is inhibited or its abundance is reduced by double stranded RNA sharing sequence similarity to the endogenous gene (Dietz-Pfeilstetter, 2010; Hammond *et al*., 2001).

Mechanism of Gene Silencing by RNA Interference

Double-stranded RNA (dsRNA) is the initiator of RNAi (Fire *et al*., 1998; Hamilton and Baulcombe, 1999). Double-stranded RNA is produced upon transcription of endogenous genes with inverted repeats, or transgenes with DNA in sense and anti-sense orientation. Silencing of mRNA has initiation and effector steps (Cerutti, 2003). In the initiator step DICER-like (DCL) proteins, which are dsRNA specific RNaseIII family ribonucleases, cleave the dsRNA into small interfering RNA (siRNA) of 21-26 nucleotides in length (Bernstein *et al*., 2001; Hamilton and Baulcombe, 1999; Zamore *et al*., 2000). Each strand of the siRNA has 5’ phosphate and 3’hydroxyl ends and is paired in a way to leave a 2 bp overhang at the 3’ end (Elbashir *et al*., 2001b). In the effector step the siRNAs are assembled into the multi protein RNA induced silencing complex (RISC) (Hammond *et al*., 2000). The siRNA binding proteins such as, R2D2 in *Drosophila* and RDE-4 in *Caenorhabditis*, facilitate the transfer of siRNA to RISC (Liu *et al*.,

2003); RISC contains Argonaute proteins, characterized by the presence of N-terminal, PAZ, mid and C-terminal PIWI domains. The enzymatic activity of the PIWI domain cleaves the non-active strand of the double stranded siRNA (passenger strand). The RNA strand, which has low thermodynamic stability at its 5' end, will remain attached to the RISC (guide strand) (Schwarz *et al*., 2003). The PAZ domain of Argonaute proteins binds to the 3' end of the siRNA and the 5' end of siRNA binds to the mid domain of Argonaute proteins (Hutvagner and Simard, 2008). Bases 2-6 of the siRNAs comprise the 'seed region', which is critical for target recognition. The RISC is guided to the target mRNA by the siRNAs. Once a complementary mRNA is identified, bases at positions 2-8 of the guide siRNA in the Argonaute protein form Watson-Crick base pairing with target mRNA, which is held in the Argonaute protein complex. The RNAse activity of the PIWI domain cleaves the mRNA 10 nucleotides from the 5' end of the siRNA. (Elbashir *et al*., 2001a; Song *et al.*, 2004).

Effect of Environmental Factors on Genome Stability and Expression of Transgenes

Crop improvement by a transgenic approach requires stable transgene expression across generations in different environments. Successful development of an agronomically superior transgenic crop demands the stability of the introduced gene under various endogenous and exogenous stimuli. Exogenous stimuli are environmental factors namely temperature, light intensity, light duration, fertility and soil conditions. At the molecular level stability of the transgene can be influenced by methylation of promoter or coding region (Matzke *et al*., 1989), ectopic gene pairing, copy number and expression level of the transgene (Chareonpornwattana *et al*., 1999), chromatin structure and sequences near the transgene integration site (Dietz-Pfeilstetter, 2010; Iglesias *et al*., 1997).

Salmon red colored flowers were produced in *Petunia* transformed with a maize *A1* gene under the control of the CaMV 35S promoter. However, in field-grown plants developed from the cross RL01-17 (homozygous red) X RL01 (homozygous white), approximately 60% of plants showed light red flowers compared with 5% in greenhouse-grown plants. This was seen after a 3-week period of high temperature (36ºC) and high light intensity, and as shown by molecular analysis, methylation of the transgene promoter resulted in light colored flowers. This effect was less pronounced in plants grown from seeds recovered from early crosses (Meyer *et al*., 1992). The age of plants at flowering and season of seed production were also suggested to be reasons for changes in expression of the transgene along with the environmental effects.

Agronomic practices can also affect transgene expression as seen in studies of transgenic tobacco carrying the mutant *A. thaliana* acetohydroxyacid synthase (AHAS) gene (*csr1-1*), which imparts resistance to the herbicide chlorsulfuron. Transgenic tobacco plants carrying the *csr1-1* gene expressed from the CaMV 35S promoter, showed susceptibility to chlorsulfuron after field transplantation, although greenhouse-grown plants were resistant to chlorsulfuron. In greenhouse studies, with plants grown with and without transplanting, herbicide susceptibility was shown only by transplanted seedlings (Brandle *et al*., 1995).

Transgenic rice (#354) carrying multiple copies of a rice chitinase gene under control of the CaMV 35S promoter showed silencing of the transgene in 23% of the progeny in the T_3 generation, although intact transgene expression was seen in T_0, T_1 and T_2 generations. Plant #174 which had the same transgene copy number as plant #354, but 5-10-fold reduced expression of the transgene showed no silencing. Thus in this study, an RNA threshold model was proposed to explain the silencing of the transgene (Chareonpornwattana *et al*., 1999).

In *Gladiolus*, expression of GUS protein under *UBQ3*, CaMV 35S and *rolD* promoters showed considerable variation under greenhouse and field conditions in a two year trial. Expression of GUS protein was significantly higher under outdoor conditions with CaMV 35S and *rolD* promoters, whereas it was low with the *UBQ3* promoter. Expression of GUS varied in the field between the two years with higher expression seen in the second year except for GUS under CaMV 35S, which showed lower expression in field during second year. GUS expression of one transgenic line containing the *UBQ3* promoter was always higher in greenhouse grown plants. The glasshouse had a higher lower limit of temperature (19-22ºC) when compared to field conditions (6-17ºC) and temperature was suggested to have influenced GUS expression under field conditions along with endogenous factors, namely number of transgenes and transgene insertion site in the genome (Kamo, 2008).

Nicotiana benthamiana plants carrying a GFP transgene showed varied silencing response under high and low light intensities (Kotakis *et al*., 2010). The frequency of GFP silencing was increased under high light intensities when the temperature regime was kept constant. The difference in systemic silencing was more pronounced than short-range silencing between the two light intensities. Quantitative RT-PCR showed higher expression of *DCL1* and *DCL3*, the genes involved in silencing pathways, under high light intensities. Northern blot analysis showed higher amounts of GFP transcripts in silenced and non-silenced lines under low light intensity in comparison to higher light intensity.

Contrasting environmental stresses like high temperature of 31°C, water stress and nutrient stress produced no significant change of *uid*A and *bar* gene expression in transgenic barley (T_3#30), carrying inverted repeats of the transgene (Meng *et al*., 2006). Molecular analysis showed no difference in transgene methylation pattern between transgenic lines grown under stress and

normal conditions. However, methylation of promoter and chromatin modifications were observed in transgene methylation assay and DNase I assay in transgene silenced lines produced by *in vitro* culture of (T_3#30).

Sulfur Metabolism in Plants

Plant roots take up sulfur as sulfate, the oxidized form. Multiple and highly regulated trans-membrane sulfate transporters are involved in its transfer from the roots to leaves (Hawkesford, 2000). All the metabolic reactions that reduce sulfate to sulfide for cysteine biosynthesis mainly take place in the leaves. Sulfate entering the leaves is stored in the vacuole or transferred to the chloroplast for reduction. In the chloroplast, sulfate is activated by linking to 5'-AMP to form adenosine 5'-phosphosulfate (APS), catalyzed by ATP-sulfurylase (ATPS). APS reductase reduces APS to sulfite, which is converted to sulfide by sulfite reductase. Inorganic sulfide is incorporated into cysteine by sulfhydrylation of *O*-acetyl serine (OAS) catalyzed by OAS-thiol lyase (OASTL). *O*-acetyl serine is formed upon acetylation of serine by acetyl-COA and the enzyme serine acetyl transferase (SATase). *O*-aetylserine thiol lyase and SATase together form the cysteine synthase complex (Bick and Leustek, 1998; Droux, 2004; Hawkesford, 2000; Hell, 1997).

Cysteine is the first reduced sulfur compound in plants and serves as the source of various sulfur compounds essential for plant metabolism. Methionine, an essential sulfur-containing amino acid is derived from cysteine. Other important cysteine derivatives are glutathione (GSH), a sulfur reservoir and transporter, vitamins such as thiamine and biotin, co-factors and iron-sulfur proteins. Methionine is the precursor of S-adenosyl-methionine (SAM), which serves as the primary methyl group donor in a multitude of biological processes including cell wall development, DNA replication and synthesis of secondary metabolites (Droux, 2004).

Methionine is an essential amino acid as it cannot be synthesized by humans and farm animals and has to be obtained from plants. Among the essential amino acids lysine, tryptophan (in cereals) and methionine (in legumes) are the most limiting (Droux, 2004). Enhancing the protein quality by increasing the essential amino acid content is important, with regard to human nutrition and in plant improvement. The free amino acids in the cytoplasm of seed cells are the source for protein synthesis, and the proportion of the sulfur containing amino acids within the plant can be manipulated by:

a) Increasing sulfur availability to plants which in turn alters the type of protein subunits synthesized;
b) Developing transgenic plants carrying sulfur-rich seed storage proteins.

Plant Sulfur Nutrition and its Effect on Seed Protein Content

Seed protein profile variation and physiological response of plants to surplus or deficiency of sulfur have been studied in multiple crops and in transgenic plants carrying an added sulfur sink. In soybean the major seed storage proteins (SP) are glycinin (11S) and *β*–conglycinin (7S). Of the total seed proteins, glycinin and *β*–conglycinin account for 52% and 35%, respectively (Kinsella, 1979). Sulfur amino acids in the five subunits of glycinin range from 3 to 4.5% (Fukushima, 1991). The *β* –conglycinin is composed of *α* (72 kDa), *α'* (76 kDa) and *β* (53 kDa) subunits. The sulfur amino acid content of *β*–conglycinin is less than 1% (Sebastiani *et al*., 1990). Higher 11S/7S ratio is indicative of higher protein quality in soybean. In *in vitro* studies with soybean cotyledons grown with 0, 0.017, 1.5 and 7.5 mM concentrations of SO_4, it was observed that the 11S/7S ratio of protein was >1 when the sulfate concentration was 7.5 mM and between 0.6 and 0.7 for the other three sulfate concentrations, suggestive of the role of sulfur in improving the protein quality. Supplementing the soybean cotyledon growth medium with L-

methionine was also found to increase the 11S/7S ratio to >1 under all sulfate levels indicative of equivalence of high sulfur to methionine (Holowach *et al*., 1984).

Greenhouse and field studies conducted to study the relationship between the sulfur-rich glycinin and sulfur-poor *β*–subunit of *β*–conglycinin in soybeans under varying sulfur levels showed that the relative amounts of the 11S and 7S protein fractions was related to the N:S ratio, which in turn was dependent on soil sulfur status. The N:S ratios in cotyledons decreased and 11S/7S ratios increased with increased sulfur concentrations (Sexton *et al*., 1998).

Nutrient studies in soybean showed that despite sulfur or potassium deficiencies, the total protein per seed at maturity was almost similar to the controls. In sulfur-deficient seeds the amount of sulfur-rich glycinin was reduced by half that of the control with an increase in sulfur-poor *β*-conglycinin. The *β*-subunit of *β*-conglycinin increased 3-fold though *α* and α'-subunits remained unaffected. The differential response of *β*-conglycinin subunits to stress shows their specificity to stress and separate regulatory effects as each subunit is specified by a separate mRNA (Gayler and Sykes, 1985; Sexton *et al*., 1998).

The ratio of total nitrogen to sulfur increased in whole seed and total protein when *Lupinus angustifolius* was grown under low sulfur concentrations. The high nitrogen to sulfur ratio was associated with lower proportions of conglutin *α* and *γ*, which are sulfur-containing amino acids. Also there was a higher content of sulfur poor conglutin- *β* (Blagrove, 1976).

Studies by Randall *et al.* (1979) showed a similar trend in *Pisum.* In *Pisum sativum* seeds, the globulin protein fraction consists of legumin (11S) and vicilin (7S). Legumins have a higher proportion of sulfur containing amino acids than vicilin. Under sulfur deficiency there was a relative decrease in legumin along with an increase in vicilin. Low sulfur was also seen to reduce

an unknown cotyledonary protein fraction of 22 kDa thus affecting the total protein profile (Randall, 1979).

Total protein extract from barley grown under sulfur poor (no sulfur) and sulfur sufficient (36 ppm sulfur as $MgSO_4 \cdot 7H_2O$) conditions showed considerable differences in protein profiles and amino acid content. Gel electrophoresis of the hordein fraction showed a reduction of sulfur-rich 'B' and 'C' hordein polypeptides and an increase in sulfur poor 'C' polypeptide in the low sulfur treatments. A high proportion of the sulfur-poor 'C' polypeptide increased the glutamic acid, glutamine, proline and phenylalanine content of the total hordein fraction with a reduction in cysteine and methionine content (Shewry *et al*., 1983).

Proteins from legumes are poor in the sulfur-containing amino acid methionine whereas those from cereals are poor in lysine and tryptophan. A method to improve the essential amino acid content in these crops is the heterologous expression of proteins rich in essential amino acids (Beauregard and Hefford, 2006). Nutritional quality of legume proteins can be improved by expressing a methionine-rich protein from other plant sources in the legume seed. Sulfur amino acid rich plant proteins include 18 kDa Zein with 37% methionine (Chui, 2003), 10 kDa Zein with 30% methionine (Kirihara, 2001), 2S albumin from Brazil nut with 25% methionine (Ampe *et al*., 1986), sunflower seed albumin with 16% methionine and 8% cysteine residues (Kortt *et al*., 1991).

Transgenic petunia harboring the *β*-subunit of *β*-conglycinin from soybean showed enhanced accumulation of the *β*-subunit when grown under sulfur deficiency. Exogenous methionine repressed the accumulation of the *β*-subunit in *in vitro* cultured seeds (Fujiwara *et al*., 1992). Petunia seed storage proteins were unaffected by the transgene and the treatments. Gene

regulation in response to sulfur nutrition is conserved even in the transformed plant as evidenced from this experiment.

Field trials with transgenic lupin plants (*Lupinus angustifolius* L.), carrying sunflower seed albumin protein with a seed specific pea vicilin promoter, showed low levels of oxidized sulfur and high levels of reduced sulfur (amino acids) when compared with non-transgenic controls. Methionine levels in seeds were increased by 94% and cysteine content was reduced by 12% in transgenics, although no difference was found in total seed sulfur or nitrogen content when compared to non-transgenics. Increase in methionine with a decrease in cysteine in transgenics suggests a change in other protein and non-protein fractions of seeds (Molvig *et al*., 1997).

Studies with chickpea (Chiaiese *et al*., 2004) and lupin (Tabe and Droux, 2002) plants, carrying the sunflower seed albumin gene, have shown that the sulfur amino acid content of transgenic seeds is higher than controls, but less than the expected levels, as the cysteine levels are found to be low in these seeds. The uptake and total seed sulfur increased with increased sulfur supply, both in transgenic plants carrying the sunflower seed albumin gene, and in non-transgenic controls. The added transgenic sink had no effect in increasing sulfur uptake although it increased the partitioning of oxidized sulfur to reduced sulfur in the transgenic seeds. The activity of cystathionine-γ-synthase was observed to increase in transgenic lupin compared with various enzymes in the sulfur assimilation and methionine biosynthetic pathways (Tabe and Droux, 2002). This is the first enzyme in the methionine biosynthetic pathway and would account for the increased methionine concentration in transgenic seeds. Lack of increase in cysteine in transgenic plants can possibly be due to the down regulation of endogenous proteins, rich in cysteine, such as conglutin-δ in lupins (Tabe and Droux, 2002) or a cysteine-rich trypsin inhibitor in chick pea (Chiaiese *et al*., 2004) .

Conclusion

Breeding and selecting the plants with the desired genotype and phenotype can be expedited by the application of molecular techniques for crop improvement. Application of RNAi, a reverse genetics technique, was successful in developing a peanut line with reduced allergens. Advancement of this line for varietal release or for further breeding depends on the stability of silencing. Stability of transgene efficacy has to be tested in multiple generations and multiple environments, which is done in the current study. Also any crop improvement program should integrate the various fields of research such as development of genetic markers, identifying traits linked to markers, developing suitable genetic engineering techniques for each crop, incorporating recent sequencing approaches and testing with various agronomic approaches for sustainable and economic crop production. The use of a peanut transposon, *AhMITE1*, for peanut sequencing by target enrichment, gene identification and marker development in peanut is also attempted in this study.

References

Altshuler D.L., Durbin R.M., Abecasis G.R., Bentley D.R., Chakravarti A., Clark A.G., Collins F.S., De la Vega F.M., Donnelly P., Egholm M., Flicek P., Gabriel S.B., Gibbs R.A., Knoppers B.M., Lander E.S., Lehrach H., Mardis E.R., McVean G.A., Nickerson D., Peltonen L., Schafer A.J., Sherry S.T., Wang J., Wilson R.K., Deiros D., Metzker M., Muzny D., Reid J., Wheeler D., Li J., Jian M., Li G., Li R., Liang H., Tian G., Wang B., Wang J., Wang W., Yang H., Zhang X., Zheng H., Ambrogio L., Bloom T., Cibulskis K., Fennell T.J., Jaffe D.B., Shefler E., Sougnez C.L., Gormley N., Humphray S., Kingsbury Z., Koko-Gonzales P., Stone J., McKernan K.J., Costa G.L., Ichikawa J.K., Lee C.C., Sudbrak R., Borodina T.A., Dahl A., Davydov A.N., Marquardt P., Mertes F., Nietfeld

W., Rosenstiel P., Schreiber S., Soldatov A.V., Timmermann B., Tolzmann M., Affourtit J., Ashworth D., Attiya S., Bachorski M., Buglione E., Burke A., Caprio A., Celone C., Clark S., Conners D., Desany B., Gu L., Guccione L., Kao K., Kebbel A., Knowlton J., Labrecque M., McDade L., Mealmaker C., Minderman M., Nawrocki A., Niazi F., Pareja K., Ramenani R., Riches D., Song W., Turcotte C., Wang S., Dooling D., Fulton L., Fulton R., *et al.* (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061-1073.

Ampe C., Vandamme J., Decastro L.A.B., Sampaio M., Vanmontagu M., Vandekerckhove J. (1986) The amino-acid-sequence of the 2S sulfur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* HBK). Eur. J. Biochem. 159:597-604.

Assaad F.F., Tucker K.L., Signer E.R. (1993) Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis.* Plant Mol. Biol. 22:1067-1085.

Ayliffe M.A., Pryor A.J. (2009) Transposon-based activation tagging in cereals. Funct. Plant Biol.36:915-921.

Beauregard M., Hefford M.A. (2006) Enhancement of essential amino acid contents in crops by genetic engineering and protein design. Plant Biotechnol. J. 4:561-574.

Beguiristain T., Grandbastien M.A., Puigdomenech P., Casacuberta J.M. (2001) Three *Tnt1* subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol. 127:212-221.

Bernstein E., Caudy A.A., Hammond S.M., Hannon G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363-366.

Bick J.A., Leustek T. (1998) Plant sulfur metabolism - the reduction of sulfate to sulfite. Curr. Opin. Plant Biol. 1:240-244.

Blagrove R.J., Gillespie, M. and Randall, P.J. (1976) Effect of sulfur supply on the seed globulin composition of *Lupinus angustifolius*. Aust. J. Plant Physiol. 3:173-184.

Brandle J.E., McHugh S.G., James L., Labbe H., Miki B.L. (1995) Instability of transgene expression in-field grown tobacco carrying the CSRL-L gene for sulfonylurea herbicide resistance. Bio-Technology 13:994-998.

Bureau T.E., Wessler S.R. (1992) *Tourist* - a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283-1294.

Bureau T.E., Wessler S.R. (1994) *Stowaway* - a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6:907-916.

Bureau T.E., Ronald P.C., Wessler S.R. (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc. Nat. Acad. Sci. USA 93:8524-8529.

Cantu D., Vanzetti L.S., Sumner A., Dubcovsky M., Matvienko M., Distelfeld A., Michelmore R.W., Dubcovsky J. (2010) Small RNAs, DNA methylation and transposable elements in wheat. BMC Genomics 11:408

Casa A.M., Brouwer C., Nagel A., Wang L.J., Zhang Q., Kresovich S., Wessler S.R. (2000) The MITE family *Heartbreaker (Hbr)*: Molecular markers in maize. Proc. Nat. Acad. Sci.USA 97:10083-10089.

Casacuberta J.M., Santiago N. (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1-11.

Cerutti H. (2003) RNA interference: traveling in the cell and gaining functions? Trends Genet. 19:39-46.

Chang R.Y., O'Donoughue L.S., Bureau T.E. (2001) Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. Theor. Appl. Genet. 102:773-781.

Chiaiese P., Ohkama-Ohtsu N., Molvig L., Godfree R., Dove H., Hocart C., Fujiwara T., Higgins T.J.V., Tabe L.M. (2004) Sulphur and nitrogen nutrition influence the response of chickpea seeds to an added, transgenic sink for organic sulphur. J. Exp. Bot. 55:1889-1901.

Chareonpornwattana S., Thara K.V., Wang L., Datta S.K., Panbangred W., Muthukrishnan S. (1999) Inheritance, expression, and silencing of a chitinase transgene in rice. Theor. Appl. Genet. 98:371-378.

Charrier B., Foucher F., Kondorosi E., d'Aubenton-Carafa Y., Thermes C., Kondorosi A., Ratet P. (1999) *Bigfoot*: a new family of MITE elements characterized from the *Medicago* genus. Plant J.18:431-441.

Chui C.F.C., Falco, S.C., Rice, J.A. and Knowlton, S. (2003) High sulphur seed protein gene and method for increasing the sulfur amino acid content of plants. Canadian Patent CA 2104022.

Cogoni C., Irelan J.T., Schumacher M., Schmidhauser T.J., Selker E.U., Macino G. (1996) Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. EMBO J.15:3153-3163.

Courtial B., Feuerbach F., Eberhard S., Rohmer L., Chiapello H., Camilleri C., Lucas H. (2001) *Tnt1* transposition events are induced by in vitro transformation of *Arabidopsis thaliana,* and transposed copies integrate into genes. Mol. Genet. Genomics 265:32-42.

Cronn R., Knaus B.J., Liston A., Maughan P.J., Parks M., Syring J.V., Udall J. (2012) Targeted enrichment strategies for next-generation plant biology. Am. J. Bot. 99:291-311.

d'Erfurth I., Cosson V., Eschstruth A., Lucas H., Kondorosi A., Ratet P. (2003) Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*. Plant J.34:95-106.

Dietz-Pfeilstetter A. (2010) Stability of transgene expression as a challenge for genetic engineering. Plant Sci. 179:164-167.

Droux M. (2004) Sulfur assimilation and the role of sulfur in plant metabolism: a survey. Photosynth. Res. 79:331-348.

Du C., Hoffman, C., He L., Caronna J., Dooner H. K. (2011) The complete *Ac/Ds* transposon family of maize. BMC Genomics 12:588

Elbashir S.M., Lendeckel W., Tuschl T. (2001a) RNA interference is mediated by 21-and 22-nucleotide RNAs. Genes & Dev.15:188-200.

Elbashir S.M., Martinez J., Patkaniowska A., Lendeckel W., Tuschl T. (2001b) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J.20:6877-6888.

Fagard M., Vaucheret H. (2000) (Trans)gene silencing in plants: How many mechanisms? Ann.Rev. Plant Physiol. Plant Mol. Biol. 51:167-194.

Feng Q., Zhang Y.J., Hao P., Wang S.Y., Fu G., Huang Y.C., Li Y., Zhu J.J., Liu Y.L., Hu X., Jia P.X., Zhang Y., Zhao Q., Ying K., Yu S.L., Tang Y.S., Weng Q.J., Zhang L., Lu Y.,

Mu J., Lu Y.Q., Zhang L.S., Yu Z., Fan D.L., Liu X.H., Lu T.T., Li C., Wu Y.R., Sun T.G., Lei H.Y., Li T., Hu H., Guan J.P., Wu M., Zhang R.Q., Zhou B., Chen Z.H., Chen L., Jin Z.Q., Wang R., Yin H.F., Cai Z., Ren S.X., Lv G., Gu W.Y., Zhu G.F., Tu Y.F., Jia J., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Chen W., Lan L.F., Lai Y., Cheng Z.K., Gu M.H., Jiang J.M., Li J.Y., Hong G.F., Xue Y.B., Han B. (2002) Sequence and analysis of rice chromosome 4. Nature 420:316-320.

Feschotte C., Mouches C. (2000) Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a *Pogo*-like DNA transposon. Mol. Biol. Evol. 17:730-737.

Feschotte C., Pritham E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. Annual Review of Genetics 41: pp. 331-368.

Feschotte C., Jiang N., Wessler S.R. (2002) Plant transposable elements: Where genetics meets genomics. Nat. Rev. Genet. 3:329-341.

Fire A., Albertson D., Harrison S.W., Moerman D.G. (1991) Production of antisense RNA leads to effective and specific-inhibition of gene-expression in *C. elegans* muscle. Development 113:503-514.

Fire A., Xu S.Q., Montgomery M.K., Kostas S.A., Driver S.E., Mello C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806-811.

Fraser M.J., Coszczon T., Elick T., Bauser C. (1996) Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. Insect Mol. Biol. 5:141-151.

Fu Y., Springer N.M., Gerhardt D.J., Ying K., Yeh C.-T., Wu W., Swanson-Wagner R., D'Ascenzo M., Millard T., Freeberg L., Aoyama N., Kitzman J., Burgess D., Richmond T., Albert T.J., Barbazuk W.B., Jeddeloh J.A., Schnable P.S. (2010) Repeat subtraction-mediated sequence capture from a complex genome. Plant J. 62:898-909.

Fujiwara T., Hirai M.Y., Chino M., Komeda Y., Naito S. (1992) Effects of sulfur nutrition on expression of the soybean seed storage protein genes in transgenic *Petunia*. Plant Physiol. 99:263-268.

Fukushima D. (1991) Recent progress of soybean protein foods: Chemistry, technology, and nutrition. Food Reviews International 7:323-351.

Gayler K.R., Sykes G.E. (1985) Effects of nutritional stress on the storage proteins of soybeans. Plant Physiol. 78:582-585.

Gowda M.V.C., Bhat R.S., Sujay V., Kusuma P., Varshakumari, Bhat S., Varshney R.K. (2011) Characterization of *AhMITE1* transposition and its association with the mutational and evolutionary origin of botanical types in peanut (*Arachis spp*.). Plant Syst. Evol. 291:153-158.

Grandbastien M.A., Spielmann A., Caboche M. (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant-cell genetics. Nature 337:376-380.

Grandbastien M.A., Lucas H., Morel J.B., Mhiri C., Vernhettes S., Casacuberta J.M. (1997) The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. Genetica 100:241-252.

Grover C.E., Salmon A., Wendel J.F. (2012) Targeted sequence capture as a powerful tool for evolutionary analysis. Am. J. of Bot. 99:312-319

Hamilton A.J., Baulcombe D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950-952.

Hammond S.M., Caudy A.A., Hannon G.J. (2001) Post-transcriptional gene silencing by double-stranded RNA. Nat. Rev.Genet. 2:110-119.

Hammond S.M., Bernstein E., Beach D., Hannon G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. Nature 404:293-296.

Han S.J. (2010) Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mob. DNA 1:15

Hancock C.N., Zhang F., Floyd K., Richardson A.O., LaFayette P., Tucker D., Wessler S.R., Parrott W.A. (2011) The rice miniature inverted repeat transposable element *mPing* is an effective insertional mutagen in soybean. Plant Physiol. 157:552-562.

Haun W.J., Hyten D.L., Xu W.W., Gerhardt D.J., Albert T.J., Richmond T., Jeddeloh J.A., Jia G., Springer N.M., Vance C.P., Stupar R.M. (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol. 155:645-655.

Hawkesford M.J. (2000) Plant responses to sulphur deficiency and the genetic manipulation of sulphate transporters to improve S-utilization efficiency. J. Exp. Bot. 51:131-138.

Hell R. (1997) Molecular physiology of plant sulfur metabolism. Planta 202:138-148.

Hirochika H. (1993) Activation of tobacco retrotransposons during tissue culture. EMBO J. 12: 2521-2528

Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H., Kanda M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Nat. Acad. Sci. USA 93:7783-7788.

Holowach L.P., Thompson J.F., Madison J.T. (1984) Storage protein-composition of soybean cotyledons grown-invitro in media of various sulfate concentrations in the presence and absence of exogenous l-methionine. Plant Physiol. 74:584-589.

Huang J., Zhang K., Shen Y., Huang Z., Li M., Tang D., Gu M., Cheng Z. (2009) Identification of a high frequency transposon induced by tissue culture, *nDaiZ*, a member of the *hAT* family in rice. Genomics 93:274-281.

Hutvagner G., Simard M.J. (2008) Argonaute proteins: key players in RNA silencing. Nat. Rev.Mol.Cell Biol. 9:22-32.

Iglesias V.A., Moscone E.A., Papp I., Neuhuber F., Michalowski S., Phelan T., Spiker S., Matzke M., Matzke A.J.M. (1997) Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. Plant Cell 9:1251-1264.

Ingelbrecht I., Vanhoudt H., Vanmontagu M., Depicker A. (1994) Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. Proc. Nat. Acad. Sci. USA 91:10502-10506.

Izsvak Z., Ivics ZUSA., Shimoda N., Mohn D., Okamoto H., Hackett P.B. (1999) Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol.Evol. 48:13-21.

Jarvik T., Lark K.G. (1998) Characterization of *Soymar1*, a mariner element in soybean. Genetics 149:1569-1574.

Jiang N., Wessler S.R. (2001) Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. Plant Cell 13:2553-2564.

Jiang N., Feschotte C., Zhang X.Y., Wessler S.R. (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr. Opin. Plant Biol. 7:115-119.

Jiang N., Bao Z.R., Zhang X.Y., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. (2003) An active DNA transposon family in rice. Nature 421:163-167.

Johns M.A., Mottinger J., Freeling M. (1985) A low copy number, *copia*-like transposon in maize. EMBO J. 4:1093-1101.

Kamo K. (2008) Transgene expression for *Gladiolus* plants grown outdoors and in the greenhouse. Scientia Horticulturae 117:275-280.

Kapitonov V.V., Jurka J. (1999) Molecular paleontology of transposable elements from *Arabidopsis thaliana.* Genetica 107:27-37.

Kikuchi K., Terauchi K., Wada M., Hirano H.Y. (2003) The plant MITE *mPing* is mobilized in anther culture. Nature 421:167-170.

Kinsella J.E. (1979) Functional-properties of soy proteins. J. Am. Oil Chem.Soc. 56:242-258.

Kirihara J.A., Hibberd, K.A. and Janice, A. (2001) Method for altering the nutritional content of plant seed. US patent 6,326,527.

Kortt A.A., Caldwell J.B., Lilley G.G., Higgins T.J.V. (1991) Amino-acid and cDNA sequences of a methionine-rich 2Sprotein from sunflower seed (*Helianthus annuus* L). Eur. J. Biochem.195:329-334.

Kotakis C., Vrettos N., Kotsis D., Tsagris M., Kotzabasis K., Kalantidis K. (2010) Light intensity affects RNA silencing of a transgene in *Nicotiana benthamiana* plants. BMC Plant Biol. 10:220.

Law J.A., Jacobsen S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11:204-220.

Liu Q.H., Rand T.A., Kalidas S., Du F.H., Kim H.E., Smith D.P., Wang X.D. (2003) R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. Science 301:1921-1925.

Liu Z.L., Han F.P., Tan M., Shan X.H., Dong Y.Z., Wang X.Z., Fedak G., Hao S., Liu B. (2004) Activation of a rice endogenous retrotransposon *Tos17* in tissue culture is accompanied by cytosine demethylation and causes heritable alteration in methylation pattern of flanking genomic regions. Theor. Appl. Genet. 109:200-209.

Lyons M., Cardle L., Rostoks N., Waugh R., Flavell A.J. (2008) Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. Mol. Genet. Genomics 280:275-285.

Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E., Shendure J., Turner D.J. (2010) Target-enrichment strategies for next-generation sequencing. Nat. Methods 7:111-118.

Mardis E.R. (2008) Next-generation DNA sequencing methods. Ann. Rev. Genom. Human Genet. 9: 387-402.

Matzke M., Aufsatz W., Kanno T., Daxinger L., Papp I., Mette A.F., Matzke A.J.M. (2004) Genetic analysis of RNA-mediated transcriptional gene silencing. Biochem. Biophys. Acta. 1677 :129-141.

Matzke M.A., Primig M., Trnovsky J., Matzke A.J.M. (1989) Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. EMBO J. 8:643-649.

Mazier M., Botton E., Flamain F., Blouchet J.-P., Courtial B., Chupeau M.-C., Chupeau Y., Maisonneuve B., Lucas H. (2007) Successful gene tagging in lettuce using the *Tnt1* retrotransposon from tobacco. Plant Physiol. 144:18-31.

McClintock B. (1951) Chromosome organization and genic expression. Cold Spring Harbor Symposia on Quantitative Biology 16:13-47.

Meng L., Ziv M., Lemaux P.G. (2006) Nature of stress and transgene locus influences transgene expression stability in barley. Plant Mol.Biol. 62:15-28.

Mertes F., ElSharawy A., Sauer S., van Helvoort J.M.L.M., van der Zaag P.J., Franke A., Nilsson M., Lehrach H., Brookes A.J. (2011) Targeted enrichment of genomic DNA regions for next-generation sequencing. Brief. Funct.Genomics 10:374-386.

Meyer P., Linn F., Heidmann I., Meyer H., Niedenhof I., Saedler H. (1992) Endogenous and environmental-factors influence 35S promoter methylation of a maize *A1* gene construct in transgenic petunia and its color phenotype. Mol. Gen. Genet. 231:345-352.

Meyers B.C., Tingley S.V., Morgante M. (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. 11:1660-1676.

Miyao A., Tanaka K., Murata K., Sawaki H., Takeda S., Abe K., Shinozuka Y., Onosato K., Hirochika H. (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell 15:1771-1780.

Molvig L., Tabe L.M., Eggum B.O., Moore A.E., Craig S., Spencer D., Higgins T.J.V. (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L) expressing a sunflower seed albumin gene. Proc. Nat. Acad. Sci. USA. 94:8393-8398.

Morgan G.T. (1995) Identification in the human genome of mobile elements spread by DNA-mediated transposition. J. Mol. Biol. 254:1-5.

Nakazaki T., Okumoto Y., Horibata A., Yamahira S., Teraishi M., Nishida H., Inoue H., Tanisaka T. (2003) Mobilization of a transposon in the rice genome. Nature 421:170-172.

Napoli C., Lemieux C., Jorgensen R. (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in *trans*. Plant Cell 2:279-289.

Oki N., Yano K., Okumoto Y., Tsukiyama T., Teraishi M., Tanisaka T. (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa ssp japonica*. Genes Genet.Syst. 83:321-329.

Oosumi T., Garlick B., Belknap W.R. (1996) Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. J. Mol.Evol. 43:11-18.

Patel M., Jung S., Moore K., Powell G., Ainsworth C., Abbott A. (2004) High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theor. Appl. Genet.108:1492-1502.

Prols F., Meyer P. (1992) The methylation patterns of chromosomal integration regions influence gene activity of transferred DNA in *Petunia hybrida*. Plant J. 2:465-475.

Randall P.J., Thomson, J.A. and Schroeder, H.E. . (1979) Cotyledonary storage proteins in *Pisum sativum.* IV. Effects of sulfur, phosphorus, potassium and magnesium deficiencies. Aust. J. Plant Physiol. 6:11-24.

Rocheleau C.E., Downs W.D., Lin R.L., Wittmann C., Bei Y.X., Cha Y.H., Ali M., Priess J.R., Mello C.C. (1997) *Wnt s*ignaling and an APC-related gene specify endoderm in early *C.elegans* embryos. Cell 90:707-716.

Rubin E., Lithwick G., Levy A.A. (2001) Structure and evolution of the *hAT* transposon superfamily. Genetics 158:949-957.

Sabot F., Guyot R., Wicker T., Chantret N., Laubin B., Chalhoub B., Leroy P., Sourdille P., Bernard M. (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Mol. Genet. Genomics 274:119-130.

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

Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., Zhang J., Fulton L., Graves T.A., Minx P., Reily A.D., Courtney L., Kruchowski S.S., Tomlinson C., Strong C., Delehaunty K., Fronick C., Courtney B., Rock S.M., Belter E., Du F., Kim K., Abbott R.M., Cotton M., Levy A., Marchetto P., Ochoa K., Jackson S.M., Gillam B., Chen W., Yan L., Higginbotham J., Cardenas M., Waligorski J., Applebaum E., Phelps L., Falcone J., Kanchi K., Thane T., Scimone A., Thane N., Henke J., Wang T., Ruppert

J., Shah N., Rotter K., Hodges J., Ingenthron E., Cordes M., Kohlberg S., Sgro J., Delgado B., Mead K., Chinwalla A., Leonard S., Crouse K., Collura K., Kudrna D., Currie J., He R., Angelova A., Rajasekar S., Mueller T., Lomeli R., Scara G., Ko A., Delaney K., Wissotski M., Lopez G., Campos D., Braidotti M., Ashley E., Golser W., Kim H., Lee S., Lin J., Dujmic Z., Kim W., Talag J., Zuccolo A., Fan C., Sebastian A., Kramer M., Spiegel L., Nascimento L., Zutavern T., Miller B., Ambroise C., Muller S., Spooner W., Narechania A., Ren L., Wei S., Kumari S., Faga B., Levy M.J., McMahan L., Van Buren P., Vaughn M.W., *et al.* (2009) The B73 maize genome: Complexity, dversity, and dynamics. Science 326:1112-1115.

Schwarz D.S., Hutvagner G., Du T., Xu Z.S., Aronin N., Zamore P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. Cell 115:199-208.

Sebastiani F.L., Farrell L.B., Schuler M.A., Beachy R.N. (1990) Complete sequence of a cDNA of alpha-subunit of soybean beta-conglycinin. Plant Mol.Biol. 15:197-201.

Sexton P.J., Naeve S.L., Paek N.C., Shibles R. (1998) Sulfur availability, cotyledon nitrogen : sulfur ratio, and relative abundance of seed storage proteins of soybean. Crop Sci. 38:983-986.

Shan X.H., Liu Z.L., Dong Z.Y., Wang Y.M., Chen Y., Lin X.Y., Long L.K., Han F.P., Dong Y.S., Liu B. (2005) Mobilization of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania latifolia* Griseb.). Mol.Biol. Evol. 22:976-990.

Shewry P.R., Franklin J., Parmar S., Smith S.J., Miflin B.J. (1983) The effects of sulfur starvation on the amino-acid and protein compositions of barley-grain. J. Cereal Sci. 1:21-31.

Shirasawa K., Hirakawa H., Tabata S., Hasegawa M., Kiyoshima H., Suzuki S., Sasamoto S., Watanabe A., Fujishiro T., Isobe S. (2012a) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor. Appl Genet. 124:1429-1438.

Shirasawa K., Koilkonda P., Aoki K., Hirakawa H., Tabata S., Watanabe M., Hasegawa M., Kiyoshima H., Suzuki S., Kuwata C., Naito Y., Kuboyama T., Nakaya A., Sasamoto S., Watanabe A., Kato M., Kawashima K., Kishida Y., Kohara M., Kurabayashi A., Takahashi C., Tsuruoka H., Wada T., Isobe S. (2012b) *In silico* polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol. 12:80

Smit A.F.A., Riggs A.D. (1996) Tiggers and other DNA transposon fossils in the human genome. Proc. Nat. Acad. Sci. USA. 93:1443-1448.

Song J.J., Smith S.K., Hannon G.J., Joshua-Tor L. (2004) Crystal structure of argonaute and its implications for RISC slicer activity. Science 305:1434-1437.

Tabe L.M., Droux M. (2002) Limits to sulfur accumulation in transgenic lupin seeds expressing a foreign sulfur-rich protein. Plant Physiol.128:1137-1148.

Tadege M., Ratet P., Mysore K.S. (2005) Insertional mutagenesis: a Swiss army knife for functional genomics of *Medicago truncatula*. Trends Plant Sci. 10:229-235.

Tadege M., Wen J.Q., He J., Tu H.D., Kwak Y., Eschstruth A., Cayrel A., Endre G., Zhao P.X., Chabaud M., Ratet P., Mysore K.S. (2008) Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. Plant J.54:335-347.

Takahashi H., Akagi H., Mori K., Sato K., Takeda K. (2006) Genomic distribution of MITEs in barley determined by MITE-AFLP mapping. Genome 49:1616-1620.

Takata M., Kiyohara A., Takasu A., Kishima Y., Ohtsubo H., Sano Y. (2007) Rice transposable elements are characterized by various methylation environments in the genome. BMC Genomics 8:469

tenLohuis M., Muller A., Heidmann I., Niedenhof I., Meyer P. (1995) A repetitive DNA fragment carrying a hot spot for de novo DNA methylation enhances expression variegation in tobacco and petunia. Plant J. 8:919-932.

Tu Z.J. (1997) Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*. Proc. Nat. Acad. Sci.USA. 94:7475-7480.

Tu Z.J. (2000) Molecular and evolutionary analysis of two divergent subfamilies of a novel miniature inverted repeat transposable element in the yellow fever mosquito, *Aedes aegypti*. Mol.Biol.Evol. 17:1313-1325.

Turcotte K., Srinivasan S., Bureau T. (2001) Survey of transposable elements from rice genomic sequences. Plant J. 25:169-179.

Unsal K., Morgan G.T. (1995) A novel group of families of short interspersed repetitive elements (SINEs) in *Xenopus* - evidence of a specific target site for DNA-mediated transposition of inverted-repeat SINEs. J. Mol.Biol. 248:812-823.

van Blokland R., ten Lohuis M., Meyer P. (1997) Condensation of chromatin in transcriptional regions of an inactivated plant transgene: evidence for an active role of transcription in gene silencing. Mol. Gen. Genet. 257:1-13.

Varshney R.K., Nayak S.N., May G.D., Jackson S.A. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. 27:522-530.

Walker E.L., Eggleston W.B., Demopulos D., Kermicle J., Dellaporta S.L. (1997) Insertions of a novel class of transposable elements with a strong target site preference at the *r* locus of maize. Genetics 146:681-693.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H. (2007) A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973-982.

Williams-Carrier R., Stiffler N., Belcher S., Kroeger T., Stern D.B., Monde R.-A., Coalter R., Barkan A. (2010) Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy *Mutator* lines of maize. Plant J. 63:167-177

Yamazaki M., Tsugawa H., Miyao A., Yano M., Wu J., Yamamoto S., Matsumoto T., Sasaki T., Hirochika H. (2001) The rice retrotransposon *Tos17* prefers low-copy-number sequences as integration targets. Mol. Genet. Genomics 265:336-344.

Zamore P.D., Tuschl T., Sharp P.A., Bartel D.P. (2000) RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25-33

Zhang J., Chiodini R., Badr A., Zhang G. (2011) The impact of next-generation sequencing on genomics. J. Genet. Genomics 38:95-109.

Zhang Q., Arbuckle J., Wessler S.R. (2000) Recent, extensive, and preferential insertion of members of the Miniature Inverted-repeat Transposable Element family *Heartbreaker* into genic regions of maize. Proc. Nat. Acad. Sci. USA. 97:1160-1165.

Zhang X.Y., Feschotte C., Zhang Q., Jiang N., Eggleston W.B., Wessler S.R. (2001) *P* instability factor: An active maize transposon system associated with the amplification of *Tourist*-like MITEs and a new superfamily of transposases. Proc. Nat. Acad. Sci. USA. 98:12572-12577

CHAPTER 3

STABILITY OF SEED ALLERGEN CONTENT DUE TO ALLERGEN GENE SILENCING IN MULTIPLE GENERATIONS OF PEANUT

Abstract

The transgenic peanut line B11.1.1/11, developed by the particle bombardment method, carries an inverted repeat gene cassette that silences the allergen gene *Ara h 2* through RNA interference (RNAi). Complete silencing of Ara h 2 expression was coupled with greatly reduced levels of Ara h 6 in the T_2 generation of this line. Stability of the silencing and any collateral changes in other major allergens, due to silencing, were studied in three generations of this line under field conditions. Quantitative western blotting of proteins showed that the silencing was stable since Ara h 2 signals were completely absent and Ara h 6 signals were significantly reduced relative to wild type in the three generations studied. Also no significant quantitative changes were observed with regard to other major allergens, Ara h 1 and Ara h 3, although a trend for higher Ara h 3 was observed in transgenics. Stability of RNA silencing in three generations confirms the feasibility of applying this technique for developing a reduced allergen peanut.

Chandran, M., Chu,Y., Maleki, S.J. and Ozias-Akins, P. To be submitted to Peanut Science

Introduction

Transgene stability is the most important criterion in the development of transgenic crops as the inheritance and expression of the transgene are often unpredictable (Ahuja, 2009; Bregitzer and Tonks, 2003). Stable transgene expression is important for functional genomics and breeding. Silencing and variable expression of transgenes are major factors affecting transgene utilization (Chawla *et al*., 2006). Both stability and variable expression of transgenes are influenced by the transgene construct components, region of integration, epigenetic effects and environmental factors (Dietz-Pfeilstetter, 2010).

The choice of promoter determines the plant organ or tissue in which the transgene is expressed and also the strength of expression (Sanger *et al*., 1990). *Agrobacterium*-mediated gene transfer and particle bombardment are the two commonly employed plant transformation techniques (Akhond and Machray, 2009). Studies in maize showed that *Agrobacterium*-mediated gene transfer resulted in fewer transgene integration events and more consistent expression in comparison to particle bombardment (Zhang *et al*., 2005). Stability of transgene expression is often affected by transgene copy number in the transformed plants. Multiple copies of the transgene and promoter methylation adversely affect transgene stability as revealed by studies in sweet orange (Fan *et al*., 2011). Transgene integration into hypomethylated regions and scaffold/matrix-associated regions (S/MARs) was found to enhance stable expression (Vain *et al*., 1999). S/MARs are AT-rich eukaryotic regulatory DNA regions associated with DNA topoisomerase binding and DNA replication (Bode *et al*., 1996). Stable expression of hygromycin and chloramphenicol acetyltransferase transgenes in tobacco was seen when the transgene was present as one intact copy near telomeres and AT-rich regions, possibly S/MARs

(Iglesias *et al*., 1997). Similar effects of transgene integration were seen with transgenic rice (Vain *et al*., 1999) and *Arabidopsis* (Butaye *et al*., 2004).

Environmental factors may influence transgene expression; in *Chrysanthemum* the expression of *CmETR 1/H69A*, the ethylene receptor gene inducing male sterility, varied with temperature. Under greenhouse conditions of 10-15ºC, transgenic chrysanthemum showed reduced expression of the *CmETR 1/H69A* gene and produced pollen. Pollen was not produced at temperatures between 20°C and 35°C (Shinoyama *et al*., 2012). In *Nicotiana benthamiana*, higher light intensity led to greater silencing of green fluorescent protein (GFP) when the temperature regime was kept constant. Also quantitative RT-PCR (reverse transcription polymerase chain reaction) showed higher expression of *DICER-LIKE* genes, *DCL1* and *DCL3*, involved in silencing pathways, under high light intensities. Northern blot analysis showed higher levels of GFP transcripts in silenced and non-silenced lines under low light intensity in comparison to higher light intensity (Kotakis *et al*., 2010). Other influences of environmental factors were seen in transgenic rape (Tang *et al*., 2004) and gladiolus (Kamo, 2008). On the contrary, stable expression of transgenes for multiple years was seen in rice (Gahakwa *et al*., 2000), poplar (Hawkins *et al*., 2003; Li *et al*., 2008), plum (Hily *et al*., 2004) and white clover (Panter *et al*., 2012).

Transcriptional and post-transcriptional gene silencing are epigenetic effects influencing transgene expression (Dietz-Pfeilstetter, 2010). In transcriptional gene silencing, synthesis of mRNA is arrested and in post-transcriptional gene silencing, mRNA produced is degraded in the cytoplasm. The most common mechanism of transcriptional gene silencing is RNA-directed DNA methylation, where promoter methylation is caused by small interfering RNA (siRNA) of 24 bp. The siRNAs induce cytosine methylation in DNA regions with sequence homology to the

siRNA (Kanno and Habu, 2011). In post-transcriptional gene silencing, a double stranded RNA (dsRNA) serves as the silencing initiator signal. Cleavage of dsRNA to siRNAs is catalyzed by DICER enzymes. The siRNAs are incorporated into the RNA-induced silencing complex (RISC), along with ARGONAUTE proteins. The RISC is guided to the target mRNA by the siRNAs. Once a complementary mRNA is identified, bases at positions 2-8 of siRNA in the ARGONAUTE protein form Watson-Crick base pairing with target mRNA, which is then cleaved by ARGONAUTE proteins (Sifuentes-Romero *et al*., 2011).

Stability of transgene expression is a complex phenomenon dependent on multiple factors, and analysis of expression over generations is a crucial step in developing transgenic crops, irrespective of the mode of reproduction. The peanut transgenic line B11.1.1/11 carries an RNAi (RNA interference) construct targeted to silence Ara h 2 and Ara h 6, two major peanut allergens /proteins. Although peanut is a rich and economical protein source, consumption of peanut proteins can cause allergic reactions and is a common cause of fatal anaphylaxis (Kang *et al*., 2007; Yocum and Khan, 1994). In the United States approximately 0.8% of young children and 0.6% of adults are affected by peanut allergy (Sampson, 2004), which is becoming a major health concern in developed countries (Scurlock and Burks, 2004). Ara h 1, Ara h 2, Ara h 3 and Ara h 6 are the major allergens among the thirteen allergens identified in peanut (Burks *et al*., 1998; Flinterman *et al*., 2007; Scurlock and Burks, 2004). Developing hypoallergenic peanut varieties by conventional breeding is currently not possible as no null mutants for allergens could be identified in screening peanut accessions (Kang *et al*., 2007) with the exception of a naturally occurring genotype deficient in Ara h 1 (Krause *et al.*, 2010). The reverse genetic technique, RNAi, was employed to silence *Ara h 2* and *Ara h 6* genes (Chu *et al*., 2008), two major allergens which share 63% sequence homology (Ramos *et al*., 2006). Ara h 2 is a 17-20 kDa

protein belonging to the conglutin family and accounts for 5.9 - 9.3% of total seed protein (Koppelman *et al*., 2001; Stanley *et al*., 1997). Cultivated peanut has two homoeologs of Ara h 2, Ara h 2.01 and Ara h 2.02 (Chatel *et al*., 2003; Ramos *et al*., 2006). Ara h 6 is a 14.5 kDa protein of the conglutin family and together with Ara h 2 was the major elicitor of anaphylaxis in a mouse model system (Kulis *et al*., 2012; Wen *et al*., 2007). The RNAi line, B11.1.1/11, developed by the biolistic transformation method showed nearly complete silencing of Ara h 2 and greatly reduced levels of Ara h 6 in the T_2 generation (Chu *et al*., 2008). Quantitative immunoblotting with seed proteins from the T_2 generation showed that Ara h 2 and Arah 6 were reduced 474- and 25,868-fold, respectively, in the transgenic line. IgE binding with peanut allergic patient sera showed differences only with respect to Ara h 2 and Ara h 6 between transgenic and non-transgenic line (Chu *et al*., 2008). Further analysis of transgene stability is needed before advancing this line for commercial cultivation, further breeding or additional functional analysis. The purpose of the present study was to test the stability of a transgenic peanut line B11.1.1/11 across three generations (T_3, T_4 and T_5) under field conditions and to test for any collateral effects of silencing Ara h 2 and Ara h 6, mainly on other major peanut allergens *viz*., Ara h 1 and Ara h 3.

Materials and Methods

The RNAi construct (Figure 3.1) used to silence *Ara h 2* and *Ara h 6* genes was based on the plasmid pFGC1008 (AY310333) obtained from the *Arabidopsis* Biological Resource Center, Ohio State University. It includes a 293-bp sense fragment identical to nucleotides 122-414 of *Ara h 2.01* (GenBank ID L77197) and a 228-bp antisense fragment identical to nucleotides 192-414 (Chu *et al*. 2008). Upon transcription a 222-bp hairpin is formed by the sense and antisense fragments. The same construct also targets *Ara h 6* gene since *Ara h 2* and *Ara h 6* share 63%

sequence homology overall (Ramos *et al*., 2006) and 81% similarity across the region used for the inverted repeat (Chu *et al*., 2008). The RNAi construct was delivered by particle bombardment to nine-month-old embryogenic clusters developed from mature zygotic embryos of peanut cultivar Georgia Green. Restriction digestion of T_2 genomic DNA with *Hin*dIII, Southern blotting, and hybridization with the *gus* spacer fragment separating the inverted repeat or an *Ara h 2* probe had shown multiple copies of the transgene at ~13 kb and ~7 kb although both hybridizing bands cosegregated and were thus linked (Chu *et al*., 2008).

Seed Increase for T_4 and T_5 Generations

Bulked seeds from T_2 plants (130 T_3 seeds) of transgenic line, B11.1.1/11, were used to derive T_4 and T_5 seeds after two generations in the greenhouse (from July- 2009 to May- 2010).

Field Layout

Multiple generations were grown in the field at The University of Georgia, Tifton Campus, from June 2010 to Oct 2010. There were four replications of T_3, T_4 and T_5 generations. Three replications of non-transgenic segregants (similar to 'Georgia Green') were planted in control plots. Each plot was 1.95 m x 0.9 m in dimension leaving 0.6 m between plots. Sixty-six seeds were sown in each plot, 33 in each of two rows at 16.2 plants/m (5 seeds/ft). Seeds were coated with Vitavax fungicidal seed treatment before sowing. Non-transgenic plants bordered the entire experimental area.

Testing for the Transgenic Insert and Homozygosity of T_2 Seed Bulk

The presence of the transgene in 90 T_2 seeds, 40 T_3 and 31 T_4 greenhouse-grown plants, was tested by the polymerase chain reaction (PCR) using primers specific to promoter (Forward 5'-cccacaaaaatctgagcttaacag-3' and reverse 5'-ccctttggtcttctgagactgtat-3')- or coding region (forward 5'-accacatgaagcagcacg-3' and reverse 5'-gcgtcgcagaacattaca-3') of the inverted repeat

cassette. In the field, 33 plants per plot were labelled and tested by PCR to confirm the presence or absence of the transgene. Ten microlitres of PCR product was run on 1% agarose gels and amplification of the transgene was visualized under UV light after ethidium bromide staining (0.1 µg/ml). Seeds from the labelled plants were collected and bulked for allergen protein analysis.

Germination

Germination under field conditions was noted daily for ten days beginning with the day that germination of the first seed was observed until the final observation, which was made two weeks from the date of the first observation.

Protein Extraction

For extraction of protein for quantitative western blots, fifteen seeds were pooled separately from three replications of each transgenic generation, weighed, testa removed and ground in liquid nitrogen. The seed powder was defatted with hexane at 1:35 (w/v) for 4 hrs, keeping the sample on ice. Defatted and air-dried seed powder was passed through a 30 mesh sieve. Seed protein was extracted using Tris–buffered saline (TBS-150 mM NaCl, 50 mM Tris) at pH 7.4, containing EDTA-free protease inhibitor (Roche Diagnostics, Indianapolis, IN, USA). The defatted and air-dried peanut flour was mixed with ice-cold TBS in the ratio 1:10 (w/v) and mixed overnight at 4°C (Porterfield *et al*., 2009). The mixture was centrifuged at 16,060 x g for 15 minutes at 4°C, crude protein extract collected, aliquoted and stored at -80°C. Protein from field-grown 'Georgia Green' seeds was extracted by the same protocol from fifteen pooled seeds. Quantification of protein in the crude protein extract (CPE) was done with the Pierce Bicinchoninic acid protein assay kit (Thermo Scientific, Rockford, IL, US) according to the

manufacturer's protocol. Total protein in CPE was calculated by averaging the results from three quantifications of each replication, with each sample replicated twice per quantification.

Quantitative Western Blots

Seed protein was denatured at 95°C for 5 min and resolved by running on 15% polyacrylamide gels for 1.5 hrs and blotted on to Amersham Hybond-P PVDF membrane (GE Healthcare, Piscataway, New Jersey, USA) by electro -transfer at 100 V for 1hr. The PVDF membrane was dried at 37°C overnight. To test for equal loading of proteins, the membrane was stained with SYPRO Ruby (Invitrogen, Grand Island, NY, USA) as per the manufacturer's protocol. Briefly proteins on PVDF membrane were fixed with a solution containing 10% methanol and 5% acetic acid, washed with deionized water four times for 5 min each, stained with SYPRO Ruby and washed. Fluorescence detection and imaging was done with a STORM molecular imager (Amersham Biosciences, Piscataway, NJ, USA) at 450 nm excitation/520 nm emission. For western blotting the membranes were blocked with 5% non-fat dry milk in TBST for 1 hour, followed by incubation with anti-Ara h 1 (1:5000), -Ara h 2 (1:8000), -Ara h 3 (1:5000) or -Ara h 6 (1:5000) primary antibodies custom manufactured by Sigma Immunosys (Woodlands, TX). Subsequent to three washes with TBST the membranes were reacted with Alkaline Phosphatase (AP)-labeled rabbit anti-chicken/turkey secondary antibody (Invitrogen) @ 1:10000 for ½ hr. Signal detection was done with Enhanced Chemifluorescence substrate (ECF) (GE Healthcare, Piscataway, New Jersey, USA) treatment for 5 min. Fluorescence was detected with a STORM molecular imager (Amersham Biosciences, USA) at 450 nm excitation/520 nm emission.

Quantification of Allergens in Seed Protein Extract

The standard curve for quantification was derived from 'Georgia Green' protein samples loaded at 20,000, 10,000, 5,000 and 2,500 ng per well for each gel, along with protein from the three

transgenic and one non-transgenic generations. Of the four replicate field plots of each transgenic generation, only three were used for analysis. For Ara h 1 and Ara h 3, quantification was done using 5,000 ng total protein (1,000 ng/µl) of transgenics. For Ara h 2 and Ara h 6, quantification was done with 20,000 ng total protein (4,000 ng/µl) of transgenic seeds. Volume quantification for each band was done with Image Quant software after background correction by object average method. The standard curve is developed with band volume for each allergen protein in 'Georgia Green' and concentration of protein in 1 µl of 'Georgia Green' dilutions (4 µg, 2 µg, 1 µg and 0.5 µg). Sample band volume calculated was normalized with 'Georgia Green' protein concentrations and hence the concentration corresponding to every band of sample will be equivalent to the total protein in 1 µl of 'Georgia Green' protein. The concentration of individual allergens will be proportional to the total protein concentration. Hence allergen is quantified relative to total protein concentration and not the absolute value. Quantification of three biological replicates was done in duplicate, and the mean concentration was compared by Tukey's t-test at $P \leq 0.05$ using SAS software (SAS Institute Inc., 2009).

Results

The germination percentage of the transgenic generations and the non-transgenic control (Table 3.1) varied from 82.95 % to 93.2 %, although no statistically significant variation was observed ($P \leq 0.05$). The maximum germination was recorded with the T_3 seeds (93.2 %), while the lowest germination was observed among T_4 seeds (82.95%). Percentage of the total seed extractable protein (Table 3.2) showed no significant variation among treatments ($P \leq 0.05$) and varied from 6.09% (T_5) to 9.01% (T_4) among transgenic generations. Non-transgenic seeds contained the highest extractable seed protein content of 10.4% in this field study.

Table 3.3 summarizes the relative concentrations of allergen proteins, Ara h 1, Ara h 3, Ara h 2 and Ara h 6 in the three transgenic generations and non-transgenic (null) segregant, which are the means from two technical replications of three biological replicates (Figures 3.2 and 3.3). Ara h 1 and Ara h 3 levels in T_3, T_4 and T_5 generations were not significantly different from non-transgenics with *P*-values of 0.73 and 0.17, respectively, although transgenics consistently showed a trend for higher Ara h 3 than the null segregant. Ara h 2 proteins were not detectable in RNAi lines, which was statistically significant ($P \leq 0.0001$). The concentration of Ara h 6 in all three transgenic generations was significantly lower than in the null segregant ($P \leq 0.0001$). The T_5 generation showed a higher relative level of Ara h 6 at 1265.4 than T_4 and T_3 generations, which showed ~ 872 and 880 relative concentrations, although the difference between transgenic generations was not statistically significant.

Discussion

This study was conducted to analyze the stability of allergen gene silencing in the peanut transgenic line B11.1.1/11 across three generations under field conditions. Also the collateral effect of silencing one protein/allergen on the expression of other allergen proteins was tested. The stability of transgene expression can be studied by assaying RNA, e.g., the steady-state transcription level of transgene mRNA (Bonadei *et al*., 2012), reverse-transcription PCR (RT-PCR) (Hily *et al*., 2004), or protein, e.g., fluorometric assay as for GUS (β-glucuronidase), thin layer chromatography for enzymatic activity, western blotting (Gahakwa *et al*., 2000) and enzyme-linked immunosorbent assay (ELISA) (Li *et al*., 2008) for direct protein detection. In the present study, western blotting was combined with fluorometry for quantification of allergens relative to non-transgenic 'Georgia Green'.

The phenotype of transgenic seeds and plants was similar to non-transgenic seeds and plants, and seeds showed good germination under field conditions. As in the present study, in soybean (Schmidt *et al*., 2011; Takahashi *et al*., 2003) and rice (Ashida *et al*., 2006), lines with storage protein knock-out or deficiency were similar to wild-type in phenotype, seed weight and developmental patterns.

Peanut seed protein ranged from 20.7% to 28.1% in analysis of 64 peanut genotypes (Dwivedi *et al*., 1990) and was reported as 21% in a comparative study of edible seed nuts (Venkatachalam and Sathe, 2006). In the present study, protein was extracted with Tris-buffered saline at pH 7.4 containing 150 mM NaCl, according to a previously published peanut allergen extraction protocol (Porterfield *et al*., 2009), and the average protein concentration in the various samples varied from 6.0-10% (Table 3.2). When various buffers and meal to buffer ratios were tested for peanut protein extraction efficiency, it was found that 1M NaCl-20 mM sodium phosphate buffer at pH 7.0 and 1:18 buffer ratio gave the maximum globulin recovery along with albumins (Mahaboobbasha and Cherry, 1976). A study by Poms *et al.* (2004) showed extraction efficiencies are more dependent on buffer pH, with buffer at higher pH yielding more proteins. Accordingly TBS at pH 8.2 yielded 35% more peanut protein than TBS at pH 7.4. and maximum protein was extracted with sodium borate buffer at pH 9.0, with salt concentrations at 7 mM and 15 mM showing no variation in protein extraction efficiency. Higher protein extraction efficiency was observed using 6 M urea at pH 8.0, but the buffer is highly denaturing. Hence selection of buffer would also depend on the requirements of subsequent assays (Poms *et al*., 2004). Peanut seed protein fractions were grouped as albumins, arachin and conarachin (Johns, 1916). Arachin and conarachin are salt-soluble globulins which account for 87% of the seed protein (Irving, 1945). Lower protein content per unit mass of defatted meal observed in this

study compared to other studies could possibly be due to the low salt concentration (150 mM) and pH of the buffer employed. Nevertheless, the proportion within each class of protein extracted should be comparable across samples.

Inheritance and expression of the transgene was found to be stable in T_3, T_4 and T_5 generations as no quantifiable signal (complete silencing) corresponding to Ara h 2 protein was seen by the quantitative western blot assay. No Ara h 2 signals were previously observed in the T_2 generation by ECL detection and by the IgE binding capacity of protein from the T_2 generation using sera from patients allergic to peanut (Chu *et al*., 2008). Partial silencing of Ara h 6, as observed in the T_2, also remained stable in all three generations providing further evidence for the stable integration, inheritance, and expression of the silencing transgene under the field conditions tested. Ara h 6 was reduced by 3.6-5.2-fold in the three transgenic generations studied, compared to null segregants. Numerous reports of stable integration and inheritance of transgenes across multiple generations have been published including in *Populus* (Bonadei *et al*., 2012; Hawkins *et al*., 2003; Li *et al*., 2008), rice (Chawla *et al*., 2006; Gahakwa *et al*., 2000; Vain *et al*., 1999), *Arabidopsis* (Butaye *et al*., 2004), citrus (Pons *et al*., 2012) and plum (Hily *et al*., 2004). On the contrary, in barley, multiple transgene copies were associated with higher cytosine methylation of GFP, the CaMV 35S promoter and the selectable marker gene, *nptII*, leading to variability in transgene expression (Fan *et al*., 2011). Transgene-induced resistance to papaya ring-spot virus, effected by a viral coat protein gene, was reduced in the T_2 generation compared to the T_1 generation, suggesting the unstable nature of the transgene expression (Retuta *et al*., 2012). The results in this study are promising as the transgene expression is stable even with multiple copies of the transgene in B11.1.1/11 (Chu *et al*., 2008).

The transgenic insert was present and showed stable expression in T_4 and T_5 generations, the seeds of which were raised in the greenhouse from T_3 and T_4 generations. Hence any environmental effects on silencing were not observed in this study, between greenhouse-raised plants and field-grown plants. This is in agreement with studies in white clover (Panter *et al*., 2012). Similar results were seen in transgenic citrus in which stability of transgenes was unaffected under field conditions (Pons *et al*., 2012). On the contrary transgene expression varied between field and glasshouse for *Gladiolus* (Kamo, 2008) and was influenced by temperature in *Chrysanthemum* (Shinoyama *et al*., 2012) and by light intensity in *Nicotiana* (Kotakis *et al*., 2010). Since the transgene silencing of Ara h 2 was found to be stable across four generations (this study and Chu *et al*., 2008) it is not unreasonable to speculate that the multiple integration events may be near S/MARs or hypomethylated chromosome regions, but molecular evidence for this would need to be provided.

The total protein was lower in the transgenic generations but the difference with the non-transgenic sample did not reach statistical significance. In maize *opaque2* mutants, the total protein was reduced significantly compared to wild-type possibly due to suppression of zein biosynthesis, which accounts for 60% of maize seed proteins (Hasjim *et al*., 2009). However in soybean (Schmidt *et al*., 2011) and *Phaseolus* (Hartweck and Osborn, 1997), silencing of major seed proteins resulted in altered protein composition, but the total protein was found to be constant which is consistent with this study. Of the four major allergens addressed in this study, Ara h 1 and Ara h 2 constitute approximately 12-16% and 5.9 – 9.3% of the total seed protein, respectively (Koppelman *et al*., 2001; Stanley *et al*., 1997). Ara h 2, although a potent allergen, accounts for only a minor proportion in the total seed protein. Thus the silencing of Ara h 2 and

partial silencing of Ara h 6 would not have impacted the total protein content to a significant level.

In soybean, the major seed proteins 11S glycinin and 7S β-conglycin account for ~70% of the total seed protein (Thanh and Shibasaki, 1976). Lines deficient in β-conglycin showed an increase in the glycinin fraction (Ogawa *et al*., 1989). Mutant line QR2, lacking both glycinin and β-conglycin was agronomically similar to the wild-type and showed no difference in total nitrogen, total amino acids and oil contents. However other protein fractions viz., lipoxygenase, sugar-binding proteins, lectin and other 7S subunits were elevated, concomitant with an increase in free amino acids, especially arginine (Takahashi *et al*., 2003). In rice, absence of a 26 kDa globulin increased free amino acid content (Ashida *et al*., 2006).

In *Phaseolus vulgaris*, proteins phaseolin and phytohemagglutinin constitute 40-60% and 6-12% of the total seed protein and the protein arcelin can comprise up to 50% of seed protein in Sanilac backcross lines (Osborn, 1988; Osborn *et al.,* 1988). Research on *Phaseolus vulgaris* has shown that genetically removing the phaseolin gene led to an increased concentration of other seed proteins viz., arcelin and phytohemagglutinin (Hartweck and Osborn, 1997). Reducing phaseolin, arcelin and phytohemagglutinin concentrations in *P. vulgaris* led to an upregulation of sulfur-containing proteins like legumin, albumin-1 and albumin-2 and increased activity of starch and raffinose metabolic enzymes (Liao *et al*., 2012; Marsolais *et al*., 2010).

Screening peanut accessions and breeding lines for reduced allergens or null mutants for allergen genes showed that reduction in one allergen may be compensated by other allergen/allergens (Kang *et al*., 2007). Thus altering one protein was found to alter the total protein profile and other metabolic pathways. As observed in the current study, Ara h 2 silencing and reduction in Ara h 6 concentrations in the transgenic line B11/1.1/11 had no effect on the concentration of

Ara h 1 and a minor effect on Ara h 3, the other two allergens quantified. However, mass spectrometry analysis of proteins from the same line showed a significant increase in 13-lipoxygenase and 11S Ahy-3 (Ara h 3-related) proteins (Stevenson *et al*., 2009). This variation in detection efficiency is due to the ability of mass spectrometry to differentiate products of multiple gene families. Thus the data from quantitative western should be combined with more global proteomics platforms to develop a more comprehensive profile of differences between transgenic and non-transgenic lines.

Conclusion

The stability of transgene expression, manifested as silencing of Ara h 2 and Ara h 6 allergens in peanut, is demonstrated in this study. Silencing was found to be stable when three generations of transgenic peanut were grown under field conditions. Transgenic lines also showed no significant differences from wild-type with regard to, total protein, total protein profile and composition of two other major allergens, Ara h 1 and Ara h 3. Ara h 2, although a minor protein component, functions as a major peanut allergen. Although complete elimination of allergic reactions in individuals is not feasible with the transgenic line, B11.1.1/11, it can be considered as having potential to reduce allergenicity. Stability observed in this single year and single location trial provides evidence that this reduced allergen line would be useful as a source of peanut protein for clinical laboratory studies. Further multiyear and multilocation trials would be needed to evaluate the longer-term stability of expression and agronomic performance, before this line could be released as a hypoallergic peanut or used for further crop improvement work to stack multiple silencing constructs.

References

Ahuja M.R. (2009) Transgene stability and dispersal in forest trees. Trees 23:1125-1135.

Akhond M.A.Y., Machray G.C. (2009) Biotech crops: technologies, achievements and prospects. Euphytica 166:47-59.

Ashida K., Iida S., Yasui T. (2006) Lack of 26 kDa globulin accompanies increased free amino acid content in rice (*Oryza sativa* L.) grains. J. Cereal Sci. 43:387-392.

Bode J., StengertIber M., Kay V., Schlake T., DietzPfeilstetter A. (1996) Scaffold/matrix-attached regions: Topological switches with multiple regulatory functions. Crit. Rev. Eukaryot. Gene Expr. 6:115-138.

Bonadei M., Zelasco S., Giorcelli A., Gennaro M., Calligari P., Quattrini E., Carbonera D., Balestrazzi A. (2012) Transgene stability and agronomical performance of two transgenic Basta ®-tolerant lines of *Populus alba* L. Plant Biosystems 146:33-40.

Bregitzer P., Tonks D. (2003) Inheritance and expression of transgenes in barley. Crop Sci. 43:4-12.

Butaye K.M.J., Goderis I., Wouters P.F.J., Pues J., Delaure S.L., Broekaert W.F., Depicker A., Cammue B.P.A., De Bolle M.F.C. (2004) Stable high-level transgene expression in *Arabidopsis thaliana* using gene silencing mutants and matrix attachment regions. Plant J. 39:440-449

Chatel J.M., Bernard H., Orson F.M. (2003) Isolation and characterization of two complete Ara h 2 isoforms cDNA. Int. Arch. Allergy Immunol. 131:14-18.

Chawla R., Ariza-Nieto M., Wilson A.J., Moore S.K., Srivastava V. (2006) Transgene expression produced by biolistic-mediated, site-specific gene integration is consistently inherited by the subsequent generations. Plant Biotechnol. J. 4:209-218

Chu Y., Faustinelli P., Laura Ramos M., Hajduch M., Stevenson S., Thelen J.J., Maleki S.J., Cheng H., Ozias-Akins P. (2008) Reduction of Ige binding and nonpromotion of *Aspergillus flavus* fungal growth by simultaneously silencing Ara h 2 and Ara h 6 in Peanut. J. Agric. Food Chem. 56:11225-11233.

Dietz-Pfeilstetter A. (2010) Stability of transgene expression as a challenge for genetic engineering. Plant Sci. 179:164-167.

Dwivedi S.L., Jambunathan R., Nigam S.N., Raghunath K., Shankar K.R., Nagabhushanam G.V.S. (1990) Relationship of seed mass to oil and protein contents in peanut *Arachis hypogaea* L. Peanut Sci.17:48-52.

Fan J., Liu X., Xu S.X., Xu Q., Guo W.W. (2011) T-DNA direct repeat and 35S promoter methylation affect transgene expression but do not cause silencing in transgenic sweet orange. Plant Cell Tiss. Organ Cult. 107:225-232.

Gahakwa D., Maqbool S.B., Fu X., Sudhakar D., Christou P., Kohli A. (2000) Transgenic rice as a system to study the stability of transgene expression: multiple heterologous transgenes show similar behaviour in diverse genetic backgrounds. Theor. Appl. Genet. 101:388-399.

Hartweck L.M., Osborn T.C. (1997) Altering protein composition by genetically removing phaseolin from common bean seeds containing arcelin or phytohemagglutinin. Theor. Appl. Genet. 95:1012-1017.

Hasjim J., Srichuwong S., Scott M.P., Jane J.L. (2009) Kernel composition, starch structure, and enzyme digestibility of *opaque-2* maize and quality protein maize. J. Agric. Food Chem. 57:2049-2055.

Hawkins S., Leple J.C., Cornu D., Jouanin L., Pilate G. (2003) Stability of transgene expression in poplar: A model forest tree species. Ann. For. Sci. 60:427-438.

Hily J.M., Scorza R., Malinowski T., Zawadzka B., Ravelonandro M. (2004) Stability of gene silencing-based resistance to Plum pox virus in transgenic plum (*Prunus domestica* L.) under field conditions. Transgenic Res. 13:427-436.

Iglesias V.A., Moscone E.A., Papp I., Neuhuber F., Michalowski S., Phelan T., Spiker S., Matzke M., Matzke A.J.M. (1997) Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. Plant Cell 9:1251-1264.

Irving G.W., Fontaine, T. D. and Warner, R. C. (1945) Electrophoretic investigation of peanut proteins. I. Peanut meal extract, arachin and conarachin. Arch. Biochem. Biophys. 7.

Johns C.O.a.J., D. B. (1916) The proteins of the peanut, *Arachis hypogaea*. I. the globulins arachin and conarachin. J. Biol. Chem. 28:77-87.

Kamo K. (2008) Transgene expression for *Gladiolus* plants grown outdoors and in the greenhouse. Scientia Horticulturae 117:275-280.

Kang I.-H., Gallo M., Tillman B.L. (2007) Distribution of allergen composition in peanut (*Arachis hypogaea* L.) and wild progenitor (*Arachis*) species. Crop Sci.47:997-1003.

Kanno T., Habu Y. (2011) siRNA-mediated chromatin maintenance and its function in *Arabidopsis thaliana*. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms 1809:444-451.

Koppelman S.J., Vlooswijk R.A.A., Knippels L.M.J., Hessing M., Knol E.F., van Reijsen F.C., Bruijnzeel-Koomen C. (2001) Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties Runner, Spanish, Virginia, and Valencia, bred in different parts of the world. Allergy 56:132-137.

Kotakis C., Vrettos N., Kotsis D., Tsagris M., Kotzabasis K., Kalantidis K. (2010) Light intensity affects RNA silencing of a transgene in *Nicotiana benthamiana* plants. BMC Plant Biol. 10:220.

Krause S., Latendorf T., Schmidt H., Darcan-Nicolaisen Y., Reese G., Petersen A., Janssen O., Becker W.-M. (2010) Peanut varieties with reduced Ara h 1 content indicating no reduced allergenicity. Mol. Nutr. Food Res. 54:381-387.

Kulis M., Chen X., Lew J., Wang Q., Patel O.P., Zhuang Y., Murray K.S., Duncan M.W., Porterfield H.S., Burks A.W., Dreskin S.C. (2012) The 2S albumin allergens of *Arachis hypogaea*, Ara h 2 and Ara h 6, are the major elicitors of anaphylaxis and can effectively desensitize peanut-allergic mice. Clin. Exp. Allergy 42:326-336.

Li J.L., Meilan R., Ma C., Barish M., Strauss S.H. (2008) Stability of herbicide resistance over 8 years of coppice in field-grown, genetically engineered poplars. West. J. App. For. 23:89-93.

Liao D.Q., Pajak A., Karcz S.R., Chapman B.P., Sharpe A.G., Austin R.S., Datla R., Dhaubhadel S., Marsolais F. (2012) Transcripts of sulphur metabolic genes are co-ordinately regulated in developing seeds of common bean lacking phaseolin and major lectins. J. Exp. Bot. 63:6283-6295.

Mahaboobbasha S.M., Cherry J.P. (1976) composition, solubility, and gel electrophoretic properties of proteins isolated from Florunner (*Arachis hypogaea* L.) peanut seeds. J. Agric. Food Chem. 24:359-365.

Marsolais F., Pajak A., Yin F.Q., Taylor M., Gabriel M., Merino D.M., Ma V., Kameka A., Vijayan P., Pham H., Huang S.Z., Rivoal J., Bett K., Hernandez-Sebastia C., Liu Q.A., Bertrand A., Chapman R. (2010) Proteomic analysis of common bean seed with storage

protein deficiency reveals up-regulation of sulfur-rich proteins and starch and raffinose metabolic enzymes, and down-regulation of the secretory pathway. J. Proteomics 73:1587-1600.

Ogawa T., Tayama E., Kitamura K., Kaizuma N. (1989) Genetic-improvement of seed storage proteins using 3 variant alleles of 7S globulin subunits in soybean (*Glycine-max* L). J. J. Breed. 39:137-147.

Osborn T.C. (1988) Genetic-control of bean seed protein. Crit. Rev. Plant Sci. 7:93-116.

Osborn T.C., Alexander D.C., Sun S.S.M., Cardona C., Bliss F.A. (1988) Insecticidal activity and lectin homology of arcelin seed protein. Science 240:207-210.

Panter S., Chu P.G., Ludlow E., Garrett R., Kalla R., Jahufer M.Z.Z., Arbiza A.D., Rochfort S., Mouradov A., Smith K.F., Spangenberg G. (2012) Molecular breeding of transgenic white clover (*Trifolium repens* L.) with field resistance to Alfalfa mosaic virus through the expression of its coat protein gene. Transgenic Res. 21:619-632.

Poms R.E., Capelletti C., Anklam E. (2004) Effect of roasting history and buffer composition on peanut protein extraction efficiency. Mol. Nutr. Food Res. 48:459-464.

Pons E., Peris J.E., Pena L. (2012) Field performance of transgenic citrus trees: Assessment of the long-term expression of *uidA* and *nptII* transgenes and its impact on relevant agronomic and phenotypic characteristics. BMC Biotechnol. 12.

Porterfield H.S., Murray K.S., Schlichting D.G., Chen X., Hansen K.C., Duncan M.W., Dreskin S.C. (2009) Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants. Clin. Exp. Allergy 39:1099-1108.

Ramos M.L., Fleming, G., Chu, Y., Akiyama Y., Gallo, M., Ozias-Akins, P. . (2006) Chromosomal and phylogenetic context for conglutin genes in *Arachis* based on genomic sequence. . Mol.Genet. Genomics 275:578-592.

Retuta A.M.O., Magdalita P.M., Aspuria E.T., Espino R.R.C. (2012) Evaluation of selected transgenic papaya (*Carica papaya* L.) lines for inheritance of resistance to papaya ringspot virus and horticultural traits. Plant Biotechnol. 29:339-349.

Sampson H.A. (2004) Update on food allergy. J. Allergy Clin. Immunol. 113:805-819.

Sanger M., Daubert S., Goodman R.M. (1990) Characteristics of a strong promoter from figwort mosaic-virus - comparison with the analogous-35S promoter from cauliflower mosaic-virus and the regulated mannopine synthase promoter. Plant Mol. Biol.14:433-443.

SAS Institute Inc. (2009). Administering SAS® Enterprise Guide® 4.2. Cary, NC:SAS Institute Inc.

Schmidt M.A., Barbazuk W.B., Sandford M., May G., Song Z.H., Zhou W.X., Nikolau B.J., Herman E.M. (2011) Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. Plant Physiol. 156:330-345.

Scurlock A.M., Burks A.W. (2004) Peanut allergenicity. Ann. Allergy Asthma Immunol. 93:S12-S18.

Shinoyama H., Sano T., Saito M., Ezura H., Aida R., Nomura Y., Kamada H. (2012) Induction of male sterility in transgenic chrysanthemums (*Chrysanthemum morifolium* Ramat.) by expression of a mutated ethylene receptor gene, *Cm-ETR1/H69A*, and the stability of this sterility at varying growth temperatures. Mol. Breeding 29:285-295.

Sifuentes-Romero I., Milton S.L., Garcia-Gasca A. (2011) Post-transcriptional gene silencing by RNA interference in non-mammalian vertebrate systems: Where do we stand? Mutat. Res. 728:158-171.

Stanley J.S., King N., Burks A.W., Huang S.K., Sampson H., Cockrell G., Helm R.M., West C.M., Bannon G.A. (1997) Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. Arch. Biochem. Biophys. 342:244-253.

Stevenson S.E., Chu Y., Ozias-Akins P., Thelen J.J. (2009) Validation of gel-free, label-free quantitative proteomics approaches: Applications for seed allergen profiling. J. Proteomics 72:555-566.

Takahashi M., Uematsu Y., Kashiwaba K., Yagasaki K., Hajika M., Matsunaga R., Komatsu K., Ishimoto M. (2003) Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. Planta 217:577-586.

Tang J.H., Scarth R., McVetty P.B.E. (2004) Stability of the expression of acyl-ACP thioesterase transgenes in oilseed rape doubled haploid lines. Crop Sci.44:732-740.

Thanh V.H., Shibasaki K. (1976) Major proteins of soybean seeds - straightforward fractionation and their characterization. J. Agric. Food Chem.24:1117-1121.

Vain P., Worland B., Kohli A., Snape J.W., Christou P., Allen G.C., Thompson W.F. (1999) Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J. 18:233-242.

Wen H.-W., Borejsza-Wysocki W., DeCory T.R., Durst R.A. (2007) Peanut allergy, peanut allergens, and methods for the detection of peanut contamination in food products. Comp.Rev.Food Sci. Food Safety 6:47-58.

Yocum M.W., Khan D.A. (1994) Assessment of patients who have experienced anaphylaxis - a 3-year survey. Mayo Clinic Proceedings 69:16-23.

Zhang Y.S., Yin X.Y., Yang A.F., Li G.S., Zhang J.R. (2005) Stability of inheritance of transgenes in maize (*Zea mays* L.) lines produced using different transformation methods. Euphytica 144:11-22.

Table 3.1 Germination % of seeds [†]

Generations	Seed Germination (%)
T_3	93.20±2.62[a*]
T_4	82.57±9.13[a]
T_5	88.63±5.17[a]
Non-Transgenics	90.00±12.88[a]

[†]Germination was recorded for two weeks from the date of observation of first germination. Values represent average ± standard deviation of germination percentage in 4 replications of transgenic and 3 replications of non-transgenic generations, each replication having a total of 66 seeds. Means compared by Tukey's *t*-test at $P \leq 0.05$. Observed $P=0.32$.

*Values within a column represented by same letter are not statistically different.

Table 3.2 Extractable seed protein content †

Generations	Seed Protein (%)
T_3	8.20±3.40[a*]
T_4	9.01±1.30[a]
T_5	6.09±1.15[a]
Non-Transgenics	10.40±0.36[a]

†Seed protein extracted from ground, defatted seed meal from 15 seeds pooled per replication with Tris buffered saline (150 mM NaCl, 50 mM Tris) at pH 7.4. Each value represents the average ± standard deviation of 2 technical replicates of each of 3 biological replicates and means compared by Tukey's *t*-test at $P \leq 0.05$. Observed P=0.12.

*Values within a column represented by same letter are not statistically different.

Table 3.3 Relative allergen levels in transgenic generations and non-transgenic null segregants.[†]

Generation	Ara h 1[+]	Ara h 3[+]	Ara h 2[‡]	Ara h 6[‡]
T_3	1451.09 ± 105.90[a*]	1697.83 ± 240.50[a]	0[b]	880.12 ± 229.07[b]
T_4	1204.37 ± 214.88[a]	1672.91 ± 102.91[a]	0[b]	872.80 ± 170.19[b]
T_5	1290.54 ± 126.17[a]	1664.09 ± 225.57[a]	0[b]	1265.47 ± 302.60[b]
NTg	1158.00 ± 619.26[a]	1232.34 ± 404.24[a]	3819.32± 237.73[a]	4593.34 ± 381.78[a]
Observed *P* value	0.73	0.17	<0.0001	<0.0001

[†]Relative allergen levels calculated by quantitative western of each allergen and analyzed by Tukey's *t*-test at $P \leq 0.05$. Band volume calculated for each transgenic (Tg) or non-transgenic (NTg) band was normalized with the protein concentration in 'Georgia Green' standards. Each allergen value is expressed as an equivalent in ng/µl of GG total protein.

Each value represents the average ± standard deviation of 3 biological replicates.

+ Loaded 5 µl of protein at 1,000 ng/ µl, ‡Loaded 5 µl of protein at 4,000 ng/ µl

*Values within a column represented by same letter are not statistically different

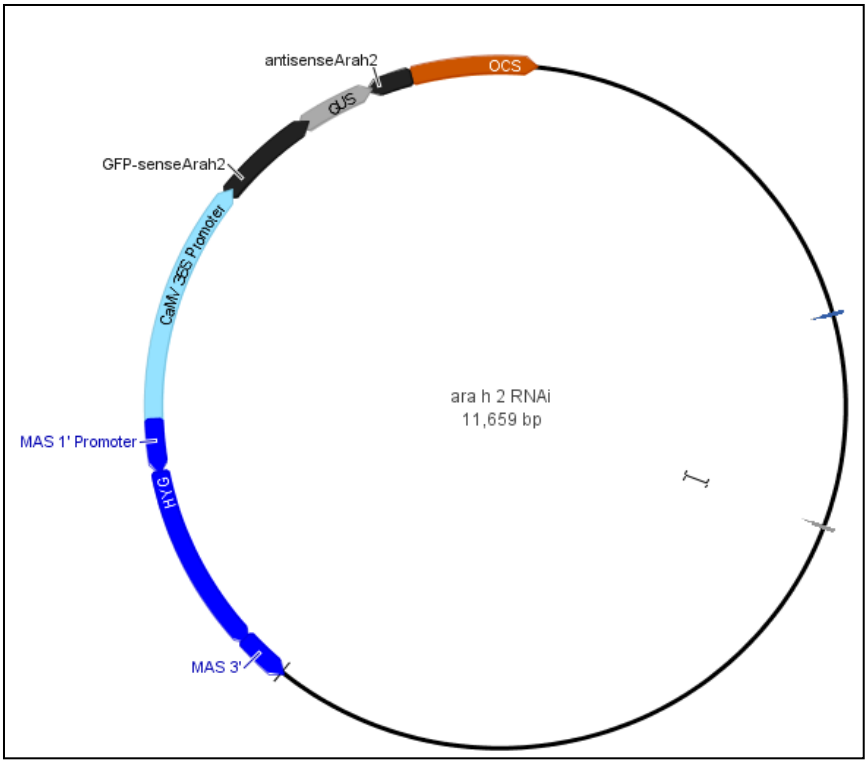

Figure 3.1 Ara h 2 RNAi construct (Chu *et al*., 2008)

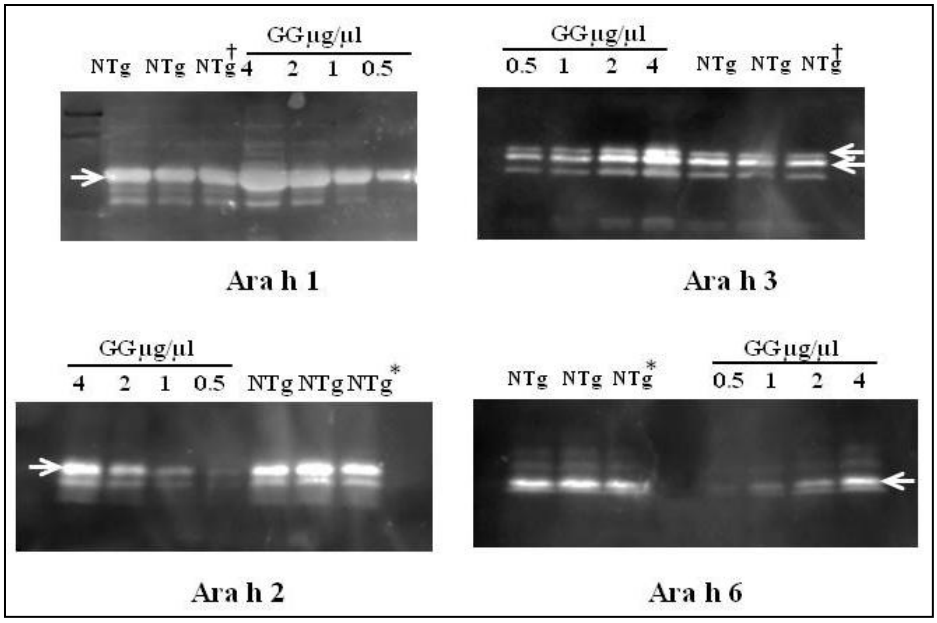

Figure 3.2 Western blot for allergen quantification in non-transgenic control.

†Non-transgenic protein samples at 1µg/µl. *Non-transgenic protein samples at 4µg/µl.

Loaded 5 µl of samples and standards.

Allergen band selected for quantification indicated by arrows. Band volume was normalized with protein concentration in unit volume of 'Georgia Green' standards. Each value represents the average ± standard deviation of 3 biological replicates.

GG-Georgia Green, NTg-Non-transgenic

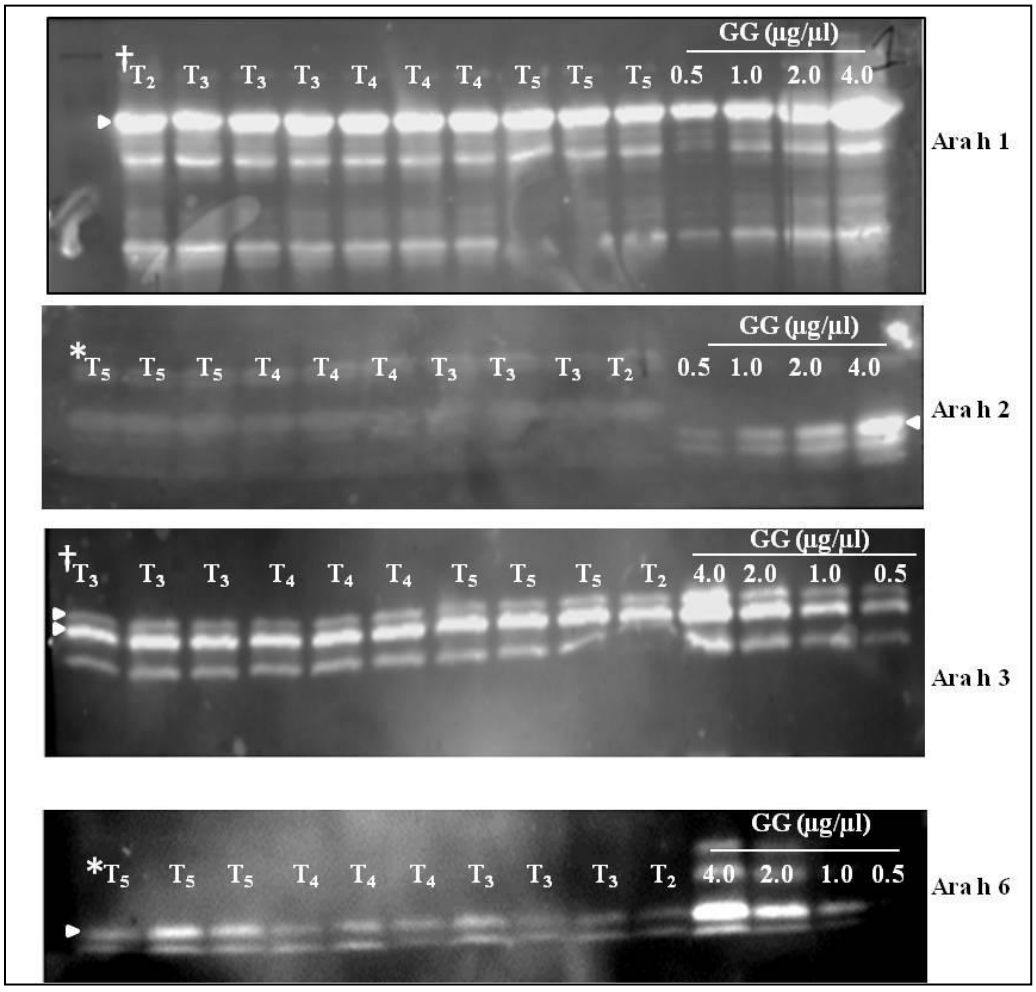

Figure 3.3 Western blot for allergen quantification in transgenic generations T_3, T_4 and T_5.

†Transgenic protein samples at 1 µg/µl. *Transgenic protein samples at 4 µg/µl.

Loaded 5 µl of samples and standards.

Allergen band selected for quantification indicated by arrows.

Band volume was normalized with protein concentration per unit volume of 'Georgia Green' standards. Each value represents the average ± standard deviation of 3 biological replicates.

GG - Georgia Green, Tg - transgenic, NTg - Non-transgenic.

CHAPTER 4

EFFECT OF SOIL SULFUR ON EXPRESSION OF ALLERGEN GENES IN TRANSGENIC PEANUT SILENCED FOR ARA H 2 AND ARA H 6

Abstract

Transgenic peanut line B11.1.1/11, developed by transformation with an RNA interference (RNAi) construct, shows complete silencing of Ara h 2 and greatly reduced levels of Ara h 6 allergens. These two allergens are sulfur-rich proteins. Studies on the effect of soil sulfur on peanut allergens Ara h 1, Ara h 3, Ara h 2 and Ara h 6 in this transgenic line showed that Ara h 1 levels remained unaltered at 0.012, 0.3 and 3.0 mM nutrient solution concentrations. Ara h 3 levels were significantly lower in transgenics and non-transgenic segregants grown with 0.012 mM sulfur-containing nutrient solution, than with 0.3 mM or 3 mM sulfur. Ara h 2 and Ara h 6 levels also were low in null segregant plants grown at 0.012 mM sulfur levels, and increases in Ara h 2 and Ara h 6 levels in null segregant seeds at 0.3 mM and 3.0 mM sulfur levels showed the effect of sulfur on the expression of these sulfur-rich proteins. Ara h 2 was undetectable in transgenic lines at all sulfur levels indicating the stability of transgene silencing. Low Ara h 6 concentrations in transgenic plants at the different sulfur nutrition levels is indicative of the stability of a partially silenced gene to added sulfur.

Chandran, M., Chu, Y., Maleki, S.J. and Ozias-Akins, P. To be submitted to Peanut Science

Introduction

Proteins function as enzymes, antibodies, transporters and structural components in the plant and furnish essential nutrients for humans and other animals. Amino acids are the building blocks of proteins. Amino acids that can be synthesized by humans are the non-essential amino acids. The amino acids that the human body cannot synthesize have to be obtained from the diet and constitute the essential amino acids (EAA). They include methionine, leucine, isoleucine, lysine, phenylalanine, threonine, tryptophan and valine (Jeor *et al*., 2001). A balanced diet must have all essential amino acids and enough nitrogen to synthesize the non-essential amino acids (Berdanier, 2000). The nutritional value of a protein is determined by its essential amino acid content.

Plant proteins, which comprise a major proportion of the human diet, especially in developing countries, are economical to produce, but are nutritionally incomplete as they may be deficient in one or a few EAA. Cereal proteins are low in lysine, tryptophan and threonine. Sulfur-rich amino acids, methionine and cysteine, often are deficient in legume proteins. Mixing foods of complementary amino acid composition, such as cereals and beans, is a means to achieve a high quality diet containing all essential amino acids (Jeor *et al*., 2001; Sun and Liu, 2004).

Plants accumulate proteins in seeds. Seed proteins are classified as housekeeping proteins and storage proteins. Housekeeping proteins comprise small amounts of numerous proteins required for normal cell metabolism. Storage proteins consist of relatively few proteins in large quantities and serve as an energy source during seed germination (Higgins, 1984). Based on solubility, Osborne (1909) classified seed proteins as albumins (extractable with water), globulins (extractable with dilute salt solutions), prolamins (extractable with aqueous alcohol) and glutelins (extractable with acid or alkali solutions) (Higgins, 1984; Osborne, 1909).

Predominantly legumes are rich in albumins and globulins and cereals contain prolamins and glutelins (Mandal and Mandal, 2000).

All storage proteins are synthesized on the rough endoplasmic reticulum (ER), and most undergo post-translational modifications such as glycosylation, chaperone-aided folding, disulfide bridge formation and isomerization. The storage proteins are deposited in membrane-bound organelles called protein bodies derived from the ER or protein storage vacuoles (Higgins, 1984; Mandal and Mandal, 2000; Muntz *et al*., 2001).

Legume seeds are major sources of protein in the human diet as the protein content is higher than in cereals (Bewley, 1994). Demand for non-meat sources of proteins with balanced amino acid profiles is increasing worldwide. Peanut (*Arachis hypogaea* L.) is an important crop legume containing 24-29% protein and 45-53% oil (Cherry, 1977; Kang *et al*., 2007b). In a comparative study including edible tree nuts and a high-oleic Virginia peanut cultivar VA 98R (Venkatachalam and Sathe, 2006), the highest protein content was found in peanuts. Peanut proteins account for 11% of the world's protein supply (Kottapalli *et al*., 2008). Although an economic source of protein, one of the limiting factors of peanut protein quality is the low content of methionine, threonine and lysine (Andersen *et al*., 1998; Basha and Pancholy, 1981; Pancholy, 1978). Work by Venkatachalam and Sathe (2006) found threonine to be a limiting essential amino acid along with sulfur-rich methionine and cysteine.

Conventional breeding to enhance a limiting amino acid was successful in maize with the identification of high lysine *opaque2* mutants (Gevers, 1992; Glover, 1992; Mertz *et al*., 1964). Breeding to increase the sulfur amino acid content in legumes was not feasible as very little variation in amino acid content was shown among the germplasm (Pancholy, 1978; Tabe and Higgins, 1998).

Initial efforts to improve sulfur-rich proteins in plants included sulfur nutrient supplementation. *In vitro* and *in vivo* studies in pea (Randall, 1979), soybean (Gayler and Sykes, 1985; Sexton *et al*., 1998b), and barley (Shewry *et al*., 1983) showed an increased proportion of sulfur-containing protein fractions with an increase in sulfur nutrition. Advancements in the field of molecular technology have enabled the manipulation of the methionine biosynthetic pathway (Tabe and Higgins, 1998) and production of transgenic plants expressing heterologous sulfur-rich proteins (Ufaz and Galili, 2008). Free threonine and methionine contents were increased in transgenic tobacco seeds expressing the feedback-insensitive aspartate kinase enzyme, an enzyme involved in the biosynthesis of methionine, threonine, leucine and isoleucine. Protein-incorporated methionine showed a slight increase while protein-incorporated threonine remained unaffected (Karchi *et al*., 1993).

Sulfur-rich proteins targeted for protein improvement by transgenic technology include 18 kDa Zein with 37% methionine (Chui, 2003), 10 kDa Zein with 30% methionine (Kirihara, 2001), 2S albumin from Brazil nut with 25% methionine (Ampe *et al*., 1986), sunflower seed albumin with 16% methionine and 8% cysteine residues (Kortt *et al*., 1991). This approach was successful in improving the seed sulfur amino acid content of soybean (Altenbach *et al*., 1989; Nordlee *et al*., 1996; Townsend and Thomas, 1994), lupin (Molvig *et al*., 1997; Tabe and Droux, 2002), rice (Hagan *et al*., 2003) and chickpea (Chiaiese *et al*., 2004). However, in transgenic soybean expressing the Brazil nut 2S albumin, the allergenic properties of the transgene-encoded protein were still retained. In radioallergosorbent testing, IgE from 77.78% of subjects allergic to Brazil nut bound to the transgenic soybean protein corresponding to Brazil nut 2S albumin and no IgE binding was detected in non-transgenic soybean (Nordlee *et al*., 1996).

Increasing the sulfur amino acid content by modifying the enzyme biosynthetic pathway or by expressing heterologous sulfur-rich proteins has not been attempted in peanut. Any attempt to alter the seed storage proteins in peanut also should consider the problem of peanut seed allergy. Peanut allergy is caused by consumption of seed storage proteins and is a common cause of fatal anaphylaxis (Kang *et al*., 2007a; Yocum and Khan, 1994). In the United States, approximately 0.8% of young children and 0.6% of adults are affected by peanut allergy (Sampson, 2004) and is a major health concern in developed countries (Scurlock and Burks, 2004). Peanut allergy has low outgrowth rates, and the risk can be reduced only by avoiding peanut products (Wood, 2003). Hence developing peanut varieties with hypoallergenic properties is as important as improving the protein content in seeds.

Thirteen allergens have been identified in peanut (www.allergome.org), Ara h 1 to Ara h 13. Ara h 1, Ara h 2 and Ara h 6 are considered as major allergens since they are recognized by more than 95% of people with peanut allergy (Burks *et al*., 1998; Flinterman *et al*., 2007; Scurlock and Burks, 2004). Ara h 2 and Ara h 6 allergies are found to be severe in children (Flinterman *et al*., 2007). Ara h 1 is a 63.5 kDa vicilin-type glycoprotein (7S) comprising 12-16% of the seed protein (Burks *et al*., 1998; Koppelman *et al*., 2001). Ara h 2 is a 17-20 kDa protein belonging to the conglutin family and accounts for 5.9-9.3% of total seed protein (Koppelman *et al*., 2001; Stanley *et al*., 1997). Cultivated peanut has two genes of Ara h 2, Ara h 2.01 and Ara h 2.02 (Chatel *et al*., 2003; Ramos *et al*., 2006). Ara h 6 is a 14.5 kDa protein of the conglutin family and together with Ara h 2 was the major elicitor of anaphylaxis in a mouse model system (Kulis *et al*., 2012; Wen *et al*., 2007). The cDNAs of Ara h 2 and Ara h 6 share 63% sequence homology (Ramos *et al*., 2006; Stanley *et al*., 1997) and the two proteins are found to cross react in experiments with peanut allergic patient serum (Koppelman *et al*., 2005). The ability of Ara h

3 to cause allergic reactions is less than Ara h 2 or Ara h 6, but it also is considered to be a major allergen (Koppelman *et al*, 2004) since around 50% of peanut allergic patients recognized the recombinant protein (Rabjohn *et al*, 1999). Other peanut allergens have low sensitizing rates and hence are considered as minor allergens at present (Wen *et al*., 2007).

Analysis of 60 peanut accessions and 88 Florida breeding lines identified multiple breeding lines and peanut cultivars ('Georgia Red' and 'Florunner') that were missing a 36 kDa Ara h 3 isoform, but none were null mutants for major peanut allergen families and the low content of one allergen was compensated by increased levels of another allergen (Kang *et al*., 2007a). Although the 66 kDa Ara h 1 protein was absent in two Indonesian varieties (Bali-1 and Bali-2), no reduction in allergenicity was detected in a mediator release assay of rat basophilic leukemia cells (Krause *et al*., 2010). Any allergenicity reduction due to elimination of Ara h 1 protein might be compensated for by other allergens (Krause *et al*., 2010). Hence, the potential for using natural diversity for conventional breeding to reduce allergens in seeds is limited and reverse genetics technologies need to be adopted. RNA interference (RNAi) was used to silence Ara h 2 (Dodo *et al*., 2008) and the two allergens which share sequence homology, Ara h 2 and Ara h 6 (Chu *et al*., 2008). In the simultaneous silencing experiment by Chu *et al.* (2008), western blotting of the seed protein from transgenic line B11.1.1/11 detected no signal for Ara h 2 and a faint signal for Ara h 6. Thus this line shows essentially complete silencing of Ara h 2 and greatly reduced levels of Ara h 6. Also no IgE binding of the transgenic seed protein was detected to sera from subjects allergic to multiple peanut proteins.

Ara h 2 and Ara h 6 belong to the sulfur-rich conglutin family of proteins. The number of methionine residues in seed storage proteins of peanut varies from three in Ara h 2.01 (1.88%)(GenBank ID EF609641.1), four each in Ara h 2.02 (2.33%) (GenBank ID AY158467)

and Ara h 3 (0.78%) (GenBank ID HM640250), eight in Ara h 1 (1.28%) (GenBank ID AF432231), and nine in Ara h 6 (6.21%) (GenBank ID EF609643). The number of cysteine residues varies from six in Ara h 3 (1.17%) (GenBank ID HM640250.1), seven in Ara h 1 (1.12%) (GenBank ID AF432231.1), eight in Ara h 2.01 (5.0%) (GenBank ID EF609641.1) and Ara h 2.02 (4.65%) (GenBank ID AY158467), and ten in Ara h 6 (6.9%) (GenBank ID EF609643) (Table 4.1). Silencing of Ara h 2 did not alter the phenotypic and agronomic traits in the transgenics (Chu *et al*., 2008; Dodo *et al*., 2008). The protein profile of other allergens remained unaffected based on Coomassie staining and western blotting in Ara h 2-silenced transgenic lines developed by Chu *et al*., (2008) but was seen to vary considerably from wild-type in Ara h 2-silenced transgenic lines developed by Dodo *et al*., (2008). Silencing seed proteins should always consider the corresponding changes in other proteins associated with silencing. Although the protein profile of transgenic line B11.1.1 appeared to be the same as wild-type in western blotting (Chu *et al*., 2008), collateral changes, such as up-regulation of 13-lipoxygenase and 11S Ahy-3 proteins were detected by mass spectrometry analysis (Stevenson *et al*., 2009).

Soil sulfur levels can be depleted as a consequence of cultivation of high yielding cultivars, use of high analysis fertilizers and reduced industrial emission of SO_2 into the atmoshpere (Scherer, 2001). Hence deficiency of sulfur, which is a major plant nutrient, can limit agricultural production and also affect produce quality (McGrath, 1996). Studies have shown increases in the sulfur-containing proteins in seeds, protein profile changes and yield increases by external sulfur application (Randall, 1979; Gayler and Sykes, 1985; Sexton *et al*., 1998). In peanut, the major seed storage proteins are also allergens, and increasing the seed quality by altering sulfur nutrition may also affect the content of seed allergens. The effect of variable sulfur nutrition on

expression of the incompletely silenced *Ara h 6* genes, the more completely silenced *Ara h 2* genes, and other major peanut allergens in the RNAi line B11.1.1/11 was tested in the current study.

Materials and Methods

The effect of soil sulfur levels on the concentrations of Ara h 1, Ara h 2 and Ara h 3, and Ara h 6 in peanut seeds was tested in the $T_{3:4}$ generation of the transgenic (Tg) line B11.1.1/11. Non-transgenic null segregants (NTg) were grown as the controls (C). Standard curves for protein quantification were developed with proteins extracted from field-grown 'Georgia Green' seeds.

Experimental Design and Planting

Greenhouse planting was done during December 2009. Seeds were planted in polyethylene containers 90 cm x 60 cm x 20 cm in dimension filled with a mixture of sand, vermiculite and perlite in the ratio 2:1:1. The planting mixture was irrigated with deionized water before planting. There were fifteen containers total, five for each sulfur nutrient level, each containing 3 Tg and 3 control NTg plants. The positions of containers and plants within the containers were randomized.

The nutrient solution for irrigation was prepared in deionized water with constant concentrations of macro- and micro-nutrients except for sulfur (Table 4. 2). Three concentrations of sulfur (S) were provided as $MgSO_4$ at 3 mM (High S), 0.3 mM (Medium S) and 0.012 mM (Low S). The concentration of Mg was kept constant with $MgCl_2$ at 2.97 mM (Low S), 2.7 mM (Medium S) and nil for high S solutions. The pH of the solution was adjusted to 5.8. There were 5 replications of each S treatment concentration (0.012, 0.3, and 3 mM). The plants were watered

on alternate days with nutrient solution at 500 ml per plant. The potting medium was flushed with deionized water after every three applications of nutrient solution to prevent salt accumulation in the root zone.

Testing for Presence of the *Ara h 2* Transgene in T_3 Seeds

Seeds from T_2 plants (T_3 seeds) were tested by the polymerase chain reaction (PCR) for the presence of the *Ara h 2* silencing cassette (Chu *et al.,* 2008) (Figure 4.1) before planting. DNA was extracted from a small portion of the cotyledon by the CTAB method (Chu *et al*., 2009) and PCR amplified with primers specific to the CaMV 35S promoter of the RNAi construct to test for the presence of the transgene. PCR amplification was performed in a 20 µl reaction mix containing 2 µl DNA, 1X PCR buffer, and 0.2 mM dNTPs, 0.2 µM of each primer, 1% PVP, 0.01% bovine serum albumin (BSA) and 0.16 U *Taq* DNA polymerase. The thermal cycling conditions were 5 min denaturation at 94ºC, 38 cycles of 30 s denaturation at 94°C, 30 s annealing at 57°C and 1 min extension at 72°C with a final 7 min extension at 72°C. The presence of the 975 bp insert was confirmed by running 10 µl of the PCR product on 1% agarose gels in TBE buffer (Tris-borate-EDTA) containing ethidium bromide (0.1 µg/1 ml). Ninety seeds were tested for the presence of the insert. Control plants, which were non-transgenic segregants from the T_2 generation, were also tested by the same procedure.

Seed Protein Extraction

Seeds were selected from each plant for seed protein extraction. Two seeds per plant were selected from low S treatment (0.012 mM) and pooled for each replication to obtain six seeds for each Tg and NTg genotype per replication. For medium (0.3 mM) and high (3 mM) S treatments five seeds per plant were selected to obtain fifteen seeds for each Tg and NTg genotype per replication. Thus there were 5 pooled replicates from each sulfur treatment. Homogeneity of

seed maturity within treatments was maintained by choosing the most mature seeds of nearly uniform size (Figure 4.2). Field-grown controls were fifteen 'Georgia Green' seeds from a pooled sample.

Seeds were weighed, testa removed, embryos and cotyledons crushed in a mortar and pestle then ground with liquid nitrogen and defatted with hexane at 1:35 (w/v) for 4 h, keeping the sample on ice. The defatted samples were lyophilized and stored at -80°C.

Seed powder was passed through a 30 mesh sieve and seed protein was extracted using Tris–buffered saline (TBS-150 mM NaCl, 50 mM Tris) at pH 7.4, containing EDTA-free protease inhibitor (Roche Diagnostics, Indianapolis, IN, USA) (Porterfield *et al*., 2009). The defatted and lyophilized peanut flour was suspended in ice-cold Tris-buffered saline at a ratio of 1:10 (w/v) and mixed overnight at 4°C. The mixture was centrifuged at 16,060 x g for 15 min at 4°C. Crude protein extract was collected, aliquoted and stored at -80°C.

Seed protein was quantified with Pierce Bicinchoninic acid (BCA) protein assay kit (Thermo Scientific, Rockford, IL, US) as per the manufacturer's protocol. The results present the average of three technical replications for each of four biological replicates.

Quantitative Western Blots

The standard for quantification was derived from 'Georgia Green' protein samples loaded at 20,000, 10,000, 5,000 and 2,500 ng per well for each gel, along with Tg and NTg protein samples. For Ara h 1 and Ara h 3, samples were loaded at 5,000 ng of total protein per well. For Ara h 2 and Ara h 6, samples were loaded at 20,000 ng of total protein per well. Five microlitres of standard and sample proteins were loaded per well.

Seed proteins were denatured at 95°C for 5 min and resolved by running on 15% sodium dodecyl sulfate polyacrylamide gels (SDS-PAGE) for 1.5 h at 100 V in a Bio-Rad Mini-PROTEAN 3

Cell apparatus (Bio-Rad, Hercules, CA, USA) with running buffer containing 0.3% Tris-base, 1.44% glycine, 1% SDS and devoid of methanol. Proteins were blotted onto Amersham Hybond-P PVDF membrane (GE Healthcare, Pittsburgh, PA, USA) by electro-transfer at 100 V for 1 h in a Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell apparatus with transfer buffer containing 0.3% Tris base and 1.44% glycine. The PVDF membrane was placed between blotting papers and dried at 37°C overnight. To test for equal loading of proteins, the membrane was stained with SYPRO Ruby (Invitrogen, Grand Island, NY, USA) as per the manufacturer's protocol. Briefly, proteins on PVDF membrane were fixed with a solution containing 10% methanol and 5% acetic acid, washed four times with deionized water for 5 minutes each, stained with SYPRO Ruby and washed. Fluorescence detection and imaging were done with a STORM scanner (Amersham Biosciences, Piscataway, NJ, USA) at 450 nm/520 nm excitation/emission. For western blotting, the membranes were blocked with 5% non-fat dry milk in TBST (Tris buffered saline with Tween), containing 0.02 M Tris-HCl, 0.8% NaCl and 0.1% Tween-20 for 1 h. This was followed by incubation with anti-Ara h 1 (1:5000), -Ara h 2 (1:8000), -Ara h 3 (1:5000) or -Ara h 6 (1:5000) primary antibodies custom manufactured by Sigma Immunosys (Woodlands, TX). Subsequent to three washes with TBST the membranes were reacted with Alkaline Phosphatase (AP)-labeled rabbit anti-chicken/turkey secondary antibody (Invitrogen) at 1:10,000 for 30 min. Signal detection was done with Enhanced Chemifluorescence substrate (ECF) (GE Healthcare, Piscataway, NJ, USA) treatment for 5 min. Fluorescence was detected by imaging at 450/520 nm excitation/emission.

Quantification of Allergens in Seed Protein Extract

The standards for quantification were derived from 'Georgia Green' protein samples loaded at 20,000, 10,000, 5,000 and 2,500 ng per well for each gel, along with Tg and NTg protein samples. For Ara h 1 and Ara h 3, quantification was done using 5,000 ng total protein (1,000 ng/µl). Ara h 2 and Ara h 6 were quantified with 20,000 ng protein (4,000 ng/µl) from each Tg and NTg sample. Band volume quantification was done with Image Quant software with object average background correction for each band. Allergen band volume calculated for the samples was normalized with 'Georgia Green' protein concentrations. The standard curve is developed with band volume for each allergen protein in 'Georgia Green' and concentration of protein in 1 µl of 'Georgia Green' dilutions (4 µg, 2 µg, 1 µg and 0.5 µg). Sample band volume calculated was normalized with 'Georgia Green' protein concentrations and hence the concentration corresponding to every band of sample will be equivalent to the total protein in 1 µl of 'Georgia Green' protein. The concentration of individual allergens will be proportional to the total protein concentration. Hence allergen is quantified relative to total protein concentration and not the absolute value. The concentrations were compared by two- way ANOVA and mean separation done with Tukey's *t*-test at $P \leq 0.05$ SAS software (SAS Institute Inc. 2009)

Carbon-Nitrogen-Sulfur (CNS) Analysis

Aliquots of seed meal prepared for protein extraction were used for CNS analysis. Replicate seed samples, four each from low, medium and high sulfur treatments were used. The four replicates selected were from the same samples as used for quantitative western blots. Analysis was done with the dry combustion method in a Leco CNS 2000 analyzer at the Laboratory for

Environmental Analysis, University of Georgia, Athens. Carbon and sulfur were detected as CO_2 and SO_2, respectively, by infra-red absorption measurement. Nitrogen was determined by thermal conductivity.

Amino Acid Analysis

Free and total amino acid analysis was done in one pooled sample each of seeds from low, medium and high sulfur plants of transgenic and non-transgenic genotypes. Seed samples (0.05 g each) were pooled from four pooled replicates (same as the ones used for quantitative western blots) from each sulfur treatment. Amino acid analysis was done at the Molecular Structure Facility, UC Davis, CA. Briefly, for total amino acid detection, approximately 10 mg of seed meal was dissolved in 1N HCl and heated to 56°C for 15 min. The mixture was defatted with CH_2Cl_2 (2-chloro methane), supernatant collected, dried, and hydrolyzed with 6N HCl at 110ºC for 24 h. NorLeu dilution buffer was added and samples analyzed on an L-8800 Hitachi instrument. For analysis of cysteine and methionine, performic acid oxidation was done prior to acid hydrolysis. For free amino acid determination, approximately 10 mg of seed meal was dissolved in 1N HCl and heated to 56°C for 10 min, proteins were precipitated with 10% sulfo-salicylic acid, supernatant was collected, dried, dissolved in AE-cysteine buffer and analyzed. on a L-8800 Hitachi instrument.

Results

Seed Weight

Plants that received the low sulfur nutrition (0.012 mM S) showed considerable yellowing, reduced vegetative growth and yield when compared to the plants grown under medium (0.3 mM) and high (3 mM) sulfur nutrition (Figures 4.3 and 4.4). The 100 seed weight (Table 4.3) of Tg and NTg plants was lowest with low S treatment and seed weight at low S was significantly

different from medium and high S treatments. The 100 seed weight of Tg and NTg plants under medium and high S conditions varied from 33.39 g to 39.97 g and showed no statistically significant difference between either S treatment or genotype. Seed weight was not influenced by genotype and no interaction was seen between genotype and sulfur. Plants grown under low S levels produced fewer seeds (2.7-4 seeds/plant), when compared with medium sulfur (27.5-42 seeds/plant) and high sulfur (53-57.2 seeds/plant).

Total Protein

Total protein in seed was seen to be more dependent on genotype (P=0.005) than sulfur levels (P=0.012). No genotype X sulfur interaction was observed in total seed protein in two-way ANOVA analysis. Total seed protein showed significant variation ($P \leq 0.05$) between transgenic plants (5.27%) and non-transgenic plants (4.39%) and between low S (4.35%) and high S (5.47%) treatments in mean separation by Tukey's *t*-test (Table 4.4).

Relative Concentrations of Major Peanut Allergens Ara h 1, Ara h 2, Ara h 3 and Ara h 6

The relative concentrations of the peanut proteins Ara h 1, Ara h 2, Ara h 3 and Ara h 6, in the transgenic and non-transgenic seeds at different S levels are shown in Table 4.5. As shown by two-way ANOVA relative levels of Ara h 1 was not influenced by genotype or sulfur nutrition. Sulfur nutrition influenced seed Ara h 3 levels (Observed P<0.0001), but no genotype or genotype x sulfur influence was noticed. Effect of genotype (Observed P<0.0001), sulfur (Observed P<0.0001) and genotype x sulfur (Observed P<0.0001) was noticed for seed Ara h 2 levels. Similarly for Ara h 6, significant influence of genotype (Observed P<0.0001), sulfur (Observed P<0.0001) and genotype x sulfur interaction (Observed P<0.0001) was observed. There was no significant variation in the relative concentration of Ara h 1 (Figure 4.5) between transgenic and non-transgenic T_4 seeds grown at three different sulfur levels (P=0.36). The mean

value of Ara h 3 did not vary between transgenic and non-transgenic genotypes at all three sulfur levels tested (Figure 4.6), although relative Ara h 3 protein concentration in seeds of plants grown under low sulfur status was significantly lower by 9-11 fold than plants grown with medium or high sulfur nutrition ($P \leq 0.0001$). Sulfur levels had no influence on the Ara h 2 protein concentration in Ara h 2-silenced transgenic plants as no Ara h 2 signal was detected in quantitative western blot experiments (Figure 4.7). Non-transgenic plants grown under low sulfur nutrition showed significantly lower concentrations of Ara h 2 and Ara h 6 proteins when compared to non-transgenic plants in medium and high sulfur levels. The amounts of Ara h 2 and Ara h 6 were lowered by ~ 6 times ($P \leq 0.0001$) and ~ 22 times ($P \leq 0.0001$), respectively, in the non-transgenics at low S nutrition compared to the two higher S levels. Quantitative westerns detected Ara h 6 in transgenics grown under medium and high sulfur nutrition, but at significantly lower amounts than in non-transgenics (Figure 4.8).

Seed CNS Content

The percentage carbon in dry defatted meal varied from 49% in low S to 56% in high S treatment (Table 4.6). No significant variation in carbon was seen between transgenic and non-transgenic plants; however, seed carbon was influenced by sulfur (P=0.00012) and no genotype x sulfur effect was observed. The percentage carbon in seeds from low S treatment was significantly lower than in seeds in medium and high S treatments. Seed nitrogen content varied from 6.13-9.12% among the three S treatments between genotypes (Table 4.6). Both genotype (Observed P<0.0001) and soil sulfur (P=0.00022) levels showed an effect on seed nitrogen in two-way ANOVA. Seed nitrogen content was significantly higher in the low S treatment (8.1-9.12%) and no significant variation was shown between medium and high sulfur. Significant variation was shown between transgenic and non-transgenic plants in seed nitrogen content with non-transgenics having higher nitrogen than transgenics. Seed sulfur was influenced by sulfur

nutrition (Observed $P<0.0001$) and genotype ($P=0.00015$) and no genotype x sulfur interaction was observed, Sulfur content in seeds ranged from 0.16-0.32 % in the three S nutrient regimes among genotypes (Table 4.6). Non-transgenic plants under three S treatments showed significantly higher seed S content than transgenics, and seeds from medium and high S soils had significantly higher S than seeds from low S. Seed N:S ratio was influenced by sulfur alone (Observed $P<0.0001$). Nitrogen:sulfur ratios varied from 51.30 in low S transgenic to 25.35 in medium S non-transgenic plants. Significant variation was seen between the low S treatment and the two higher levels. The N:S ratio was higher in low S treatment and no difference was seen between transgenic and non-transgenic plants within a S level (Table 4.6).

Free and Total Amino Acids

The total and free amino acids per unit seed mass (Table 4.7) and proportion of each amino acid to the total amino acids under various S treatments (Tables 4.8 and 4.9) were determined to assess any differences in amino acid profile among treatments. These data are helpful in deriving the trend in amino acid profile changes under different S levels in genotypes tested since statistical analysis was not possible given that replicate samples were not analyzed. Accordingly in the present study, the highest total amino acid content was shown by low S treatment, probably attributable to the increase in the free amino acid pool. At low S, both free and total arginine levels were higher than in medium and high S treatments. Arginine represented 68% of the free amino acid pool and ~18% of the total amino acid pool. Under higher S levels it represented ~5-9% and ~7-9% of free and total amino acid pools, respectively. When compared among low, medium and high S levels, and between transgenics and non-transgenics, free amino acid percentages showed little variation between genotypes within an S level. Wide variation was seen between low S and the two higher S levels for all amino acids (Table 4.8). Under low S, the

majority of amino acids were found in low proportions, except for six, namely histidine, arginine, asparagine, ornithine, hydroxyproline and glutamine. Among these six, histidine, ornithine, hydroxyproline and glutamine were measurable only under low S conditions. Major noticeable changes under low S were the high proportions of arginine (6-14 times higher than in medium and high S in the total free amino acid pool) and asparagine (~2 times higher in low S vs. medium and high S transgenics in the total free amino acid pool). Citrulline showed differences between transgenics and non-transgenics in both medium and high S treatments. Total amino acids (Table 4.9) did not vary much, for most amino acids, between S levels compared with the variation seen for free amino acids. Low S treatments resulted in approximately 2-fold higher levels of arginine. Arginine levels also were higher in non-transgenics under medium and high S levels. Cysteine and methionine levels were higher under medium and high S treatments, as well as higher in non-transgenics compared with transgenics under these two levels.

Discussion

Low sulfur nutrition led to a significant reduction in seed weight compared with seeds from plants grown under medium and high sulfur nutrition. Sulfur is an essential macronutrient for plants and the deficiency is associated with chlorosis, growth retardation and reduced yield (Nikiforova *et al*., 2004; Zhao *et al*., 1999). Sulfur impacts CO_2 assimilation, Rubisco enzyme activity and photosynthesis in crops including wheat (Gilbert *et al*., 1997) and rice (Lunde *et al*., 2008). Reproductive growth was more affected than vegetative growth by sulfur deficiency in wheat (Zhao *et al*., 1999). Sulfur participates in several redox reactions, is essential for synthesis of amino acids methionine and cysteine, several proteins, co-enzymes and iron-sulfur clusters. Adequate soil sulfur concentration is essential for nitrogen fixation (Scherer *et al*., 2008). Poor

growth, reduced seed weight and low yield under low sulfur observed in this study and other crops viz., barley (Shewry *et al*., 1983), *Brassica nap*us (Eppendorfer and Eggum, 1992) rice (Khurana *et al*., 1999), cotton (Yin *et al*., 2011), bishop's weed (Ahmad *et al*., 2010) and white lupin (Cazzato *et al*., 2012), may possibly be due to the changes in various metabolic activities involving sulfur. Plants under medium and high sulfur showed vigorous growth as seen in studies with *Lupinus angustifolius* (Blagrove *et al*., 1976).

In the current study, variation was observed in total seed protein between transgenic seeds grown under high sulfur and non-transgenic seeds grown under low sulfur. Both genotype and S levels influenced seed protein as shown by ANOVA analysis. This may possibly be due to the higher levels of Ara h 3 in transgenic plants under high S compared to low S plants. In studies on nitrogen and sulfur fertilization of *Camelina sativa*, it was observed that although total seed protein increased under high sulfur nutrition, the effect was not significantly different from low sulfur treatment (Losak *et al*., 2011). Crude protein percentage in lupin (Cazzato *et al*., 2012) and *Brassica* (Malhi *et al*., 2007) increased significantly when sulfur was applied at 30 kg/ha compared with no sulfur (0 kg/ha) application, but percent protein was not significantly increased at sulfur doses higher than 30 kg/ha. As opposed to these studies varying sulfur levels showed no significant influence on seed protein content in field studies with oil seed rape or in lupin grown with S for the first six weeks and later with or without S (Asare and Scarisbrick, 1995; Govahi and Saffari, 2006).

Peanut seed protein ranged from 20.7% to 28.1% in analysis of 64 peanut genotypes (Dwivedi *et al*., 1990) and was reported as 21% in a comparative study of edible seed nuts (Venkatachalam and Sathe, 2006). In the present study, the average protein concentration in the various samples was lower than previously reported since proteins were extracted with Tris buffered saline at pH

7.4 containing 150 mM NaCl. When various buffers and meal to buffer ratio were tested for peanut protein extraction efficiency it was found that 1M NaCl-20 mM sodium phosphate buffer at pH 7.0 and 1:18 buffer ratio gave the maximum globulin recovery along with albumins (Mahaboobbasha and Cherry, 1976). A study by Poms *et al.* (2004) showed extraction efficiencies were more dependent on buffer pH, with buffer at higher pH yielding more proteins. Accordingly TBS at pH 8.2 yielded 35% more peanut protein than TBS at pH 7.4, and maximum protein was extracted with sodium borate buffer at pH 9.0, with salt concentrations at 7 mM and 15 mM showing no variation in protein extraction efficiency. Higher protein extraction efficiency was observed using 6 M urea at pH 8.0, but the buffer is highly denaturing. Hence selection of buffer would also depend on the requirements of subsequent assays (Poms *et al*., 2004). Peanut seed protein fractions were grouped as albumins, arachin and conarachin (Johns, 1916). Arachin and conarachin are salt soluble globulins which account for 87% of the seed protein (Irving, 1945). Lower protein content per unit mass of defatted meal recovered in this study compared to other studies could possibly be due to the low salt concentration (150 mM) and pH of the buffer employed.

The levels of Ara h 1 in low, medium and high sulfur treatments among transgenic and non-transgenic plants showed no significant variation. This was expected, as Ara h 1 is not categorized as a sulfur-rich protein. On the other hand, low levels of sulfur caused an approximately 9-11-fold reduction in Ara h 3 levels, compared to high sulfur levels, irrespective of the plant genotype. Ara h 3 belongs to the glycinin family of proteins; glycinin is the major 11S storage protein in soybeans (Rabjohn *et al*., 1999). Glycinin is an oligomer of six subunits (Badley *et al*., 1975). Each subunit, encoded by a separate gene, is composed of acidic and basic polypeptides linked by a single disulfide bond between highly conserved cysteines at positions

112 (in the acidic sub unit) and 390 (in the basic subunit). The conserved cysteines in the acidic subunit at positions 36 and 69 are involved in intra-chain disulfide linkages (Nielsen, 1984; Nielsen *et al*., 1989; Staswick *et al*., 1984). Disulfide bonding in glycinin is important for stabilizing protein conformation (Hoshi and Yamauchi, 1983) and in protein denaturation and aggregation by heat (Zarins and Marshall, 1990). Low sulfur levels in soil are found to reduce the cysteine and methionine contents in barley (Shewry *et al*., 1983), *Arabidopsis* (Nikiforova *et al*., 2006), pea and cauliflower (Eppendorfer and Eggum, 1995). A similar reduction of cysteine content in seeds of peanut under low sulfur nutrition was observed, thus lowering the concentration of the glycinin Ara h 3, which also has conserved cysteine residues in acidic and basic polypeptides.

Lines silenced for Ara h 2 showed no chemiflourescence signal in the T_4 generation seeds in western blotting experiments. Similar results were shown by Chu *et al.* (2008) in chemiluminescent detection of Ara h 2 proteins from the T_2 generation seeds. These results confirm the stability of RNAi silencing in the T_3 generation and that variation in S nutrition does not alter the stability of transgene expression. Transgene expression stability was shown to be influenced by temperature in *Chrysanthemum* transformed with the male sterile gene *CmETR1/H69A* (Shinoyama *et al*., 2012) and by light in *Nicotiana benthamiana* containing a GFP transgene (Kotakis *et al*., 2010). However in studies with barley carrying the ubiquitin-driven *bar* gene, transgene stability was not affected by heat shock or water and nutrient stress (Meng *et al*., 2006). Statistically significant low levels of Ara h 6 signal detected in transgenics at medium and high S levels, compared to non-transgenics showed the stability of transgene expression to added S. At low S levels, the available S was insufficient for the synthesis of Ara h 6 in the transgenic lines since no chemifluorescent signal was detected. Increased legumin

mRNA levels and higher legumin mRNA stability has been reported under high S conditions in pea (Beach *et al*., 1985; Evans *et al.*, 1985). Transcription and transcript stability were not tested in the current study. Even if mRNA levels of *Ara h 6* increased with S nutrition, the low level of Ara h 6 in transgenic peanut at medium and high sulfur levels should be the consequence of efficient post-transcriptional mRNA degradation through RNAi, which probably is unaffected by sulfur nutrition.

Seed sulfur contents in the current study showed no significant variation between transgenics and non-transgenics under different sulfur regimes (S X genotype), although there was a trend for lower S in transgenics. Similarly no significant variation in total seed sulfur was noticed between non-transgenic and transgenic lupin overexpressing sunflower seed albumin (SSA) when grown under similar conditions, but the amount of reduced S (methionine) was higher in transgenics compared to wild-type (Molvig *et al.*, 1997). In an independent study, seed sulfur in wild-type and transgenic lupin overexpressing SSA showed higher total S and sulfur-containing amino acids with increasing soil S status (Tabe and Droux, 2002); however, transgenic plants carrying the SSA transgene showed higher seed S than non-transgenics, a result expected to be contrary to our study given the difference between overexpressing and silencing S-rich proteins. The total seed S-containing amino acids was found to increase with increasing S levels in this study, and as expected RNAi lines showed less S-containing amino acids than non-transgenic at all S levels. A similar response to S was seen in non-transgenic lupin, but transgenic lupin with added SSA showed higher methionine than cysteine (Molvig *et al*., 1997; Tabe and Droux, 2002), which was not similar to the transgenic status in this study where S-rich protein was silenced.

Seed nitrogen increased under soil S deficiency. Seed N varied significantly between transgenic and wild type. Under abiotic stress nitrogen-containing compounds accumulate, the most

common being amino acids such as arginine, proline, amides (glutamine and asparagine), citrulline and ornithine, which are formed as products of detoxification of excess ammonium produced under stress (Karmoker *et al*., 1991; Rabe, 1990; Rai, 2002). In barley grown under low sulfur, a high proportion of aspartic acid and asparagine and low levels of methionine and cysteine were found (Shewry *et al*., 1983). The difference in N% between transgenic and non-transgenic plants can possibly be due to the differences in amino acid content in the two genotypes, with the non-transgenic controls having high total amino acid content compared to transgenics. Low S and high nitrogen content of seeds resulted in high and unfavorable N/S ratios under S deficiency and varied significantly from plants under medium and high S nutrition. This is in accordance with earlier studies where higher seed and plant S was seen with higher S nutrition, when S was combined with both high and low N levels (Chiaiese *et al*., 2004; Sexton *et al*., 1998b).

As developing seeds can modulate the amino acid and protein composition based on nutrient availability (Tabe *et al*., 2002) the amino acid profile of the plants grown at different sulfur levels was tested. Members of a serine-arginine rich protein family are induced during abiotic stress (Duque, 2011). Accumulation of arginine-rich stress proteins might account for the increased arginine in the free and total amino acid pools under low S conditions, which can also lead to high nitrogen content as seen in this study.

Free amino acids (FAA) are critical for the roasted flavor of peanut. Free amino acids react with the sugars in seed to produce organic compounds, which impart the typical roasted peanut flavor through the Maillard reaction (Basha and Young, 1985; Mason *et al*., 1967). Typical roasted flavor is attributed by glutamic acid, glutamine, asparagine, phenylalanine, aspartic acid and histidine. Free arginine, tyrosine, lysine and threonine are responsible for the atypical flavors

(Mason *et al*., 1967; Newell *et al*., 1967). Arginine, although beneficial to human health in multiple ways (Wu *et al.*, 2009), can be detrimental at high concentrations with regard to peanut roasting quality. Hence sulfur nutrition is important for maintaining the quality of peanut since low S levels showed higher arginine irrespective of the plant genotype.

Sulfur is a component of chloroplast Fe-S cluster proteins and low S concentrations lead to reduced chlorophyll, Rubisco level, photosynthesis and CO_2 fixation in rice and barley (Karmoker *et al*., 1991; Lunde *et al.*, 2008). Low seed carbon content in seeds of low S treatment in this study may be a consequence of the reduced growth and photosynthesis under S deficiency. Carbon, nitrogen and sulfur contents in seeds result from interactions between carbon, nitrogen and sulfur metabolic pathways. When S is limited, activity of nitrate reductase enzyme is reduced leading to reduced nitrogen assimilation (Jamal *et al*., 2006; Ruiz *et al*., 2005) as the enzymatic activity of nitrate reductase depends on S-Mo biochemistry (Cerqueira *et al*., 2009). Nitrogenase activity and nitrogen fixation in roots were reduced under S deficiency (Varin *et al*., 2010) as S is an essential component of the Fe-Mo cofactor of the nitrogenase protein (Chan *et al*., 1993). The unbalanced nitrogen metabolism results in the accumulation of nitrogen compounds, especially nitrogen-rich amino acids (Hoefgen and Nikiforova, 2008; Varin *et al.*, 2010). Reduced S also results in low sulfolipids and S-adenosyl methionine, required for photosynthesis. Thus S limitation leads to reduced metabolism by its influence on photosynthesis and the carbon assimilative pathway (Hoefgen and Nikiforova, 2008).

Conclusion

Studies with the transgenic line B11.1.1/11, with silenced Ara h 2 and greatly reduced Ara h 6 protein/allergen levels, have shown the stability of transgene silencing at varying levels of sulfur. The expression of Ara h 1 was not increased by high sulfur nutrition and did not differ between

genotypes. Low S reduced Ara h 3 levels compared to medium and high S and no genotypic effect was seen. High S treatments showed higher Ara h 2 in non-transgenics compared to transgenics but, Ara h 2 remained silenced at all three S levels in transgenics showing the stability of transgene. Medium and high sulfur levels increased Ara h 6 proteins in non-silenced lines and silenced lines compared to low S nutrient concentration. However the concentration of Ara h 6 was significantly lower in partially silenced transgenic lines, compared to non-transgenics. The stability of allergen gene silencing is promising for using these materials to test for allergenicity changes due to allergen reduction. At present allergenic reactions to peanut can only be tackled by avoiding peanut and peanut products. Reducing the allergenic properties in food crops is a contemporary focus of research. As evidenced from this study even partial silencing of an allergen gene by molecular techniques should be effective in greatly reducing the allergen content of peanut.

References

Ahmad S., Fazili I.S., Khan S.N., Abdin M.Z. (2010) Standardization and estimation of nitrate reductase activity in the leaves of *Ammi majus* L. (Bishops weed) in relation to sulphur deficiency and seed yield. Aust. J. Crop Sci. 4:515-522.

Altenbach S.B., Pearson K.W., Meeker G., Staraci L.C., Sun S.S.M. (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol. Biol. 13:513-522.

Ampe C., Vandamme J., Decastro L.A.B., Sampaio M., Vanmontagu M., Vandekerckhove J. (1986) The amino-acid-sequence of the 2S sulfur-rich proteins from seeds of brazil nut (*Bertholletia excelsa* HBK). Eur. J. Biochem. 159:597-604.

Andersen P.C., Hill K., Gorbet D.W., Brodbeck B.V. (1998) Fatty acid and amino acid profiles of selected peanut cultivars and breeding lines. J. Food Compost. Anal. 11:100-111.

Asare E., Scarisbrick D.H. (1995) Rate of nitrogen and sulfur fertilizers on yield, yield components and seed quality of oilseed rape (*Brassica napus* L). Field Crops Res. 44:41-46.

Badley R.A., Atkinson D., Hauser H., Oldani D., Green J.P., Stubbs J.M. (1975) Sructure, physical and chemical properties of soy bean protein glycinin. Biochimica Et Biophysica Acta 412:214-228.

Basha S.M., Young C.T. (1985) Changes in the polypeptide composition of peanut (*Arachis hypogaea* L) seed during oil roasting. J. Agric. Food Chem. 33:350-354.

Basha S.M.M., Pancholy S.K. (1981) Identification of methionine-rich polypeptides in peanut (*Arachis hypogaea* L) seed. J. Agric. Food Chem. 29:331-335.

Beach L.R., Spencer D., Randall P.J., Higgins T.J.V. (1985) Transcriptional and post-transcriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. Nucleic Acids Res. 13:999-1013.

Berdanier C. (2000) Proteins. in Advanced Nutrition: Macronutrients, 2nd ed., Boca Raton, FL: CRC Press, 130-196.

Bewley J.D.a.B., M. (1994) Seeds. Physiology of development and germination, 2nd edn. London, Nwe York: Plenum Press.

Blagrove R.J., Gillespie J.M., Randall P.J. (1976) Effect of sulfur supply on the seed globulin composition of *Lupinus angustifolius*. Aust.J. Plant Physiol. 3:173-184.

Burks W., Sampson H.A., Bannon G. (1998) Peanut allergens. Allergy 53:725-730.

Cazzato E., Laudadio V., Stellacci A.M., Ceci E., Tufarelli V. (2012) Influence of sulphur application on protein quality, fatty acid composition and nitrogen fixation of white lupin (*Lupinus albus* L.). Eur. Food Res. Technol. 235:963-969.

Cerqueira N.M.F.S.A., Gonzalez P.J., Brondino C.D., Romao M.J., Romao C.C., Moura I., Moura J.J.G. (2009) The effect of the sixth sulfur ligand in the catalytic mechanism of periplasmic nitrate reductase. J. Comput.Chem. 30:2466-2484.

Chan M.K., Kim J.S., Rees D.C. (1993) The nitrogenase femo-cofactor and p-cluster pair - 2.2-angstrom resolution structures. Science 260:792-794.

Chatel J.M., Bernard H., Orson F.M. (2003) Isolation and characterization of two complete Ara h 2 isoforms cDNA. Int. Arch. Allergy Immunol. 131:14-18.

Cherry J.P. (1977) Potential sources of peanut seed proteins and oil in genus *Arachis*. J. Agric. Food Chem. 25:186-193.

Chiaiese P., Ohkama-Ohtsu N., Molvig L., Godfree R., Dove H., Hocart C., Fujiwara T., Higgins T.J.V., Tabe L.M. (2004) Sulphur and nitrogen nutrition influence the response of chickpea seeds to an added, transgenic sink for organic sulphur. J. Exp. Bot. 55:1889-1901.

Chu Y., Holbrook C.C., Ozias-Akins P. (2009) Two alleles of *ahFAD2B* control the high oleic acid trait in cultivated peanut. Crop Sci. 49:2029-2036.

Chu Y., Faustinelli P., Laura Ramos M., Hajduch M., Stevenson S., Thelen J.J., Maleki S.J., Cheng H., Ozias-Akins P. (2008) Reduction of IgE binding and nonpromotion of *Aspergillus flavus* fungal growth by simultaneously silencing Ara h 2 and Ara h 6 in peanut. J. Agric. Food Chem.56:11225-11233.

Chui C.F.C., Falco, S.C., Rice, J.A. and Knowlton, S. (2003) High sulphur seed protein gene and method for increasing the sulfur amino acid content of plants. Canadian Patent CA 2104022.

Dodo H.W., Konan K.N., Chen F.C., Egnin M., Viquez O.M. (2008) Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen Ara h 2 leads to its significant reduction and a decrease in peanut allergenicity. Plant Biotechnol.J. 6:135-145.

Duque P. (2011) A role for SR proteins in plant stress responses. Plant signal.Behav. 6:49-54.

Dwivedi S.L., Jambunathan R., Nigam S.N., Raghunath K., Shankar K.R., Nagabhushanam G.V.S. (1990) Relationship of seed mass to oil and protein contents in peanut *Arachis hypogaea* L. Peanut Sci.17:48-52.

Eppendorfer W.H., Eggum B.O. (1992) Dietary fiber, sugar, starch and amino acid content of kale, ryegrass and seed of rape and field beans as influenced by S-fertilization and N-fertilization. Plant Foods Hum. Nutr.42:359-371.

Eppendorfer W.H., Eggum B.O. (1995) Sulfur amino acid content and nutritive value of pea and cauliflower crude protein as influenced by sulfur deficiency. Zeitschrift Fur Pflanzenernahrung Und Bodenkunde 158:89-91.

Evans I.M., Gatehouse J.A., Boulter D. (1985) Regulation of storage protein synthesis in pea (*Pisum sativum* L.) cotyledons under conditions of sulfur deficiency. Biochem. J. 232:261-265.

Flinterman A.E., van Hoffen E., Jager C.F.D., Koppelman S., Pasmans S.G., Hoekstra M.O., Bruijnzeel-Koomen C.A., Knulst A.C., Knol E.F. (2007) Children with peanut allergy

recognize predominantly Ara h2 and Ara h6, which remains stable over time. Clin. Exp. Allergy 37:1221-1228.

Gayler K.R., Sykes G.E. (1985) Effects of nutritional stress on the storage proteins of soybeans. Plant Physiol. 78:582-585.

Gevers H.O.a.L., J.K. (1992) Development of modified Opaque-2 maize in South Africa. In E.T. Mertz, ed, Quality protein maize. American Association of Cereal Chemists. St. Paul, MN.:49-78.

Gilbert S.M., Clarkson D.T., Cambridge M., Lambers H., Hawkesford M.J. (1997) SO_4^{2-} deprivation has an early effect on the content of ribulose-1,5-bisphosphate carboxylase/oxygenase and photosynthesis in young leaves of wheat. Plant Physiol. 115:1231-1239.

Gillespie J.M., Blagrove R.J., Randall P.J. (1978) Effect of sulfur supply on seed globulin composition of various species of lupin. Aust.J. Plant Physiol. 5:641-650.

Glover D.V. (1992) Corn protein genetics, breeding, and value in foods and feeds. In E.T. Mertz, ed, Quality protein maize. American Association of Cereal Chemists. St. Paul, MN.:49-78.

Govahi M., Saffari M. (2006) Effect of potassium and sulphur fertilizers on yield, yield components and seed quality of spring canola (*Brassica napus* L.) seed. J. Agron. 5:577-582.

Hagan N.D., Upadhyaya N., Tabe L.M., Higgins T.J.V. (2003) The redistribution of protein sulfur in transgenic rice expressing a gene for a foreign, sulfur-rich protein. Plant J. 34:1-11.

Higgins T.J.V. (1984) Synthesis and regulation of major proteins in seeds. Ann. Rev. Plant Physiol. Plant Mol. Biol. 35:191-221.

Hoefgen R., Nikiforova V.J. (2008) Metabolomics integrated with transcriptomics: assessing systems response to sulfur deficiency stress. Physiol. Plant. 132:190-198.

Hoshi Y., Yamauchi F. (1983) Determination of sulfhydryl and disulfide contents of soybean-11S globulin and their change by lyophilization. Agric. Biol. Chem. 47:2435-2440.

Irving G.W., Fontaine, T. D. and Warner, R. C. (1945) Electrophoretic investigation of peanut proteins. I. Peanut meal extract, arachin and conarachin. Arch. Biochem. and Biophys. 7.

Jamal A., Fazli I.S., Ahmad S., Kim K.-T., Oh D.-G., Abdin M.Z. (2006) Effect of sulfur on nitrate reductase and ATP sulfurylase activities in groundnut (*Arachis hypogea* L.). Journal. Plant Biol. 49:513-517.

Jeor S.T.S., Howard B.V., Prewitt T.E., Bovee V., Bazzarre T., Eckel R.H., Comm A.H.A.N. (2001) Dietary protein and weight reduction - A statement for healthcare professionals from the nutrition committee of the council on nutrition, physical activity, and metabolism of the American Heart Association. Circulation 104:1869-1874.

Johns C.O.a.J., D. B. (1916) The proteins of the peanut, *Arachis hypogaea*. I. the globulins arachin and conarachin. J. Biol. Chem. 28:77-87.

Kang I.-H., Gallo M., Tillman B.L. (2007a) Distribution of allergen composition in peanut (*Arachis hypogaea* L.) and wild progenitor (*Arachis*) species. Crop Sci. 47:997-1003.

Kang I.-H., Srivastava P., Ozias-Akins P., Gallo M. (2007b) Temporal and spatial expression of the major allergens in developing and germinating peanut seed. Plant Physiol. 144:836-845.

Karchi H., Shaul O., Galili G. (1993) Seed specific expression of a bacterial desensitized aspartate kinase increases the production of seed threonine and methionine in transgenic tobacco. Plant J.3:721-727.

Karmoker J.L., Clarkson D.T., Saker L.R., Rooney J.M., Purves J.V. (1991) Sulfate deprivation depresses the transport of nitrogen to the xylem and the hydraulic conductivity of barley (*Hordeu vulgare* L.) roots. Planta 185:269-278.

Khurana N., Gupta J., Chatterjee C. (1999) Sulphur deficiency influences metabolism and grain quality of rice. Ind. J. Plant Physiol. 4:277-281.

Kim H., Hirai M.Y., Hayashi H., Chino M., Naito S., Fujiwara T. (1999) Role of O-acetyl-L-serine in the coordinated regulation of the expression of a soybean seed storage protein gene by sulfur and nitrogen nutrition. Planta 209:282-289.

Kirihara J.A., Hibberd, K.A. and Janice, A. (2001) Method for altering the nutritional content of plant seed. US patent 6,326,527.

Koppelman S.J., Vlooswijk R.A.A., Knippels L.M.J., Hessing M., Knol E.F., van Reijsen F.C., Bruijnzeel-Koomen C. (2001) Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties Runner, Spanish, Virginia, and Valencia, bred in different parts of the world. Allergy 56:132-137.

Koppelman S.J., Wensing M., Ertmann M., Knulst A.C. and Knol E.F. (2004) Relevance of Ara h 1, Ara h 2 and Ara h 3 in peanut allergic patients, as determined by immunoglobin E western blotting, basophil-histamine release and intracutaneous testing: Ara h 2 is the most important peanut allergen. Clin. Exp.Allergy 34:583-590

Koppelman S.J., de Jong G.A.H., Laaper-Ertmann M., Peeters K., Knulst A.C., Hefle S.L., Knol E.F. (2005) Purification and immunoglobulin E-binding properties of peanut allergen Ara h 6: evidence for cross-reactivity with Ara h 2. Clin. Exp.Allergy 35:490-497.

Kortt A.A., Caldwell J.B., Lilley G.G., Higgins T.J.V. (1991) Amino acid and cDNA sequences of a methionine-rich 2Sprotein from sunflower seed (*Helianthus annuus* L). Eur. J. Biochem.195:329-334.

Kotakis C., Vrettos N., Kotsis D., Tsagris M., Kotzabasis K., Kalantidis K. (2010) Light intensity affects RNA silencing of a transgene in *Nicotiana benthamiana* plants. BMC Plant Biol. 10:220.

Kottapalli K.R., Payton P., Rakwal R., Agrawal G.K., Shibato J., Burow M., Puppala N. (2008) Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. Plant Sci. 175:321-329.

Krause S., Latendorf T., Schmidt H., Darcan-Nicolaisen Y., Reese G., Petersen A., Janssen O., Becker W.-M. (2010) Peanut varieties with reduced Ara h 1 content indicating no reduced allergenicity. Mol. Nutr. Food Res. 54:381-387.

Kulis M., Chen X., Lew J., Wang Q., Patel O.P., Zhuang Y., Murray K.S., Duncan M.W., Porterfield H.S., Burks A.W., Dreskin S.C. (2012) The 2S albumin allergens of *Arachis hypogaea*, Ara h 2 and Ara h 6, are the major elicitors of anaphylaxis and can effectively desensitize peanut-allergic mice. Clin. Exp.Allergy 42:326-336.

Losak T., Hlusek J., Martinec J., Vollmann J., Peterka J., Filipcik R., Varga L., Ducsay L., Martensson A. (2011) Effect of combined nitrogen and sulphur fertilization on yield and

qualitative parameters of *Camelina sativa* L. Crtz. (false flax). Acta Agriculturae Scandinavica Section B-Soil and Plant Science 61:313-321.

Lunde C., Zygadlo A., Simonsen H.T., Nielsen P.L., Blennow A., Haldrup A. (2008) Sulfur starvation in rice: the effect on photosynthesis, carbohydrate metabolism, and oxidative stress protective pathways. Physiol. Plant. 134:508-521.

Macritchie F., Gupta R.B. (1993) Functionality-composition relationships of wheat flour as a result of variation in sulfur availability. Aust. J. Agric. Res. 44:1767-1774.

Mahaboobbasha S.M., Cherry J.P. (1976) Composition, solubility, and gel electrophoretic properties of proteins isolated from florunner (*Arachis hypogaea* L.) peanut seeds. J. Agric. Food Chem.24:359-365.

Malhi S.S., Gan Y., Raney J.P. (2007) Yield, seed quality, and sulfur uptake of *Brassica* oilseed crops in response to sulfur fertilization. Agron. J.99:570-577.

Mandal S., Mandal R.K. (2000) Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. Current Sci.79:576-589.

Mason M.E., Johnson B., Hamming M.C. (1967) Volatile components of roasted peanuts. Major monocarbonyls and some noncarbonyl components. J. Agric. Food Chem.15:66

McGrath S.P., Zhao, F.J. and withers, P.J.A. (1996) Development of sulphur deficiency in crops and its treatment. Proceedings of Fertilizer Society No.379. Peterborough, The Fertilizer Society.

Meng L., Ziv M., Lemaux P.G. (2006) Nature of stress and transgene locus influences transgene expression stability in barley. Plant Mol. Biol. 62:15-28.

Mertz E.T., Nelson O.E., Bates L.S. (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145:279-&.

Molvig L., Tabe L.M., Eggum B.O., Moore A.E., Craig S., Spencer D., Higgins T.J.V. (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L) expressing a sunflower seed albumin gene. Proc. Nat.Acad. Sci. USA.94:8393-8398.

Muntz K., Belozersky M.A., Dunaevsky Y.E., Schlereth A., Tiedemann J. (2001) Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. J. Exp. Bot. 52:1741-1752.

Newell J.A., Mason M.E., Matlock R.S. (1967) Precursors of typical and atypical roasted peanut flavor. J. Agric. Food Chem. 15:767

Nielsen N.C. (1984) The chemistry of legume storage proteins. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 304:287

Nielsen N.C., Dickinson C.D., Cho T.J., Thanh V.H., Scallon B.J., Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B. (1989) Characterization of the glycinin gene family in soybean. Plant Cell 1:313-328.

Nikiforova V.J., Gakiere B., Kempa S., Adamik M., Willmitzer L., Hesse H., Hoefgen R. (2004) Towards dissecting nutrient metabolism in plants: a systems biology case study on sulphur metabolism. J. Exp. Bot. 55:1861-1870.

Nikiforova V.J., Bielecka M., Gakiere B., Krueger S., Rinder J., Kempa S., Morcuende R., Scheible W.R., Hesse H., Hoefgen R. (2006) Effect of sulfur availability on the integrity of amino acid biosynthesis in plants. Amino Acids 30:173-183.

Nordlee J.A., Taylor S.L., Townsend J.A., Thomas L.A., Bush R.K. (1996) Identification of a Brazil-nut allergen in transgenic soybeans. N. Engl. J. Med.334:688-692.

Osborne T.B. (1909) The vegetable proteins, Longmans Green, London.

Pancholy S.K., Despande, A.S. and Krall, S. (1978) Amino acids, oil and protein content of some selected peanut cultivars. Proceedings of American Peanut Research and Education Association 10:30-37.

Poms R.E., Capelletti C., Anklam E. (2004) Effect of roasting history and buffer composition on peanut protein extraction efficiency. Mol. Nut. Food Res. 48:459-464.

Porterfield H.S., Murray K.S., Schlichting D.G., Chen X., Hansen K.C., Duncan M.W., Dreskin S.C. (2009) Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants. Clin. Exp. Allergy 39:1099-1108.

Rabe E. (1990) Stress physiology - the functional significance of the accumulation of nitrogen-containing compounds. J. Hort. Sci. 65:231-243.

Rabjohn P., Helm E.M., Stanley J.S., West C.M., Sampson H.A., Burks A.W., Bannon G.A. (1999) Molecular cloning and epitope analysis of the peanut allergen Ara h 3. J. Clin. Invest.103:535-542.

Rai V.K. (2002) Role of amino acids in plant responses to stresses. Biol.Plant. 45:481-487.

Ramos M.L., Fleming G., Chu Y., Akiyama Y., Gallo M., Ozias-Akins P. (2006) Chromosomal and phylogenetic context for conglutin genes in *Arachis* based on genomic sequence. Mol.Genet. Genomics 275:578-592.

Randall P.J., Thomson J.A., Schroeder H.E. (1979) Cotyledonary storage proteins in *Pisum sativum* .4. effects of sulfur, phosphorus, potassium and magnesium deficiencies. Aust. J. Plant Physiol.6:11-24.

Ruiz J.M., Rivero R.M., Romero L. (2005) Regulation of nitrogen assimilation by sulfur in bean. J. Plant Nutr. 28:1163-1174.

Sampson H.A. (2004) Update on food allergy. J. Allergy and Clin. Immunol, 113:805-819.

SAS Institute Inc. (2009). Administering SAS® Enterprise Guide® 4.2. Cary, NC:SAS Institute Inc.

Scherer H.W. (2001) Sulphur in crop production. Eur. J. Agron.14:81-111.

Scherer H.W., Pacyna S., Spoth K.R., Schulz M. (2008) Low levels of ferredoxin, ATP and leghemoglobin contribute to limited N(2) fixation of peas (*Pisum sativum* L.) and alfalfa (*Medicago sativa* L.) under S deficiency conditions. Biol. Fertil. Soils 44:909-916.

Scurlock A.M., Burks A.W. (2004) Peanut allergenicity. Ann. Allergy Asthma Immunol. 93:S12-S18.

Sexton P.J., Paek N.C., Shibles R.M. (1998a) Effects of nitrogen source and timing of sulfur deficiency on seed yield and expression of 11S and 7S seed storage proteins of soybean. Field Crops Res. 59:1-8.

Sexton P.J., Naeve S.L., Paek N.C., Shibles R. (1998b) Sulfur availability, cotyledon nitrogen : sulfur ratio, and relative abundance of seed storage proteins of soybean. Crop Sci. 38:983-986.

Shewry P.R., Franklin J., Parmar S., Smith S.J., Miflin B.J. (1983) The effects of sulfur starvation on the amino acid and protein compositions of barley grain. J. Cereal Sci. 1:21-31.

Shinoyama H., Sano T., Saito M., Ezura H., Aida R., Nomura Y., Kamada H. (2012) Induction of male sterility in transgenic chrysanthemums (*Chrysanthemum morifolium* Ramat.) by expression of a mutated ethylene receptor gene, *Cm-ETR1/H69A*, and the stability of this sterility at varying growth temperatures. Mol. Breeding 29:285-295.

Stanley J.S., King N., Burks A.W., Huang S.K., Sampson H., Cockrell G., Helm R.M., West C.M., Bannon G.A. (1997) Identification and mutational analysis of the

immunodominant IgE binding epitopes of the major peanut allergen Ara h 2. Arch. Biochem. Biophys. 342:244-253.

Staswick P.E., Hermodson M.A., Nielsen N.C. (1984) Identification of the cystines which link the acidic and basic components of the glycinin subunits. J. Biol.Chem. 259:3431-3435.

Stevenson S.E., Chu Y., Ozias-Akins P., Thelen J.J. (2009) Validation of gel-free, label-free quantitative proteomics approaches: Applications for seed allergen profiling. J. Proteomics 72:555-566.

Sun S.S.M., Liu Q.Q. (2004) Transgenic approaches to improve the nutritional quality of plant proteins. In Vitro Cell. Dev.Biol.-Plant 40:155-162.

Tabe L., Higgins T.J.V. (1998) Engineering plant protein composition for improved nutrition. Trends Plant Sci. 3:282-286

Tabe L., Hagan N., Higgins T.J.V. (2002) Plasticity of seed protein composition in response to nitrogen and sulfur availability. Curr.Opin. Plant Biol. 5:212-217.

Tabe L.M., Droux M. (2002) Limits to sulfur accumulation in transgenic lupin seeds expressing a foreign sulfur-rich protein. Plant Physiol.128:1137-1148.

Townsend J.A., Thomas L.A. (1994) Factors which influence the *Agrobacterium*-mediated transformation of soybean. J. Cell. Biochem. 56(S18A):78-78.

Ufaz S., Galili G. (2008) Improving the content of essential amino acids in crop plants: Goals and opportunities. Plant Physiol. 147:954-961.

Varin S., Cliquet J.B., Personeni E., Avice J.C., Lemauviel-Lavenant S. (2010) How does sulphur availability modify N acquisition of white clover (*Trifolium repens* L.)? J. Exp.Bot. 61:225-234.

Venkatachalam M., Sathe S.K. (2006) Chemical composition of selected edible nut seeds. J. Agric. Food Chem. 54:4705-4714.

Wen H.-W., Borejsza-Wysocki W., DeCory T.R., Durst R.A. (2007) Peanut allergy, peanut allergens, and methods for the detection of peanut contamination in food products. Comp. Rev. Food Sci. Food Safety 6:47-58.

Wood R.A. (2003) The natural history of food allergy. Pediatrics 111:1631-1637.

Wrigley C.W., Ducros D.L., Fullington J.G., Kasarda D.D. (1984) Changes in polypeptide composition and grain quality due to sulfur deficiency in wheat. J. Cereal Sci. 2:15-24.

Wu G.Y., Bazer F.W., Davis T.A., Kim S.W., Li P., Rhoads J.M., Satterfield M.C., Smith S.B., Spencer T.E., Yin Y.L. (2009) Arginine metabolism and nutrition in growth, health and disease. Amino Acids 37:153-168.

Yin X.H., Gwathmey O., Main C., Johnson A. (2011) Effects of sulfur application rates and foliar Zinc fertilization on cotton lint yields and quality. Agron.J.103:1794-1803.

Yocum M.W., Khan D.A. (1994) Assessment of patients who have experienced anaphylaxis - A 3-year survey. Mayo Clinic Proceedings 69:16-23.

Zarins Z.M., Marshall W.E. (1990) Thermal denaturation of soy glycinin in the presence of 2-mercaptoethanol studied by differential scanning calorimetry. Cereal Chem. 67:35-38.

Zhao F.J., Hawkesford M.J., McGrath S.P. (1999) Sulphur assimilation and effects on yield and quality of wheat. J. Cereal Sci. 30:1-17.

Table 4.1 Cysteine and methionine residues in major peanut allergens

Allergen	GenBank ID	Protein ID	Number of amino acids	Number of Cysteine residues	Number of Methionine residues	Cysteine (%)	Methionine (%)
Ara h 1	AF432231	AAL27476	626	7	8	1.12	1.28
Ara h 3	HM640250	ADQ53859	512	6	4	1.17	0.78
Ara h 2.01	EF609641.1	ABQ96212	160	8	3	5.00	1.88
Ara h 2.02	AY158467	AAN77576	172	8	4	4.65	2.33
Ara h 6	EF609643	ABQ96214	145	10	9	6.90	6.21

Table 4.2 Composition of nutrient solution

Chemical	Molarity (mM)
KNO_3	1
$Ca(NO_3)$	4
NaH_2PO_4	1
$MnCl_2$	0.025
$ZnSO_4\ 7H_2O$	0.006
$CuSO_4\ 5H_2O$	0.006
H_3BO_3	0.1
Na_2MoO_4	0.001
$CoCl_2\ 6H_2O$	0.1 µM
FeEDTA	0.02
	Low S-0.012
	Medium S- 0.3
$MgSO_4\ 7H_2O$	High S-3.0
	Low S-2.97
	Medium S-2.7
$MgCl_2\ 6H_2O$	High S-0

Table 4.3 100 seed weight

			100 seed weight (g) ‡	Observed *P*
Sulfur (S)	0.012 mM		23.66 ± 3.45[b]	0.0001
	0.3 mM		34.27 ± 8.3[a]	
	3.0 mM		37.76 ± 4.6[a]	
Genotype	Tg		31.84 ± 4.19[a]	0.94
	NTg		31.97 ± 6.7[a]	
S X Genotype	0.012 mM†	Tg	20.40 ± 2.15[a]	0.12
		NTg	26.97 ± 4.76[a]	
	0.3 mM*	Tg	35.16 ± 6.63[a]	
		NTg	33.39 ± 10.0[a]	
	3.0 mM*	Tg	39.97 ± 3.79[a]	
		NTg	35.55 ± 5.44[a]	

‡Seed weight represents the average from 5 replications ± standard deviation. 100 seed weight values with same letters within a row (S, genotype, S X genotype) are not significantly different at $P \leq 0.05$ by Tukey's *t*-test.

† Hundred seed weight under low S (0.012 mM) calculated from weight of 5-6 seeds per replication and averaged.

* Hundred seed weight under medium (0.3 mm) and high (3.0 mM) S calculated from 33-42 seeds per replication and averaged.

Table 4.4 Extractable seed protein

			Total protein (%)†	Observed *P*
Sulfur (S)	0.012 mM		4.35 ± 0.47[b]	0.012
	0.3 mM		4.66 ± 0.83[ab]	
	3.0 mM		5.47 ± 0.955[a]	
Genotype	Tg		5.27 ± 0.81[a]	0.0059
	NTg		4.39 ± 0.69[b]	
S X Genotype	0.012 mM	Tg	5.00 ± 0.65 [a]	0.33
		NTg	3.68 ± 0.29[a]	
	0.3 mM	Tg	4.80 ± 0.76[a]	
		NTg	4.50 ± 0.91[a]	
	3.0 mM	Tg	5.98 ± 1.03[a]	
		NTg	4.96 ± 0.88[a]	

†Non-transgenic and transgenic seed protein was extracted from six pooled seeds of low S and fifteen pooled seeds of medium and high S from each of the five replications and quantified by BCA assay.

Total protein values represent average ± standard deviation from two technical replications of the five biological replicates.

Values with same letters within a row (S, genotype, S X genotype) are not significantly different at $P \leq 0.05$ by Tukey's *t*-test.

Table 4.5 Relative concentrations of allergens in extractable seed protein‡

			Ara h 1†	Ara h 3†	Ara h 2*	Ara h 6*
Sulfur (S)	0.012 mM		1161.2 ± 313.35[a]	92.0 ± 70.04[b]	335.5 ± 63.28[b]	101.1 ± 60.41[b]
	0.3 mM		1082.6 ± 281.12[a]	789.1 ± 276.06[a]	1995.5 ± 187.70[a]	2382.4 ± 674.93[a]
	3.0 mM		1413.6 ± 332.33[a]	913.8 ± 183.75[a]	1962.1 ± 206.13[a]	2203.4 ± 392.65[a]
Genotype	Tg		1250.8 ± 310.44[a]	675.48 ± 192.02[a]	0[b]	158.3 ± 195.98[b]
	NTg		1187.5 ± 307.43 [a]	521 08 ± 161.22[a]	2862.1 ± 304.74[a]	2966.3 ± 556.01[a]
S X Genotype	0.012 mM	Tg	1166.51 ± 378.10[a]	107.95 ± 119.22[a]	0.0[b]	0[b]
		NTg	1155.94 ± 248.60[a]	76.00 ± 20.87[a]	671.07 ± 126.56[b]	202.28 ± 120.83[b]
	0.3 mM	Tg	1207.32 ± 360.57[a]	944.78 ± 324.89[a]	0.0[b]	228.46 ± 263.95[b]
		NTg	957.89 ± 201.68[a]	633.33 ± 227.24[a]	3991.08 ± 375.39[a]	4536.33 ± 1085.91[a]
	3.0 mM	Tg	1378.63 ± 192.65[a]	973.70 ± 131.95[a]	0.0[b]	246.45 ± 324.00[b]
		NTg	1448.57 ± 472.02[a]	853.88 ± 235.56[a]	3924.10 ± 412.27[a]	4160.36 ± 461.31[a]

‡Allergens quantified by quantitative western blotting with 'Georgia Green' protein as standard. Band volumes calculated for allergens in samples were normalized with protein concentrations from the standard 'Georgia Green'. Allergen values are average ± standard deviation from two technical replications of four biological replicates.

Values of each allergen with same letter within a row (S, genotype, S X genotype) are not significantly different at $P \leq 0.05$ by Tukey's *t*-test.

† Loaded 5 µl of protein at 1000 ng/µl, * Loaded 5 µl of protein at 4000 ng/µl.

Tg - transgenic, NTg - Non-transgenic.

Table 4.6 Carbon, nitrogen and sulfur (CNS) in seeds‡.

			Carbon (%)	Nitrogen (%)	Sulfur (%)	N:S
Sulfur (S)	0.012 mM		49.89[b]	8.61[a]	0.175[b]	49.67[a]
	0.3 mM		53.52[a]	7.13[b]	0.26[a]	28.07[b]
	3.0 mM		56.54[a]	6.92[b]	0.27[a]	25.67[b]
Genotype	Tg		52.80[a]	6.80[b]	0.20[b]	35.90[a]
	NTg		53.84[a]	8.31[a]	0.27[a]	33.04[a]
S X Genotype	0.012 mM	Tg	49.83[a]	8.11[a]	0.16[a]	51.30[a]
		NTg	49.96[a]	9.12[a]	0.19[a]	48.04[a]
	0.3 mM	Tg	51.62[a]	6.13[a]	0.20[a]	30.79[a]
		NTg	55.42[a]	8.13[a]	0.32[a]	25.35[a]
	3.0 mM	Tg	56.95[a]	6.16[a]	0.24[a]	25.61[a]
		NTg	56.14[a]	7.69[a]	0.30[a]	25.74[a]

‡CNS values represent average of four biological replicates analyzed using 0.2 g defatted and lyophilized seed meal in a Leco CNS analyzer.

Values of carbon, nitrogen, sulfur, N:S with same letter within a row are not significantly different at $P \leq 0.05$ by Tukey's *t*-test.

Table 4.7 Total and free amino acids per unit mass of defatted seed meal‡

Sulfur	Genotype	TAA (nmoles/mg)	FAA (nmoles/mg)
0.012 mM	Transgenic	1644.5	269.2
	Non-transgenic	1833.7	237.16
0.3 mM	Transgenic	836.3	23.43
	Non-transgenic	1275.9	25.7
3.0 mM	Transgenic	1006.2	22.46
	Non-transgenic	1552.1	25.55

‡Total amino acids (TAA) and free amino acids (FAA) per unit mass of defatted seed mass represent single measurement of 4 pooled samples, analyzed with L-8800 Hitachi analyzer.

Table 4.8 Percentage of free amino acids in seeds under different soil sulfur levels‡

Amino acids	L[a]-Tg[d]	L-NTg[e]	M[b]-Tg	M-NTg	H[c]-Tg	H-NTg
Aspartic acid	0.2	0.2	9.2	12.7	8.7	11.8
Threonine	0.4	0.3	1.3	1.2	1.4	1.1
Serine	0.5	0.5	1.5	1.4	1.4	1.4
Glutamic acid	5.5	6.4	38.8	36.6	35.5	30.4
Glycine	0.8	0.7	7.4	6.7	6.4	6.0
Alanine	1.6	1.6	7.1	6.7	6.1	6.9
Valine	0.3	0.3	2.1	2.1	2.1	2.4
Isoleucine	0.1	0.1	0.7	0.7	0.7	0.7
Leucine	0.1	0.1	0.9	0.9	0.9	1.1
Tyrosine	0.1	0.1	0.7	0.6	0.7	0.9
Phenylalanine	0.3	0.5	4.6	3.5	6.7	3.2
Lysine	2.7	2.4	1.8	1.6	1.8	1.8
Histidine	1.6	1.2	0.0	0.0	0.0	0.0
Arginine	68.4	68.2	6.0	4.7	9.1	8.7
Proline	0.8	0.6	3.5	4.0	2.9	4.2
Asparagine	14.5	13.7	6.4	10.0	6.3	14.6
Ornithine	0.7	0.8				
Hydroxyproline	0.2	0.2				
Citrulline	0.4	0.6	8.0	6.6	9.2	4.7
Glutamine	0.9	1.5				

‡Percentage of each amino acid calculated based on free amino acids (nmoles) in 50 µl final volume analyzed in L-8800 Hitachi analyzer, in single analysis of 4 pooled samples.

[a] L-Low S (0.012 mM)

[b] M-Medium S (0.30 mM),

[c] H-High S (3.0 mM),

[d] Tg-transgenic plant,

[e] NTg-Non-transgenic plant

Table 4.9 Percentage of total amino acids in seeds under different soil sulfur levels‡

Amino acids	L[a]-Tg[d]	L-NTg[e]	M[b]-Tg	M-NTg	H[c]-Tg	H-NTg
Asparagine/aspartic acid	13.5	13.1	10.3	11.8	10.7	12.0
Threonine	2.5	2.5	3.1	2.6	3.0	2.7
Serine	4.9	5.2	6.0	6.0	5.8	6.1
Glutamine/glutamic acid	14.4	14.9	15.2	18.7	15.9	18.2
Glycine	12.3	12.8	20.0	14.0	18.7	12.9
Alanine	5.0	5.0	5.4	4.9	5.6	5.1
Valine	3.6	3.6	4.1	3.7	4.2	4.1
Isoleucine	2.8	2.8	2.8	2.8	2.9	3.0
Leucine	5.0	5.2	5.2	6.1	5.4	6.3
Tyrosine	2.8	3.0	2.9	2.7	2.8	2.9
Phenylalanine	3.3	3.5	3.7	3.3	3.8	3.6
Lysine	3.8	3.6	4.9	3.8	4.3	3.7
Histidine	2.2	2.1	2.2	2.0	2.2	2.0
Arginine	18.9	17.4	7.7	9.5	7.9	9.5
Proline	4.4	4.5	4.2	4.5	4.5	4.6
Cysteic acid	0.7	0.7	1.7	2.6	1.6	2.3
Methionine sulfone	0.0	0.0	0.6	1.1	0.6	1.1

‡Percentage of each amino acid calculated based on amino acids (nmoles) in 50 µl final volume analyzed in L-8800 Hitachi analyzer, in single analysis of 4 pooled samples.

[a] L-Low S (0.012 mM)

[b] M-Medium S (0.30 mM),

[c] H-High S (3.0 mM),

[d] Tg-transgenic plant,

[e] NTg-Non-transgenic plant

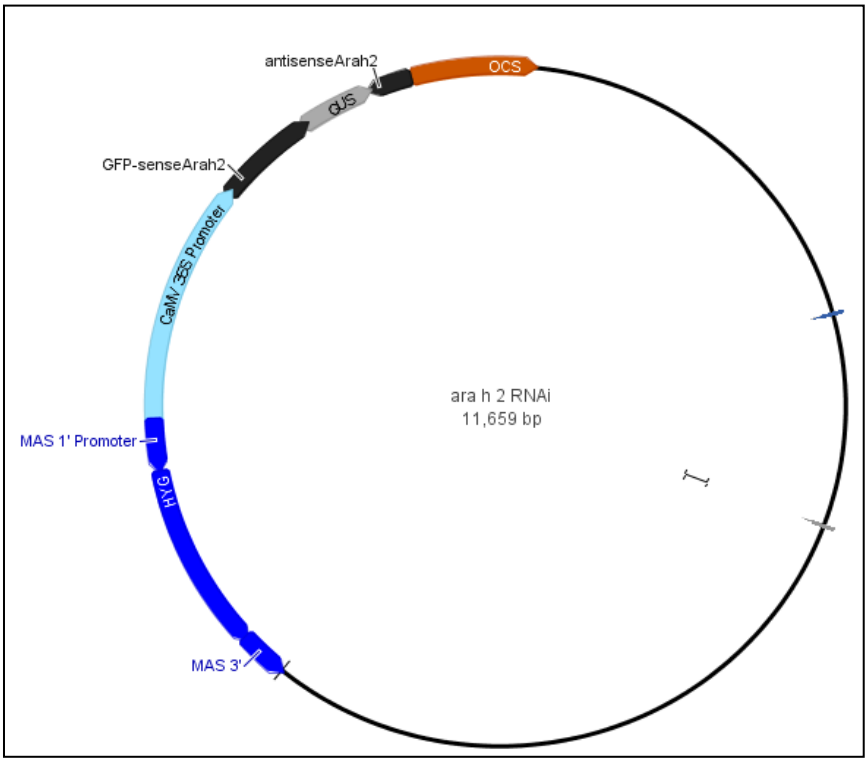

Figure 4.1 Ara h 2 RNAi construct (Chu *et al*., 2008)

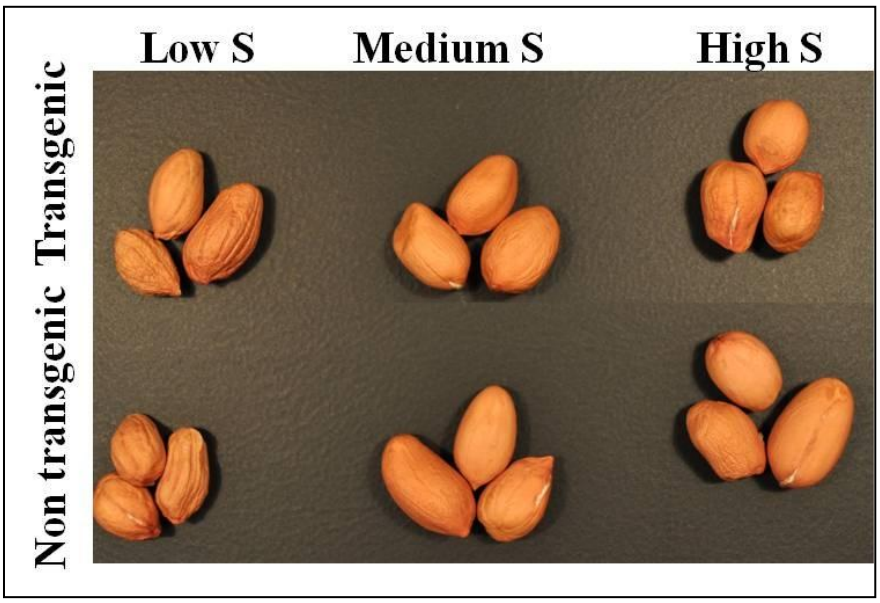

Figure 4.2 Seed samples used for soluble protein extraction and quantification of relative amounts of peanut allergens.

For low S two seeds were collected per plant, seeds from 3 plants per replication pooled, to get 6 seeds per replication of transgenic/non-transgenic seeds. For medium and high S five seeds were collected per plant, seeds from 3 plants per replication pooled, to get 15 seeds per replication of transgenic/non-transgenic seeds.

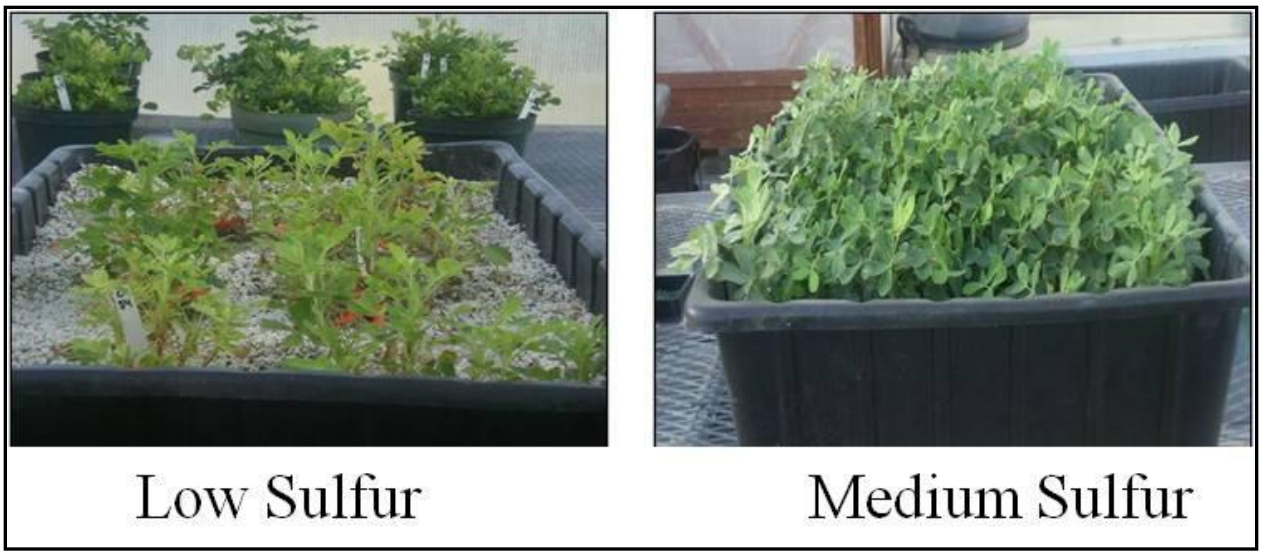

Figure 4.3 Plant vegetative growth three months after planting under two different soil nutrition levels.

Figure 4.4 Plants at harvest under various sulfur nutrient levels

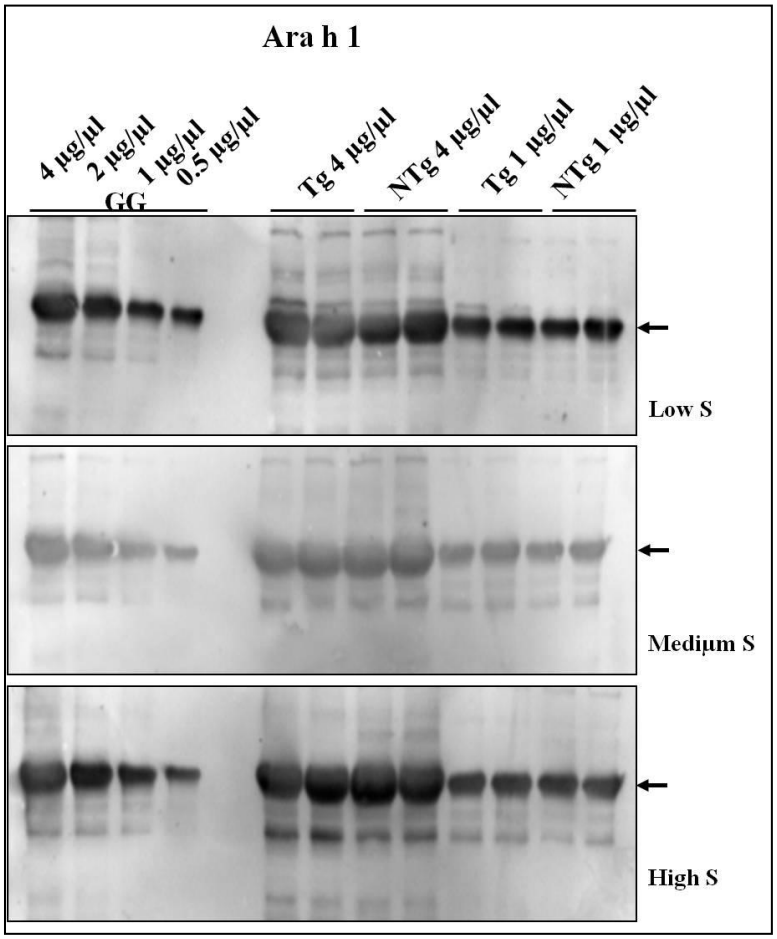

Figure 4.5 Western blotting of Ara h 1

Quantitative western of Ara h 1 done in each of two transgenic (Tg) and non-transgenic (NTg) replicates along with 'Georgia Green' (GG) standards are shown. Total volume of each sample loaded was 5 µl. Arrows represent bands quantified. Band volume calculated for each Tg or NTg band was normalized with protein concentration in'Georgia Green' standards.

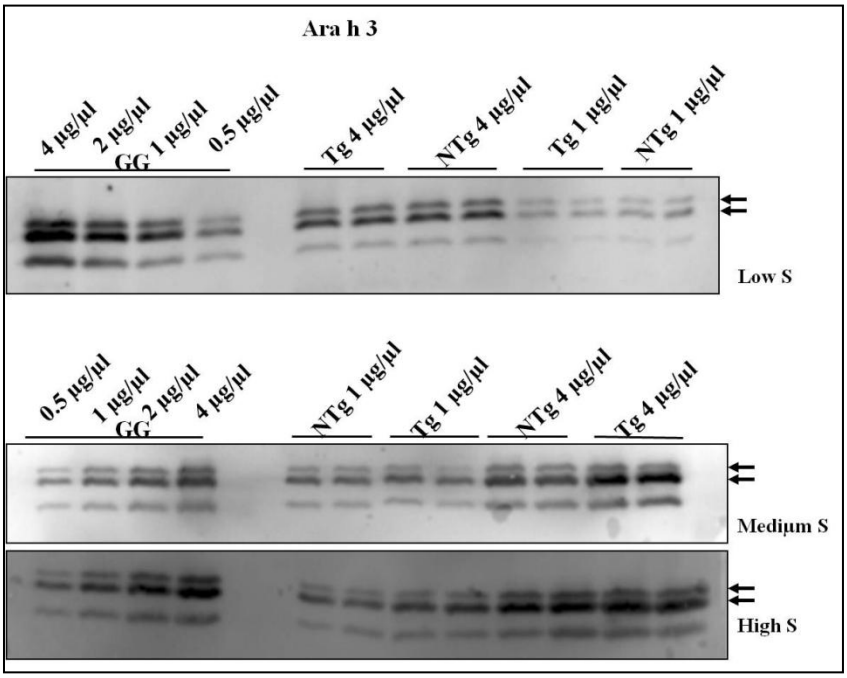

Figure 4.6 Western blotting of Ara h 3

Quantitative western of Ara h 3 done in each of two transgenic (Tg) and non-transgenic (NTg) replicates along with 'Georgia Green' (GG) standards are shown. Total volume of each sample loaded was 5 µl. Arrows represent the two Ara h 3 bands quantified. Band volume calculated for each Tg or NTg band was normalized with protein concentration in'Georgia Green' standards.

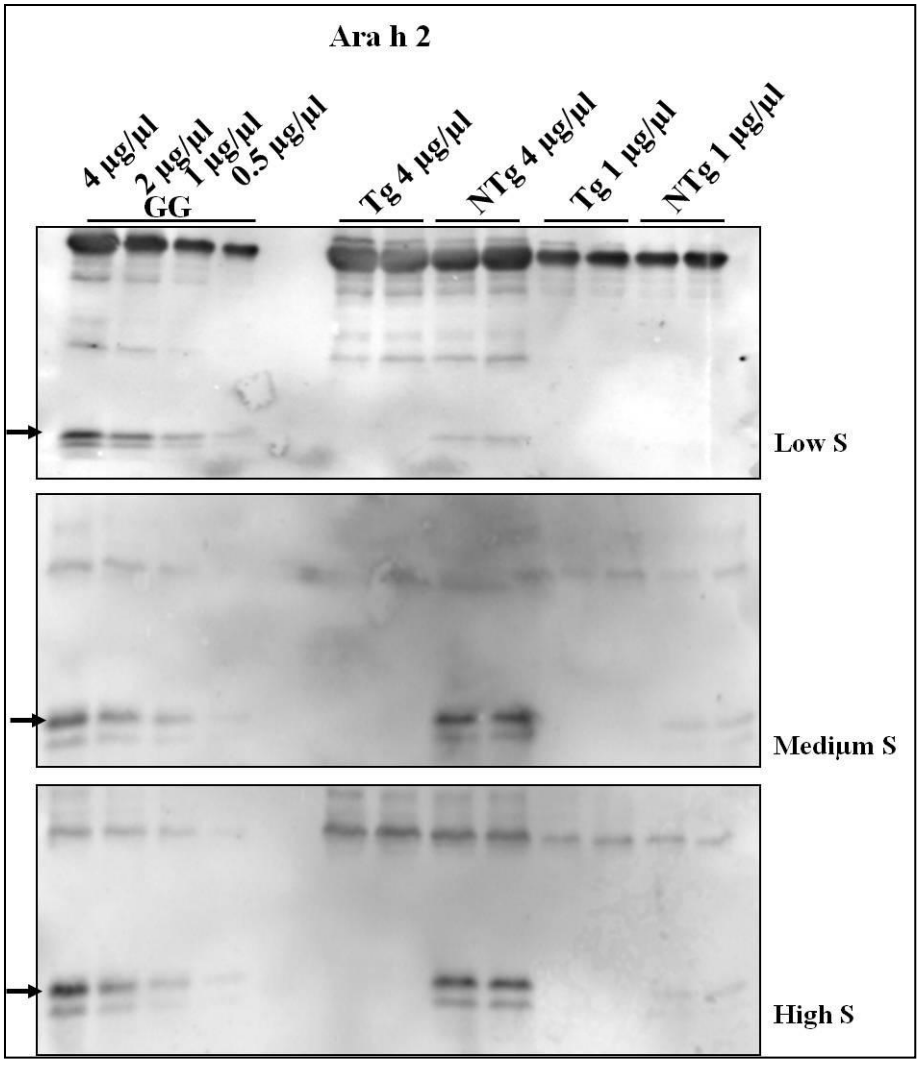

Figure 4.7 Western blotting of Ara h 2

Quantitative western of Ara h 2 done in each of two transgenic (Tg) and non-transgenic (NTg) replicates along with 'Georgia Green' (GG) standards are shown. Total volume of each sample loaded was 5 µl. Arrow represents the Ara h 2 band quantified. Bands above Ara h 2 correspond to Ara h 1. Band volume calculated for each Tg or NTg band was normalized with protein concentration in 'Georgia Green' standards.

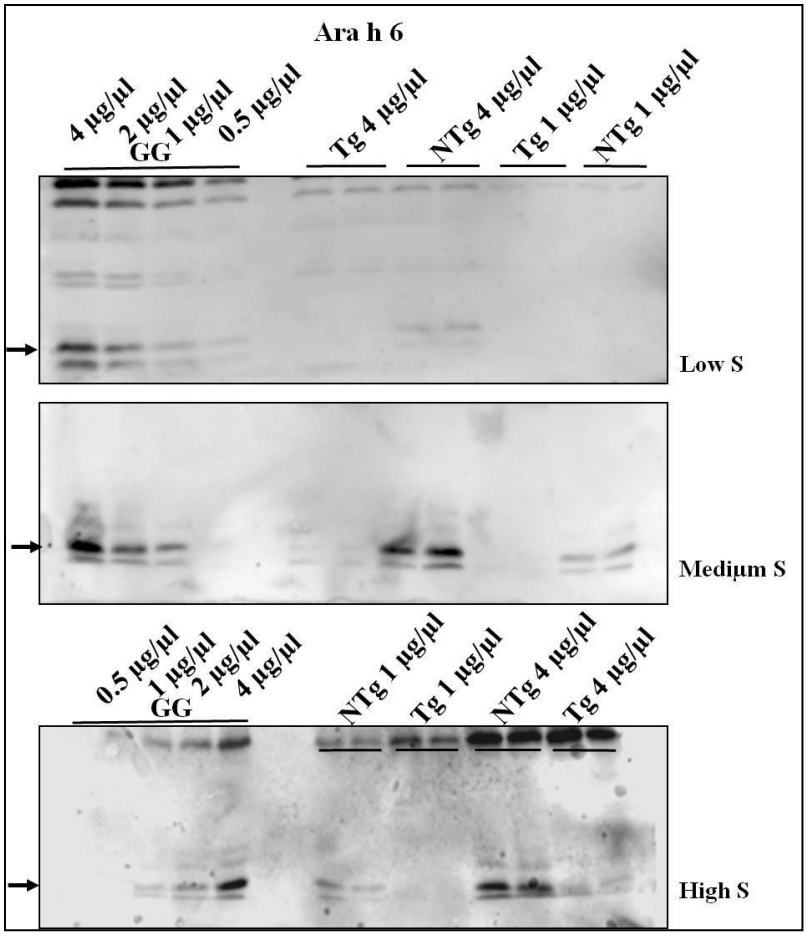

Figure 4.8 Western blotting of Ara h 6

Quantitative western of Ara h 6 done in each of two transgenic (Tg) and non-transgenic (NTg) replicates along with 'Georgia Green' (GG) standards are shown. Total volume of each sample loaded was 5 µl. Arrow represents the Ara h 6 band quantified. Bands above Ara h 6 correspond to Ara h 3. Band volume calculated for each Tg or NTg band was normalized with protein concentration in 'Georgia Green' standards.

CHAPTER 5

ACTIVATION OF MINIATURE INVERTED REPEAT TRANSPOSABLE ELEMENTs UNDER STRESS AND POTENTIAL FOR UTILIZATION IN GENE TAGGING IN PEANUT *(Arachis hypogaea L.)*

Abstract

Miniature Inverted-repeat Transposable Elements (MITEs) are DNA transposons found in plants and animals. They typically are found near coding regions in the genome and can be activated under a variety of stresses to which the organisms are subjected. The activation of a particular transposable element is dependent on the genotype, the duration and type of stress. This study showed that the peanut MITE, *AhMITE1,* is activated by different stresses such as ethylmethane sulfonate (EMS) and tissue culture. Activation of the peanut MITE was found to vary under chemical mutation and tissue culture with transposition frequencies of 1.24% and 6.25% respectively for cultivar 'Tifrunner'. Genotypic differences may also influence MITE activation as preliminary results indicate. Sequencing the transposed fragments and similarity searches of flanking sequences showed hits to peanut genomic clones. Since MITEs are expected to transpose to gene rich regions of the genome they can be utilized for mutation and gene tagging. This study has shown that the mutations in the EMS and tissue cultured populations are heritable and these populations can be further utilized for gene tagging in mutants identified by transposon display.

Chandran, M., Conner, J.A., and Ozias-Akins, P. To be submitted to The Plant Genome

Introduction

Predicting the coding sequences in a genome and the functional analysis of predicted genes is an ultimate goal of sequencing projects (Parinov and Sundaresan, 2000). Once a gene is annotated, its putative function must be validated by multiple methods. Although the number of genes in rice was estimated to be around 50,000 following initial draft sequencing (Goff *et al.*, 2002; Yu *et al.*, 2002), many of these are transposable elements (TEs) and more recent evidence-based estimates of non-TE genes identified approximately 30,000 (Gojobori, 2007; Jiang and Ramachandran, 2010). With developments in molecular technology, reverse genetics approaches are becoming more efficient for functional analysis of annotated genes. In a reverse genetics approach, mutants with disrupted genes, the sequence of which is known, serve as tools to study the function of a gene (Maes *et al.*, 1999). The key metabolic or developmental processes affected by modifying a gene or its expession are correlated to the gene fuction. The common methods for developing mutants are chemical or ionizing radiation mutagenesis, RNAi and insertional mutagenesis (An *et al.*, 2005; Parinov and Sundaresan, 2000; Ramachandran and Sundaresan, 2001). In RNAi, dsRNA silences genes with sequence identity to the dsRNA employed (Baulcombe, 1996; Fire *et al.*, 1991; Fire *et al.*, 1998). Gene inactivation by RNAi requires the production of a large number of transgenic lines, targets only one gene or closely related genes in a single experiment (Ramachandran and Sundaresan, 2001), and can produce off target effects (Hardy *et al.*, 2010, Fitzgerald *et al.*, 2012). Gene mutations, deletions or insertions by zinc finger nucleases, transcription activator-like effector nucleases (TALENs), and meganucleases target specific genes (gene targeting), but have been accomplished up to now only with model species *viz.*, maize, tobacco, *Arabidopsis* and rice. These nucleases are engineered proteins consisting of a DNA binding domain, which is site-specific, and an

endonuclease domain which causes the double strand break in a chromosome for integration of foreign DNA by homologous recombination (Curtin *et al*., 2012; Tzfira e*t al*., 2012).

Insertional mutagenesis, with *Agrobacterium* T-DNA or transposons, is the most common approach for gene function analysis. T-DNA and transposons insert at random sites in the chromosome and most often create loss of function mutants. The gene in which the insertion has occurred is identified and recovered by PCR-based strategies using the known sequences of the inserted element, which functions as a tag. T-DNA insertions are stable through multiple generations and developing a large population of mutants increases the chance of identifying any gene of interest, as in *Arabidopsis* and rice with small genome sizes (Krysan *et al.*, 2002). T-DNA insertional mutagenesis has been successfully employed in *Arabidopsis* to develop mutant populations (Alonso *et al.*, 2003; Kempin *et al.*, 1997; Krysan *et al.*, 1999). In *A.thaliana*, near saturation of the gene space has been achieved with 225,000 *Agrobacterium* T-DNA insertion events and mutations identified in more than 21,700 of the approximately 29,454 predicted genes (Alonso *et al.*, 2003). Similarly, T-DNA insertional mutant populations created in rice by multiple groups are large enough to identify gene knock-outs (Hsing *et al.*, 2007; Jeong *et al.*, 2002; Jiang and Ramachandran, 2010; Wu *et al.*, 2003). Insertional mutagenesis also has been applied to other crops, *viz*., the temperate grass *Brachypodium* (Vain *et al.*, 2008), *Lotus japonicus* (Martirani *et al.*, 1999), chinese cabbage (Yu *et al.*, 2010), a wild relative of tomato, *Solanum pennellii* (Atares *et al.*, 2011) and strawberry (Ruiz-Rojas *et al.*, 2010). Although widely used, T-DNA insertional mutagenesis requires an efficient *in vitro* transformation and plant regeneration protocol, may produce multiple insertion events, and can result in rearrangements such as deletions, direct or inverted repeats (Ramachandran and Sundaresan, 2001).

As with T-DNA, transposons are another commonly employed tool for insertional mutagenesis. Transposons are mobile genetic elements, discovered initially in maize by Barbara McClintock during genetic analysis of kernel pigmentation (McClintock, 1951). Transposons are grouped as DNA/class II transposons and Retrotransposon/class I transposons. Class I transposons transpose by a copy and paste mechanism through an RNA intermediate. Class II transposons move as DNA elements by a cut and paste mechanism (Bennetzen, 2000; Feschotte *et al.*, 2002). Autonomous transposons encode the transposase enzyme required for transposition. Non-autonomous transposons carry sequences that are recognized by a transposase enzyme but no transposase of their own, being dependent on the autonomous transposons within the same family for the transposase enzyme. Transposons, which account for major proportions of eukaryotic genomes (Cantu *et al.*, 2010; Feschotte *et al.*, 2002; Schnable *et al.*, 2009), can be activated to move from one chromosomal location to another by tissue culture (Hirochika, 1993; Kikuchi *et al.*, 2003), chemical mutagens (Gowda *et al.*, 2010; Patel *et al.*, 2004), ionising radiation (Nakazaki *et al.*, 2003), hybridization shock (Liu *et al.*, 2004; Shan *et al.*, 2005) and by any stress to which an organism is subjected. In transposon tagging, transposons with known sequences serve as tags in a similar manner to T-DNA. Transposons can excise from a mutated region of the chromosome thus restoring the wild-type phenotye and confirming the mutation (Meissner *et al.*, 1999).

The commonly used transposons for gene tagging are *Activator (Ac)/Dissociation (Ds), Mutator (Mu), Suppressor-Mutator/Enhancer (Sp-Mu/En)* in maize, Transposable elements *Antirrhinnum majus (Tam)* in *A. majus*, *dTph* elements from *Petunia hybrida*, *Tos17* (transposon of *Oryza sativa*) and *mPing* (miniature Ping) from rice (Hancock *et al.*, 2011; Ramachandran and Sundaresan, 2001). Genes may be disrupted by endogenous or heterologous transposons (May

and Martienssen, 2003). Endogenous transposon-tagged alleles are at the *DAG* (Chatterjee and Martin, 1997), *centroradialis* (Bradley *et al.*, 1996), *PHANTASTICA* (Waites *et al.*, 1998) loci in *Antirhinnum; Knotted1* (Hake *et al.*, 1989), *silky1* (Ambrose *et al.*, 2000), *Golden2* (Cribb *et al.*, 2001), *Tangled1* (Smith *et al.*, 2001) and *fasciated ear2* (Taguchi-Shiobara *et al.*, 2001) loci in maize. Examples of genes cloned by heterologous tagging experiments are *aintegumenta* (Elliott *et al.*, 1996), *LRX1* (Baumberger *et al.*, 2001) and *sucrose uncoupled-6* (Huijser *et al.*, 2000).

Mobilization of transposons has been observed upon tissue culture (Hirochika, 1993; Hirochika *et al.*, 1996; Huang *et al.*, 2009; Kikuchi *et al.*, 2003; Liu *et al.*, 2004; Shan *et al.*, 2005), physical and chemical mutation (Gowda *et al.*, 2010; Nakazaki *et al.*, 2003; Patel *et al.*, 2004), hybridization shock (Shan *et al.*, 2005), abiotic stress (Kalendar *et al.*, 2000) and pathogen infection (Grandbastien *et al.*, 1997). Transposon activation can be dependent on the genotype also as evidenced from studies in rice (Liu *et al.*, 2004).

Peanut (*Arachis hypogaea* L.) is a recent allotetraploid (AABB) with a genome size of ~2800 Mbp. Peanut genome sequencing has been initiated (http://www.peanutbioscience.com/peanutgenomeproject.html) and once the assembly is complete the next steps would be gene prediction and functional annotation of the predicted genes. Currently expressed sequence tags (ESTs) from peanut and its wild relatives are used for identification of genes that may be associated with a trait. Numerous genes have been identified in peanut by analyzing ESTs and among them are arachin and conglutin protein encoding genes (Yan *et al*., 2005), oxalate oxidase (Chen *et al*., 2008), auxin repressed protein, cytokinin oxidase, and metallothionein type 2 (Guimarães *et al*., 2010), stearoyl-acyl carrier protein desaturase (Shilman *et al.*, 2011) and chalcone isomerase (Zhang *et al*., 2011). Large datasets now have been used to generate peanut EST assemblies

some of which have implemented preliminary annotation (Nagy *et al*., 2012; Zhang *et al*., 2012a).

Generating T-DNA and transposon-tagged lines for gene identification and cloning has not been exploited in peanut on a large scale, although they have been employed in other legumes (Tadege *et al.*, 2009). The first transposon discovered in peanut, *AhMITE1*, is a DNA/Class II transposon, and was identified in a mutant exhibiting the high oleic acid trait. The MITE is 205 bp long with 25 bp terminal inverted repeats (TIRs) and 9 bp target site duplication (TSD). The high oleic acid trait is caused by the insertion of the MITE in the *ahFAD2B* gene causing a truncated protein, along with a point mutation in the duplicate *ahFAD2A* gene (Chu *et al*., 2009). The mutated genes, *ahFAD2A* and *ahFAD2B*, are oleoyl-PC desaturase genes that convert oleic to linoleic acid in non-photosynthetic tissues (Ray *et al.*, 1993). This catalytic conversion is prevented in the double mutants.

Southern blot hybridization experiments in peanut cultivars (Patel *et al*., 2004; Shirasawa *et al*., 2012a) revealed multiple copies of *AhMITE1* in the peanut genome. Genomic and EST database searches have identified sequences similar to *AhMITE1* in peanut. MITEs are generally associated with the gene rich regions of the genome (Sabot *et al.*, 2005; Wessler *et al.*, 1995; Zhang *et al.*, 2000), hence the peanut MITE can be used as a potential tool for gene tagging. A collection of gene-tagged lines is becoming more relevant for peanut as the sequencing project is underway and gene functional analysis has not been done on an extensive scale.

Ethylmethane sulfonate mutants developed from peanut variety VL1 showed wide variation in plant morphology and response to rust and late leaf spot (LLS). The parent variety VL1 is resistant to rust and susceptible to LLS, but some mutants were resistant to LLS and susceptible to rust. Analysis of mutants by PCR with primers specific to *AhMITE1* and flanking sequence of

MITE identified earlier (Bhat *et al.*, 2008) found that the mutants had a MITE insertion at the flanking regions tested. Later generations showed occasional variants too, due to MITE excision (Gowda *et al*., 2010). Activation of *AhMITE1* was also seen in studies with mutants developed by EMS, gamma irradiation and tissue culture. The morphological variation seen in these mutants as in branching habit, disease incidence and changes resembling alternate botanical types, were found to be associated with *AhMITE1* insertion or excision at certain genomic locations (Gowda *et al*., 2011). The objective of the present study was to investigate the activation of *AhMITE1* by stress and its potential utility as a tool for endogenous gene tagging in peanut. Transposition activation was studied in the M_2 generation of an EMS mutated population with 'Tifrunner' background. Peanut has a well-developed protocol for *in vitro* regeneration (Ozias-Akins, 1989) and the differential transposition of *AhMITE1* under stress situations was compared under prolonged tissue culture and after chemical mutagenesis. As transposon activation can vary with the genotype (Liu *et al.*, 2004), activation under prolonged tissue culture was studied with two runner market types. The technique of transposon display (TD) employed in this study is a component of the reverse genetics approach to identify mutants even in duplicated genes, as initial mutant screening is not based on phenotypic characters. The study provides information about the efficiency of different mutagenic techniques and genotypic differences in MITE activation. Also the potential utility of EMS-induced and tissue-cultured populations for transposon tagging of genes was assessed by sequencing the transposed fragments and determining the similarity of MITE flanking regions to sequences in published databases.

Materials and Methods

M_2 generation plants derived from 'Tifrunner' seeds, treated with 0.4% EMS for 12 h, were used to study the effect of this chemical mutagen on MITE transposition (Knoll *et al*., 2011). The M_2 generation of mutants was cultivated in the field on the Tifton Campus, University of Georgia, using standard cultural practices for peanut. Transposon display was conducted using DNA extracted from 401 M_2 plants. M_2 plants were individually harvested.

Transposition of *AhMITE1* under tissue culture was tested with two runner cultivars of peanut, namely, 'Tifrunner' and 'Florunner', separated into its three component lines (F439-16-10-3-1, F439-16-10-3-2 and F439-16-10-3-3). 'Florunner' was selected for study as the peanut MITE, *AhMITE1,* was initially reported in the mutant 'Mycogen-Flavo', which was generated in a 'Florunner' background. This high oleic acid peanut mutant was obtained by EMS mutagenesis of 'Florunner' (U.S. patent 5948954). 'Tifrunner' was included as it is the reference genome for peanut and also has 'Florunner' in its pedigree.

Tissue Culture

Mature peanut seeds were surface sterilized with 70% ethanol for 5 min, followed by 20% Clorox for 20 min with continuous shaking. Following 3 rinses in sterile deionized water, the plumule region of the embryo was excised under a stereomicroscope and cultured on embryo induction medium. Embryo induction medium consisted of MS salts and vitamins, 3 mg/L picloram (Dow-Agrosciences, IN, USA), 3% sucrose (Fisher, NJ, USA), 8 g/L agar, pH adjusted to 5.8 and sterilized for 20 min at 121°C, followed by the addition of 1 g/L of filter sterilized L-glutamine (Acros-Organics ,NJ, USA). The cultures were kept in the dark at 25 +/- 2 ºC for up to 10 subcultures, with sub-culturing done every 4 weeks, before being transferred to regeneration media at 16 hr light and 8 hr dark photoperiod. The clusters were labeled as separate clusters

during the ninth and sixth subcultures for ‘Tifrunner’ and ‘Florunner’ component lines, respectively. A total of 16 clusters of ‘Tifrunner’ were advanced through regeneration. The number of clusters in F439-16-10-3-1, F439-16-10-3-2 and F439-16-10-3-3 were 7, 7 and 5, respectively. Initial regeneration was on basal MS medium with 2% sucrose and 8 g/L agar, followed by basal MS medium with 1 mg/L thidiazuron (TDZ) (Sigma, St. Louis, MO, USA), 2% sucrose, 8 g/L agar, and pH adjusted to 5.8. From the second regeneration medium the cultures were transferred to basal MS medium with 2% sucrose and 8 g/L agar and then to MS with 0.5 mg/L kinetin (Sigma), 0.25 mg/L benzyl amino purine (Sigma) and 8 g/L agar. Regenerated shoots were rooted in MS medium containing 0.1 mg/L NAA (Sigma), 2% sucrose and 8 g/L agar in Magenta vessels. Rooted plants were transplanted to potting mix and maintained in the greenhouse. DNA was extracted from five plants per cluster for analysis by TD.

DNA Extraction

Young unopened leaves were collected from the tissue culture-derived and mutant plants into 96 well collection plates (Qiagen Inc. USA, Valencia, CA) and stored at -80ºC or processed immediately. Leaves were frozen in liquid nitrogen and ground with a blue pestle. DNA was extracted using the DNeasy 96 Plant Kit (Qiagen Inc. USA, Valencia, CA) according to the manufacturer's instructions. Fluorometric quantitation of DNA was done using Hoechst 33258 dye and fluorescence was read with a FluoroCount (Packard/Perkin-Elmer, Waltham, MA) microplate reader.

Transposon Display

Transposon display was conducted according to the protocol developed by Casa *et al.* (2000) which is a modification of Amplified Fragment Length Polymorphism (AFLP) (Vos *et al.*, 1995). Restriction of genomic DNA with *Eco*RI and *Mse*I enzymes followed by ligation of restriction enzyme specific adapters was done at 37^0 C for 4 h in a 50 µl reaction mixture containing genomic DNA at 10 ng/µl or 4 ng/µl, *Eco*RI and *Mse*I at 5 U each (New England Biolabs, Beverly, MA, USA), 50 pmol *Mse*I adapter, and 5 pmol *Eco*RI adapter (Table 5.1) in restriction ligation (RL) buffer (10 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 5 mM DTT, 0.05 µg/µl BSA), 2 U of T4 DNA ligase (Promega, Madison, WI, USA) and 10 mM ATP. Pre-amplification was done with 5 ul of 1:10 dilution of restricted and ligated DNA in a 50 ul reaction volume containing 1X PCR Buffer (New England Biolabs, Beverly, MA), 1.5 ng each of *Mse*I adapter primer and MITE-specific primer (Table 5.1), 1 U Taq DNA polymerase (New England Biolabs) and 10 mM dNTPs. The pre-amplification was carried out with one cycle of 45 s at 94°C, 30 s at 63°C, 1 min at 72°C, followed by 12 cycles of a touch-down profile in which the annealing temperature was decreased at a rate of 1°C/cycle from 62°C, and 20 cycles at a constant annealing temperature of 49°C for 30 s. A final extension step at 72°C for 7 min was included.

DNA selective amplification was conducted with a nested MITE-specific primer plus the *Mse*I primer with one selective nucleotide (A, T, G, or C) at the 3' end (Table 5.1). The pre-amplified DNA was diluted at 1:20 and 5 µl was used in a 20 µl amplification reaction containing 1X Jump Start PCR buffer, 30 ng each of selective and MITE-specific primers, 0.2 mM dNTP, and 0.2 µl of JumpStart Taq DNA polymerase (Sigma). The touch-down PCR cycle in selective

amplification differed from pre-amplification for annealing temperatures, 58°C, 57°C and 46°C in the three consecutive cycles, respectively.

Selectively amplified PCR products were denatured at 94°C for 5 min after adding 10 µl of blue stop solution (*LI-COR* Bioscience, Lincoln, Nebraska, USA), and then immediately placed on ice. A 6.5% polyacrylamide gel was pre-run at 1500 V for 20 min on a *LI-COR* 4300 DNA Analyzer. Samples (0.5 µl each) were loaded onto the gel and then run for 3.5 h under the same conditions as for the pre-run.

In the EMS mutant population, TD was initiated with 500 ng and 200 ng total DNA (Figures 5.1 and 5.2). The lower concentration of 200 ng was tested to rule out the possibility of any inhibitors that might affect restriction digestion or PCR at higher DNA concentrations. Both concentrations gave the same result with regard to banding patterns and hence 200 ng of DNA was used for TD in the tissue-cultured samples.

Reproducibility of Variant Banding Patterns in the M_3 Generation of EMS Mutants, T_1 Generation of Tissue-Cultured Plants and Segregation Analysis

To exclude the possibility that changes seen in TD, with tissue-cultured lines compared to parental genotypes, were transient and non-heritable, TD was repeated on DNA from T_1 plants, which also permitted segregation analysis in a small progeny of maximum eight plants for each line. Repeatability of the results observed with the T_0 generation, with regard to new bands identified, was tested in the T_1 generation. In clusters where all five plants showed the same altered banding, a maximum of eight seeds from one plant was used to grow the T_1 generation. Fewer seeds were used when T_1 seeds were limited as in 'Cluster 5' of 'Tifrunner' and 'Cluster 4' of 'F439-16-10-3-3'. If only a single plant out of five plants in a cluster showed variation, the T_1 generation was grown from the respective plant. Based on seed availability, the T_1 generation

was grown from three clusters of 'Tifrunner', two clusters of F439-16-10-3-2 and one cluster each of F439-16-10-3-1 and F439-16-10-3-3.

Similarly altered banding identified in the M_2 generation by TD, for EMS mutated 'Tifrunner' lines, was tested for stability and segregation in the M_3 generation with a maximum of 10 seeds per line. As seed was limited for plant 24.3, only 4 seeds were used to grow the M_3 generation. A Chi-square test was done to assess the goodness-of-fit for 3:1 segregation at a significance level of 5%.

Gel Excision of Transposed Fragments for Sanger Sequencing

PCR products from TD selective amplification were run on 8% polyacrylamide gels. Fragments were excised using an Odyssey infrared imaging system (*LI-COR*). The gel slices were put in low TE (10 mM Tris, pH 7.5 and 0.1 mM EDTA) and subjected to three freeze-thaw cycles to extract the DNA. The gel slice was separated by centrifuging at 16,060 x g for 20 min and supernatant was collected. The MITE-specific fragments were separated from any non-MITE fragments that may have been trapped in gel slices by hybridizing to biotin-labeled MITE-specific oligos and pull down with streptavidin beads (Table 5.1), followed by streptavidin separation (Carrier *et al.*, 2012). The recovered MITE-specific fragments were amplified according to the selective amplification PCR profile using a non-fluorescent MITE-specific primer (Table 5.1). The amplified products were cloned into the PCR®4-TOPO® vector (Invitrogen, USA), and transformed into One Shot TOP10 Chemically Competent *E. coli* as per manufacturer's protocol. Bacterial colonies were grown in LB (Luria-Bertani) freeze broth (Woo *et al*., 1994) supplemented with kanamycin at 50 µg/ml and were screened for desired inserts using selective amplification and gel separation as described above to confirm the correct fragment size. After identifying colonies with desired fragments, an aliquot from freeze broth

was grown in LB medium with kanamycin and plasmids were extracted with a Qiagen Miniprep kit and sent for sequencing to the Georgia Genomics Facility, Athens, GA. Sequence analysis was done after trimming vector sequences

Testing TD Across Peanut Cultivars

To test whether variation in banding shown by TD is due to EMS mutagen and not due to seed contamination in 'Tifrunner' seed lot used for initial mutation 16 peanut cultivars grown in plots adjacent to 'Tifrunner' plot was tested by transposon display (TD). The cultivars tested were Marc I, AP4, Florida Fancy, Carver, McCloud, GA01R, AlNorden, Florida-07, Tifguard, Georgia Green, GA02C, Georganic, DP1, Tifrunner, York and C99R.

Results

Detection of Novel Banding Patterns in Individuals from the EMS-Mutagenized Population

All four primer combinations, MITE-specific primer 1756 + *Mse*I+A/T/G/C, amplified varying numbers of bands in 'Tifrunner', 26, 20, 10 and 16, respectively. The size ranges of the amplified fragments were 100-500 bp, 100-400 bp, 100-350 bp and 50-400 bp, respectively. Of the 401 EMS-mutagenized M_2 plants tested by TD, five (1.24%) showed additional or missing bands with the four primer pairs employed (Table 5.2, Figures 5.1 and 5.2). The M_2 mutant plant 25.3 showed new or missing bands with all of the primer combinations and plant 36.8 showed new bands with three primer combinations. The number of altered fragments ranged from one to four, in the size range of 100 to 255 bp. Selective amplification with *Mse*I+C detected changes among the maximum number of plants (4/401). The primer *Mse*I+T detected only additional bands but no deleted bands. The primer *Mse* I+T identified eight new fragments in three plants, thus showing the maximum number of additional bands. The four primer combinations identified a total of twenty-three additional or deleted fragments from 401 plants. Among the twenty-three

fragments, 70% (16/23) were not present in the 'Tifrunner' background. All of these fragment patterns were confirmed by two technical and two biological replications.

Some common variant bands were detected in different M_2 plants with the same primer combinations. These include bands at 139 bp in M_2 plants 14.7 and 17.6 and bands at 182 bp in plants 25.3 and 36.8 with primers 1756 + *Mse*I/C. Similarly, plants 24.3, 25.3 and 36.8 showed a new band at 219 bp with primers 1756 + *Mse*I/T. The same primer pair also amplified similar bands at 114 bp for plants 25.3 and 36.8. None of the variant plants had banding patterns that were identical to one another with different primers tested (Figures 5.1 and 5.2). Although plants 14.7 and 17.6 showed similar banding patterns in the M_2 generation (1756 + *Mse*I/C), they differed according to segregation analysis (Figure 5.5, Table 5.2).

M_3 progeny were grown from each of the M_2 plants that showed an altered banding pattern (plant numbers-14.7, 17.6, 24.3, 25.3 and 36.8) and TD was repeated. This permitted segregation and stability analysis of the altered bands in a small number of progeny of 4 to 10 M_3 plants (Figures 5.3, 5.4 and 5.5). TD analysis showed 75% (12/16) of altered bands to be homozygous and present in all respective progeny. While 12.5% (2/16) of altered bands showed Mendelian 3:1 segregation; the remaining 12.5 % were segregating in non-Mendelian ratio. Chi-square analysis showed that the segregation ratios were not significantly different (observed $P>0.05$) from 3:1 segregation (Table 5.2).

While MITE transposition was one explanation for the additional and missing fragments in the five plants, another explanation could be outcrossing or admixture in the seed lot used for mutagenesis. In order to test the possibility of seed admixture, 15 peanut varieties and seeds from 5 different 'Tifrunner' seed lots were tested by TD to compare banding patterns with those from the M_2 plants that showed patterns different from 'Tifrunner'. No similarity was shown by

the variant plants to any of the cultivars tested as evidenced from the TD banding patterns (Figure 5.6). Plants from all five 'Tifrunner' seed lots showed identical banding patterns by TD. It was also possible to analyze MITE insertional polymorphism in these lines in the same experiment. When three primer pairs were tested (1756 + *Mse*I A/T/G), the percentage of bands that were polymorphic among the 16 genotypes ranged from 33.3 to 61.2 %. The primer pair 1756 + *Mse*I/T showed the highest percentage of polymorphic bands and the least was by 1756 + *Mse*I/G (Table 5.3)

Detection of Novel Banding Patterns in Tissue Culture-Derived Plants

The potential for *AhMITE1* transposition as a consequence of tissue culture was tested by screening five T_0 plants regenerated from each 10-mo-old embryogenic cluster. In 'Tifrunner', of the 16 T_0 generation clusters from which plants were tested, plants from clusters 1, 5 and 49 (3/16 clusters) showed different banding patterns. T_0 plants from 3, 6 and 1 clusters of 'Florunner' component lines, F439-16-10-3-1, F439-16-10-3-2 and F439-16-10-3-3, respectively, showed altered banding patterns (Table 5.4).

Transmission of variant bands was tested in the T_1 generation to the extent possible depending on seed availability. In progeny testing with seeds from one plant each from the three 'Tifrunner' clusters (1, 5, 49) to study transmission of altered banding patterns in the T_1 generation, only cluster 1 exhibited reproducible variant bands. Progeny of plant # 3 from cluster 1 showed six additional bands with two primer combinations; only plant 3 was tested since all five T_0 plants had shown the same fragment pattern. The primers 1756 + *Mse*I/T and 1756 + *Mse* I/C produced four and two new bands, respectively (Table 5.5, Figure 5.7). Among the three component lines of 'Florunner', the variant fragment pattern was transmitted from T_0 to T_1 only for plants from component line F439-16-10-3-2. Between the two clusters (11 and 38) of F439-16-10-3-2

tested, two primer combinations showed four new bands; two additional bands shown by primer pairs 1756 + *Mse*I/A and one additional band shown by primers 1756 + *Mse*I/T in 'Cluster 11' and one additional band with primer 1756 + *Mse*I/T 'Cluster 38' were reproducible in the T_1 generation (Table 5.5, Figure 5.8). When variation was observed in only one out of five plants derived from a cluster (e.g. 5, 23, and 49), the variation was not detected in progeny and may have been due to technical artifact since biological replicate DNA samples were not run for these plants in T_0 generation. For the other case of non-heritable variation (cluster 4 of F439-16-10-3-3), only two progeny were available for testing which by chance could have been two that did not inherit the variable fragments. Of the four clusters where all five regenerated plants showed the same variant bands ('Tifrunner' cluster 1, F439-16-10-3-2 clusters 11 and 38, F439-16-10-3-3 cluster 4), three (75%) were the source of heritable variation.

Of the six bands showing stable variation in 'Tifrunner', one band segregated 3:1 and one band showed no segregation in the M_2; none of the bands showed segregation that was significantly different from 3:1 at $P \leq 0.05$ based on the Chi square test, although *P*-values are not accurate at low sample numbers. In 'Florunner/ F439-16-10-3-2', three bands did not segregate indicating homozygosity of the T_0 plant from cluster 11, an unexpected result for three mutations that arose in tissue culture. Additional progeny from this line would need to be screened to confirm homozygosity. Mendelian 3:1 segregation was observed for one variant fragment in progeny from the cluster 38 plant.

Comparison of Variants from EMS Mutation and Tissue-Culture

Variation in *AhMITE1* banding patterns was tested in plants derived from tissue-culture and EMS treatment in 'Tifrunner'. Comparison of banding patterns showed that primer 1756 + *MseI*/T amplified a heritable variant band at 219 bp in M_2 plants 24.3, 25.3, 36.8 and T_1 progeny

from a cluster 1 plant. A band of the same size is likely to but may not always represent the same sequence, which would require further investigation to determine. 'Tifrunner' showed altered banding in 1.24 % (5/401) of plants by EMS mutation and in 6.25% (1/16) of tissue-cultured clusters from which T_1 plants were tested. Overall mutation frequency among 'Florunner' lines was 10.5%. Although small numbers of tissue-cultured clusters per genotype were analyzed, differences in MITE activation/altered banding were observed since only one of the three component lines of Florunner showed heritable variation in the T_1 generation.

Sequence Characterization of Transposon Junctions

Seven variant bands greater than 190 bp in length were excised from TD gels on which amplified T_1 DNA had been run, but clones containing only four of them were obtained. Colony screening by TD confirmed the expected sizes of inserts for these four. Three clones representing each amplicon were sequenced and shown to be identical. All fragments had the MITE sequence, *Mse*I adapter sequence with the selective nucleotide and the flanking sequences (Table 5.6). The length of flanking sequences varied between 96 bp and 121 bp. Similarity search (BLASTn) of the full-length sequence to expressed sequence tags (ESTs) in NCBI showed the best alignment to the same *A. hypogaea* root cDNA clone (FS978055) for all the four sequences. All aligned regions corresponded to the MITE and not to the flanking sequences. BLASTn of the flanking sequences alone to ESTs showed no significant hits. Similarity search of the full-length fragment (MITE + flanking sequences) to genomic survey sequences (GSS) showed hits to multiple *A. hypogaea* clones. Sequence alignment showed that the cloned fragments aligned with *A. hypogaea* GSS sequences mainly at the MITE regions. Alignment of flanking sequences was

seen only to GSS sequences, DH968695 and DH965962, which were the top hits for cloned fragments 1 and 2, respectively. No significant protein hits were found by BLASTx of the initial BLASTn identified sequences (Table 5.7).

Discussion

The current study reports variation in insertion sites for peanut AhMITE1 putatively induced by the stress of tissue culture and chemical mutagenesis. This variation potentially can be used to create novel transpositions for gene tagging. These results are in agreement with previous observations that MITEs can be activated by stress (Casa *et al.*, 2000; Jiang *et al.*, 2003; Kikuchi *et al.*, 2003; Momose *et al.*, 2010). Activation of transposable elements is dependent on the transposon and the nature of stress. Rice retrotransposon, *Tos17 (*Takeda *et al.*, 1999), and MITE, *mPing* (Kikuchi *et al*., 2003), show transposition in response to tissue culture. Peanut MITEs have been reported to be activated by stress (Gowda *et al*., 2011; Patel *et al*., 2004), although a study to compare activation in the same genotype exposed to different stresses has not been conducted. The variation in banding observed in this study can be explained by 1) MITE transposition due to the effect of mutagen or prolonged tissue culture or 2) inherent variation in the seed sources, which cannot be disregarded at this point. We concluded that seed contamination as the cause for the variation seen in the EMS population was unlikely after testing for similar variant bands in multiple genotypes that were known to have been grown in the same year. The observation that different M_2 plants show varied bands at same position with the same primers (the band at 219 bp in plants 24.3, 25.3 and 36.8 with primer 1756 + *Mse*I/T) suggests an inherent insertion site in multiple genotypes; however, each of the plants with this common band had additional unique variant bands which supports transposition. Identical variant bands across different plants could also be due to the insertion preference of MITEs to certain

'hot-spots' in the genome. Recurrent EMS mutagenesis of peanut mutant 'VL1', a Valencia-type peanut, resulted in a mutation frequency of 0.026 (1 in 38.5 plants) for late leaf spot resistance associated with a specific MITE-containing locus (Gowda *et al*., 2010). While we did not analyze our population for specific phenotypes, evidence for insertion at the same locus (new band at 219 bp) in multiple mutants was obtained at a frequency of 0.007 (1 in 134 plants)

Both chemical mutagen treatment and tissue culture resulted in the recovery of plants showing variant banding patterns at slightly different frequencies, 1.25% (Tifrunner) vs 6.25% (Tifrunner) -10.5% (Florunner), respectively. Frequencies of variant recovery cannot be directly compared between treatments given that seed treatment would affect only up to two reproductive meristems in each seed versus continuous meristematic growth of tissue cultures due to repetitive embryogenesis over a time span of 10 mo. Nevertheless, prolonged culture of tissues resulted in the recovery of variants at a higher frequency. Similar results were obtained in tobacco (Hirochika, 1993), wherein the copy number of *Tto1* increased up to 10-fold in cell lines and to a lesser extent in regenerated plants. The increased transposition in cell lines was attributed to the prolonged culture (up to 3 years) compared with plants regenerated only 3-4 months after culture initiation. As opposed to tobacco retrotransposon *Tto1,* which is activated by tissue culture and protoplast culture, the rice retrotransposon *Tos 1,* was activated by tissue culture alone and not by protoplast culture (Hirochika *et al.*, 1996). In the rice variety 'Matasumae', transposition of *Tos17* was observed upon introgression of wild rice DNA, but not as a consequence of tissue culture (Liu *et al.*, 2004). The differential response of transposons to stress, as seen in our study, is in agreement with these earlier reports. Peanut *AhMITE1* is a non-autonomous transposon, which depends on transposases of similar families for transposition. Identifying the *trans*-acting transposase and the regulatory factor or factors for transposition would enable the differential

response to chemical mutagenesis and tissue culture to be more thoroughly characterized. In 'Florunner', of the three component lines, only F439-16-10-3-2 showed MITE activation in response to tissue culture and heritable variation was observed to originate in two out of seven clusters of this genotype tested suggesting that a transposase in this genotype may be more easily activated. However, three variant bands originating from one cluster did not show any segregation in the T_1 generation. Homozygosity of a T_0 plant would be unexpected given that mutation at the same position in homologous chromosomes is unlikely to occur. It is possible that these variant banding patterns were preexisting in the seed lot used to establish the cultures.

We included four genotypes in our tissue culture study although three of them were closely related to one another and the numbers of clusters analyzed per genotype were too small to conclude that genotypic differences contributed to the frequency of variant banding patterns. To more thoroughly compare the response of genotype, an experiment designed with larger numbers of initial explants, identity tracking throughout culture, and regeneration shortly after culture initiation as well as after prolonged culture would be necessary. Differences between genotypes for transposon activation in other species have been observed, e.g., activation of *Tos17* in rice under tissue culture was dependent on the genotype (Liu *et al.*, 2004); transposition of *Tos17* was seen only in the rice line RZ35 during tissue culture of genetically similar introgression lines RZ2 and RZ35. Differences in transposon activation were documented between barley varieties where mutated loci identified by SSAP reactions were 15.9% and 24.4% respectively in varieties Brenda and Scarlett (Polok, 2011). Activation of rice MITE *mPing* under tissue culture was higher for C5924 (*indica)* cell lines than for Nipponbare (*japonica*) (Jiang *et al.*, 2003).

In barley, chemical mutation by sodium azide (NaN_3) and N-methyl-N-nitrosourea resulted in more insertions than deletions with respect to the transposons *Tpo-1-like* and *BARE-1* (Polok,

2011). We also observed that 70% of the MITE-specific variant bands in the EMS mutated population of 'Tifrunner' were additional bands rather than missing bands.

Stability and segregation analysis in the M_3 generation with 5 plants (14.7, 17.6, 24.3, 25.3 and 36.8) that exhibited new bands in the M_2 generation showed that 75% of alterations were homozygous in nature. Given that M_2 seeds are recovered from the mutagenized M_1 after sexual reproduction, recovery of homozygous mutations in some lines is expected. The frequency of homozygous mutations seen in our study was similar to that observed in MITE-insertion mutants reported by Gowda *et al.*, (2010) where two out of four families bred true in the M_3 generation.

BLASTn of the flanking sequences obtained from additional TD fragments showed no hits to any coding regions. This may be due to the short flanking sequences, low number of sequences tested, and likelihood that insertions are in regulatory regions of a gene rather than coding regions. Only a limited number of bands were cloned and sequenced (4 of 9 altered bands from tissue-culture regenerants and none from the mutant population). The two genomic clones, DH968695 and DH965962, which showed similarity to flanking sequences of cloned fragments 1 and 2 were derived from peanut cultivars 'Satanoka' and 'Kintoki', by capturing and cloning MITE sequences (Shirasawa *et al.*, 2012b), suggesting that these insertion sites are not unique within the cultivated peanut germplasm.

Conclusion

Transposon display with two restriction enzymes and four primer combinations identified variant bands in populations derived by EMS treatment and plants regenerated from tissue culture. Additional variant bands and tagged genomic sequences could be detected in the tissue cultured lines and chemical mutants if different restriction enzymes and primer combinations are tested. Testing TD with a larger population of mutants or tissue-cultured lines is needed to differentiate

between induced mutation vs. inherent mutation. Sequencing technologies have advanced in delivering large volumes of data at reduced cost. Since the peanut MITE is activated by stress, the peanut mutant populations and wild types can also be utilized for high throughput sequencing of MITE-rich regions to obtain information about the transposition sites, insertion in gene-rich regions and polymorphic insertions on a larger scale. Transposon Display can be used for initial screening of members of the mutated population, to identify potential mutants. Further these mutants can be processed through a sequencing and bioinformatics pipeline for gene tagging analysis. This approach would bypass the restriction-digestion step with multiple enzymes, and the time consuming steps of gel excision, cloning, and sequencing of individual variant bands.

References

Alonso J.M., Stepanova A.N., Leisse T.J., Kim C.J., Chen H.M., Shinn P., Stevenson D.K., Zimmerman J., Barajas P., Cheuk R., Gadrinab C., Heller C., Jeske A., Koesema E., Meyers C.C., Parker H., Prednis L., Ansari Y., Choy N., Deen H., Geralt M., Hazari N., Hom E., Karnes M., Mulholland C., Ndubaku R., Schmidt I., Guzman P., Aguilar-Henonin L., Schmid M., Weigel D., Carter D.E., Marchand T., Risseeuw E., Brogden D., Zeko A., Crosby W.L., Berry C.C., Ecker J.R. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653-657.

Ambrose B.A., Lerner D.R., Ciceri P., Padilla C.M., Yanofsky M.F., Schmidt R.J. (2000) Molecular and genetic analyses of the *Silky1* gene reveal conservation in floral organ specification between eudicots and monocots. Mol. Cell 5:569-579.

An G.H., Jeong D.H., Jung K.H., Lee S. (2005) Reverse genetic approaches for functional genomics of rice. Plant Mol. Biol. 59:111-123.

Atares A., Moyano E., Morales B., Schleicher P., Garcia-Abellan J.O., Anton T., Garcia-Sogo B., Perez-Martin F., Lozano R., Flores F.B., Moreno V., Bolarin M.D., Pineda B. (2011) An insertional mutagenesis programme with an enhancer trap for the identification and tagging of genes involved in abiotic stress tolerance in the tomato wild-related species *Solanum pennellii*. Plant Cell Rep. 30:1865-1879.

Baulcombe D.C. (1996) RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol. Biol. 32:79-88.

Baumberger N., Ringli C., Keller B. (2001) The chimeric leucine-rich repeat/extensin cell wall protein LRX1 is required for root hair morphogenesis in *Arabidopsis thaliana*. Genes & Dev.15:1128-1139.

Beetham P.R., Kipp P.B., Sawycky X.L., Arntzen C.J., May G.D. (1999) A tool for functional plant genomics: Chimeric RNA/DNA oligonucleotides cause in vivo gene-specific mutations. Pro. Nat. Acad.Sci.USA 96:8774-8778.

Bennetzen J.L. (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol. Biol. 42:251-269.

Bhat R.S., Patil V.U., Chandrashekar T.M., Sujay V., Gowda M.V.C., Kuruvinashetti M.S. (2008) Recovering flanking sequence tags of a Miniature Inverted-repeat Transposable Element by thermal asymmetric interlaced-PCR in peanut. Current Science 95:452-453.

Bradley D., Carpenter R., Copsey L., Vincent C., Rothstein S., Coen E. (1996) Control of inflorescence architecture in *Antirrhinum.* Nature 379:791-797.

Cantu D., Vanzetti L.S., Sumner A., Dubcovsky M., Matvienko M., Distelfeld A., Michelmore R.W., Dubcovsky J. (2010) Small RNAs, DNA methylation and transposable elements in wheat. BMC Genomics 11:408.

Carrier G., Le Cunff L., Dereeper A., Legrand D., Sabot F., Bouchez O., Audeguin L., Boursiquot J.-M., This P. (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. PLoS ONE 7(3): e32973.

Casa A.M., Brouwer C., Nagel A., Wang L.J., Zhang Q., Kresovich S., Wessler S.R. (2000) The MITE family *Heartbreaker* (*Hbr*): Molecular markers in maize. Proc. Nat. Acad. Sci. USA 97:10083-10089.

Chatterjee M., Martin C. (1997) *Tam3* produces a suppressible allele of the *DAG* locus of *Antirrhinum majus* similar to *Mu*-suppressible alleles of maize. Plant J. 11:759-771.

Chen X., Brenneman T., Culbreath A., Holbrook C., Guo B. (2008) Characterization of two peanut oxalate oxidase genes and development of peanut cultivars resistant to stem rot (*Sclerotium rolfsii*). Phytopathology 98:S36-S36.

Chu Y., Holbrook C.C., Ozias-Akins P. (2009) Two alleles of *ahFAD2B* control the high oleic acid trait in cultivated peanut. Crop Sci. 49:2029-2036.

Cribb L., Hall L.N., Langdale J.A. (2001) Four mutant alleles elucidate the role of the G2 protein in the development of C-4 and C-3 photosynthesizing maize tissues. Genetics 159:787-797.

Curtin S.J., Voytas D.F., Stupar R.M. (2012) Genome engineering of crops with designer nucleases. Plant Genome 5:42-50.

Elliott R.C., Betzner A.S., Huttner E., Oakes M.P., Tucker W.Q.J., Gerentes D., Perez P., Smyth D.R. (1996) *AINTEGUMENTA*, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. Plant Cell 8:155-168.

Feschotte C., Jiang N., Wessler S.R. (2002) Plant transposable elements: Where genetics meets genomics. Nat. Rev. Genet. 3:329-341.

Fire A., Albertson D., Harrison S.W., Moerman D.G. (1991) Production of antisense RNA leads to effective and specific inhibition of gene expression in *C.elegans* muscle. Development 113:503-514.

Fire A., Xu S.Q., Montgomery M.K., Kostas S.A., Driver S.E., Mello C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806-811.

Fitzgerald T.L., Kazan K., Manners J.M. (2012) The application of reverse genetics to polyploid plant species. Crit. Rev. Plant Sci. 31:181-200.

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., Hadley D., Hutchinson D., Martin C., Katagiri F., Lange B.M., Moughamer T., Xia Y., Budworth P., Zhong J.P., Miguel T., Paszkowski U., Zhang S.P., Colbert M., Sun W.L., Chen L.L., Cooper B., Park S., Wood T.C., Mao L., Quail P., Wing R., Dean R., Yu Y.S., Zharkikh A., Shen R., Sahasrabudhe S., Thomas A., Cannings R., Gutin A., Pruss D., Reid J., Tavtigian S., Mitchell J., Eldredge G., Scholl T., Miller R.M., Bhatnagar S., Adey N., Rubano T., Tusneem N., Robinson R., Feldhaus J., Macalma T., Oliphant A., Briggs S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. *ssp japonica*). Science 296:92-100.

Gojobori T. (2007) Curated genome annotation of *Oryza sativa* ssp *japonica* and comparative genome analysis with *Arabidopsis thaliana* - The Rice Annotation Project. Genome Res.17:175-183.

Gowda M.V.C., Bhat R.S., Motagi B.N., Sujay V., Kumari V., Sujatha B. (2010) Association of high frequency origin of late leaf spot resistant mutants with *AhMITE1* transposition in peanut. Plant Breed. 129:567-569.

Gowda M.V.C., Bhat R.S., Sujay V., Kusuma P., Varshakumari, Bhat S., Varshney R.K. (2011) Characterization of *AhMITE1* transposition and its association with the mutational and evolutionary origin of botanical types in peanut (*Arachis spp*.). Plant Syst. Evol. 291:153-158.

Grandbastien M.A., Lucas H., Morel J.B., Mhiri C., Vernhettes S., Casacuberta J.M. (1997) The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses. Genetica 100:241-252.

Guimarães P B.A., Proite K, de Araújo A, Leal-Bertioli S, Pic-Taylor A, da Silva F, Morgante C, Ribeiro S, Bertioli D. (2010) A study of gene expression in the nematode resistant wild peanut relative, *Arachis stenosperma*, in response to challenge with *Meloidogyne arenaria*. Trop. Plant. Biol. 3:183-192.

Hake S., Vollbrecht E., Freeling M. (1989) Cloning *Knotted*, the dominant morphological mutant in maize using *DS2* as a transposon tag. EMBO J. 8:15-22.

Hancock C.N., Zhang F., Floyd K., Richardson A.O., LaFayette P., Tucker D., Wessler S.R., Parrott W.A. (2011) The rice miniature inverted repeat transposable element *mPing* is an effective insertional mutagen in soybean. Plant Physiol. 157:552-562.

Hardy S., Legagneux V., Audic Y., Paillard L. (2010) Reverse genetics in eukaryotes. Biol. Cell 102:561-580.

Hirochika H. (1993) Activation of tobacco retrotransposons during tissue-culture. EMBO J. 12:2521-2528.

Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H., Kanda M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc.Nat. Acad. Sci.USA 93:7783-7788.

Hsing Y.I., Chern C.G., Fan M.J., Lu P.C., Chen K.T., Lo S.F., Sun P.K., Ho S.L., Lee K.W., Wang Y.C., Huang W.L., Ko S.S., Chen S., Chen J.L., Chung C.I., Lin Y.C., Hour A.L., Wang Y.W., Chang Y.C., Tsai M.W., Lin Y.S., Chen Y.C., Yen H.M., Li C.P., Wey C.K., Tseng C.S., Lai M.H., Huang S.C., Chen L.J., Yu S.M. (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol. Biol. 63:351-364.

Huang J., Zhang K., Shen Y., Huang Z., Li M., Tang D., Gu M., Cheng Z. (2009) Identification of a high frequency transposon induced by tissue culture, *nDaiZ*, a member of the *hAT* family in rice. Genomics 93:274-281.

Huijser C., Kortstee A., Pego J., Weisbeek P., Wisman E., Smeekens S. (2000) The *Arabidopsis SUCROSE UNCOUPLED-6* gene is identical to *ABSCISIC ACID INSENSITIVE-4*: involvement of abscisic acid in sugar responses. Plant J. 23:577-585.

Jeong D.H., An S.Y., Kang H.G., Moon S., Han J.J., Park S., Lee H.S., An K.S., An G.H. (2002) T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. 130:1636-1644.

Jiang N., Bao Z.R., Zhang X.Y., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. (2003) An active DNA transposon family in rice. Nature 421:163-167.

Jiang S.Y., Ramachandran S. (2010) Natural and artificial mutants as valuable resources for functional genomics and molecular breeding. Int. J. Biol. Sci. 6:228-251.

Kalendar R., Tanskanen J., Immonen S., Nevo E., Schulman A.H. (2000) Genome evolution of wild barley (*Hordeum* spontaneum) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. Pro. Nat. Acad. Sci. USA. 97:6603-6607.

Kempin S.A., Liljegren S.J., Block L.M., Rounsley S.D., Yanofsky M.F., Lam E. (1997) Targeted disruption in *Arabidopsis*. Nature 389:802-803.

Kikuchi K., Terauchi K., Wada M., Hirano H.Y. (2003) The plant MITE *mPing* is mobilized in anther culture. Nature 421:167-170.

Knoll J.E., Ramos M.L., Zeng Y., Holbrook C.C., Chow M., Chen S., Maleki S., Bhattacharya A., Ozias-Akins P. (2011) TILLING for allergen reduction and improvement of quality traits in peanut (*Arachis hypogaea* L.). BMC Plant Biol. 11:81

Krysan P.J., Young J.C., Sussman M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283-2290.

Krysan P.J., Young J.C., Jester P.J., Monson S., Copenhaver G., Preuss D., Sussman M.R. (2002) Characterization of T-DNA insertion sites in *Arabidopsis thaliana* and the implications for saturation mutagenesis. OMICS 6:163-174.

Liu Z.L., Han F.P., Tan M., Shan X.H., Dong Y.Z., Wang X.Z., Fedak G., Hao S., Liu B. (2004) Activation of a rice endogenous retrotransposon *Tos17* in tissue culture is accompanied by cytosine demethylation and causes heritable alteration in methylation pattern of flanking genomic regions. Theor. Appl. Genet. 109:200-209.

Maes T., De Keukeleire P., Gerats T. (1999) Plant tagnology. Trends Plant Sci.4:90-96.

Martirani L., Stiller J., Mirabella R., Alfano F., Lamberti A., Radutoiu S.E., Iaccarino M., Gresshoff P.M., Chiurazzi M. (1999) T-DNA tagging of nodulation and root-related genes in *Lotus japonicus:* Expression patterns and potential for promoter trapping and insertional mutagenesis. Mol. Plant-Microbe Interact. 12:275-284.

May B.P., Martienssen R.A. (2003) Transposon mutagenesis in the study of plant development. Crit. Rev. Plant Sci. 22:1-35.

McClintock B. (1951) Chromosome organization and genic expression. Cold Spring Harbor Symposia on Quantitative Biology 16:13-47.

Meissner R.C., Jin H.L., Cominelli E., Denekamp M., Fuertes A., Greco R., Kranz H.D., Penfield S., Petroni K., Urzainqui A., Martin C., Paz-Ares J., Smeekens S., Tonelli C., Weisshaar B., Baumann E., Klimyuk V., Marillonnet S., Patel K., Speulman E., Tissier A.F., Bouchez D., Jones J.J.D., Pereira A., Wisman E., Bevan M. (1999) Function search in a large transcription factor gene family in *Arabidopsis*: Assessing the potential of reverse genetics to identify insertional mutations in *R2R3 MYB* genes. Plant Cell 11:1827-1840.

Momose M., Abe Y., Ozeki Y. (2010) Miniature inverted-repeat transposable elements of *Stowaway* Are active in potato. Genetics 186:59-66.

Nagy E.D., Guo Y., Tang S., Bowers J.E., Okashah R.A., Taylor C.A., Zhang D., Khanal S., Heesacker A.F., Khalilian N., Farmer A.D., Carrasquilla-Garcia N., Penmetsa R.V., Cook D., Stalker H.T., Nielsen N., Ozias-Akins P., Knapp S.J. (2012) A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics. 13:469.

Nakazaki T., Okumoto Y., Horibata A., Yamahira S., Teraishi M., Nishida H., Inoue H., Tanisaka T. (2003) Mobilization of a transposon in the rice genome. Nature 421:170-172.

Ozias-Akins P. (1989) Plant regeneration from immature embryos of peanut. Plant Cell Rep. 8:217-218.

Parinov S., Sundaresan V. (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. Curr. Opin. Biotechnol. 11:157-161.

Patel M., Jung S., Moore K., Powell G., Ainsworth C., Abbott A. (2004) High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theor. Appl. Genet. 108:1492-1502.

Polok K.a.Z., R. (2011) Mutagenic treatment induces high transposon variation in barley (*Hordeum vulgare* L.). Acta Agriculturae Slovenica 97:179-188.

Ramachandran S., Sundaresan V. (2001) Transposons as tools for functional genomics. Plant Physiol. Biochem. 39:243-252.

Ray T.K., Holly S.P., Knauft D.A., Abbott A.G., Powell G.L. (1993) The primary defect in developing seed from the high oleate variety of peanut (*Arachis hypogaea* L.) is the absence of beta(12)-desaturase activity. Plant Sci. 91:15-21.

Ruiz-Rojas J.J., Sargent D.J., Shulaev V., Dickerman A.W., Pattison J., Holt S.H., Ciordia A., Veilleux R.E. (2010) SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca* L. Theor. Appl. Genet.121:449-463.

Sabot F., Guyot R., Wicker T., Chantret N., Laubin B., Chalhoub B., Leroy P., Sourdille P., Bernard M. (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Mol. Genet. Genomics 274:119-130.

Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., Zhang J., Fulton L., Graves T.A., Minx P., Reily A.D., Courtney L., Kruchowski S.S., Tomlinson C., Strong C., Delehaunty K., Fronick C., Courtney B., Rock S.M., Belter E., Du F., Kim K., Abbott R.M., Cotton M., Levy A., Marchetto P., Ochoa K., Jackson S.M., Gillam B., Chen W., Yan L., Higginbotham J., Cardenas M., Waligorski J., Applebaum E., Phelps L., Falcone J., Kanchi K., Thane T., Scimone A., Thane N., Henke J., Wang T., Ruppert

J., Shah N., Rotter K., Hodges J., Ingenthron E., Cordes M., Kohlberg S., Sgro J., Delgado B., Mead K., Chinwalla A., Leonard S., Crouse K., Collura K., Kudrna D., Currie J., He R., Angelova A., Rajasekar S., Mueller T., Lomeli R., Scara G., Ko A., Delaney K., Wissotski M., Lopez G., Campos D., Braidotti M., Ashley E., Golser W., Kim H., Lee S., Lin J., Dujmic Z., Kim W., Talag J., Zuccolo A., Fan C., Sebastian A., Kramer M., Spiegel L., Nascimento L., Zutavern T., Miller B., Ambroise C., Muller S., Spooner W., Narechania A., Ren L., Wei S., Kumari S., Faga B., Levy M.J., McMahan L., Van Buren P., Vaughn M.W., *et al.* (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326:1112-1115.

Shan X.H., Liu Z.L., Dong Z.Y., Wang Y.M., Chen Y., Lin X.Y., Long L.K., Han F.P., Dong Y.S., Liu B. (2005) Mobilization of the active MITE transposons *mPing* and *Pong* in rice by introgression from wild rice (*Zizania latifolia* Griseb.). Mol. Biol. Evol. 22:976-990.

Shilman F., Brand Y., Brand A., Hedvat I., Hovav R. (2011) Identification and molecular characterization of homeologous *delta 9-stearoyl acyl carrier protein desaturase 3* genes from the allotetraploid Peanut (*Arachis hypogaea*). Plant Mol.Biol. Rep. 29:232-241.

Shirasawa K., Hirakawa H., Tabata S., Hasegawa M., Kiyoshima H., Suzuki S., Sasamoto S., Watanabe A., Fujishiro T., Isobe S. (2012a) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor. Appl. Genet. 124:1429-1438.

Shirasawa K., Koilkonda P., Aoki K., Hirakawa H., Tabata S., Watanabe M., Hasegawa M., Kiyoshima H., Suzuki S., Kuwata C., Naito Y., Kuboyama T., Nakaya A., Sasamoto S., Watanabe A., Kato M., Kawashima K., Kishida Y., Kohara M., Kurabayashi A., Takahashi C., Tsuruoka H., Wada T., Isobe S. (2012b) *In silico* polymorphism analysis

for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol. 12:80.

Smith L.G., Gerttula S.M., Han S.C., Levy J. (2001) *Tangled1*: A microtubule binding protein required for the spatial control of cytokinesis in maize. J. Cell Biol. 152:231-236.

Tadege M., Wang T.L., Wen J., Ratet P., Mysore K.S. (2009) Mutagenesis and beyond! tools for understanding legume biology. Plant Physiol. 151:978-984.

Taguchi-Shiobara F., Yuan Z., Hake S., Jackson D. (2001) The *fasciated ear2* gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. Genes Dev. Development 15:2755-2766.

Takeda S., Sugimoto K., Otsuki H., Hirochika H. (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J. 18:383-393.

Tzfira T., Weinthal D., Marton I., Zeevi V., Zuker A., Vainstein A. (2012) Genome modifications in plant cells by custom-made restriction enzymes. Plant Biotechnol. J.10:373-389.

Vain P., Worland B., Thole V., McKenzie N., Opanowicz M., Fish L.J., Bevan M.W., Snape J.W. (2008) *Agrobacterium*-mediated transformation of the temperate grass *Brachypodium distachyon* (genotype Bd21) for T-DNA insertional mutagenesis. Plant Biotechnol. J. 6:236-245.

Vos P., Hogers R., Bleeker M., Reijans M., Vandelee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M. (1995) AFLP - a new technique for DNA-fingerprinting. Nucleic Acids Res. 23:4407-4414.

Waites R., Selvadurai H.R.N., Oliver I.R., Hudson A. (1998) The *PHANTASTICA* gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in *Antirrhinum.* Cell 93:779-789.

Wessler S.R., Bureau T.E., White S.E. (1995) LTR-retrotransposons and MITES - Important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5:814-821.

Woo S.S., Jiang J.M., Gill B.S., Paterson A.H., Wing R.A. (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res. 22:4922-4931.

Wu C.Y., Li X.J., Yuan W.Y., Chen G.X., Kilian A., Li J., Xu C.G., Li X.H., Zhou D.X., Wang S.P., Zhang Q.F. (2003) Development of enhancer trap lines for functional analysis of the rice genome. Plant J. 35:418-427.

Yan Y.S., Lin X.D., Zhang Y.S., Wang L., Wu K.Q., Huang S.Z. (2005) Isolation of peanut genes encoding arachins and conglutins by expressed sequence tags. Plant Sci. 169:439-445.

Yu J. G., Lee G.H., Kim J.S., Shim E.J., Park Y.D. (2010) An insertional mutagenesis system for analyzing the Chinese cabbage genome using *Agrobacterium* T-DNA. Mol. Cells 29:267-275.

Yu J., Hu S.N., Wang J., Wong G.K.S., Li S.G., Liu B., Deng Y.J., Dai L., Zhou Y., Zhang X.Q., Cao M.L., Liu J., Sun J.D., Tang J.B., Chen Y.J., Huang X.B., Lin W., Ye C., Tong W., Cong L.J., Geng J.N., Han Y.J., Li L., Li W., Hu G.Q., Huang X.G., Li W.J., Li J., Liu Z.W., Liu J.P., Qi Q.H., Liu J.S., Li T., Wang X.G., Lu H., Wu T.T., Zhu M., Ni P.X., Han H., Dong W., Ren X.Y., Feng X.L., Cui P., Li X.R., Wang H., Xu X., Zhai W.X., Xu Z., Zhang J.S., He S.J., Zhang J.G., Xu J.C., Zhang K.L., Zheng X.W., Dong J.H., Zeng

W.Y., Tao L., Ye J., Tan J., Ren X.D., Chen X.W., He J., Liu D.F., Tian W., Tian C.G., Xia H.G., Bao Q.Y., Li G., Gao H., Cao T., Zhao W.M., Li P., Chen W., Wang X.D., Zhang Y., Hu J.F., Liu S., Yang J., Zhang G.Y., Xiong Y.Q., Li Z.J., Mao L., Zhou C.S., Zhu Z., Chen R.S., Hao B.L., Zheng W.M., Chen S.Y., Guo W., Li G.J., Liu S.Q., Tao M., Zhu L.H., Yuan L.P., Yang H.M. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 296:79-92.

Zhang J., Liang S., Duan J., Wang J., Chen S., Cheng Z., Zhang Q., Liang X., Li Y. (2012a) De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). BMC Genomics 13:90.

Zhang Q., Arbuckle J., Wessler S.R. (2000) Recent, extensive, and preferential insertion of members of the Miniature Inverted-repeat Transposable Element family *Heartbreaker* into genic regions of maize. Proc. Nat. Acad.Sci. of the USA. 97:1160-1165.

Zhang Y., Xia H., Yuan M., Zhao C.Z., Li A.Q., Wang X.J. (2011) Cloning and expression analysis of peanut (*Arachis hypogaea* L.) CHI gene. Electronic Journal Biotechnology North America 15:9-12

Table 5.1 Adapter and Primer Sequences

Primer	Description	Sequence (5'→ 3')
2060	MITE-specific biotinylated oligo	GGTGGATACTACAATGAAGATGGCATAATTGTCTTCATATGAG
2061	MITE-specific biotinylated oligo	GGTGGGTACTCCCATGAAGATRTTATAATTRTCTTCATGTGRT
1072	*Eco*RI-Adapter 1	CTCGTAGACTGCGTACC
1073	*Eco*RI-Adapter 2	AATTGGTACGCAGTC
1074	*Mse*I-Adapter 1	GACGATGAGTCCTGAG
1075	*Mse*I-Adapter 2	TACTCAGGACTCAT
1077	*Mse*I adapter primer	GATGAGTCCTGAGTAA
1808	MITE-specific pre-amplification primer	AAAACAACACTTTTATAAATATCAA
1756	Fluorescently labeled MITE-specific selective-amplification primer	CTAACCATACAATCC
2663	Unlabeled MITE-specific selective-amplification primer for cloning	CTAACCATACAATCC
1853	Selective amplification primer specific to *Mse*I-*Mse* I+A	GATGAGTCCTGAGTAAA
1854	Selective amplification primer specific to *Mse*I-*Mse* I+T	GATGAGTCCTGAGTAAT
1079	Selective amplification primer specific to *Mse*I-*Mse* I+G	GATGAGTCCTGAGTAAG
1855	Selective amplification primer specific to *Mse*I-*Mse* I+C	GATGAGTCCTGAGTAAC

Table 5.2 Variant bands and segregation of new bands in M_3 generation of EMS mutants

		M_2 generation		M_3 generation	
		~ Band position (bp)			
Plant[†] number	Primer[‡]	New band	Missing band	Segregation Present:absent	*P*-value[*]
14.7	+C	139		6:3	0.56
17.6	+C	139		10:0	0.06
24.3	+T	219		4:0	0.24
	+G	115	123	3:1	1
		166		3:1	1
25.3	+A	260	198	9:0	0.08
			177		
	+T	114		9:0	0.08
		127		9:0	0.08
		139		9:0	0.08
		219		9:0	0.08
	+G		166		
	+C	182	123	9:0	0.08
			198		
			249		
36.8	+T	114		10:0	0.06
		219		6:4	0.27
		240		10:0	0.06
	+G	174		10:0	0.06
	+C	200		10:0	0.06

†Plants are identified by the column number in Qiagen 96 well collection plate followed by well number in a column, each column having 8 wells. Columns are serially numbered across 96 well plates used for sample collection.

‡Selective amplification primer combinations were 1756 + *Mse*I/A/T/G/C, with extensions alone shown in table

*Tested for 3:1 segregation in M_3 generation by Chi square test at $P \leq 0.05$

Table 5.3 MITE insertional polymorphism in 16 peanut cultivars

Primers	Total number of bands	Total number of polymorphic bands	% polymorphism
1756+*Mse*I/A*	34	19	55.88
1756+*Mse*I/T	31	19	61.29
1756+*Mse*I/G	12	4	33.33

* Primer 1756 corresponds to the MITE and *Mse*I corresponds to the *Mse*I adapter with a single nucleotide addition at 3' end.

Table 5.4 Clusters showing variant banding pattern in T_0 generation

	Cluster number	Primer*	Plant showing variation	New bands	Missing Bands
'Tifrunner' ($16\ T_0$ clusters)	1	+T	All 5 plants	4	
		+C	All 5 plants	2	
	5	+T	2	1	
	49	+C	5	1	1
'Florunner/ F439-16-10-3-1' ($7\ T_0$ clusters)	1	+C	1	1	
	4	+T	1		1
		+G	1		1
		+C	4	1	
	23	+C	3	1	
'Florunner/ F439-16-10-3-2' ($7\ T_0$ clusters)	8	+C	All 5 plants		1
	11	+A	All 5 plants	2	
		+T	All 5 plants	1	1
	12	+A	All 5 plants	2	
		+T	All 5 plants	1	1
		+G	1	1	
	27	+A	All 5 plants	2	
		+T	All 5 plants	1	1
		+C	1	1	
	29	+A	All 5 plants	2	
		+T	All 5 plants	1	1
	38	+T	All 5 plants	1	
		+C	All 5 plants		1
'Florunner/ F439-16-10-3-3' ($5\ T_0$ clusters)	4	+C	All 5 plants	1	

*Selective amplification primer combinations were 1756 + *Mse*I/A/T/G/C, with extensions alone shown in table.

Table 5.5 Clusters showing variant banding in T_1 generation

Genotype	Cluster number /Number of plants tested	Cluster number exhibiting stable variation	Primer	~Band position (bp)	Segregation (Present : absent)	*P*-value‡
Tifrunner	1/8* 5/1 49/7	1	+T	115	7:1	0.41
				219	5:3	0.41
				385	4:4	0.1
				406	8:0	0.1
			+C	198	5:3	0.41
				217†	6:2	0.1
Florunner/ F439-16-10-3-1	23/7	0				
Florunner/ F439-16-10-3-2	11/8* 38/8*	11	+A	198	8:0	0.1
				219†	8:0	0.1
			+T	206†	8:0	0.1
		38	+T	198†	6:2	0.1
Florunner/ F439-16-10-3-1	4/2	0				

‡ Tested for 3:1 segregation in Chi square test at $P \leq 0.05$

*Indicates clusters showing heritable variation

†Indicates bands cloned and sequenced.

Table 5.6 Sequences of variant bands detected in tissue-cultured lines

Fragment	Sanger sequence reads (5'→3')*	Description
1	gatgagtcctgagtaaA tgcaatggtccctaaa ataaccaatgcccatatcttgttcaatgattgcgtt tgtgctgtttgtgtaaatggttctaatgaattggtttt gattctacttagtttttatggaatttattttt ggtggata ctacaatgaagatggcataattgtcttcatatgag tatatttctttttgacctttggatgatggattgtatggttag	Derived from tissue-cultured 'Florunner' Cluster #11, Primers 1756+MseI/A
3	gatgagtcctgagtaaT gtgcgtgtgtgtgtattt atatagagagagggggtattattgtgaaagtgtg tgttagtgggttgaggcaagttgttgaggcaagttg caaaagctataattagtaaaaggt ggatactac aatgaagatggcataattgtcttcatatgagtatattt ctttttgacctttggatgatggattgtatggttag	Derived from tissue-cultured 'Florunner' Cluster #11, Primers 1756+MseI/T
4	gatgagtcctgagtaaT atcgacttgaagtgtagaac tagatcatctgtattcactattcagtacaaacaggagtgt gaacaagtgggtataagtaatcctgccaacaaacaa ggt ggatactacaatgaagatggcataattgtcttcatatgagtata tttctttttgacctttggatgatggattgtatggttag	Derived from tissue-cultured 'Florunner' Cluster #38, Primers 1756+MseI/T
7	gatgagtcctgagtaaC taaaatcattcctccctctctctcc tttcttcaaccccccttgctttctctttctcaaactttcttttctatt cagcattcatccttctcactgctacttactcacatttctttttt ggtggatactacaatgaagatggcataattgtcttcatatgagt atatttctttttgacctttggatgatggattgtatggttag	Derived from tissue-cultured 'Tifrunner' Cluster #1, Primers 1756+MseI/C

*Sanger sequences of four variant bands excised from the gel. *Mse*I adapter sequences and MITE sequences are underlined. Selective nucleotide is capitalized.

Table 5.7 BLASTn of cloned fragments to GSS database

	BLASTn	Query coverage	Identity (%)	E-Value	Description
Fragment 1 (flanking sequence only)	DH968695	100	98	7.00E-53	Arachis *hypogaea* 'Satonoka' library genomic clone SATE0881
Fragment 2 (flanking sequence only)	DH965962	100	97	4.00E-43	Arachis *hypogaea* 'Kintoki' library genomic clone KIAC09E12
Fragment 1 (flanking and MITE sequence)	DH968695	93	99	9.00E-100	*Arachis hypogaea* 'Satonoka' library genomic clone SATE0881
Fragment 2 (flanking and MITE sequence)	DH965962	52	97	9.00E-43	*Arachis hypogaea* 'Kintoki' library genomic clone KIAC09E12
Fragment 4 (flanking and MITE sequence)	DH967905	45	100	7.00E-38	*Arachis hypogaea* 'Satonoka' library genomic clone SACT08J22

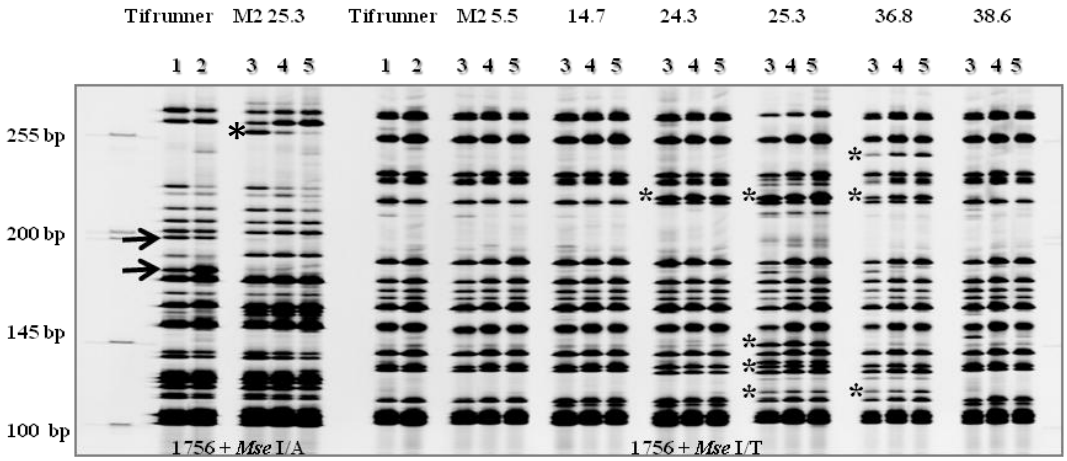

Figure 5.1 Partial TD profile of the M_2 generation of EMS mutagenized population of 'Tifrunner'.

TD with primers 1756+*Mse* I/A and 1756+*Mse* I/T.

1)'Tifrunner' with 500ng total DNA for restriction-Ligation, 2) 'Tifrunner' with 200ng total DNA for restriction-ligation, 3) M_2 generation DNA from first biological replicate with 500ng total DNA for restriction-ligation, 4) M_2 generation DNA from second biological replicate with 500ng total DNA for restriction-ligation, 5) M_2 generation DNA from second biological replicate with 200ng total DNA for restriction-ligation. Arrows indicate putative deletions - bands present in wild type and absent in mutants. Asterisks (*) indicate putative insertions - bands absent in wild type and present in mutants.

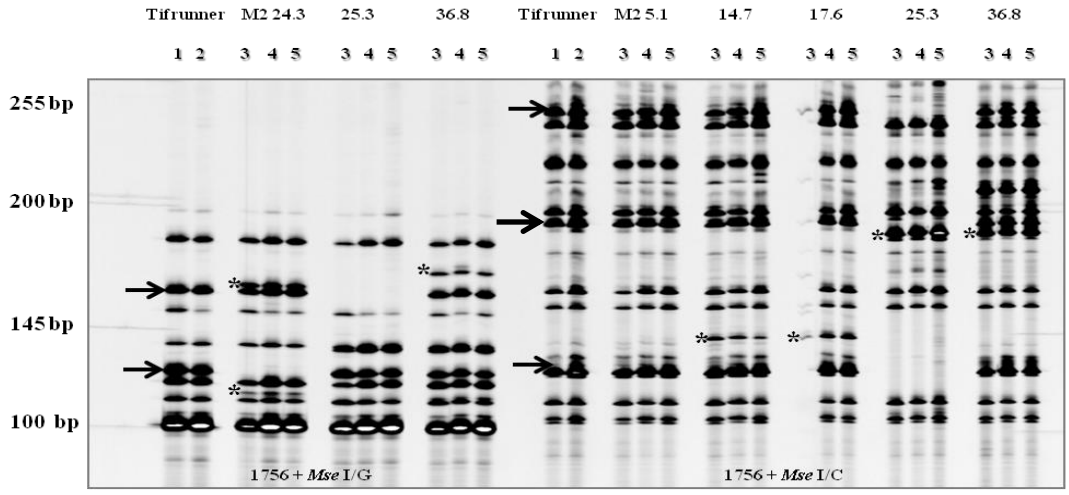

Figure 5.2 Partial TD profile of the M_2 generation of EMS mutagenized population of 'Tifrunner'

TD with primers 1756+*Mse* I/G and 1756+*Mse* I/C

1)'Tifrunner' with 500ng total DNA for restriction-Ligation, 2) 'Tifrunner' with 200ng total DNA for restriction-ligation, 3) M_2 generation DNA from first biological replicate with 500ng total DNA for restriction-ligation, 4) M_2 generation DNA from second biological replicate with 500ng total DNA for restriction-ligation, 5) M_2 generation DNA from second biological replicate with 200ng total DNA for restriction-ligation. Arrows indicate putative deletions - bands present in wild type and absent in mutants. Asterisks (*) indicate putative insertions - bands absent in wild type and present in mutants.

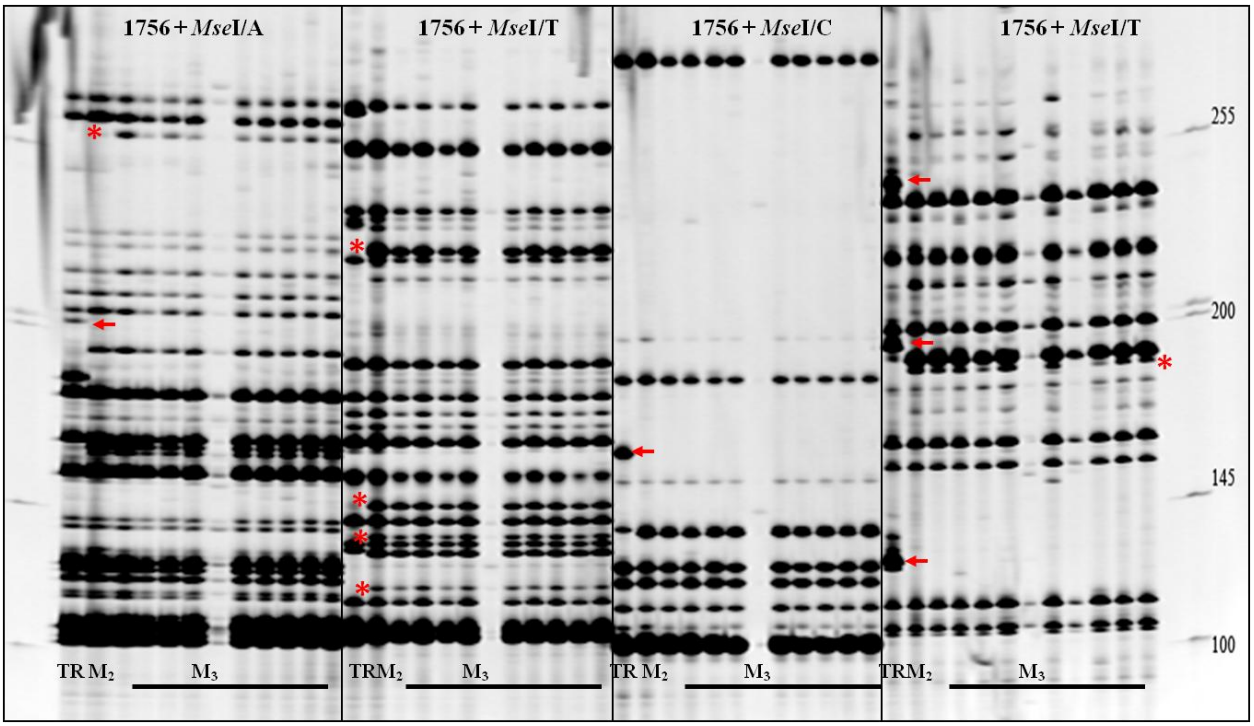

Figure 5.3- Partial TD profile of MITE transposition in the M_3 generation of M_2 plant 25.3

TR-'Tifrunner', M_2-M_2 generation, M_3-M_3 generation.

* New bands

→ Deletions/missing bands

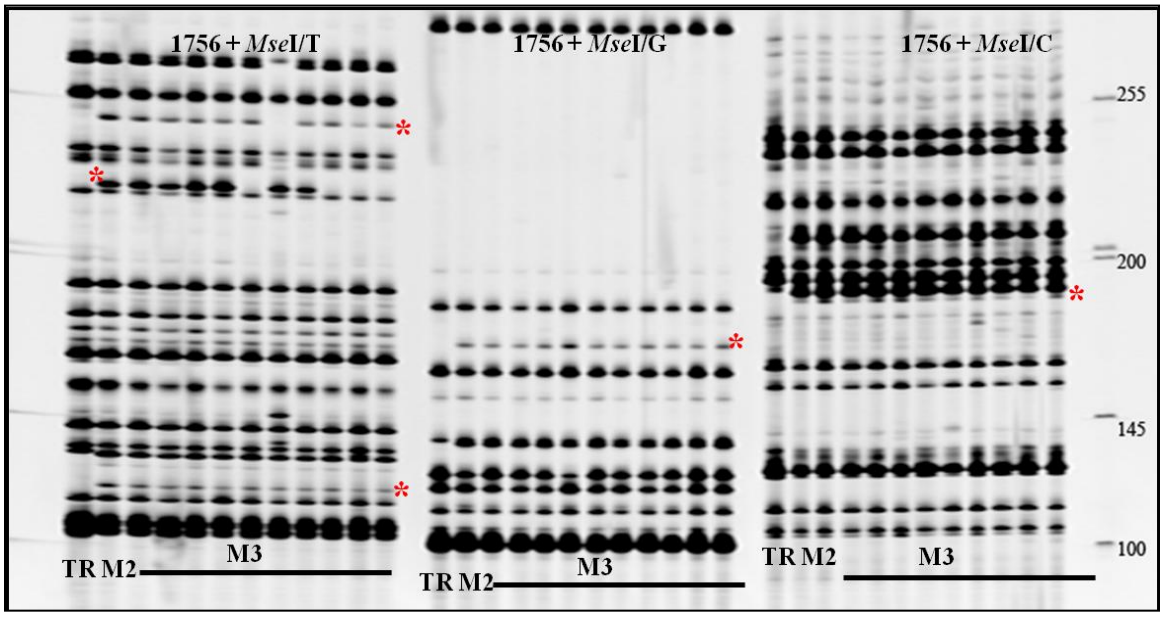

Figure 5.4- Partial TD profile of MITE transposition in the M_3 generation of M_2 plant 36.8 TR-'Tifrunner', M_2-M_2 generation, M_3-M_3 generation.

*New insertions.

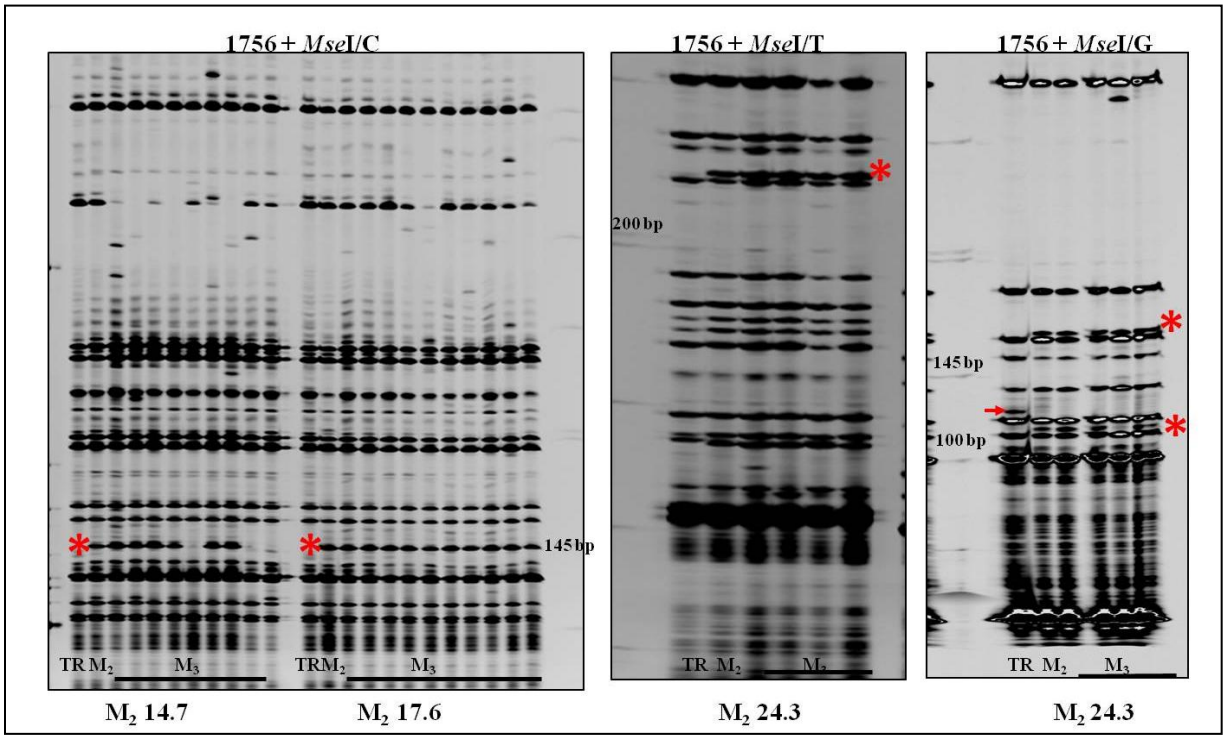

Figure 5.5-Partial TD profile of MITE transposition in the M_3 generation of M_2 plants 14.7, 17.6 and 24.3.

TR-'Tifrunner', M_2-M_2 generation, M_3-M_3 generation.

* New insertions

→ Deletions (missing bands).

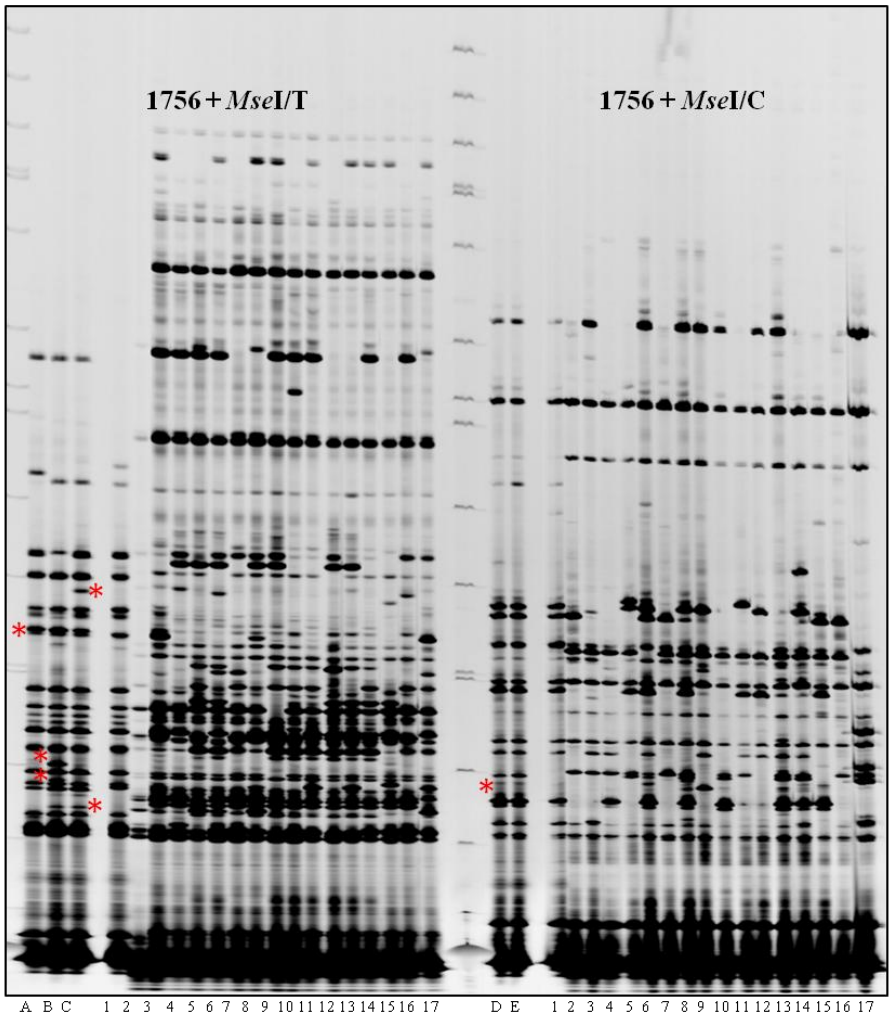

Figure 5.6- TD gel showing banding pattern of M_2 variants and different peanut agronomic types.

A, B, C, D, E – M_2 variant plants 24.3, 25.3, 36.8, 14.7 and 17.6 respectively.

1-Tifrunner, 2-Marc I, 3-AP4, 4-Carver, 5-McCloud, 6-GA01R, 7-Norden, 8-Fla07, 9-Tifguard, 10-Georgia green, 11-GA02C, 12-Georganic, 13-DP1, 14-York, 15-C99R, 16- Florunner, 17-Flfancy.

All 5 'Tifrunner' seed lots were identical in banding-one shown in gel above.

* Variant Bands shown by TD in EMS mutants.

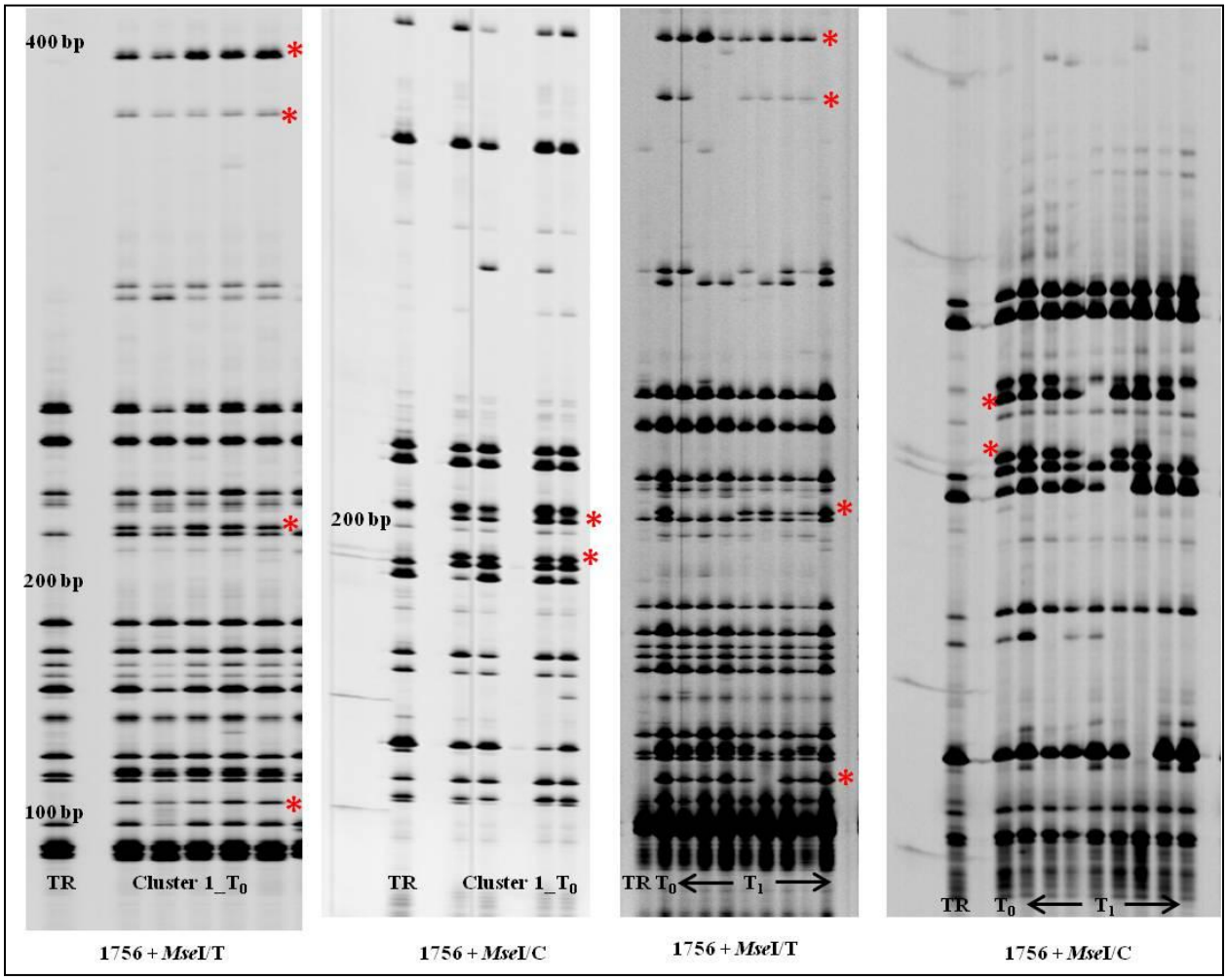

Figure 5.7 Partial TD profile of the T_0 and T_1 generations of tissue cultured 'Tifrunner', Cluster 1.

TR-'Tifrunner'

T_1 generation represented by plant number 3 of Cluster 1

* New insertions.

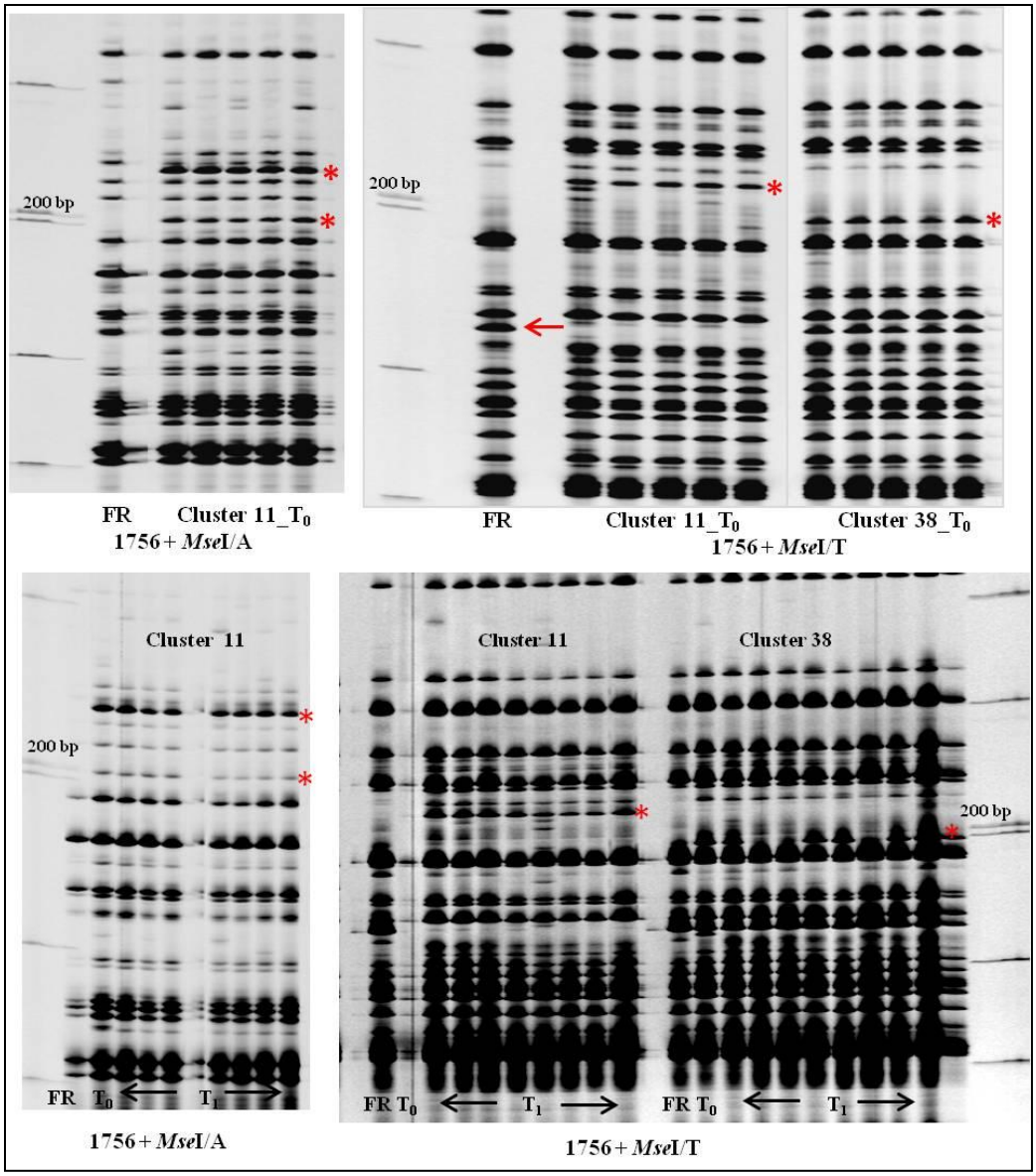

Figure 5.8- Partial TD profile of T_0 andT_1 generations of tissue cultured 'Florunner'.Clusters 11 and 38.

FR-'Tifrunner'.

T1 generation represented by Cluster 11_plant number 3 and Cluster 38_plant number 2

*New insertions

→ deletions / missing bands.

CHAPTER 6

Miniature Inverted Repeat Transposable Element Based Sequence Capture for Gene Tagging and Polymorphism Detection in Peanut (*Arachis hypogaea* L.)

Abstract

The feasibility of Miniature Inverted-repeat Transposable Element (MITE)-based sequence capture and high-throughput sequencing with the Illumina platform for marker development and gene tagging is demonstrated in this study. Sequencing of 96 'Tifrunner' clones by Sanger sequencing, prior to Illumina sequencing enabled development of 17 polymorphic markers as assayed across 29 peanut genotypes, and 12% of the Sanger reads showed similarity to protein coding regions. MITE capture and Illumina sequencing of five genotypes ('Tifrunner', 'Florunner', 'Marc I' and two ethyl methane sulfonate-generated variants) and alignment with known MITE-containing sequences revealed that reads greater than or equal to 100 counts from at least one genotype aligned with 797/840 of these. Visual inspection of alignments to 200 MITE-containing sequences identified potential polymorphism for 25% of the sequences among these five genotypes with respect to MITE insertion. MITE capture and next generation sequencing could provide a method to identify polymorphisms in a highly variable fraction of the *A. hypogaea* genome.

Chandran, M., Conner, J.A., and Ozias-Akins, P. To be submitted to The Plant Genome

Introduction

Transposons, mobile genetic elements, first discovered in maize (McClintock, 1951) are classified as Class I retrotransposons and Class II DNA transposons. Class I transposons move or transpose via RNA intermediates and Class II transposons move by excision and reinsertion of the elements. DNA transposons are characterized by short terminal inverted repeat (TIR) and can be autonomous or non-autonomous (Grzebelus *et al.*, 2009). Autonomous elements encode the transposase enzyme required for transposition. Non-autonomous DNA transposons, which are internal deletion derivatives of autonomous DNA transposons, transpose by a transposase encoded by a related autonomous transposon (Feschotte *et al.*, 2002). Computational analysis of DNA sequences led to the discovery of miniature inverted-repeat transposable elements (MITEs) which are different from other transposons by their short length (100-600 bp) and high copy number (~1000-15,000) per genome. MITEs are characterized by terminal inverted repeats (TIR), and show insertion preference to AT-rich regions of the genome creating 2-3 bp target site duplications (TSD) (Bureau and Wessler, 1992; Bureau and Wessler, 1994). Other characteristics of MITEs, which are of practical significance, are their insertion preference to genic regions and insertion polymorphism. The first active MITE described, *mPing* in rice, was shown to be associated with and transposed to genic regions (Jiang *et al*., 2003; Zhang *et al*., 2000). Similarly the *Heartbreaker* family of MITEs in maize was found to be inserted in low copy regions of the genome and was reported to be highly polymorphic between maize and teosinte lines (Casa *et al*., 2000). Insertion preference to genic regions and polymorphism make MITEs excellent tools for gene tagging and marker development.

Peanut (*Arachis hypogaea* L.), a major source of oil and protein, is a recent allotetraploid (AABB) genome of approximately 2800 Mbp. Crop improvement in peanut is aimed at

increasing yield, resistance to biotic and abiotic stress, oil quality improvement and allergen reduction. Application of molecular techniques for crop improvement will depend on identifying genes and biochemical pathways associated with a particular trait and markers associated with traits of interest. Linkage maps have been developed in peanut using RFLP (Kochert *et al*., 1991), SSR (Nagy *et al*., 2012), AFLP (He and Prakash, 1997) and transposon-based (Shirasawa *et al*., 2012b) markers. Transposon-based markers were developed with the peanut MITE, *AhMITE1*, identified in a high oleic acid mutant of peanut, Mycogen-Flavo. This high oleic mutant was developed from 'Florunner' using ethylmethane sulfonate (EMS) mutagenesis (U.S.patent 5948954). In this mutant, a MITE was found to be inserted into the coding region of a *FAD2* gene, resulting in a premature stop codon (Patel *et al.*, 2004). The Peanut MITE can also be activated by chemical and physical mutagens and tissue culture (Gowda *et al.*, 2011; Shirasawa *et al*., 2012a).

Since MITE insertions are polymorphic and new insertions may preferentially target genic regions, sequencing the MITE-containing and flanking genomic regions would be informative with regard to marker development and gene tagging. Identifying genes will also be important in annotating the sequences generated by the peanut genome initiative. In the current study MITE and flanking sequences in the peanut genome were pulled down from total DNA and sequenced using the Illumina platform. The feasibility of the technique was tested to identify MITE insertional polymorphisms in a panel of peanut cultivars, germplasm and variants created by EMS mutation for marker development and gene tagging.

Materials and methods

The peanut genotypes used for Illumina sequencing were 'Florunner' (Norden *et al*., 1969), 'Tifrunner' (Holbrook and Culbreath, 2007), 'Marc I' (Gorbet *et al*., 1992) and two variants,

25.3 and 36.8, identified by transposon display in the M_2 generation of EMS mutants developed from 'Tifrunner. 'Florunner' was selected for study since the peanut MITE, *AhMITE1*, was initially reported in the mutant 'Mycogen-Flavo', which was generated in the 'Florunner' background. This high oleic acid peanut mutant was obtained by EMS mutagenesis of 'Florunner' (U.S. patent 5948954). 'Tifrunner' is included as it was the genotype mutated with EMS to generate a TILLING (Targeting Induced Local Lesions IN Genomes) population (Knoll *et al*., 2011), is the reference genome for peanut, and also has 'Florunner' in its pedigree. Marc I is another cultivar that has 'Florunner' in its pedigree.

DNA extraction

For detecting MITE insertional polymorphisms, DNA was extracted from young unopened leaves using DNeasy 96 Plant Kit or DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA, USA) and quantified by Hoechst 33258 dye or Nanodrop (Thermo Scientific, Wilmington, DE, USA). Fluorescence was read with a FluoroCount (Packard/Perkin-Elmer, Waltham, MA, USA) microplate reader.

For Illumina library preparation, young unopened leaves were collected from the M_3 generation of EMS-generated mutant plants and from 'Florunner', 'Tifrunner' and 'Marc I' into individual Eppendorf tubes and stored at -80°C before DNA extraction. Leaves were frozen in liquid nitrogen and ground to a fine powder with blue pestles. DNA was extracted using Qiagen DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. Fluorometric quantitation of DNA was done using Picogreen dye (Invitrogen, Eugene, Oregon, USA) and fluorescence was read with a FluoroCount (Packard/Perkin-Elmer) microplate reader.

Library preparation for Illumina sequencing

Leaf genomic DNA was diluted to 100 ng/µl in low TE (10 mM Tris, pH 7.5 and 0.1 mM EDTA) and sheared with dsDNA fragmentase enzyme (M0348S, New England Biolabs, Ipswich, MA, USA) in a reaction volume of 100 µl containing 5 µg DNA, 1X fragmentase restriction buffer and 1X BSA according to the manufacturer's protocol. Reactions were stopped after 20 min by adding 0.5 µl of 0.5 M EDTA (ethylene diamine tetraacetic acid). Following shearing, DNA was cleaned using the Qiagen PCR purification kit. MITE- specific library preparation for Illumina sequencing was done according to the detailed protocol developed by Carrier *et al*. (2012). Briefly, DNA ends were repaired, 5' phosphorylated and dATP added to the 3' end.

Further adapter ligation (http://comailab.genomecenter.ucdavis.edu/index.php/Barcodes) to the DNA ends (Table 6.1) and size selection of 220-500 bp DNA fragments was done. MITE-containing fragments were separated from the remaining adapter-ligated genomic DNA by pull down with biotin-labeled MITE-specific capture oligos, each at 0.05 µM final concentration. The captured and eluted products were PCR amplified with adapter-specific primers (Table 6.2). MITE-specific capture oligos corresponded to the terminal inverted repeat region of MITE sequences and were derived after aligning 23 sequences, retrieved from BLAST searches to peanut expressed sequence tag (EST) databases showing similarity to the peanut MITE (*AhMITE1*) at E 10^{-20}(Figure 6.1). The 23 sequences showed greater consensus at the 5' TIR when aligned and the two capture oligos (Table 6.2) were designed to contain the 5' TIR. After two rounds of MITE pull down and enrichment by PCR, fragments from 220-500 bp were size selected and DNA was quantified by Picogreen dye (Invitrogen, Eugene, Oregon, USA) and with an Agilent bioanalyzer (Georgia Genomics Facility (GGF), University of Georgia, Athens) to

check for the required concentration of 10 nM for Illumina sequencing. Each library was prepared in triplicate with different adapter pairs (Table 6.1) and pooled in a single lane for PE-100 sequencing on an Illumina Hi-Seq at the Hudson-Alpha Institute of Biotechnology, Huntsville, Alabama.

Cloning and Sanger sequencing a 'Tifrunner' library sample prior to Illumina sequencing

Prior to Illumina sequencing, the feasibility of the technique was tested by sequencing 96 clones from the 'Tifrunner' library. DNA was cloned using Zero Blunt TOPO PCR Cloning kit and transformed into One Shot Topo 10 competent *E.coli* cells as per the manufacturer's instructions (K2875, Invitrogen, Carlsbad, CA, USA). Sequencing of 96 clones was done at the Georgia Genomics Facility, University of Georgia, Athens. Vector and adapter sequences were trimmed in Contig Express (Vector NTI). Sequences were screened manually for quality and those with multiple Ns were eliminated. Similarity search of the sequences to the peanut MITE or MITE capture oligo was done using BioEdit software (Hall, 2001). Primers specific to the MITE and flanking region of the cloned sequence were developed with PRIMER 3 (Rozen and Skaletsky, 2000) and Vector NTI software (Table 6.3). Insertional polymorphism was tested in 29 peanut genotypes including the suggested progenitors (*A. ipaensis, A. duranensis*), *A. hirsuta* (PI 576638), as well as peanut breeding lines, and Runner, Valencia and Spanish types. The polymerase chain reaction was carried out in a 15 µl volume containing 15 ng leaf genomic DNA, 1X PCR buffer, 0.2 mM dNTP, 0.2 mM each of primers, 1% PVP (polyvinyl pyrolidine), 0.15 µg BSA (bovine serum albumin), 2.5% DMSO (dimethyl sulfoxide) and 0.12 µl Jump Start Taq.

Processing Illumina sequencing reads

For this study adapter sequences were trimmed, sequence quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and the MITE sequences were removed by RepeatMasker (http://www.repeatmasker.org). The non-masked flanking sequences from each library were aligned using Bowtie2 (Langmead, 2010) to peanut MITE-containing sequences from the NCBI GSS database generated as described in Shirasawa *et al*., (2012b). Bowtie2 parameters used for alignment were sensitive-local preset with score-min default and –N=1. *In silico* polymorphism between the five libraries was analyzed for 200 MITE-containing contigs by manual inspection of alignments using the visualizing tool, Integrated Genomics Viewer (Thorvaldso'ttir *et al*., 2013).

Results

Cloning and Sanger sequencing from the 'Tifrunner' MITE-specific library

Of the 96 fragments cloned and sequenced, 70 (72.9%) showed good sequence quality. Similarity searches to the MITE and MITE-specific capture oligo in BioEdit software identified 63 sequences out of 70 (90%) that had MITE sequences. The length of the MITE sequences varied from 57-208 bp and the length of the flanking sequences varied from 5-506 bp (Figure 6.2). Terminal inverted repeat sequences found in the MITE sequences in 'Tifrunner' are shown in Table 6.4. Six sequences had the full MITE sequence and the flanking sequences ranged from 48-151 bp. Terminal inverted repeats found in full-length MITE sequences are given in Table 6.5 Similarity searches of the 63 flanking sequences to ESTs in NCBI by BLASTn at an E-value cut-off of 10^{-6} identified 22% of the sequences (12/63) that showed significant similarity to ESTs (Table 6.6). Hits to ESTs were obtained with sequences having 74-506 bp of flanking sequence and 12.7% (8/63) of ESTs (obtained by BLASTn of flanking sequences) showed similarity to

protein sequences in BLASTx (Table 6.7). Seventeen out of 29 (58%) primer pairs, derived from 'Tifrunner' Sanger sequence reads, amplified polymorphic bands from 29 peanut genotypes (Figure 6.3). Ten primer pairs amplified monomorphic bands and two primers did not show robust amplification. Of the 27 primers that showed PCR amplification, only 2 primers showed amplification in *A. duranensis* (1 monomorphic and 1 polymorphic). Of the 8 primers that amplified MITE inclusive bands in *A. ipaensis,* three were polymorphic.

Illumina sequencing

The quality of raw Illumina reads was checked with FastQC (Table 6.8). Post quality control, there were 40,351,402 'Floruuner'sequences, 46,729,206 'Marc I' sequences, 58,974,170 'Tifrunner' sequences, 35,011,458 sequences of mutant 25.3 and 40,449,824 sequences of mutant 36.8 from the three sets of libraries for each genotype. The percentage of sequences masked in the 15 libraries by RepeatMasker ranged from 45-59%. MITE-flanking sequences (read count $\geq$ 100) from at least one of the five genotypes aligned to 797 of the 840 peanut MITE-containing contigs in NCBI's GSS database. Of the 200 transposon-containing contigs, from DDBJ, analyzed for *in silico* insertional polymorphism, among the five Illumina libraries, fifty were polymorphic. Visual inspection of alignments to 200 of these MITE-containing sequences identified potential polymorphism for 25% of the sequences (Figure 6.4).

Discussion

The feasibility of MITE capture and sequencing for targeting genic regions of the genome and polymorphism analysis was demonstrated initially by cloning and Sanger sequencing from a 'Tifrunner' library. Efficiency of MITE sequence capture with biotin-labeled oligos was demonstrated in this study since 90% of reads contained MITE sequence with flanking regions of variable length. Gene identification with MITE flanking sequences was also possible as 12.7% of

flanking sequences from ’Tifrunner’ clones showed identity to proteins after BLASTn of flanking sequences to ESTs, followed by BLASTx of ESTs. MITE insertion to genic regions correlates with studies in other species such as rice and peas (Jiang *et al*., 2003; Macas *et al*., 2005; Yang *et al*., 2001). Insertional polymorphism, as shown by PCR with primers specific to the MITE and flanking regions from ‘Tifrunner’ Sanger reads, indicates the potential of developing MITE-specific markers that show polymorphism among elite genotypes. The high frequency of MITE insertional polymorphism observed is similar to that seen with other MITEs as in barley (Chang *et al*., 2001; Lyons *et al*., 2008), rice (Park *et al*., 2003) and wheat (Yaakov *et al*., 2012). Genic insertions and polymorphism of MITE insertions shown by Sanger sequences, suggests the possibility of applying the technique of MITE-specific sequencing by a high-throughput Illumina platform. While Illumina sequencing was underway in this work, a study on MITE-based sequence capture and sequencing in peanut reported 11.5% of sequences to be near genes (Shirasawa *et al*., 2012a) and MITE-based marker polymorphism of 31% among ‘Satonoka’ and ‘Kintoki’ genotypes and 11% between ‘Nakateyutaka’ and YI-0311. We observed a similar level of i*n silico* polymorphism among the five libraries in our initial analysis of Illumina sequence data, showing the potential for development of multiple MITE-based markers, which can be used to develop a peanut high density genetic map for molecular breeding approaches. It has to be noted that some of the markers developed based on the present assembly overlap with earlier identified transposon markers (Shirasawa *et al.*, 2012b), although we expect that this sequence information also will generate novel markers. Further analysis of these sequences can be undertaken once diploid and tetraploid genome sequences are available for *Arachis*.

Conclusion

The feasibility of targeted sequence capture for Illumina sequencing with regard to peanut MITE, *AhMITE1*, is shown in this study. Capture of MITE specific fragments followed by high through put sequencing has generated high volume of *in silico* data which shows promising results in preliminary analysis. This data set can be utilized for marker development, gene and SNP identification, for creating high-density genetic maps and gene identification in peanut.

References

Bureau T.E., Wessler S.R. (1992) *TOURIST* - A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283-1294.

Bureau T.E., Wessler S.R. (1994) *STOWAWAY* - A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6:907-916.

Casa A.M., Brouwer C., Nagel A., Wang L.J., Zhang Q., Kresovich S., Wessler S.R. (2000) The MITE family *Heartbreaker* (*Hbr*): Molecular markers in maize. Proc. Nat. Acad. Sci. USA 97:10083-10089.

Chang R.Y., O'Donoughue L.S., Bureau T.E. (2001) Inter-MITE polymorphisms (IMP): a high throughput transposon-based genome mapping and fingerprinting approach. Theor. Appl. Genet. 102:773-781.

Feschotte C., Jiang N., Wessler S.R. (2002) Plant transposable elements: Where genetics meets genomics. Nat. Rev. Genet. 3:329-341.

Gorbet D.W., Knauft D.A., Norden A.J. (1992) Registration of Marc-I peanut. Crop Sci. 32:279-279.

Gowda M.V.C., Bhat R.S., Sujay V., Kusuma P., Varshakumari, Bhat S., Varshney R.K. (2011) Characterization of *AhMITE1* transposition and its association with the mutational and evolutionary origin of botanical types in peanut (*Arachis spp*.). Plant Syst. Evol. 291:153-158.

Grzebelus D., Gladysz M., Macko-Podgorni A., Gambin T., Golis B., Rakoczy R., Gambin A. (2009) Population dynamics of Miniature Inverted-repeat Transposable Elements (MITEs) in *Medicago truncatula*. Gene 448:214-220.

Hall T. (2001) BioEdit version 5.0.6. Available at: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html/>.

He G.H., Prakash C.S. (1997) Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). Euphytica 97:143-149.

Holbrook C.C., Culbreath A.K. (2007) Registration of 'Tifrunner' Peanut. Journal of Plant Registrations 1:124-124.

Jiang N., Bao Z.R., Zhang X.Y., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. (2003) An active DNA transposon family in rice. Nature 421:163-167.

Knoll J.E., Ramos M.L., Zeng Y., Holbrook C.C., Chow M., Chen S., Maleki S., Bhattacharya A., Ozias-Akins P. (2011) TILLING for allergen reduction and improvement of quality traits in peanut (*Arachis hypogaea* L.). BMC Plant Biology 11:81

Kochert G., Halward T., Branch W.D., Simpson C.E. (1991) RFLP variability in peanut (*Arachis hypogaea*) cultivars and wild-species. Theor. Appl. Genet. 81:565-570.

Lyons M., Cardle L., Rostoks N., Waugh R., Flavell A.J. (2008) Isolation, analysis and marker utility of novel Miniature Inverted repeat Transposable Elements from the barley genome. Mol. Genet. Genomics 280:275-285.

Macas J., Koblizkova A., Neumann P. (2005) Characterization of *Stowaway* MITEs in pea (*Pisum sativum* L.) and identification of their potential master elements. Genome 48:831-839.

McClintock B. (1951) Chromosome organization and genic expression. Cold Spring Harbor Symposia on Quantitative Biology 16:13-47.

Nagy E.D., Guo Y.F., Tang S.X., Bowers J.E., Okashah R.A., Taylor C.A., Zhang D., Khanal S., Heesacker A.F., Khalilian N., Farmer A.D., Carrasquilla-Garcia N., Penmetsa R.V., Cook D., Stalker H.T., Nielsen N., Ozias-Akins P., Knapp S.J. (2012) A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13:469

Norden A.J., Lipscomb R.W., Carver W.A. (1969) Registration of 'Florunner' peanuts. Crop Sci. 9:850

Park K.C., Kim N.H., Cho Y.S., Kang K.H., Lee J.K., Kim N.S. (2003) Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. Theor. Appl. Genet. 107:203-209.

Patel M., Jung S., Moore K., Powell G., Ainsworth C., Abbott A. (2004) High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theor. Appl. Genet.108:1492-1502.

Rozen S., Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132:365-386.

Shirasawa K., Hirakawa H., Tabata S., Hasegawa M., Kiyoshima H., Suzuki S., Sasamoto S., Watanabe A., Fujishiro T., Isobe S. (2012a) Characterization of active Miniature Inverted-repeat Transposable Elements in the peanut genome. Theor. Appl. Genet. 124:1429-1438.

Shirasawa K., Koilkonda P., Aoki K., Hirakawa H., Tabata S., Watanabe M., Hasegawa M., Kiyoshima H., Suzuki S., Kuwata C., Naito Y., Kuboyama T., Nakaya A., Sasamoto S., Watanabe A., Kato M., Kawashima K., Kishida Y., Kohara M., Kurabayashi A., Takahashi C., Tsuruoka H., Wada T., Isobe S. (2012b) *In silico* polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol. 12: 80

Thorvaldso'ttir H., Robinson J.T., Mesirov J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. 14:178-192.

Yaakov B., Ceylan E., Domb K., Kashkush K. (2012) Marker utility of Miniature Inverted-repeat Transposable Elements for wheat biodiversity and evolution. Theor. Appl. Genet.124:1365-1373

Yang G., Dong J., Chandrasekharan M.B., Hall T.C. (2001) *Kiddo*, a new transposable element family closely associated with rice genes. Mol. Genet. Genomics 266:417-424.

Zhang Q., Arbuckle J., Wessler S.R. (2000) Recent, extensive, and preferential insertion of members of the Miniature Inverted-repeat Transposable Element family *Heartbreaker* into genic regions of maize. Proc. Nat. Acad. Sci. USA 97:1160-1165.

Table 6.1 List of adapter sequences used to prepare Illumina sequencing libraries

Library			Barcoded adapter sequence
†Set I	Florunner	Adap A	*P-aaatcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgattttT
	MarcI	Adap A	P-aacatAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTatgttT
	Tifrunner	Adap A	P-aaacaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgtttT
	25.3	Adap A	P-actaaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttagtT
	36.8	Adap A	P-actcgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcgagtT
†Set II	Florunner	Adap A	P-cgcgtAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTacgcgT
	MarcI	Adap A	P-ctgccAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTggcagT
	Tifrunner	Adap A	P-tgtgcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgcacaT
	25.3	Adap A	P-gtctgAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcagacT
	36.8	Adap A	P-tcggaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTtccgaT
†Set III	Florunner	Adap A	P-caaaaAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCT ttttgT
	MarcI	Adap A	P-cgcacAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgtgcgT
	Tifrunner	Adap A	P-gtttcAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTgaaacT
	25.3	Adap A	P-gacttAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTaagtcT
	36.8	Adap A	P-tgaagAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
		Adap B	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcttcaT

†Each library was constructed in triplicate. Barcodes in the adapter sequences are shown in lower case red. *- phosphorylated adapters.

(http://comailab.genomecenter.ucdavis.edu/index.php/Barcodes)

Table 6.2 Oligos used for Illumina library preparation

Oligo number	Description	Sequence (5'→3')
2060	MITE_-specific biotinylated oligo	GGTGGATACTACAATGAAGATGGCATA ATTGTCTTCATATGAG
2061	MITE_-specific biotinylated oligo	GGTGGGTACTCCCATGAAGATRTTATAA TTRTCTTCATGTGRT
2150	Primer specific to adapter_A of illumina library	CAAGCAGAAGACGGCATACGACATCGG T CTCGGCATTCCTGCTGAACCGCTCTTCC GATCT
2151	Primer specific to adapter_B of illumina library	AATGATACGGCGGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCG ATCT

Table 6.3 Primers based on 'Tifrunner' Sanger sequence reads to test for MITE insertional polymorphisms‡.

Clone ID	Forward Primers 5'→3'	Reverse Primers 5'→3'
A01	TCTAGATATATAGGCTGACTGG	ATGATGGATTGTATGGTTAG
A03	TCCATCATCCAAAGGTCAAA	TTAGGTAGGTGTAGGGTAGAGTTTAAG
A09	AAATCTAACCCTACACTCC	TAGGGAGAGAGTGCGC
B01	TTCAAAAACAGGGTCCCAAG	ACATATCAAAATCTAACCCTACACTCC
B05	TGCACTTCCAATGCAACTTT	GGTGGGTACTCCCATGAAGA
B06	CCAAACACATGGGCATAACA	TGATTTGGCCACACTTCAAA
B11	GCCCATAATTTGGCGATGTA	CACTCCATCATCTAAGGGTCAA
C02	CATACTTGTGAAGGATTTGAATGTG	CAATCCATCATCCAAGGGTAA
C09	GCCATCTTCATTGTA	GGATTTGGTGACCCT
D01	TGGCTTACTTGAAGTT	CATATGAAGACAATTATGC
D04	CACTCCATCATCTAAGGGTCAA	TACAATGCACTATATATACAA
D06	TCATGAGAAGATATTTTTGCCATC	TGTGGAAGAGTAAGAAAAGAAATTGA
D07	TGGAAACATTGATTCATCACCT	TGATTTGGCCACACTTCAAA
E02	AAAGCGTAGCCGTATGTTAATATCT	TGATTTGGCCACACTTCAAA
E07	CACCTGATTTCGCTT	GAAGATGCTTTGGTT
F03	TTGACCTTTGGATGATGGA	CGTGCCTCCAACAAAAATCT
F06	ATGAAGATGCTTTGG	GGAAAAGTATAGGTAAC
F09	GAGATTTCGCACAGATGG	GCTACTCAAATGAAGATGCA
F10	AACTACAAAACTATGTTAACCATCT	ATTGATTTGACCACACTTCAAA
F12	AAGAAGATGAAGAAGGAGGAGGA	AAACCACACTTTTAACCAAAGCA
G03	GCCTAGGCTTTCCCTCCTAC	CAATCCATCATCTAAGGGTTAAAAA
G06	GCCTCAACAAAGAGGCTGAC	CCATCTTCATTGTAGTATCCACC
G07	CCTTCCAATCACAAGACCTCA	TGATTTGGCCACACTTCAAA
G12	TTCCTTCTTTGCTGCCTGTT	TTGGCCACACTTCAAATAAAAA
H06	TGAGCATTCAACAATCAGCA	CTCTACAATCCATCATCTAAAGGTCA
H09	TACAATGTTGAGTCG	ATGATGGAGTGTAGG
H10	ATGAAGATGCTTTGG	TGGACTGAAACACAT

‡ Primers to test for MITE insertional polymorphism were developed from Sanger sequence reads corresponding to MITE and flanking sequences

Table 6.4 Terminal inverted repeat sequences of MITEs from Sanger sequence reads

5' Terminal inverted repeat sequences (5'→3')	3' Terminal inverted repeat sequences (5'→3')
GGTGGATACTACAATGAAGATGGCA	TAATATCTTCATGGGAGTACCCACC
GGTGGATACTACAATGAAGATCTTA	TAATATCTTCATGGGAGTATCCACC
GGTGGATACTACAATGAAGATTTTA	TAATATCTTCATGGGAGNACCCACC
GGTGGATACTATAATGAAGATGGCA	TGCCATCTTCATTGTAGTATCCACC
GGTGGATACTCCCATGAAGATATTA	TAAAGTCTTCATTGTAGTATCCACC
GGTGGGTACTCCCATGAAGATATTA	TACCATCTTCATTGTAGTATCCACC
GGTGGGTACTCCCATGAAGATGTTA	TTGCATCTTCATTTGAGTAGCTCCC
GGTGGGTACTCCAATGAAGATATTA	
GGGAGCTACTCAAATGAAGATGTAA	
GGGAGCTACTCAAATGAAGATGCAA	

Table 6.5 Terminal inverted repeat sequences of full-length MITEs from Sanger sequence reads.

Terminal inverted repeat sequences (5'→3')	3' Terminal inverted repeat sequences (5'→3')
GGGAGCTACTCAAATGAAGATGCAA	TAATATCTTCATGGGAGTATCCACC
GGGAGCTACTCAAATGAAGATGTAA	TAATATCTTCATGGGAGTACCCACC
GGTGGGTACTCCCATGAAGATATTA	TTGCATCTTCATTTGAGTAGCTCCC

Table 6.6 BLASTn of Tifrunner flanking sequences

Tifrunner clone ID	Flanking sequence length (bp)	BLASTn accession†	BLASTn query coverage (%)	E-value
A03	92	GW955611.1	50	8.00E-24
B05	74	GO329266.1	62	7.00E-12
C02	131	FS974606	48	4.00E-15
C06	129	GO268239.1	46	2.00E-07
C08	106	JK166240.1	64	9.00E-74
D01	268	FC887668.1	62	2.00E-38
D08	506	EG432237.1	71	1.00E-173
E02	133	GO268758.1	78	4.00E-15
F06	114	GW986985.1	38	5.00E-07
G06	254	GO268377.1	50	7.00E-12
G07	169	BW658787.1	64	9.00E-30
H06	122	GO329266.1	76	2.00E-37

† BLASTn hit from similarity searches of MITE flanking sequences from Sanger sequence reads of 'Tifrunner'clones against the EST database in NCBI at E-value cut off at $10^{-6.}$

Table 6.7 BLASTx of EST hits from BLASTn of 'Tifrunner' flanking sequences‡

Tifrunner Clone ID	BLASTn accession number	BLASTx accession number	Query coverage (%)	E value	Description
A03	GW955611.1	XP_003517862.1	68	1.00E-39	Metalloendoproteinase 1 -like-*Glycine max*
B05	GO329266.1	XP_003545067.1	52	6.00E-28	SNF1-related protein kinase regulatory subunit beta-2-like -*Glycine max*
C06	GO268239.1	XP_003550891.1	50	5.00E-29	Uncharacterized protein LOC100786268-*Glycine max*
C08	JK166240.1	XP_003605478.1	86	2.00E-53	Triterpene UDP-glucosyl transferase UGT1- *Medicago truncatula*
D01	FC887668.1	XP_002314701.1	76	6.00E-127	Predicted protein-*Populus trichocarpa*
D08	EG432237.1	NP_187312.1	44	4.00E-56	Uncharacterized protein-*Arabidopsis thaliana*
F06	GW986985.1	XP_002265831.2	56	6.00E-21	Adenosine deaminase -like protein-*Vitis vinifera*
G07	BW658787.1	XP_003524110.1	56	5.00E-58	Probable WRKY transcription factor 28-like-*Glycine max*

‡Accessions showing similarity in BLASTn of MITE flanking sequences (Table 6.6) were used to query non-redundant protein sequences in NCBI by BLASTx at E-value cut off of 10^{-6}

Table 6.8 FastQC results of raw Illumina sequence reads

FastQC Parameters	Library quality
Basic statistics	Pass
Per base sequence quality	Fail
Per sequence quality scores	Pass
Per base sequence content	Fail
Per base GC content	Fail
Per sequence GC content	Fail
Per base N content	Pass
Sequence Length Distribution	Pass
Sequence Duplication Levels	Fail
Overrepresented sequences	Warning
Kmer Content	Fail

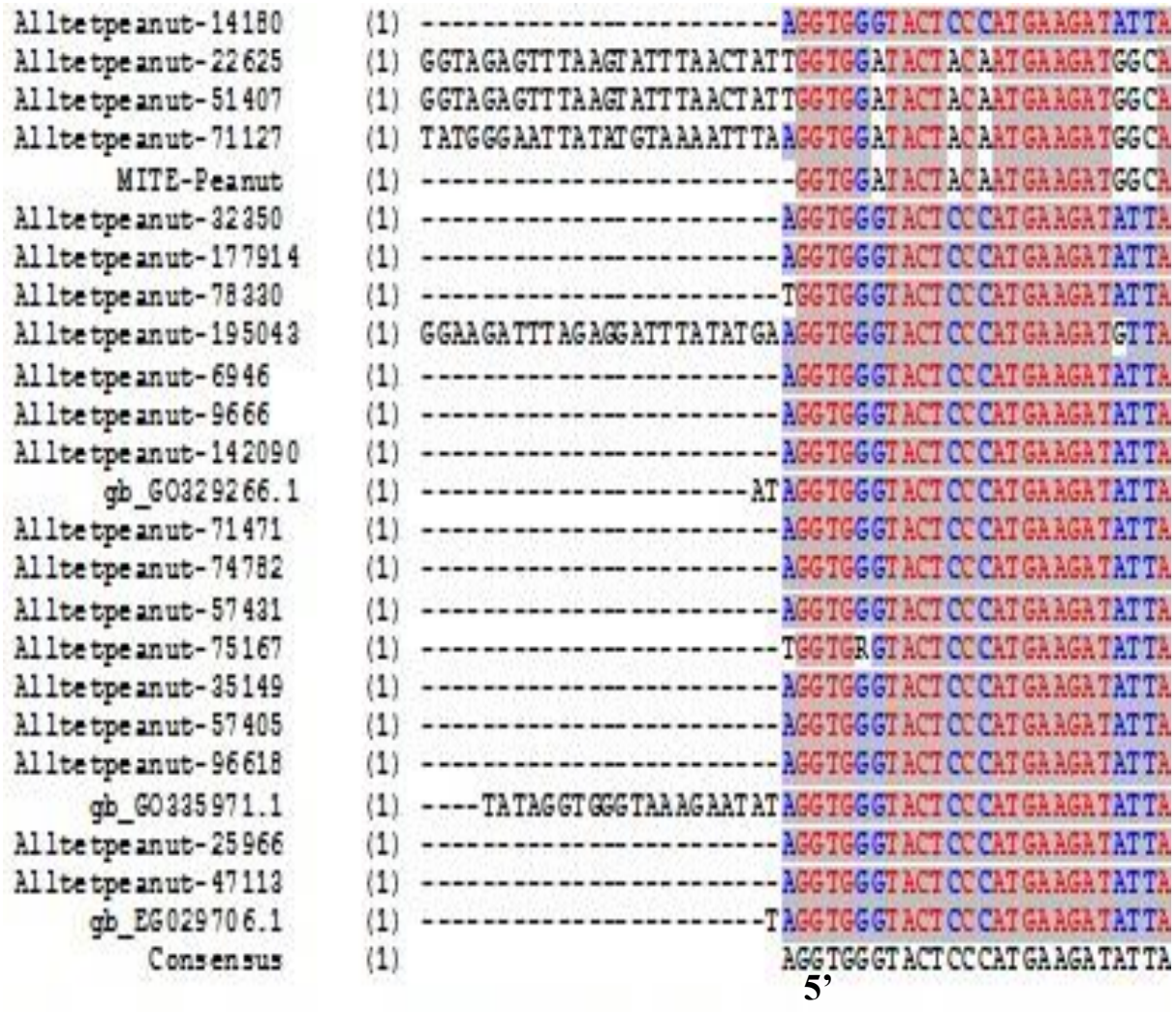

Figure 6.1 Developing Capture Oligos

Alignment continued in next page

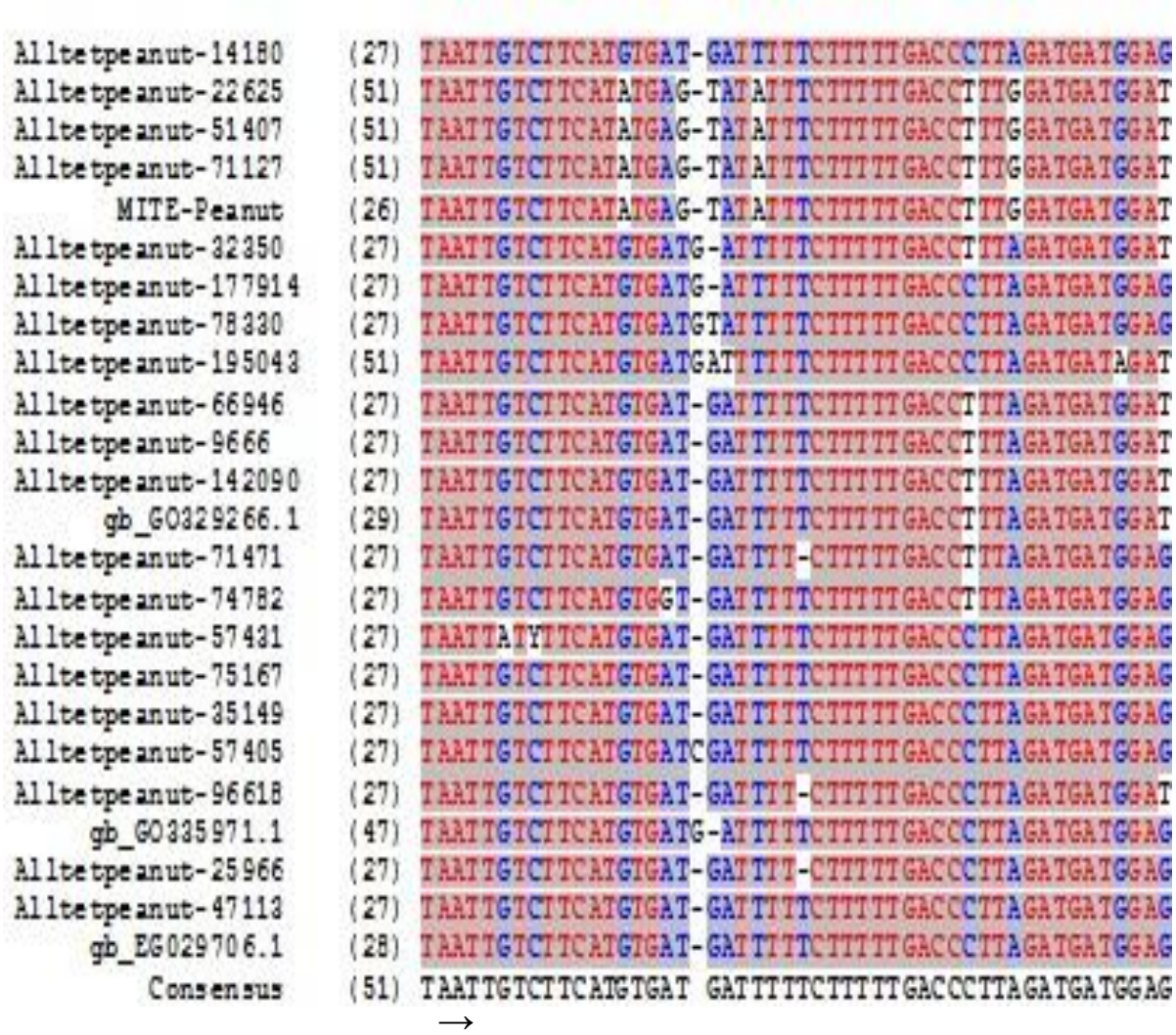

Figure 6.1 Developing Capture Oligos

Alignment of 23 MITE specific ESTs and peanut MITE used for developing MITE capture oligos (#2060 and 2061) for Illumina sequencing (partial alignment shown). The oligos are 43 bp long and developed containing the 5' region showing consensus.

The oligo # 2060 was designed grouping Alltetpeanut sequences c22625, c51407, c71127 & peanut MITE (Oligo #2060 5'-GGTGGATACTACAATGAAGATGGCATA ATTGTCTTCATATGAG-3')

Oligo # 2061 was designed grouping the remaining sequences

(Oligo #2061 5'- GGTGGGTACTCCCATGAAGATRTTATAATTRTCTTCATGTGRT-3')

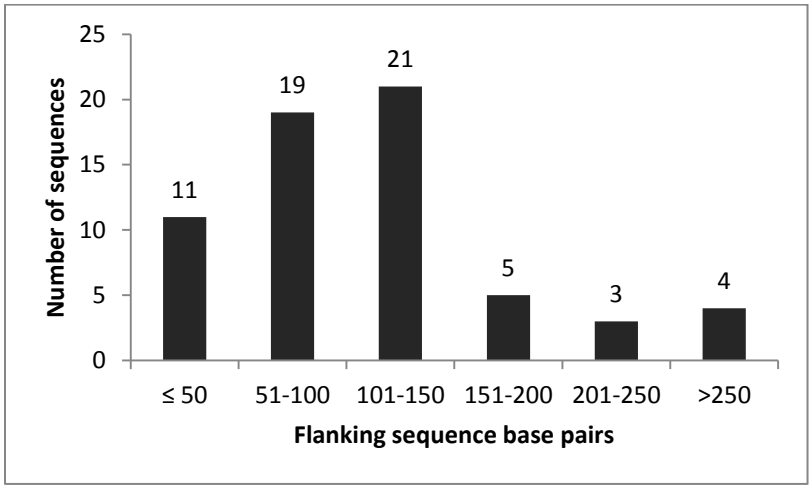

Figure 6.2 Distribution of flanking sequence lengths *vs* number of sequences

Ninety-six clones of 'Tifrunner' were Sanger sequenced. The graph represents flanking sequences of MITE-containing sequences from 63 clones.

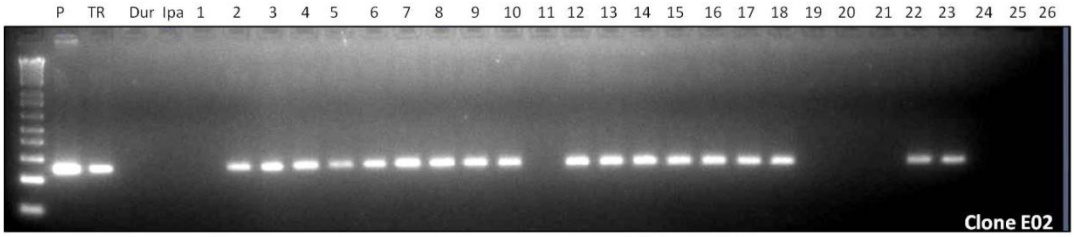

Figure 6.3 MITE insertional polymorphism

Primers were derived from 'Tifrunner' Sanger sequence reads corresponding to the MITE and flanking sequence. 10 µl of the PCR product was run on 1.5% agarose gel for 40 min at 140V and viewed under UV light after ethidium bromide staining. Clone E02 represents the sequence from which the primers corresponding to MITE and flanking regions are developed as shown in Table 6.3.

P	Plasmid	5	Carver	13	Georganic	21	NC3033
TR	Tifrunner	6	McCloud	14	DP1	22	SPT06
Dur	*A. duraensis*	7	GA01R	15	C99R	23	Bailey H/O
Ipa	*A. ipaensis*	8	Norden	16	York	24	Olin
1	*A. hirsuta*	9	Fla07	17	Gregory	25	PI203395
2	Marc I	10	Tifguard	18	Florunner	26	PI203396
3	AP4	11	Georgia Green	19	N.Mexico Valencia A		
4	FlFancy	12	GA02C	20	C76-16		

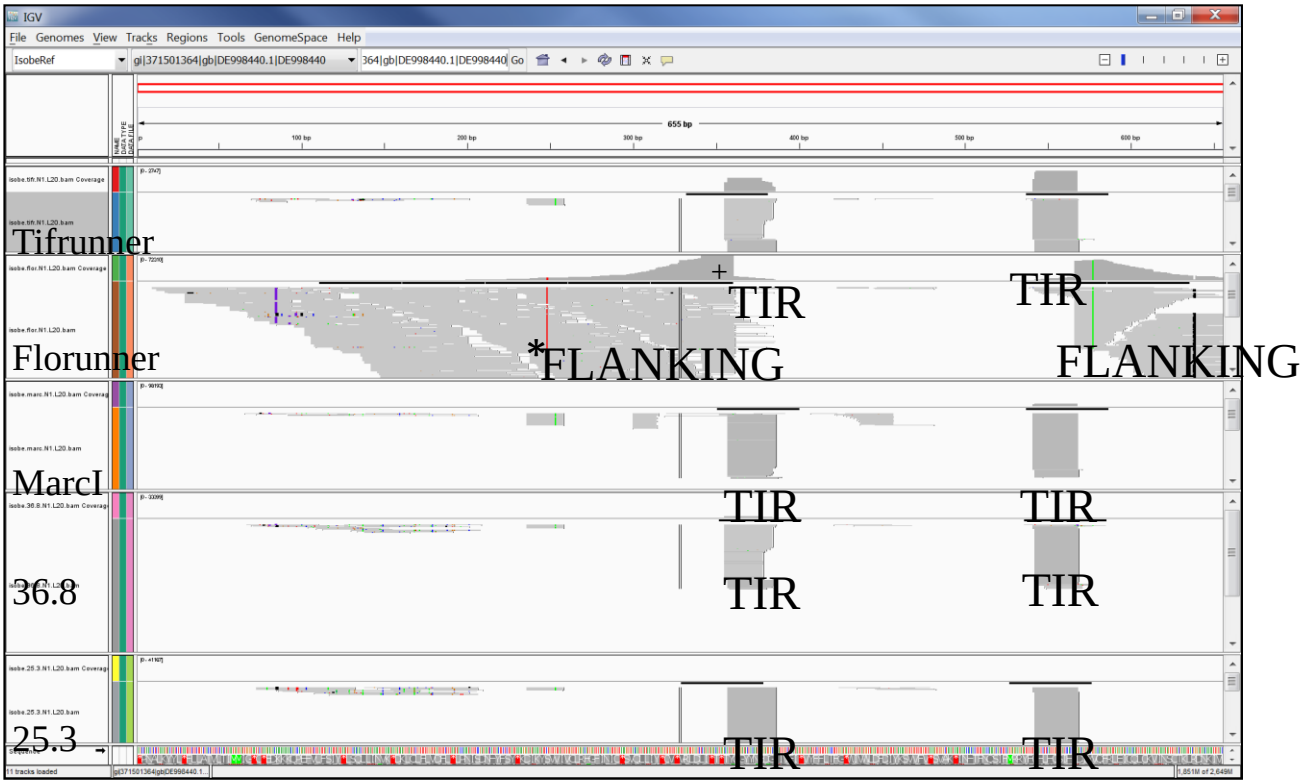

Figure 6.4 Integrated Genomics Viewer

Visualization of sequences from five peanut libraries in Integrated Genomics Viewer compared against MITE specific contigs. In figure polymorphism is identified between 'Florunner' and other cultivars studied. + Terminal inverted repeat, * Flanking sequences.

CONCLUSION

Peanut (*Arachis hypogaea* L.) seeds are a rich and economical source of protein, with 25% protein and 50% oil. The species is well adapted to the semi-arid environments. Although a rich and economical source of protein, peanut seed proteins are found to cause allergic reactions in humans. Peanut allergy is a major health concern in developing countries and in the United States alone 0.8% of young children and 0.6% of adults are affected by peanut allergy (Sampson, 2004). Developing hypo-allergenic cultivars, coupled with favorable agronomic traits, disease resistance, resistance to abiotic stress and product quality, is one of the objectives of peanut breeding.

The major peanut allergens are Ara h 1, Ara h 3, Ara h 2 and Ara h 6. Ara h 2 and Ara h 6 are considered as potent allergens, though they are found in lesser proportions in seed protein compared to Ara h 1 and Ara h 3. The potential for breeding hypo-allergenic peanut varieties is limited, as variability available in the germplasm collections is limited in this regard. Application of reverse genetics techniques has been attempted and application of RNA interference (RNAi) was successful in developing lines with reduced allergens. The line B11.1.1/11, developed by RNAi shows complete absence of Ara h 2 and reduced levels of Ara h 6 (Chu *et al.*, 2008). In any transgenic line developed, the transgene introduced should be stably expressed across multiple generations, in multiple and varied environments. In the current study the stability of transgene expression in B11.1.1/11 was studied across three generations (T_3, T_4 and T_5) under field conditions in a single year trial. Expression of Ara h 2 and Ara h 6 proteins were tested by quantitative Western blot and was found to be stably inherited across the three transgenic

generations with no statistically significant collateral changes in Ara h 1 and Ara h 3. This study has demonstrated the stability of protein silencing due to a transgene in three peanut generations. Seeds alter their protein profile in relation to nutrient availability or stress conditions to which they are subjected. The allergens, Ara h 2 and Ara h 6 are 2S albumins and have relatively higher sulfur amino acid, cysteine and methionine, content compared to Ara h 1 and Ara h 3. Sulfur is a limiting nutrient in soils and sulfur fertilization is found to increase crop yields. Sulfur nutrition to plants is also shown to increase the proportion of sulfur-containing proteins in seeds with a corresponding decrease of sulfur-poor proteins (Gayler and Sykes, 1985). The effect of three concentrations of soil sulfur (low-0.012 mM, medium-0.3 mM and high-3 mM) on Ara h 2 and Ara h 6 expression in transgenic peanut line B11.1.1/11, was studied as both are sulfur rich proteins and Ara h 6 is only partially silenced in the transgenic line. Expression of Ara h 2 and Ara h 6 tested by quantitative Western blotting showed that the transgene was stably expressed under various sulfur levels with the transgenic plants showing no Ara h 2 and very low levels of Ara h 6. The Arah 6 levels of seeds from transgenic plants grown under the three sulfur levels were significantly lower than in seeds from non-transgenic plants, thus emphasizing the stability of transgene expression under variable sulfur nutrition and an interaction between sulfur and genotype. No collateral changes in Ara h 1 were seen under different sulfur levels and between transgenic or non-transgenic plants. No genotypic effect of sulfur was observed among the Ara h 3 levels, but plants grown under low sulfur produced seeds with significantly less Ara h 3 than at the two higher levels.

Crop improvement in peanut to develop varieties with higher yield, better agronomic traits, biotic and abiotic stresses depends on the identification of genes that affect a trait or different biochemical pathways and in developing markers to expedite breeding. Presently, the most

widely used markers in peanut are simple sequence repeat (SSR) markers, although the utility of transposon-based markers recently has been demonstrated (Shirasawa *et al*., 2012). The first transposon in peanut was identified in a high oleic acid peanut mutant, developed from 'Florunner' by ethylmethane sulfonate (EMS) mutation (Patel *et al*., 2004). This transposon belongs to the Miniature Inverted-repeat Tranasposable Element (MITE) class. MITEs are common in the grass family, high in copy number, small in size, show preferential insertion to genic regions and are polymorphic across genotypes (Casa *et al*., 2000). Since MITEs are polymorphic across genotypes, the possibility of targeted sequencing in peanut with respect to peanut MITEs was studied with the aim to develop MITE specific markers in the genic regions of the peanut genome. The present study on capture of MITE-containing sequences from genomic DNA followed by Sanger sequencing has demonstrated the feasibility of this technique for marker development in the vicinity of genes. From 63 Sanger reads containing MITE and flanking DNA, 17 markers, which are polymorphic among 29 peanut genotypes, have been identified. Also 12.7% (8/63) of flanking sequences showed similarity to protein sequences by BLAST analysis. These results were promising enough to do Illumina sequencing of MITE capture libraries in five genotypes namely 'Tifrunner', 'Florunner', 'Marc I', and two EMS mutants. *In silico* analysis of Illumina sequences aligned to transposon sequences has identified polymorphism among the five libraries with respect to MITE insertion.

Since transposons are activated by stress to which the plants are subjected, the transposition of peanut MITE, *AhMITE1*, under two stress situations namely, EMS mutation and prolonged tissue-culture was studied. The effect of tissue-culture was studied in 'Tifrunner' and three component lines of 'Florunner', and the effect of EMS was tested in M_2 plants grown from EMS mutated seeds of 'Tifrunner'. The heritable mutation frequency in 'Tifrunner' under EMS

mutation was 1.24% and in tissue culture was 6.25%. Mutation frequency in tissue-cultured 'Florunner' was 10.5%. Among the three component lines of 'Florunner' only F439-16-10-3-2 showed heritable variation. Alhough mutations were observed under the two situations studied; potential genotypic differences were also observed but would require further experimentation with multiple replications to validate.

Future work is aimed at the detailed and complete analysis of the Illumina data to identify large numbers of polymorphic markers, which will enable the construction of a high-density peanut genetic maps. Identification of genic regions from BLAST searches of MITE flanking sequences will also serve as an additional source of information for insertion site preference once the peanut genome assembly is completed.

References

Casa A.M., Brouwer C., Nagel A., Wang L.J., Zhang Q., Kresovich S., Wessler S.R. (2000) The MITE family *Heartbreaker* (*Hbr*): Molecular markers in maize. Proc. Nat. Acad. Sci.USA 97:10083-10089

Chu Y., Faustinelli P., Laura Ramos M., Hajduch M., Stevenson S., Thelen J.J., Maleki S.J., Cheng H., Ozias-Akins P. (2008) Reduction of IgE binding and nonpromotion of *Aspergillus flavus* fungal growth by simultaneously silencing Ara h 2 and Ara h 6 in peanut. J.Agric.Food Chem. 56:11225-11233.

Gayler K.R., Sykes G.E. (1985) Effects of nutritional stress on the storage proteins of soybeans. Plant Physiol. 78:582-585.

Patel M., Jung S., Moore K., Powell G., Ainsworth C., Abbott A. (2004) High-oleate peanut mutants result from a MITE insertion into the *FAD2* gene. Theor. Appl. Genet.108:1492-1502.

Sampson H.A. (2004) Update on food allergy. J. Allergy Clin. Immunol. 113:805-819.

Shirasawa K., Koilkonda P., Aoki K., Hirakawa H., Tabata S., Watanabe M., Hasegawa M., Kiyoshima H., Suzuki S., Kuwata C., Naito Y., Kuboyama T., Nakaya A., Sasamoto S., Watanabe A., Kato M., Kawashima K., Kishida Y., Kohara M., Kurabayashi A., Takahashi C., Tsuruoka H., Wada T., Isobe S. (2012) *In silico* polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol. 12. :80