GENETIC BASIS OF ISOFLAVONE ACCUMULATION DURING SOYBEAN SEED DEVELOPMENT: SPECIAL FOCUS ON WATER-DEFICIT CONDITIONS

A Dissertation
Presented to
The Faculty of the Graduate School
University of Missouri-Columbia

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

By

JUAN JOSE GUTIERREZ-GONZALEZ

Dr. David A. Sleper — Dissertation Supervisor
Dr. Henry T. Nguyen — Co-advisor

DECEMBER 2009

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

GENETIC BASIS OF ISOFLAVONE ACCUMULATION DURING SOYBEAN SEED DEVELOPMENT: SPECIAL FOCUS ON WATER-DEFICIT CONDITIONS

Presented by JUAN JOSE GUTIERREZ-GONZALEZ

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance

_____

Dr. David A. Sleper

_____

Dr. Henry T. Nguyen

_____

Dr. Oliver Yu

_____

Dr. Ingolf Gruen

_____

Dr. Joseph C. Polacco

ACKNOLEDGEMENTS

"*A person who never made a mistake, never tried anything new"*
(Albert Einstein)

It was in the morning of some day on the fall of 2001, I believe. I came back to the University where I obtained my bachelor degree from, just to pick up the diploma accrediting the studies. The university was the Polytechnic University of Madrid and the degree was in Agricultural engineering. Four years had passed since the last time I was there and many things had changed. I could not recognize anybody anymore. However, walking along one of the hallways in the main building I saw a face that looked familiar. Yes! It was my former classmate Silvia. We talked as much as you are supposed to after four years. Probably, our initial conversation was trivial. It must have been because nothing comes to my mind from those initial moments. Then she told me that she was working on her PhD degree there. I must confess that my first reaction was of envy towards her. That was what I always had wanted! Why she and not me! She must have noticed my face turning into an envy-looking color because the next thing she told me was: listen! There is a professor from a university in the USA spending his sabbatical in my lab, why don't you go and talk to him? There could be a chance for you to pursue your PhD in the USA. The next thing that came to my mind was…Oh boy!! This PhD thing is affecting her so bad... She just lost her mind!! Nobody is offering me a scholarship in my own country and this girl wants me to try in the country of science, where everybody wants to go!!...Well, that professor was Joe Polacco, and this thesis is the following up from that first encounter with him.

The first acknowledgment should be given to God (in the form of destiny) that put me in the way of Silvia and then Joe that morning in 2001. Probably, I would never have reached this point without that fortuitous meeting with Joe and his help until I met my other big supporter, Dr David Sleper, during the summer of 2004.

I wish to express my sincere gratitude to my main advisor Dr. David Sleper. I must say that he is better mentor than fly-fisherman (and he is an exceptional fly-fisherman). He knows better than anyone else the obstacles I had to overcome to reach this point. He trusted me when he offered me to join his program and, in spite of ups and downs, he never stopped supporting me. There wouldn't be enough words to convey my gratefulness. My appreciation is extended to my co-advisor Dr Henry Nguyen, and committee members Dr Oliver Yu and Dr Ingolf Gruen. Without them this work would never have seen light.

I would also like to thank Kerry Clark, Dr Grover Shannon, and Dr Jeong Lee for their help in the field experiments; all the collaborators in the different projects; my lab-mates for creating a friendly easy-to-work-in environment; and the numerous friends I've met during this time in Columbia for these and yet-to-come years of friendship.

Finally, nothing of the above and below would have been possible without my parents, Cipri and Marisa. They are my longest-time supporters. My heartfelt tribute goes for them, my sister Ana and brother Marcos.

"*Once we accept our limits, we go beyond them*"
(Albert Einstein)

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

Soybean (*Glycine max* [L] Merr.) seed isoflavones have long been considered a desirable trait to target in selection programs for their contribution to human health and plant defense systems. A continuing effort is being made to assess the beneficial or prejudicial aspects derived from their consumption by humans, and is reflected in the numerous articles about the topic published every year. Within the plant, isoflavones are synthesized in response to several biotic and abiotic stresses. As a consequence, isoflavone seed content is known to be tremendously influenced by environmental conditions being temperature and water regime the two main factors, also likely associated with global climate change. Attempts to modify seed isoflavone contents have not always produced the expected results because their genetic basis is polygenic and complex. Without any doubt, the tremendous variability that isoflavones display over locations and years has hindered the genetic basis underlying their synthesis and accumulation in seeds. Our findings suggest that isoflavone seed concentration is controlled by a complex network of multiple minor-effect loci interconnected by a dense epistatic map of interactions. We hypothesize that the magnitude and significance of the main and epistatic effects of the nodes in the network will vary depending on the genetic background and environmental conditions. In an attempt to unravel the genetic architecture underlying the traits studied, we searched on a genome-wide scale for genomic regions homologous to the most important identified isoflavone biosynthetic genes. Temperature is by far the most well studied factor affecting seed isoflavone accumulation but not much is known about accumulation under drought conditions other than it is significantly elevated by irrigation. Drought is an increasing problem in many

regions of the world, thus knowing the factors affecting and the mechanisms regulating the seed isoflavone accumulation will be of vital importance for development of soybean varieties with increasing isoflavone seed content under low soil moisture levels. Overall, we advanced the knowledge of the genetics underlying isoflavone synthesis.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

"*If we knew what it was we were doing, it would not be called research, would it?"*
(Albert Einstein)

Soybean as a crop.

Soybean (*Glycine max* L. Merrill) is the world's leading oil seed crop. Nearly 57% of the vegetable oil that is produced worldwide is obtained from soybean (year 2007 data). It is also the primary source of protein for human consumption. On average, soybeans contain 40% protein, 35% carbohydrate, 20% oil, and 5% ash (Liu 1997). Asian cultures have long utilized soy-derived products in their cuisines, for example, soy milk and sauce, tofu, vegetable soybean (*edamame*), soybean paste (*miso*), soybean sprouts (*moyashi*), and fermented soybean. Soybean is also considered as the principal source of protein for livestock feed. Apart from the traditional food uses, soybean is also employed in other non-conventional or not- as-well-known applications, such as biodiesel, building materials (carpets, floors), papers, plastics, cosmetics, lubricants, pesticides, pharmaceuticals, and paints. Despite the origin of the crop being located in China nearly 5000 years ago (Gai and Guo, 2001), nowadays, USA, Brazil, and Argentina together accounted for 82 % of world production in 2006 (www.soystats.com). USA alone produces slightly more than one third of the soybean.

Generalities of isoflavonoids, uses and functions

Dietary intake of isoflavonoids has been largely related to human therapeutic and preventive medical properties, and in the past three decades numerous articles have been published aiming at garnering more knowledge (reviews Messina et al, 1999; Yu and McGonigle, 2005; Zhang and Yu, 2009). Among the characteristics attributed to them are a certain estrogenic activity, anticarcinogenic effect, antioxidant properties, inhibition of reactive nitrogen species, and osteoporosis and heart diseases prevention (Akiyama et al., 1987, 1991; Jing et al., 1993; Thorburn et al., 1994; Adlercreutz et al., 1997). Nevertheless, the accredited health benefits are not exempted from certain controversy. The estrogenic activity, which able genistein and daidzein to compete with endogenous estrogens (sex hormone 17β-estradiol) for binding to the estrogen receptors, is likely the source of most of the beneficial health effects (Kuiper et al 1997; Pedersen et al 1999; Mueller et al 2002). It is precisely their role as a hormone that also makes scientists conscious about the utilization of phytoestrogens (isoflavones) in the diet of certain human groups such as infants or pregnant women (Yu and McGonigle, 2005).

One of the main biological functions of isoflavones in plants involves plant-microbe interactions. They are the major precursors of phytoalexins and phytoanticipins in legumes, and thus acting as inducible or preformed antimicrobial compounds to protect plants against insects and pathogens (McClure, 1975). Importantly, there is a positive correlation between the production of isoflavones and disease resistance (Dixon, 2001), and the activation of enzymes in the pathway following pathogen infection is broadly reported (Peters et al. 1986; Lawson et al. 1996; Stafford 1997; McKhann et al. 1998; Dhaubhadel et al. 2003; Subramanian et al, 2005; Naoumkina et al. 2007). Somehow

related to the function in plant defense, is their role in symbiosis as chemo-attractants and in the induction of nodulation genes (Barbour et al., 1991), establishing for example the symbiotic relationship between leguminous plants and Rhizobium (Subramanian et al, 2004, 2006, 2007). The specific binding of the isoflavonoids excreted from the roots to the nodD protein factors produced by the bacteria allows this mutual unambiguous recognition and it is the main determinant of host specificity of rhizobia (Rhijn, 1995). Moreover, isoflavones are not only required for the establishment of the association but also for the perdurability of the nodule (Subramanian et al, 2007). Other described biological functions in plants include inhibition of some enzymes involved in signal transduction, antioxidant properties and *in vitro* antimicrobial activities (Graham, 1996), protection against UV light damage (Beggs et al 1985), and attraction of pollinating insects as determinants of flower color (McClure, 1975).

Soybean seed isoflavones. Synthesis and accumulation

Isoflavonoids are important secondary plant metabolites that accumulate mainly in seeds and leaves of legumes, although they can also be found in roots, stems, seed coat, embryos, and pods (Graham et al., 1991; Dhaubhadel et al., 2003). Outside the legume world, the presence of isoflavones is merely anecdotal. Isoflavones are variable metabolites, in the sense that different isoflavone-containing plant species usually have a distinct set of these compounds. In soybean, three are the known isoflavones: genistein, daidzein, and glycitein. They can be present as free, unmodified molecules, but more frequently they are found conjugated with glucose or malonyl-glucose and stored in vacuoles (Yu and McGonigle 2005).

The levels of isoflavonoids present in seeds are highly variable, and regulated by genetic and environmental factors, with an important genetic × environment interaction component (Hoeck et al. 2000; Mebrahtu et al. 2004; Gutierrez-Gonzalez et al, 2009; Murphy et al. 2009). Within the seeds, nearly 90% of the total isoflavones are located in the cotyledons and the remainder in the hypocotyls (Tsukamoto et al. 1995). On average, total isoflavones in soybean seeds ranges from less than 300 μg/g to more than 3000 μg/g. The strong environmental influence is reflected for example in that a cultivar grown in a given environment can differ up to 1.7-fold (Hoeck et al., 2000). In addition, a single cultivar grown in different years at the same location can vary up to 3-fold on its isoflavone content (Eldridge et al., 1983; Wang 1994). Conversely, a single cultivar sowed in different locations in the same year can vary in isoflavone content up to 1.5 times (Joseph et al., 2000; Seguin et al., 2004). Importantly for breeding purposes, no associations were found between isoflavone levels and other important seed quality traits such as oil content, seed quality, and weight (Primomo et al., 2005; Yin et al., 2005), which leaves room for cultivar improvement without interfering with any other desirable or non-desirable trait.

Seed isoflavone concentrations are highly dependable on the environment because many external factors affect their synthesis and accumulation. Frequently, the function of specific compounds within an organism can be assessed or guessed by looking at the factors that trigger their accumulation. For example, the induction of isoflavonoids by environmental stresses is well known, being heat and drought the major affecting abiotic factors. The best-documented abiotic interaction is caused by temperature (Tsukamoto et al., 1995; Nelson et al, 2002; Caldwell et al., 2005; Lozovaya et al, 2005), although they

also include soil fertility (Vyn et al., 2002), UV light (Beggs et al, 1985), and γ radiation (Variyar et al., 2004). In one field assay, Caldwell et al. (2005) found that increasing the temperature from 18 °C to 23 °C during seed development decreased the total isoflavone content by about 65 %. However, little is known about the accumulation of these compounds in seeds as the consequence of varying soil moisture levels. Nelson et al., (2002) showed that there is a correlation between water stress and low isoflavone seed content. In another study, Bennett et al. (2004) found that accumulation of isoflavones daidzein, genistein, and total isoflavone content was increased by irrigation as much as 2.5 fold. Among the affecting biotic environmental factors, the activation of the phenylpropanoid pathway following rhizobial infection has also been broadly reported (Lawson et al., 1996; Stafford 1997; and McKhann et al., 1998; Subramanian et al, 2004, 2006, 2007). Key enzymes in the phenylpropanoid pathway were activated upon contact with pathogens. For example, Dhaubhadel et al (2003) reported the induction of *IFS* after infection of soybean hypocotyls with *Phytophthora sojae*. There are also indications that isoflavones might play an important role in the defense against soybean cyst nematode infection (Subramanian et al. 2004). Part of the seed isoflavonoids may be produced due to direct contact with pathogens (Dakora and Philips, 1996). However, most infections occur at distant areas from the pods, such as leaves and roots. The specific reason why isoflavones are directed towards accumulation in seeds remains elusive.

A part from certain enzyme(s) involved in the glycitein synthesis, the pathway leading to the synthesis of isoflavonoids is mostly known. Isoflavones are synthesized predominantly in legumes from a branch of a plant ubiquitous pathway, the phenylpropanoid pathway, which in addition leads to other important groups of

secondary compounds: tannins, flavones, flavonols and anthocyanins (Winkel-Shirley, 2001; Yu and McGonigle, 2005; Lepiniec et al., 2006). In soybean seeds, the majority of the isoflavonoids produced are isoflavones, present as free, aglycones (genistein, daidzein and glycitein), or as glycoside derivatives (genistin, diadzin, and glycitin) malonyl-esters (6''-O-malonylgenistin, 6''-O-malonyl-diadzin, and 6''-O-malonyl-glycitin), or acetyl-esters (6''-O-acetylgenistin, 6''-O-acetyldiadzin, and 6''-O-acetylglycitin). Soybean plants contain one of the highest levels of isoflavones, more than 100-fold higher than other dietary legumes. Genistein and daidzein are the two principal isoflavones in soybean, while glycitein is present in much lesser amounts, and is unique for soy plants. They are stored as glucosyl- and malonyl-glucosyl conjugates in vacuoles (Yu and McGonigle, 2005). In the path to isoflavone production, phenylalenine ammonia-lyase (*PAL*) uses the amino acid phenylalanine as substrate in the first step. However, the key enzyme for isoflavone biosynthesis is isoflavone synthase (*IFS*), of which there are two copies in the soybean genome differing in only 14 amino acids, *IFS1* and *IFS2*. Both catalyze the reaction to convert the flavanones naringenin and liquiritigenin, within the phenylpropanoid pathway, to the isoflavones genistein and daidzein, respectively (Jung et al., 2000). Importantly, soybean *IFS1* and *IFS2* were shown to be differentially regulated, thus, while *IFS1* is expressed in a variety of plant organs, *IFS2* may be strongly induced in response to stress (Dhaubhadel et al., 2003). Chalcone isomerase (*CHI*) and chalcone reductase (*CHR*) are other two important enzymes of the pathway. *CHI*, which catalyzes the first step in the branch of the phenylpropanoid pathway towards isoflavonoid production, is divided into two highly family specific groups, type I and II, being the second group exclusive from legumes. *CHR* it is found only in leguminous plants and is

needed for the production of daidzein and glycitein. It exists as a multigene family in soybean (Dixon et al., 2002) synthesizing the precursor for daidzein and genistein. There are at least three and ten homologous sequences to *CHR* and *CHI* respectively in the soybean database (Yu et al., 2005). The biochemical reactions to form glycitein have not been entirely revealed, although isoliquiritigenin is likely a precursor (Yu and McGonigle, 2005).

Although seed isoflavone content is greatly dependent on the environment, the production is largely under genetic control (Eldridge et al., 1983; Wang et al., 1994; Hoeck et al., 2000; Nelson et al., 2002). It has been suggested that the total amount of isoflavonoids accumulated in seeds comes from two different sources (Dhaubhadel et al., 2003). First, developing soybean embryos can synthesize isoflavonoids *de novo* inasmuch as the legume-specific isoflavone synthase genes, *IFS1* and *IFS2*, are expressed in this tissue. As a second contribution of the seed isoflavone content is its transport from maternal tissues. Curiously, there seems not to be a close positive correlation of the amounts of genistin, daidzin, and glycitin present in soybean tissues with *IFS* transcript accumulation patterns. Despite most of the biosynthetic pathway being known, there is a big gap in the knowledge of how the route is transcriptionaly and post-transcriptionaly regulated. The only promoter reported to be cloned is an alfalfa *IFR* promoter (Oommen et al., 1994), and no transcription factor specific to the isoflavone synthesis pathway has been reported. To investigate this transcriptional control, Oliver Yu (personal communication) has cloned the soy *IFS1* promoter, as well as promoters from eleven other genes, probing that the *IFS1* promoter is induced upon nematode and fungal infection.

Objectives of the present thesis.

Because of implications for human health and fitness of the soybean plant *per se*, there is an increasing interest in altering the isoflavone content of soybean commercial varieties. Attempts to conduct metabolic engineering of plants to produce high-isoflavone-content lines have not always succeeded (Zernova et al., 2002; Yu, unpublished results). For example, efforts to improve isoflavone content by overexpressing three key enzymes: PAL, CHS, and IFS, either independently or combined, have failed to enhance isoflavone accumulation (Koes et al. 1994; Zernova et al 2002). Apparently, one of the most limiting obstacles for isoflavone pathway engineering is the existence of metabolic channels in the phenylpropanoid pathway, which gives rise to a flux of metabolites through different complexes and hampers the success of transgenic plants. Transgenic tobacco (Nicotiana tabacum L.) plants expressing an *IFS*/*CHI* artificial bifunctional fusion protein accumulated genistein and daidzein in petals and leaves in more amount than *IFS* construct alone, further supporting this enzyme association (Tian et al., 2006). Contrarily to genetic engineering for increasing isoflavone levels, strategies to decrease isoflavones in seeds have been applied with more success (Yu and McGonigle, 2005). Without doubt, the extreme variability that seed isoflavones display over environments has obscured our understanding of the genetics involved.

The specific objectives for the present thesis are:

1) To identify and map soybean loci governing the heritable variation in isoflavone accumulation in seeds. This includes the confirmation of previously reported

QTLs, the possible incorporation of novel QTLs, and the interactions between loci with or without individual main effects. It also aims to select appropriate combinations of markers for marker-assisted selection (MAS) in breeding programs.

2) Because seed isoflavone content is highly dependent on the environment, one important objective is to dissect the genetic component from the environmental counterpart. This would help to understand up to what extent a particular QTL can be use for MAS. QTLs that are stable over multiple environments are more useful in breeding programs because they might contribute to a consistent increased phenotype under changing conditions.

3) Another main objective is to map candidate genes in the phenylpropanoid pathway for isoflavone synthesis onto the genetic map and compare whether they match any of the identified QTLs. This will lead to a better understanding of the pathway and its regulation.

4) Lastly, another objective is to study differences in isoflavone accumulation in soybean seeds under water deficits. This would allow establishing a developmental profiling of isoflavone compounds, in well-watered versus water-deficit conditions. A related aim is to perform gene expression profiling of the most important enzymes for isoflavone synthesis and to analyze differences in transcript accumulation.

Overall, the research work presented herein is expected to shed light on the accumulation of isoflavones in soybean seeds, and particularly in relationship to water regimes.

REFERENCES

Adlercreutz H., and Mazur W. (1997) Phyto-estrogens and western diseases. Ann. Med. **29**:95-120.

Akiyama T., Ishida J., Nakagawa S., et al. (1987) Genistein, a specific inhibitor of tyrosine-specific protein kinases. J. Biol. Chem. **262**:5592-5595.

Barbour W.M., Hattermann D.R., and Stacey G. (1991) Chemotaxis of Bradyrhizobium japonicum to soybean exudates. Appl. Environ. Microbiol. **57**: 2635-2639.

Beggs CJ, Stolzer-Jehle A, Wellmann E (1985) Isoflavonoid formation as an indicator of UV stress in bean (Phaseolus vulgaris L.) leaves: the significance of photorepair in assessing potential damage by increased solar UV-B radiation. Plant Physiol. **79(3)**:630-634.

Bennett JO, Yu O, Heatherly LG, Krishnan HB (2004) Accumulation of genistein and daidzein, soybean isoflavones implicated in promoting human health, is significantly elevated by irrigation. Agric. Food Chem. 52:7574-7579.

Caldwell CR., Britz SJ, Mirecki RM (2005) Effect of temperature, elevated carbon dioxide, and drought during seed development on the isoflavone content of dwarf soybean [Glycine max (L.) Merrill] grown in controlled environments. J. Agric. Food Chem. 53:1125-1129.

Dakora FD, and Philips DA (1996) Diverse functions of isoflavonoids in legumes transcend anti-microbial definitions of phytoalexins. Physiol. Mol. Plant Pathol. 49:1-20.

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol. Biol. **53**:733-743.

Dixon R.A. (2001) Natural products and plant disease resistance. Nature **411**: 843-847.

Dixon R.A., Achnine L., Kota P., Liu C., Reddy M.S.S., and Wang L.J. (2002) The Phenylpropanoid pathway and plant defense: A genomics perspective. Mol. Plant Pathol. **3**: 371-390.

Eldridge A, Kwolek W (1983) Soybean isoflavones: Effect of the environment and variety on composition. J. Agric. Food Chem. **31**:394-396.

Gai J, and Guo W (2001) History of maodou production in China. In: T.A. Lumpkin and S. Shanmugasundaram (Eds), Proceedings Second International Vegetable Soybean Conference, Washington State University, Pullman, WA., pp,41-47.

Graham T.L. (1991) Flavonoid and Isoflavonoid Distribution in Developing Soybean Seedling Tissues and in Seed and Root Exudates. Plant Physiol. **95**:594-603.

Graham T.L., and Graham M.Y. (1996) Signaling in soybean Phenylpropanoid responses. Plant Physiol. **110**: 1123-1133.

Hoeck JA, Fehr WR, Murphy PA, Welke GA (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci. **40**:48-51.

Jing Y., Nakaya K., and Han R. (1993) Differentiation of promyelocytic leukemia cells HL-60 induced by daidzein in vitro and in vivo. Anticancer Res.**13**:1049-1054.

Joseph A.H.M., Fehr W.R., Murphy P.A., and Welke G.A. (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci. **40**:48-51.

Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggbled J, Nilsson S, and Gustafsson JA (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. Endocrinology **138**:862-339.

Lawson C.G.R., Rolfe B.G., and Djordjevic M.A. (1996) Rhizobium inoculation induces condition-dependent changes in the flavonoid composition of root exudates from Trifolium subterraneum. Australian Journal of Plant Physiology. **23**:93-101.

Liu K (1997) Soybeans: Chemistry, technology and utilization. Aspen publishers, Gaithersburg, Maryland, USA. pp. 532.

Lozovaya VV, Lygin AV, Ulanov AV, Nelson RL, Dayde J, and Widhohm JM (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci. **45**:1934-1940.

McClure J.W. (1975) Physiology and functions of flavonoids. Harborne J.B., Mabry I.J., Mabry H. Editors. Academic Press. New York, 971-1055.

McKhann HI, Paiva NL, Dixon RA, Hirsch AM (1998) Expression of genes for enzymes of the flavonoid biosynthetic pathway in the early stages of the Rhizobium-legume symbiosis. Adv. Exp. Med. Biol. **439**:45-54.

Mebrahtu T, Mohamed A, Wang CY, Andebrhan T (2004) Analysis of isoflavone contents in vegetable soybeans. Plant Foods for Human Nutrition **59**:55-61.

Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. Am. J. Clin. Nutr. 70(suppl):439S-450S.

Mueller SO, Clark JA, Myers PH and Korach KS (2002) Mammary gland development in adult mice requires epithelial and stromal estrogen receptor α. Endocrinology 143:2357-2365.

Murphy SE, Lee EA, Woodrow L, Seguin P, Kumar J, Rajcan I, and Ablett GR (2009) Genotype × Environment interaction and stability for isoflavone content in soybean. Crop Sci. **49**:1313-1321.

Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc. Natl. Acad. USA. **104**:17909-17915.

Nelson R., Lygin A., Lozovaya V., Ulaov A., and Widholm J. (2002) Genetic and environmental control of soybean seed isoflavone levels and composition. Proceedings of the 9th biennial Conf. of Cel. and Mol. Bio. of Soybean.

Oommen A, Dixon R.A., and Paiva N.L. (1994) The Elicitor-Inducible Alfalfa Isoflavone Reductase Promoter Confers Different Patterns of Developmental Expression in Homologous and Heterologous Transgenic Plants. Plant Cell. **6**:1789-1803.

Pedersen SB, Hansen PS, Lund S, Andersen PH, Odgaard A, and Richelsen B (1996) Identification of oestrogen receptors and oestrogen receptor mRNA in human adipose tissue. Eur. J. Clin. Invest. 26:262-268.

Peters NK, Frost JW, Long SR (1986) A plant flavones, luteolin, induces expression of Rhizobium meliloti nodulation genes. Science **233**:977-980.

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M, Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. **45**:2454-2462.

Seguin P., Zheng W., Smith D. L., and Deng W. (2004) Isoflavone content of soybean cultivars grown in eastern Canada. J Sci. of Food and Agriculture. **84**:1327-1332.

Stafford HA (1997) Roles of flavonoids in symbiotic and defense functions in legume roots. Bot. Rev. **63**:27-39.

Subramanian S, Hu X, Lu G, Odelland JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol. Biol. 54:623-639.

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to Phytophthora sojae. Plant Physiol. 137:1345-1353.

Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and Bradyrhizobium japonicum. Plant J. 48:261-273.

Subramanian S, Stacey G, Yu O (2007) Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci. 12:282-285.

Thorburn J. and Thorburn T. (1994) The tyrosine kinase inhibitor, genistein, prevents α-adrenergic-induced cardiac muscle cell hypertrophy by inhibiting activation of the Ras-MAP kinase signaling pathway. Biochem. Biophys. Res. Commun. **202**:1586-1591.

Tsukamoto C, Shimada S, Igita K, Kudou S, Kokubun M, Okubo K, Kitamura K (1995) Factors affecting isoflavone content in soybean seeds: Changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J. Agric. Food Chem. 43:1184-1192.

Variyar PS, Limaye A, Sharma A (2004) Radiation-induced enhancement of antioxidant contents of soybean (Glycine max Merrill). J. Agric. Food Chem. **52**:3385-3388.

Vyn TJ, Yin X, Bruulsema TW, Jackson CC, Rajcan I, Brouder SM (2002) Potassium Fertilization Effects on Isoflavone Concentrations in Soybean (*Glycine max* L. Merr.). J. Agric. Food Chem. **50**: 3501-3506.

Wang H, Murphy PA (1994) Isoflavone Composition of American and Japanese Soybeans in Iowa: Effects of Variety, Crop Year, and Location. J. Agric. Food Chem. **42**:1674-1677.

Winkel-Shirley B. (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Phys. **126**:485-493.

Yin X., and Vyn T.J. (2005) Relationships of Isoflavone, Oil, and Protein in Seed with Yield of Soybean. Agron. J. 97:1314-1321.

Yu O and McGonigle B (2005) Metabolic engineering of isoflavone biosynthesis. Advances in Agronomy. **86**:147-190.

Zhang J, and Yu O (2009) Metabolic engineering of isoflavone biosynthesis in seeds. *In* Modification of seed composition to promote health and nutrition. Hari Krishnan (ed). Agronomy Monograph Series, pp151-177.

CHAPTER 2

Genetic Control of Soybean Seed Isoflavone Content: Importance of Statistical Model and Epistasis in Complex Traits

"*Things should be made as simple as possible, but not any simpler"*
(Albert Einstein)

ABSTRACT

A major objective for geneticists is to decipher genetic architecture of traits associated with agronomic importance. However, a majority of such traits are complex, and their genetic dissection has been traditionally hampered not only by the number of minor-effect QTL but also by genome-wide interacting loci with little or no individual effect. Soybean (*Glycine max* [L.] Merr.) seed isoflavonoids display a broad range of variation, even in genetically stabilized lines that grow in a fixed environment, because their synthesis and accumulation are affected by many biotic and abiotic factors. Due to this complexity, isoflavone QTL mapping has often produced conflicting results especially with variable growing conditions. Herein, we comparatively mapped soybean seed isoflavones genistein, daidzein, and glycitein by using several of the most commonly used mapping approaches: interval mapping, composite interval mapping, multiple interval mapping and a mixed-model based composite interval mapping. In total, twenty six QTLs, including many novel regions, were found bearing additive main

effects in a population of RILs derived from the cross between Essex and PI 437654. Our comparative approach demonstrates that statistical mapping methodologies are crucial for QTL discovery in complex traits. Despite a previous understanding of the influence of additive QTL on isoflavone production, the role of epistasis is not well established. Results indicate that epistasis, although largely dependent on the environment, is a very important genetic component underlying seed isoflavone content, and suggest epistasis as a key factor causing the observed phenotypic variability of these traits in diverse environments.

INTRODUCTION

A great proportion of the agronomic traits of economic importance are complex in nature as being controlled by many genes, each with little individual effects and highly influenced by environment. Soybean seed isoflavones, namely genistein, daidzein and glycitein, have long been implicated in human health (see Messina 1999; Ososki and Kennelly 2003; Usui 2006; Rochfort and Panozzo 2007, for reviews), and more recently in plant defense and nodulation (Subramanian et al. 2004, 2005, 2006 and 2007; Zhang et al. 2007). They are considered complex traits because their levels are highly variable and regulated by multiple environmental and genetic factors (Hoeck et al. 2000; Mebrahtu et al. 2004). Induction of the phenylpropanoid pathway, a branch of which synthesizes isoflavonoids, by environmental factors has been extensively documented. For example, a particular genotype planted in different years in the same location can vary up to 3-fold for its seed isoflavone content (Eldridge and Kwolek 1983; Wang and Murphy 1994).

Temperature is the most broadly studied influencing abiotic factor (Tsukamoto et al. 1995; Caldwell et al. 2005; Lozovaya et al. 2005), although water regime (Bennett et al. 2004; Caldwell et al. 2005; Lozovaya et al. 2005), soil fertility (Vyn et al. 2002), UV light (Beggs et al. 1985), and γ radiation (Variyar et al. 2004) are also influential factors. Among the biotic elicitors, the activation of enzymes in the pathway following rhizobial and pathogen infection is also well known (Peters et al. 1986; Lawson et al. 1996; Stafford 1997; McKhann et al. 1998; Dhaubhadel et al. 2003; Naoumkina et al. 2007).

Distribution of isoflavone compounds within segregating populations suggests a quantitative genetic inheritance component (Supplemental Fig. 1). To elucidate the genomic regions underlying polygenic traits, or quantitative trait loci (QTL), several statistical methodologies have been used. In single-marker locus analysis (SMA), association between a QTL and a single marker is determined by analysis of variance (ANOVA) with significance tests or simple linear regression. SMA is easy to implement because there is no requirement for a linkage map; however, it tends to confound QTL effect and distance in such a way that a QTL with little effect closely linked to a marker may be treated equal to a major QTL located distant from the marker. Simple interval mapping (IM) was proposed by Lander and Botstein (1989) to overcome the limitations of SMA. IM increases the probability of detection by considering two consecutive markers at each time and assessing the presence of a QTL by performing likelihood ratio tests to assign a probability (LOD score) at fixed points within the interval. A multiple regression variant of IM was proposed by Haley and Knot (1992) as a less computationally demanding approach. Composite interval mapping (CIM) (Zeng 1993, 1994) goes a step further and incorporates multiple regression analysis in IM, by

considering some other key markers as cofactors (co-variables) to control external genetic variation when testing a particular interval. Neither IM nor CIM are able to detect multiple interacting QTL. Finally, multiple interval mapping (MIM) (Kao et al. 1999; Zeng et al. 2000) fits several marker intervals and their epistatic interactions simultaneously in a mixed model approach by a multidimensional search. Every time a new factor is incorporated into a model the residual variance or error component (ε) decreases, thereby the model is able to explain more phenotypic variation, and the power and precision of QTL detection improve. Deciphering the genetic architecture of complex traits requires inclusion of as many of the factors accounting for their variation as possible. Conversely, lacking of a complete model may lead to a conflicting outcome. For instance, Meksem et al. (2001) and Kassem et al. (2004) reported eight loci associated with isoflavone amounts in soybean seeds by using IM. Intriguingly, only two of them were confirmed and fourteen new ones were found when CIM was implemented running the same phenotypic data (Kassem et al. 2006).

Epistasis, or interactions between effects of alleles from two or more genetic loci, can account for a significant part of the variation for complex traits (Doerge 2001, Phillips 2008). It depends on the genetic background, and as a result, it is also believed to play an important role in certain biological processes such as adaptation, evolution and speciation (Whitlock 1995; Juenger et al. 2005). There is increasing evidence that epistatic models can improve the statistical power to detect QTL as well as the accuracy of their localization and the precision in determining the associated confidence intervals, because they might incorporate genetic variation that would not be associated with the trait otherwise (Yi and Xu 2002). Conversely, if interacting loci are not considered,

variation caused by epistasis is removed from the tested model and absorbed into residual error. Nevertheless, detecting epistasis genome-wide is cumbersome and often hampered by either computationally demanding algorithms, a lack of appropriate statistical methodology, or a reduced mapping population size (Doerge 2001; Mao and Da 2004). Conventional single-QTL mapping models, like IM and CIM, are not able to deal with interacting loci, treating epistasis as background noise. Here, we use two software packages for the comparative. First, QTL-Cartographer has been extensively used for QTL mapping as it implements IM, CIM, and MIM among other methods. Second, in addition to incorporating epistasis in the model, the power of the other package, QTLNetwork, resides in that data from different locations and within-location replication are combined to perform an integrated analysis by a mixed linear model (MCIM) (Yang et al. 2005, 2007).

Herein we used a recombinant inbred line (RIL)-population developed from a cross between an isoflavone low-content line, Essex (1236±160 and 940±227 μg/g, total converted isoflavone per gram of seed weight at Bradford Research and Extension Center, BREC, and Delta Research Center, DRC, respectively, see materials and methods) and the high-content PI 437654 (2092±54 and 1623±120) to elucidate the genetic architecture of isoflavones. The study proves that isoflavone accumulation in seeds is influenced by multiple interacting genetic loci, with epistasis playing an important role. Altogether, twenty six additive genomic regions and many additive-by-additive epistatic interactions were found by comparative mapping using existing methodologies to affect isoflavone contents. We also discuss the importance of the statistical model and of considering epistatic interactions when mapping complex traits.

MATERIALS AND METHODS

Plant Material and Growing Conditions

Eight soybean (*Glycine max* [L.] Merr.) accessions: RCAT-Angora, AC-756, PI 437654, Williams 82, PI 438489B, Magellan, Essex and Forest, were chosen for a comparative study to select appropriate parental lines for isoflavone studies. These were planted in 19-liter pots following a randomized complete block design (RCBD) with four replications and grown between April and September 2005 in a greenhouse (GH) at the University of Missouri. Pots were weighted and adjusted to the same substrate amount. Growth conditions were 30ºC/20ºC day/night temperatures, and 16/8 hr day/night lengths. All but PI 438489B were also planted in April 2005 in two-row plots following a RCBD with three replications in two field locations: University of Missouri Bradford Research and Extension Center (BREC, 38º 9'N), located near Columbia, MO, and Grand Pass (GP, 39º 19'N), Saline County, MO. Plants were allowed to grow under irrigated conditions until maturity, after which seed samples were collected from single plants (GH) or entire plot and pooled (BREC and GP). The pooled linear statistical model was a split plot design which contained the effect of environment (GH, BREC, GP), replication within environment (4), genotype (RCAT-Angora, AC-756, PI 437654, Williams 82, Magellan, Essex and Forest), and the interaction of environment × genotype. Mean differences were determined using Fisher's Protected Least Significant Difference according to the methods described by Chew V, 1977.

A mapping population comprised of 196 F_7-derived recombinant inbred lines (RILs) developed from a cross between Essex and the high-isoflavone content plant

introduction PI 437654 were planted in two-row plots under irrigated conditions in a RCBD with three replications at two locations: BREC and the University of Missouri Delta Research Center (DRC, 36º 44'N), located near Portageville, MO. Physiologically matured seed samples were harvested for isoflavone quantification from one plant per plot at the DRC and from a pool of at least three plants at BREC. Effects were tested using PROC GLM in SAS STAT 9.1 (SAS Institute Inc., Cary, NC). The pooled linear model contained the effect of environment, replication(environment), genotype, and environment × genotype. Variance components to calculate heritability estimates were determined using PROC Mixed.

Isoflavone Extraction and Quantification

Approximately 2.5 grams (~20 seeds) of soybean seeds were ground to a fine powder using a General Electric seed grinder (model 5XBG008, New York City, NY). The powder was extracted with 7 mL of 80% methanol at 55 °C for 2 h, vortexing every 30 min. After centrifugation (6076 RCF (×g), 5 min), the supernatant was filtered using Fisherbrand 0.45 μm 25 mm nylon syringe filters (Fisher Scientific, Pittsburgh, PA). Samples were analyzed by reverse-phase HPLC on an Agilent 1100 high-performance liquid chromatography (HPLC) system (Santa Clara, CA). Separation and elution were accomplished using an 18 min linear gradient initiated with 20% methanol / 80% 10mM ammonium acetate (v/v) (pH 5.6) and completed with 100% methanol at a flow rate of 1 mL/min. Detection of metabolites was performed by photodiode array. Identification and quantification of each isoflavone component were based on available standards (Indofine Chemical Co., Somerville, NJ). Measurements are given as micrograms of isoflavone per

gram of seeds plus/minus standard deviation or standard error, when corresponds (μg/g ± SD or SE).

Linkage Map and QTL Analysis

DNA was extracted from young trifoliate leaves with the help of AutoGenprep 960 robot (AutoGen, Holliston, MA). Linkage map contained a total of 276 SSR and AFLP markers which were distributed on 26 linkage groups (LG), covering 2,406.5 cM and average distance of 8.7 cM (for more detailed information see Wu et al. 2008). Statistical analysis was performed using the SAS STAT 9.1 program, and heritability in the broad-sense over environments was calculated according to Hill et al. (1998):

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{E} + \frac{\sigma_\varepsilon^2}{ER}}$$

being σ_g^2, σ_{ge}^2, and σ_ε^2 the genotype, genotype × environment, and error components of variance, respectively; E, number of environments; and R the number of replications.

QTLNetwork v2.0 (Institute of Bioinformatics, Zhejiang University, Hangzhou, China) was used for the mixed-model based composite interval mapping (MCIM) with individual observations from three replications, and two-locations combined input data. Windows QTL Cartographer v.2.5 (Statistical Genetics, North Carolina State University) and trait averaged values for each environment were used for interval mapping (IM), composite interval mapping (CIM), and multiple interval mapping (MIM). For MCIM, critical F-value was assessed by permutation test using 1000 permutations. QTL effects were estimated using Markov chain Monte Carlo method. Candidate interval selection, epistatic effects, and putative QTL detection were calculated with an experimental-wise

type I error of α=0.05 α=0.001, and α=0.001, respectively. Significance level for candidate intervals was relaxed to allow more genomic intervals to be included in the model. Genome scan was performed using 10 cM window size and 1 cM walk speed in all models. Forward and backward regression method was chosen to select markers for CIM analysis. To create MIM models, initial search for main QTLs was done by the MIM forward search method. Epistatic network outcome from QTLNetwork was considered as a starting point in the setting of epistatic interactions. LOD-score significance threshold values were estimated for each trait and location by permutation test analysis with 1000 permutations and error α=0.05, lending for IM: LOD 3.1 and 3.3 for genistein at BREC and the DRC respectively, 3.3 and 3.1 for daidzein, 2.9 and 3.1 for glycitein, and 3.0 and 3.1 for total isoflavones. For CIM the threshold values were: 3.0 and 3.1, 3.2 and 3.0, 3.0 and 3.1, and 3.3 and 3.2, respectively. Because different models were compared, and for illustrative purposes, QTL of LOD score above 3.0 were included when they were significantly detected by other models and/or locations. R^2 (%) is the proportion of the phenotypic variation explained by the each particular QTL. The heritability of the additive effect, $h^2(a)$, or the proportion of the phenotypic variation explained by the additive component of the QTL, is calculated by the mixed linear model, and theoretically it should account for the same variance components as R^2, except for the fixed effects.

RESULTS

Genotypic quantification

Phenotypic characterization of soybean lines over years and locations suggests that isoflavones are influenced by both genes and environment (Hoeck et al. 2000; Mebrahtu et al. 2004). In our studies, lines displayed a 2.8-fold range of variation within field locations and 3.0-fold in GH (Table 2.1). Among them, PI 437654 consistently accumulated the highest levels of isoflavones across all environments: 1804±102 μg/g at BREC, 2052±215 μg/g at GP, and 2657±247 μg/g at GH. On the other hand, the lowest values were obtained for AC756 in both field locations: 652±128 μg/g (BREC) and 735±93 μg/g (GP); and for Magellan (902±121 μg/g) in the GH. Differences of least squares means between field locations of each tested genotype were not significant (t-test; $P<0.05$) (data not shown). However, PI 437654, RCAT-Angora, AC756, and Essex had significantly higher isoflavone contents in the GH than in the field ($P<0.05$), and particularly AC756 as much as 2-fold.

An analysis of variance was conducted to establish the causes of such variation. Highly significant ($P<0.0001$) genotypic and environmental effects were present for genistein, daidzein and total isoflavones. For glycitein, there were also significant genotypic ($P<0.0001$) and environmental ($P<0.05$) effects. Genotype × environment interaction (g × e) effect was an important component as well for genistein, daidzein and total isoflavone ($P<0.0001$), and for glycitein ($P<0.05$). As a whole, the results suggest that all genetic, environmental and g × e interaction effects contribute to the variation of isoflavone levels observed in genetically homogeneous lines.

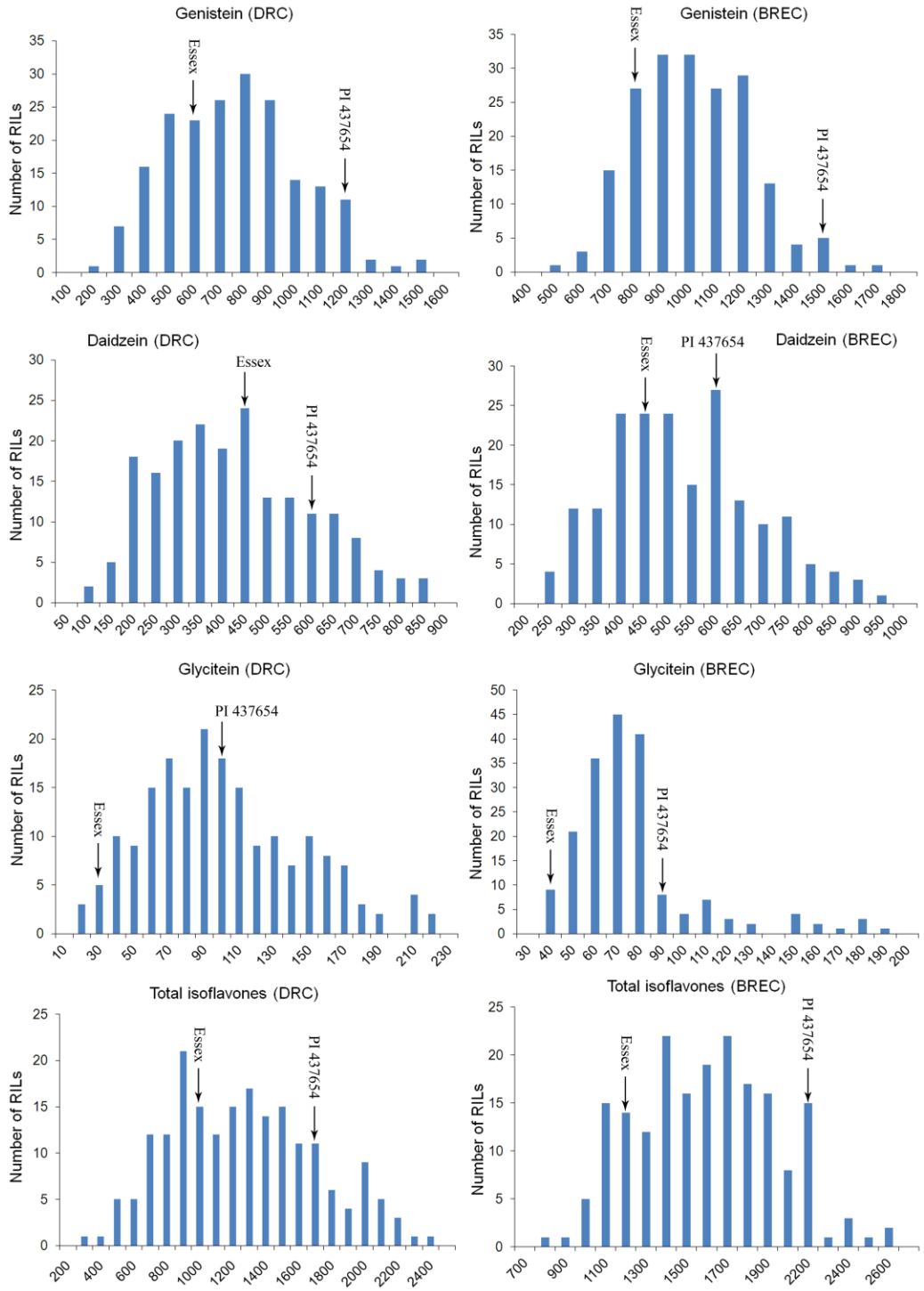

Fig. 2.1. Distribution of average genistein, daidzein, glycitein, and total isoflavones of three replications of Essex × PI 437654 RILs growing in two field locations, BREC and DRC. Arrows indicate the position of the two parental lines. Horizontal axe shows each particular isoflavone seed content in μg/g of seeds.

Genetic and Phenotypic Variation within Mapping Population

The progeny of RILs appears to follow a normal distribution (Fig. 2.1). Importantly, the transgressive segregation found suggests the presence of positive-effect alleles in both parents. Analysis of variance (ANOVA) and heritability are often conducted in mapping studies to assess sources of variation within experimental populations. Heritability (H^2, $0 \leq H^2 \leq 1$) is an estimate of the genetic contributions to the phenotypic variance, and in its broad-sense includes additive, dominant, and epistatic genetic variance components. Within Essex × PI 437654 segregating mapping population, the heritability estimates were high for genistein, daidzein, and total isoflavones, 0.83, 0.89, and 0.86, respectively; and moderate for glycitein (0.43). ANOVA conducted over locations revealed a significant ($P<0.0001$) genetic, environmental, and g × e effects (Table 2.2). The results confirm that the growing environment is crucial and that isoflavone accumulation in soybean seeds is not only controlled by both genetic and environmental factors but also there are important g × e interactions. Observations are also in agreement with other researchers, and consistent with the long-time consideration of these compounds as complex traits.

Isoflavone QTL Mapping

The apparently conflicting results obtained by some research groups when using different mapping models (Meksem et al. 2001 and Kassem et al. 2004, 2006) caused us to consider a multi-model strategy by testing some of the most broadly used mapping algorithms on the same phenotypic data. Some models, such as the one implemented in MCIM, are able to combine data sets from different locations, years, and replications by

means of a mixed model. Others, on the other hand, analyze each dataset independently. A total of seven loci displaying additive effect were found to explain genistein seed levels by a combination of all models, six of which were detected by MCIM (Table 2.3 and Fig. 2.2). Among all, there was a major QTL on linkage group (LG) A1, named *gen-A1* for genistein locus at LG A1. A QTL in the proximity was previously reported to influence genistein, daidzein and total isoflavones (Primomo et al. 2005). Other important QTL were found in linkage groups D1b (*gen-D1b*), E (*gen-E*), and M (*gen-M*). In addition, minor loci were detected in groups B2 (*gen-B2*), also reported by Kassem et al. (2004), and F (*gen-F*). Lastly, a locus in LG L (*gen-L*) was not revealed by MCIM but it was detected by CIM and MIM although only in one location. Surprisingly, out of the total seven discovered QTL, IM model was not able to localize any single locus at BREC, and only *gen-A1* at DRC.

Six additive genomic regions were also discovered for daidzein, the second most abundant soy isoflavone, one of which was exclusively identified by MIM and only at DRC (*dai-L*). The other five were: *dai-A1*, *dai-A2*, *dai-D1b*, *dai-M*, and *dai-H* (Table 2.4 and Fig. 2.2). The latter also displayed an additive × environment (a × e) interaction effect when MCIM was implemented. Importantly, only the major QTL (*dai-A1*) was detected by all models at both locations. This region was also found to explain genistein accumulation (*gen-A1*) as previously shown. The *dai-H* locus was identified by all models as well, but only at DRC.

Glycitein is the most variable isoflavone in soybean (Table 2.2). This complicates QTL discovery because it requires larger mapping populations and advanced statistical methods that can take into account most of that variation. Six loci were found to control

glycitein accumulation when considering all mapping algorithms together (Table 2.5 and Fig. 2.2). Two of them viz. one in LG C2 (*gly-C2*), and the other in LG E (*gly-E),* were revealed by MCIM model with data from BREC and DRC combined. Locus *gly-C2* was the major glycitein QTL reported herein. However, IM could not map a single QTL at BREC, and just *gly-C2* at DRC. This same region was also detected with CIM in DRC, and with MIM in both locations. Two other loci in LG D2 (*gly-D2_1* and *gly-D2_2*) were found to be significant only in BREC. On the other hand, *gly-G* and *gly-K* were detected at DRC solely. Interestingly, no one of the QTLs for glycitein discovered in the present study was previously reported.

Total seed isoflavone (sum of genistein, daidzein and glycitein) content was also considered as a trait and mapped. As a result, seven QTLs were identified. All but one were found by MCIM with two locations combined data (Table 2.6 and Fig. 2.2). The major QTL was located in LG A1 (*tot-A1*), and matched also the major QTL for genistein and daidzein, *gen-A1* and *dai-A1*. All mapping approaches identified this as a major locus. In addition to its important additive component, *tot-A1* also showed a significant a × e interaction effect revealed by MCIM when analyses were performed over both locations. The other identified QTLs were located on LGs A2 (*tot-A2*), D1b (*tot-D1b*), E (*tot-E*), H (*tot-H*), M (*tot-M*), and L (*tot-L*), also showing a × e interaction effect. IM only detected two out of the seven, *tot-A1* and *tot-H*, although the later only in one environment. With CIM, *tot-D1b* and *tot-E* regions were also added. Finally, model MIM incorporated *tot-A2* and *tot-L* at BREC and DRC, respectively. Taken together, the results of all four traits suggest that the accumulation of isoflavones in soybean seeds is largely under additive genetic control, and that there can also be an a × e interaction effect.

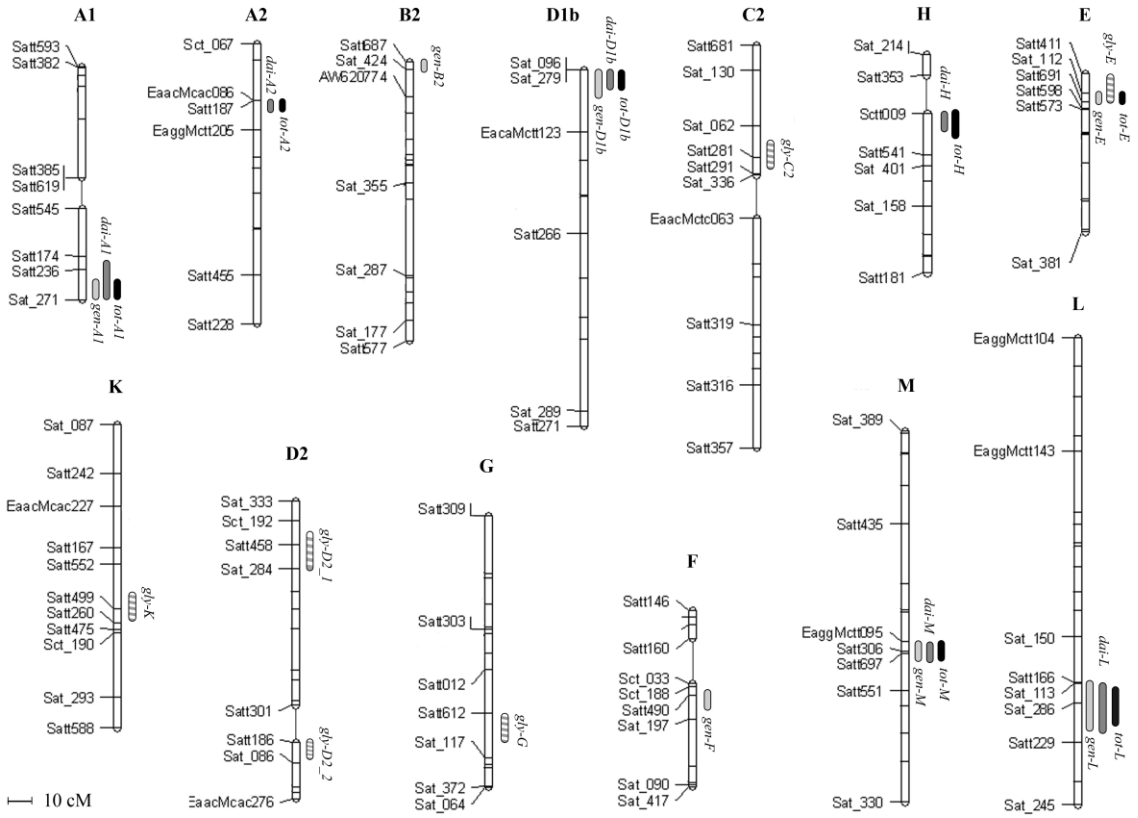

Fig. 2.2 Linkage group map summarizing QTL locations detected for all different models. Light grey ovals indicate genistein loci, dark grey ovals indicate daidzein loci, black ovals indicate total isoflavone loci, and hatched ovals indicate glycitein loci. Flanking and other key markers are depicted at the right side of the linkage group (LG). Marks along LGs that do not have a name attached correspond to other marker positions that were used to genotype the population. The name of the QTL, shown aside each oval, is compound of the influenced trait: genistein (*gen*), daidzein (*dai*), glycitein (*gly*) and total of isoflavones (*tot*), followed by the linkage group.

Epistatic Interactions

The structure of the segregating population allowed us to map only the additive-by-additive (aa) epistatic network. However, aa is not only the main interaction but also the easiest to detect because its statistical power of detection is higher than for the other components, i.e. additive-by-dominant, dominant-by-additive, and dominant-by-dominant (Mao and Da 2004). Among all mapping methods tested herein, only models MCIM and MIM include variation due to epistasis. Four epistatic interactions were found by MCIM to be significant for genistein when data were combined over both locations, and seven and five at BREC and DRC by MIM respectively (Fig. 2.3 and Table 2.7). Importantly, no single epistatic interaction was maintained over environments. Daidzein usually had more significant epistatic interactions and accounted for more variance than genistein. For example, at BREC eight (R^2=37.7%) interactions were mapped by MIM (Table 2.8), and just one of them accounted for up to 10.1 % of the variation. The importance of considering epistasis when dealing with complex traits was revealed greatest for glycitein, for which the proportion of its variance due to epistatic effects was up to 2.8 times the variance of the additive QTL. Strikingly, a single interaction was able to explain as much as 22.0% of the phenotypic variation and had an aa additive value of 22.53 μg/g (at BREC by MIM). In this environment, epistasis was responsible for 54.1% of the variance (Table 2.9). Finally, epistatic networks for the sum of all isoflavones showed ten statistically significant interacting pairs at BREC (R^2=28.3%), and nine at DRC (27.7%) (MIM, Table 2.10). The overall results show that epistasis can account for an important part of the isoflavone phenotypic variation and suggest that epistasis might

be one of the main causes of the variable isoflavone amounts observed in soybean genotypes over certain environments.

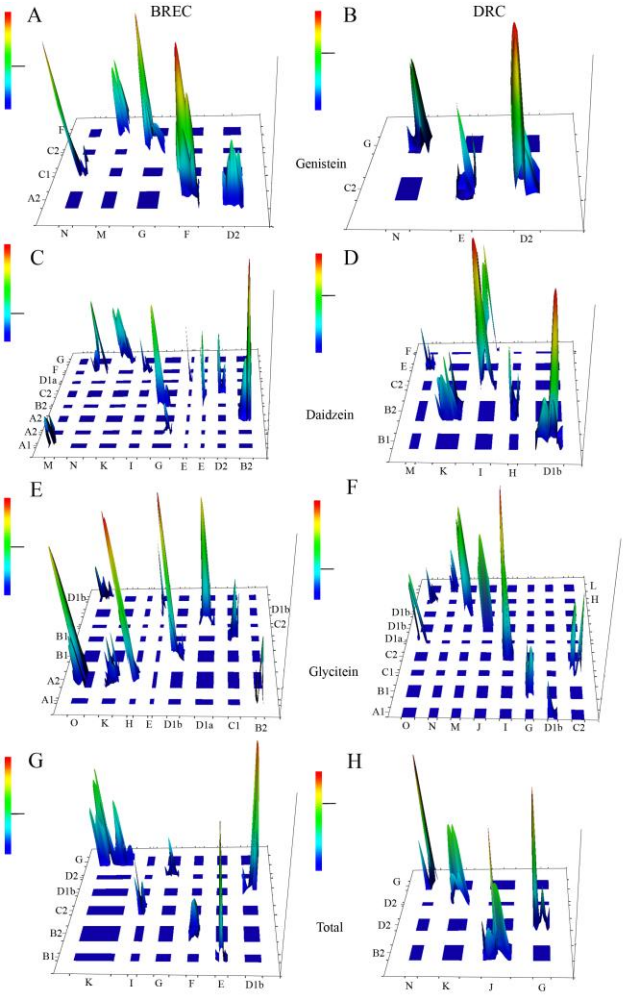

Fig. 2.3 Peaks of epistasis throughout the genome for genistein, A) at BREC and B) at DRC; for daidzein C) at BREC and D) at DRC; glycitein E) at BREC and F) at DRC; and for the sum of all isoflavones G) at BREC and H) at DRC. Linkage groups of interacting loci are depicted on the x- and y-axes. The statistic F-value is taken as the z-axe. F-values have also a color key representation along peaks. Marks at multicolor bars at the left of each figure show the threshold F-value above which the epistasis interaction is significant.

DISCUSSION

Fluctuation in Isoflavone Content and Composition

Basal levels of certain secondary metabolites may be quite variable even in genetically stabilized lines, particularly if their production is affected by a set of biotic and abiotic factors that can act as inducers or repressors. Eight genotypes were tested in field and controlled conditions to gain more knowledge on what might be the origin of the reported tremendous oscillations observed for isoflavone concentrations over environments (Eldridge et al. 1983; Wang and Murphy 1994; Tsukamoto et al. 1995; Hoeck et al. 2000; Mebrahtu et al. 2004; Caldwell et al. 2005; and Lozovaya et al. 2005). In general, plants tended to accumulate more isoflavones when growing in a controlled environment, probably because temperature can be kept in a more desired range. A series of factors, such as temperature, water, and soil nutrient condition are known to play an important role in isoflavone production, and could explain this difference. Although average values were maintained similar, 25.6ºC in GH versus 24.4ºC in both field locations, recorded maximums were 29.4, 32.2 and 32.8ºC in a three-month period (July-September) in GH, 38.7, 37.1 and 32.2ºC at BREC, and 40.0, 37.8 and 33.3ºC at DRC, suggesting that cooler temperatures exert a favorable effect in seed isoflavone accumulation, as previously reported (Tsukamoto et al. 1995; Caldwell et al. 2005; and Lozovaya et al. 2005). Not to be obviated is also the fact that the growth cycle in the greenhouse was longer than in the field, thus plants had more time to synthesize and accumulate metabolic compounds.

Genetic Control of Isoflavone Content

The variability observed for seed isoflavones in the evaluation of soybean lines was attributable not only to genetic causes but also to both the environment and g × e interaction. Highly significant environmental and g × e effects could also be found within the mapping population (Table 2.2). Remarkably, broad-sense heritability estimates (H^2) were intriguingly high (80 to 90%) for most of the isoflavone components, which indicates that genetic factors are the main contributors to the phenotypic variability. Elevated heritability values for isoflavones have been previously reported (Meksem et al. 2001; Chiari et al. 2004). With high H^2 estimates, one should not expect the variation observed on genetically identical individuals. In fact, in a population of RILs the only components contributing to H^2 are the additive and *aa* epistatic genetic variances. Thus, it is tempting to speculate that either the additive loci or the epistatic interactions or both might change their effects along with environmental changes, but then at the same time, the sum of these effects should be kept high to reach an elevated H^2. It could also be possible that the changes occur such that additive loci and epistasis counteract to affect seed isoflavone content, at least in part. Effectively, we observed that a decrease in phenotypic variance accounted for by the sum of additive QTL could be compensated with an increase in the epistatic contribution, and vice versa, as described below.

Nineteen quantitative trait loci with additive effects were found to affect genistein, daidzein and glycitein production. Among them, a major QTL was located in LG A1 (*gen-A1*, *dai-A1* and *tot-A1*) which accounted for up to 30.3% of isoflavone variation, depending on the model and trait (Tables 2.3, 2.4, 2.5 and 2.6). It was the only QTL that was detected by all the tested models and locations. Other important genomic

regions for the synthesis of genistein are *gen-E*, *gen-L*, *gen-D1b*, and *gen-M*. Daidzein shares a great number of enzymes in the phenylpropanoid metabolic pathway with genistein. Moreover, their relative seed content within the mapping population were 62% for genistein and 33% for daidzein, on average. Therefore, it is not surprising that four total-isoflavone genomic regions were coincident for both genistein and daidzein, i.e. at linkage groups A1, D1b, L, and M. Glycitein, however, is very distinct in many aspects, suggesting that a different branch of the pathway is probably involved in its synthesis, with distinct enzymes and/or differences in activation (Yu and McGonigle 2005). Glycitein is present in small amounts but displays more variability, the additive QTL are less numerous and usually with very little individual effects, and epistatic contributions are consistently high (Table 2.5 and Table 2.9). This variability could be explained because it exists mainly in the tiny embryo section of the seed, with practically no presence in cotyledons (Kim et al. 2007). Despite those barriers for QTL mapping, a major (*gly-E*) and five minor QTLs were found, suggesting that even minor QTL can be detected with adequate combinations of prediction models and population increased size.

Epistatic Network of Interactions

The ability for detecting aa epistatic interactions depends not only on the sample size but also on the heritability of the aa effect (h^2aa) (Mao and Da 2004). Traditionally, there have been two approaches to determine interacting loci. In the first method, the whole genome must be screened for the candidate intervals and then all possible interacting pairs are tested for significance. It is comprehensive but computationally demanding. The alternative approach relies on checking only a known set of loci introduced into the model. By using both approaches we mapped epistatic connections

for genistein, daidzein, glycitein, and total isoflavones at each location separately and combined. Genistein had fewer epistatic interactions and accounted for less phenotypic variation than for any other isoflavone, but still the contribution to the total variance (R^2) was 15.2 and 27.4% at DRC and BREC, respectively. Glycitein, on the other hand, had as much as 54.1% of its variation explained by interacting loci. Daidzein and total isoflavones showed intermediate values, with about one third of the trait variability due to epistasis. Remarkably, only one epistatic interaction was significant at both locations for all traits, proving that epistatic effects must be analyzed in particular environments (Fig. 2.3). In fact, regardless of the number and the magnitude of the effect, specific epistatic interactions are rarely consistent over environments. Epistasis has been considered as an important source of genetic variation in the adaptation to a changing environment (Whitlock 1995). We actually found that epistasis was very sensitive to environment, and epistatic networks were largely changed when plants grew in variable environments.

Importance of the Mapping Model

Variability observed in complex traits has overwhelmingly hindered mapping studies, especially if the trait is affected by many genes with little individual effects. Powerful statistical methodologies and advanced software are required to precisely isolate the loci involved. Epistatic networks of interacting loci with no additive effect by themselves add even more complexity. Therefore, it is not surprising that different mapping approaches may lead to very distinct sets of isoflavone QTL, even when the same phenotypic data were used (Kassem et al. 2004, 2006; Primomo et al. 2005), because only considering all possible sources of variation together with a large population the chance of recovering QTL increases. The wide range of variation

displayed by isoflavones suggests that only models that include most sources of variation will detect minor-effect QTL. Statistical methodologies have been developed to improve detection of loci with minor effects. For example, Mayer (2005) compared MIM and regression interval mapping and he found that MIM was more powerful and produced more accurate QTL positions and effect estimates. Similar results were obtained by Kao et al. (1999) when comparing IM, CIM, and MIM. We also proved that the use of prediction models able to include most sources of variation were more effective in detecting QTL with minor effects. Every time that some variation was removed from the error component and introduced in the model, a larger number of QTL were detected. This was particularly true for the minor QTL. For example, *gen-E*, *gen-F*, *dai-M*, and *tot-L* were discovered only by MCIM and MIM, the two models which include variation due to interacting loci. Conversely, out of seven total isoflavone QTL, a single marker analysis only detected four and six significant marker-trait associations ($P<0.05$) at DRC and BREC, respectively (data not shown). More importantly, as many as ten associations were found which turned out not to be relevant by other prediction methods. The importance of the statistical method is further remarked, for example, for *tot-D1b*, which was not detected by IM at any location (Table 2.6). When CIM was implemented, it was revealed, although only at BREC. Finally, it was detected in both locations when including epistasis as a source of variation by MIM.

In addition to inclusion of epistatic effects, best results were obtained when a larger sample size with data from different locations and within-location replications were entered in the analyses (MCIM). It can be hypothesized that a QTL would remain undiscovered if it had minor effects at both locations but did not reach the statistical

significance when individual locations were considered alone. Consequently, it would only be detected because of the increased statistical power due to a larger sample size. Effectively, including replicated data from all locations as individual observations might have a major impact on QTL detection for isoflavone concentration. For example, *gen-M* was only detected by MCIM in analysis of the combined data across locations. However, MCIM run for each location independently failed to detect the QTL at either location. This was repeatedly observed for other loci (data not shown). Inversely, if a QTL has a significant small effect at one location but has no effect at the other location, the significance of the QTL might be minimized when pooling the entries of both locations. This could be the case of *dai-L*, which was significant at DRC (MIM) but no longer remained detected by the combined MCIM. Alternatively, it might also be possible that MCIM, with a model considering a × e effects, was able to discern more clearly this environmental interaction than MIM, and thus not regarding it as significant. However, caution must be taken when interpreting mapping results, especially when the analysis is aimed at determining the significance of epistatic and environmental effects, due to the limited population size inherent to a practical development of mapping populations.

A common feature of the comparison was the increasing phenotypic variance explained by the models as more parameters were included. Thus, CIM explained a higher percentage of the variance than IM. Considering epistasis in the model increased the percentage dramatically. For example, daidzein variation explained with IM at BREC was 20.3%. By switching to CIM the percentage was 25.0%. Finally, MIM was able to raise the value to 65.6% by including variation due to epistasis. It might also happen that epistatic and main effects compensate each other to the point of maintaining total

isoflavone content between certain levels. That would explain the elevated H^2 observed for the traits. In fact, the sum of the variance that is explained for the additive model for genistein, daidzein and glycitein had a similar value in both locations indicating that even though the set of QTL may differ in a changing environment, the total variance that they account for remains constant. The sum of epistatic interactions, on the other hand, tends to produce very different values when the growing conditions change, which points to epistasis as the main cause of the variation experienced by isoflavones over environments.

CONCLUSIONS

The range of values of isoflavones is overwhelming, even for homozygous genotypes growing in the same year and location, which greatly complicates genetic studies. In this sense, model selection was shown to exert a tremendous influence for the detection of main and interacting QTL for complex traits. Consistently, when more source of variance is included, the model is able to highlight more QTL. In addition, although epistasis detection requires a large sample size, our results indicate that all measurements should be taken in a given environment because of the dependency of epistatic interactions to growing conditions. On the other hand, despite that environmental interaction is also possible for main QTL and that QTL effects depend on environment, combining data sources for several locations and within-location replications under the same model (MCIM) may increase the chance for their detection due to the increased statistical power. As a final conclusion, only considering exhaustive statistical methodologies that include epistasis, together with large mapping populations

growing in different environments (years and/or locations), and a dense genetic map with evenly distributed markers, the chance of recovering QTL improves for complex traits.

Table 2.1 Isoflavone seed content of different genotypes.

Lines	MG[a]	Genistein			Daidzein			Glycitein			Total isoflavones		
		BREC	GP	GH	BREC	GP	GH	BREC	GP	GH	BREC	GP	GH
PI-438489B	IV			2264±136			348±53			146±12			2553±198
PI-437654	III	1487±67	1655±177	1909±65	214±13	268±30	491±67	102±22	127±12	143±15	1803±102	2051±215	2656±247
RCAT Angora	II	1024±36	1070±125	1395±183	170±27	181±21	236±24	31±1	37±2	39±6	1227±60	1288±145	1653±343
AC 756	II	537±103	635±158	1191±88	88±14	121±28	229±40	26±2	33±4	34±6	652±118	734±93	1455±222
Essex	V	922±8	847±63	1119±93	177±9	156±26	232±33	50±1	46±6	57±11	1150±0	1068±98	1409±184
Forrest	V	986±146	1052±72	835±152	178±46	193±29	167±32	40±3	44±5	37±7	1206±194	1290±93	1040±308
Williams82	III	752±100	798±48	803±94	118±13	126±6	107±14	41±1	45±2	35±4	903±98	970±55	946±107
Magellan	IV	648±19	674±90	759±102	102±7	112±12	92±16	46±3	51±2	49±4	797±23	838±97	901±121

Measurements in µg per g of seeds ± standard deviation (SD). BREC, Bradford Research and Extension Center; GP, Grand Pass; GH, greenhouse. The assays and number of repeats are described in Material and Methods.

[a] Maturity Group

Table 2.2 Mean and range, ANOVA, random effects, and heritability over environments of Essex × PI 437654 mapping population.

				Effects[b]					
Trait	Mean ± SD[a]	Range[a]	CV(%)	G	E	G × E	H^{2c}	Skewness	Kurtosis
Genistein	844 ± 296	107–1980	18.8	<0.0001	<0.0001	<0.0001	0.83	0.15	-0.15
Daidzein	463 ± 198	73–1384	22.7	<0.0001	<0.0001	<0.0001	0.89	0.63	0.59
Glycitein	88 ± 48	14–342	26.4	<0.0001	<0.0001	<0.0001	0.43	1.65	3.35
Total Isoflavones	1393 ± 478	220–3008	18.3	<0.0001	<0.0001	<0.0001	0.86	0.18	-0.17

[a] Mean ± Standard deviation (SD) and range units are given in µg of isoflavones per gram of seeds.
[b] P values of random effects over environments: genetic (G), environmental (E), and G × E interaction.
[c] Heritability of the traits in the broad sense over environments.

Table 2.3 QTL bearing additive effect detected for genistein seed content by different models.

	MCIM (combined)				IM				CIM				MIM			
					BREC		DRC		BREC		DRC		BREC		DRC	
QTL[a]	Interval	Range[b]	A[c]	h^2(a)	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2
gen-A1	SATT236-SAT_271	29.4-36.4	95.8±9.0	5.8			6.1	25.3	3.5	9.4	6.3	22.1			9.4	21.3
gen-B2	SATT687-SAT_424	0.0-3.0	49.0±7.7	1.5					3.1	6.0			3.2	4.5		
gen-D1b	SAT_279-EACAMCTT123	0.0-10.0	59.5±7.8	2.3					3.8	8.4			8.4	9.6		
gen-E	SAT_112-SATT691	8.0-11.0	71.0±7.7	2.9									7.6	7.7	3.0	3.5
gen-F	SATT490-SAT_197	3.4-10.0	40.4±7.0	1.7									4.5	5.4		
gen-M	SATT697-SATT551	87.4-93.4	54.8±7.0	2.6												
gen-L	SAT_113-SAT_286	141.4-160.9	-66.5								3.0	6.9			4.0	6.4
	Sum of contributions of all QTLs in the model			16.8				25.3		23.8		29.0		27.2		31.2

[a] Name given to a particular QTL, *gen* for genistein content, followed by the linkage group and a number when more than one in the same LG. [b] Interval of confidence in centiMorgans with respect to the first marker in the LG. [c] Main additive effect in mg/g plus/minus standard error when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explain by the additive component of the QTL. R^2 percentage of variance accounted for by the QTL. LOD is the Log of odds (LOD score) associated to the QTL detected by each method. Only QTLs displaying R^2 or h^2(a) values in the corresponding column were detected by that particular model. MCIM for mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC).

Table 2.4 QTL bearing additive effect detected for daidzein seed content by different models.

MCIM (combined)					IM				CIM				MIM			
					BREC		DRC		BREC		DRC		BREC		DRC	
QTL[a]	Interval	Range[b]	A[c]	h^2(a)	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2
dai-A1	SATT236-SAT_271	22.5-36.4	69.5±5.7	9.2	5.3	20.3	5.7	19.7	5.7	13.9	6.7	18.8	7.8	18.0	7.5	15.5
dai-A2	SATT187-EAGGMCTT205	23.0-27.1	-18.2±4.6	2.4					3.2	5.4						
dai-D1b	SAT_279-EACAMCTT123	0.0-8.0	37.1±4.8	2.2					3.0	5.7			5.4	5.0		
dai-H	SCTT009-SATT541	0.0-7.0	45.1±5.0	3.8			4.0	9.4			3.2	5.8			5.8	7.9
dai-M	SATT697-SATT551	87.4-94.4	42.4±4.7	3.0									4.3	4.9		
dai-L	SAT_113-SAT_286	142.8-161.9	-27.54±5.13												5.0	6.5
	Sum of contributions of all QTLs in the model			20.7		20.3		29.1		25.0		24.6		27.9		29.9

[a] Name given to a particular QTL, *dai* for daidzein content, followed by the linkage group and a number when more than one in the same LG. [b] Interval of confidence in centiMorgans with respect to the first marker in the LG. [c] Main additive effect in mg/g plus/minus standard error when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. h(a)^2 is the heritability of the additive effect or percentage of variation that is explain by the additive component of the QTL. R^2 percentage of variance accounted for by the QTL. LOD is the Log of odds (LOD score) associated to the QTL detected by each method. Only QTLs displaying R^2 or h^2(a) values in the corresponding column were detected by that particular model. MCIM for mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC).

Table 2.5 QTL bearing additive effect detected for glycitein seed content by different models.

MCIM (combined)					IM				CIM				MIM			
					BREC		DRC		BREC		DRC		BREC		DRC	
QTL[a]	Interval	Range[b]	A[c]	h^2(a)	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2
gly-C2	SATT281-SATT291	41.0-48.6	-12.2±1.3	4.4			3.0	7.8			4.1	7.6	4.0	5.6	3.3	5.7
gly-E	SAT_112-SATT691	1.0-11.0	8.9±1.4	2.9					4.3	7.9			3.1	4.2		
gly-K	SATT552-SATT499	68.6-78.3	10.48												3.1	3.6
gly-G	SATT612-SAT_117	81.0-92.5	-13.51												4.7	7.3
gly-D2_1	SATT458-SAT_284	13.6-26.8	8.9						4.1	6.8			4.4	5.8		
gly-D2_2	SATT186-SAT_086	0.0-3.2	-7.59										3.2	4.4		
	Sum of contributions of all QTLs in the model			7.3				7.8		14.7		7.6		20.0		16.6

[a] Name given to a particular QTL, *gly* for glycitein content, followed by the linkage group and a number when more than one in the same LG. [b] Interval of confidence in centiMorgans with respect to the first marker in the LG. [c] Main additive effect in mg/g plus/minus standard error when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explain by the additive component of the QTL. R^2 percentage of variance accounted for by the QTL. LOD is the Log of odds (LOD score) associated to the QTL detected by each method. Only QTLs displaying R^2 or h^2(a) values in the corresponding column were detected by that particular model. MCIM for mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC).

Table 2.6 QTL bearing additive effect detected for total isoflavone seed content by different models.

MCIM (combined)					IM				CIM				MIM			
					BREC		DRC		BREC		DRC		BREC		DRC	
QTL[a]	Interval	Range[b]	A[c]	h^2(a)	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2
tot-A1	SATT236-SAT_271	29.4-36.4	200.4±14.4	8.1	3.5	14.9	7.3	30.3	3.5	9.5	8.2	26.3	9.9	16.6	17.4	30.0
tot-D1b	SAT_279-EACAMCTT123	0.0-7.0	110.4±11.0	2.4					5.0	9.5			6.9	6.7	5.6	3.2
tot-E	SAT_112-SATT691	8.0-11.3	76.8±12.0	1.9					3.3	6.4			3.1	3.8	3.7	3.1
tot-H	SCTT009-SATT541	0.0-10.0	59.1±13.0	2.1			3.2	8.0			3.2	6.3				
tot-L	SAT_286-SATT229	143.8-159.9	-91.3±12.8	2.3											7.4	7.4
tot-M	SATT697-SATT551	87.4-93.4	102.1±11.2	3.3					4.1	8.6			3.6	4.0		
tot-A2	SATT187-EAGGMCTT205	23.0-27.1	54.13±10.75										3.1	3.4		
	Sum of contributions of all QTLs in the model			20.0		14.9		38.3		34.0		32.6		31.1		43.7

[a] Name given to a particular QTL, *tot* for total isoflavones, followed by the linkage group and a number when more than one in the same LG. [b] Interval of confidence in centiMorgans with respect to the first marker in the LG. [c] Main additive effect in mg/g plus/minus standard error when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explain by the additive component of the QTL. R^2 percentage of variance accounted for by the QTL. LOD is the Log of odds (LOD score) associated to the QTL detected by each method. Only QTLs displaying R^2 or h^2(a) values in the corresponding column were detected by that particular model. MCIM for mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC).

Table 2.7. *Epistasis interactions accounting for genistein accumulation.*

	Interval i	Pos[a]	Interval j	Pos	AA ± SE[b]	LOD	$h^2(AA)/R^2$ [c]
	SATT687-SAT_424	1.0	SAT_279-EACAMCTT123	3.0	-33.4±8.6		1.2
	AW620774-SATT534	17.4	SATT239-SATT270	14.2	-42.6±7.6		1.7
MCIM	SATT490-SAT_197	5.0	SATT697-SATT551	91.4	-48.5±7.2		2.3
	SATT674-SAT_228	43.0	BF008905-SATT259	13.9	46.6±8.9		1.4
							5.4
	SATT329-EaggMcag363	75.7	SAT_284-SATT372	30.7	-39.38	1.9	4.0
	SATT329-EaggMcag363	75.7	SATT490-SAT_197	7.0	-34.06	1.5	2.2
MIM	SATT281-SATT291	52.7	SAT_403-EaggMctt569	32.0	50.44	3.1	6.4
BREC	SAT_305-SATT071	73.5	SATT614-SATT239	7.6	42.07	2.9	3.8
	SAT_284-SATT372	30.7	SATT614-SATT239	7.6	38.43	2.0	3.2
	SATT490-SAT_197	7.0	EaacMctc247-SATT175	73.4	-48.73	3.4	5.4
	SCT_199-SATT191	102.1	EaacMctc247-SATT175	73.4	-34.14	2.1	2.4
							27.4
	EaacMcac113-EacaMctt084	48.8	EacaMctt122-SATT082	77.4	-30.74	1.0	2.3
MIM	EacaMctt122-SATT082	77.4	SATT012-SATT612	78.0	-36.91	1.0	1.0
DRC	EaacMcac113-EacaMctt084	48.8	SATT367-SATT614	0.1	-36.29	1.4	1.4
	SCT_188-Satt490	4.9	SAT_113-SAT_286	145.8	38.05	1.3	2.7
	SATT012-SATT612	78.0	SCT_195-Satt159	4.4	-69.7	3.5	7.8
							15.2

Epistatic interactions detected by MCIM using combined dataset of two locations (BREC and DRC), MIM at BREC, and MIM at DRC. [a]approximate position of the interacting loci. [b]estimated additive by additive effect ± standard deviation when corresponds, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase genistein expression while a negative sign indicate that recombinant allele combinations increase genistein values. [c]heritability of the additive by additive interaction effect (for MCIM) or percentage of the phenotypic variation explain for the corresponding epistatic interaction (MIM). Below each column is the sum of all interactions for the whole epistatic model, in percentage.

Table 2.8. *Epistasis interactions accounting for daidzein accumulation.*

	Interval i	Pos[a]	Interval j	Pos	AA ± SE[b]	LOD	$h^2(AA)/R^{2}$ [c]
MCIM	Satt236-Sat_271	34.4	Satt187-Eaggmctt205	23.1	28.1±5.7		1.9
	Satt050-Satt385	21.5	Satt159-Eaggmctt098	4.4	-20.5±4.7		0.8
	Sct_067-Satt589	4.0	Satt131-Satt303	45.5	48.0±5.7		2.8
	Eaacmctc384-Sat_177	104.0	Sat_112-Satt691	10	38.4±5.8		2.4
	Eaacmctc063-Eacamctt419	0.0	Sat_158-Satt302	45.8	42.1±5.2		3.0
	Satt082-Satt301	83.0	Satt552-Satt499	65.4	37.8±6.5		2.2
	Satt352-Sat_308	47.4	Sct_190-Sat_293	95.2	43.6±7.1		1.3
	Satt666-Satt353	8.0	Satt314-Sat_158	30.5	35.7±5.3		2.2
							16.6
MIM BREC	Satt187-Eaggmctt205	23.2	Satt411-Sat_112	1.0	-23.85	1.7	1.9
	Sat_177-Satt577	109.0	Satt411-Sat_112	1.0	32.42	2.6	4.1
	Sat_067-Satt589	3.0	Satt352-Sat_308	46.4	41.98	3.5	6.6
	Sat_177-Satt577	109.0	Satt012-Satt612	80.8	29.76	2.1	2.9
	Satt050-Satt385	29.5	Sct_190-Sat_293	102.3	52.10	2.3	10.1
	Sctt009-Satt541	0.1	Sct_190-Sat_293	102.3	-32.06	1.4	4.2
	EaggMctt569-Satt131	36.0	Sat_293-Satt588	120.0	27.08	1.6	3.8
	Sat_197-Satt522	22.0	Satt567-Satt435	22.4	-30.35	1.8	4.1
							37.7
MIM DRC	EacaMctt084-EaacMctc379	58.1	Sat_106-AW781285	45.6	-32.49	1.5	3.1
	Sat_177-Satt577	109.0	Sat_112-Satt691	8.1	24.41	1.2	2.0
	Satt050-Satt385	29.5	Sctt009-Satt541	0.1	36.35	2.1	3.1
	EacaMctt084-EaacMctc379	58.1	Sat_268-EaggMctt164	32.1	-40.96	3.3	6.6
	Satt030-Satt269	2.7	Sat_268-EaggMctt164	32.1	-25.78	1.5	3.1
	EacaMctt084-EaacMctc379	58.1	Sct_190-Sat_293	102.3	30.09	1.1	3.4
	EaacMcac086-Satt187	23.1	Sat_286-Satt229	153.9	-48.76	3.0	5.3
	Satt050-Satt385	29.5	Sct_195-Satt159	4.4	-32.22	1.7	2.5
	Satt050-Satt385	45.6	EaggMctt098-Satt009	19.7	42.37	2.2	5.4
							34.5

Epistatic interactions detected by MCIM using combined dataset of two locations (BREC and DRC), MIM at BREC, and MIM at DRC. [a]approximate position of the interacting loci. [b]estimated additive by additive effect ± standard deviation when corresponds, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase daidzein expression while a negative sign indicate that recombinant allele combinations increase daidzein values. [c]heritability of the additive by additive interaction effect (for MCIM) or percentage of the phenotypic variation explain for the corresponding epistatic interaction (MIM). Below each column is the sum of all interactions for the whole epistatic model, in percentage.

Table 2.9. *Epistasis interactions accounting for glycitein accumulation.*

	Interval i	Pos[a]	Interval j	Pos	AA ± SE[b]	LOD	$h^2(AA)/R^{2\,c}$
	Sat_140-Eaacmctc178	19.6	Satt281-Satt291	45.6	-6.6±1.6		1.8
	Sat_337-Sat_140	8.2	Satt281-Satt291	45.6	8.8±1.5		0.9
MCIM	Satt179-Sat_106	27.7	Satt549-Satt339	78.6	7.5±1.8		4.4
	Sat_106-AW781285	45.6	Satt549-Satt339	78.6	10.2±1.7		1.4
	Satt263-Eaggmcag365	35.9	Satt188-Sat_282	45.7	6.2±1.4		1.4
							9.9
	EaggMcag363-Satt455	75.7	Sat_337-Sat_140	10.5	-5.39	1.2	2.8
	EacaMctt067-Satt484	133.9	Sat_337-Sat_140	10.5	7.18	2.2	3.5
	Satt089-Satt233	51.2	Satt281-Satt291	45.6	-8.13	3.6	4.5
MIM	Satt455-Satt228	100.1	Satt281-Satt291	45.6	8.77	3.3	4.1
BREC	Satt281-Satt291	45.6	Sat_106-AW781285	55.6	-8.23	3.9	3.4
	Satt281-Satt291	45.6	Satt558-Satt634	51.4	-8.53	4.0	4.0
	Satt089-Satt233	51.2	Satt186-Sat_086	0.1	-4.42	1.1	1.0
	Satt089-Satt233	51.2	Sctt009-Satt541	15.1	8.99	4.4	5.4
	Satt455-Satt228	100.1	Satt477-EaggMctt096	66.0	-7.86	2.6	4.2
							32.9
	EaagMcat279-Satt050	8.1	Sat_247-Satt519	57.3	-6.86	1.8	1.3
	Satt396-Sat_337	0.1	Sat_130-Sat_062	10.4	-14.24	6.6	7.8
	Sat_140-EaacMctc178	19.1	Sat_062-Satt281	45.6	8.83	2.3	2.7
	Satt396-Sat_337	0.1	Sat_272_2-Satt129	83.8	-7.47	1.9	2.1
MIM	Sat_247-Satt519	57.3	Satt612-Sat_117	89.2	-13.14	4.6	6.1
DRC	EacaMctt067-Satt484	133.9	Sat_268-EaggMctt164	32.1	-9.27	2.7	0.8
	Satt316-Satt357	68.1	Sat_268-EaggMctt164	32.1	-9.62	3.2	4.2
	Sat_289-Satt271	143.1	Satt435-Satt323	54.8	22.53	8.2	22.0
	Sat_272_2-Satt129	83.8	Satt188-Sat_282	44.0	-9.24	3.0	3.9
	Sat_158-Satt302	49.3	Satt188-Sat_282	44.0	11.23	4.2	3.2
							54.1

Epistatic interactions detected by MCIM using combined dataset of two locations (BREC and DRC), MIM at BREC, and MIM at DRC. [a]approximate position of the interacting loci. [b]estimated additive by additive effect ± standard deviation when corresponds, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase glycitein expression while a negative sign indicate that recombinant allele combinations increase glycitein values. [c]heritability of the additive by additive interaction effect (for MCIM) or percentage of the phenotypic variation explain for the corresponding epistatic interaction (MIM). Below each column is the sum of all interactions for the whole epistatic model, in percentage.

Table 2.10 *Epistasis interactions accounting for total isoflavones accumulation.*

	Interval i	Pos[a]	Interval j	Pos	AA ± SE[b]		$h^2(AA)/R^{2\,c}$
	Sat_279-Eacamctt123	0	Sat_112-Satt691	10	-52.4±12.3		1.5
	Sat_265-Eaagmcat279	6.9	Sat_286-Satt229	154.9	79.3±13.3		2.3
	Eacamctt067-Satt484	134.9	Satt319-Eaacmcac113	43.8	78.4±11.7		2.3
MCIM	Eacamctt419-Satt643	19.6	Sat_299-GMLPSI2	72.5	100.7±12.9		3.7
	Sat_090-Sat_417	41.6	Satt567-Satt435	33.4	-74.9±13.4		1.4
	Sat_117-Sct_199	99.6	Satt462-Eaacmctc382	107.2	-71.3±13.7		1.4
							11.1
	EaacMcac086-Satt187	23.1	Satt687-Sat_424	0.1	57.62	1.9	0.3
	EaagMcat279	31.4	Satt319-EaacMcac113	43.8	74.06	3.0	2.1
	Satt687-Sat_424	0.1	Satt319-EaacMcac113	43.8	-84.11	3.7	4.2
MIM	EaacMctc063-EacaMctt419	18.6	Sat_279-EacaMctt123	0.1	66.15	2.9	2.5
BREC	Satt236-Sat_271	31.4	Sat_403-EaggMctt569	35.3	-66.96	2.5	3.5
	Satt687-Sat_424	0.1	Sat_403-EaggMctt569	35.3	49.77	1.6	0.7
	EaacMctc063-EacaMctt419	18.6	GMLPSI2-Sct_189	73.4	39.20	1.0	1.1
	Sat_279-EacaMctt123	0.1	GMLPSI2-Sct_189	73.4	72.91	3.1	2.5
	Sat_403-EaggMctt569	35.3	Satt260-Satt475	82.6	83.04	3.7	5.0
	EaagMcat279-Satt050	8.1	Sat_286-Satt229	150.4	96.87	5.6	6.4
							28.3
	EaacMcac086-Satt187	23.1	Sat_112-Satt691	10.0	-65.16	1.6	1.9
	Satt372-Satt002	40.8	Satt012-Satt612	80.8	99.26	2.9	2.3
	Satt372-Satt002	40.8	Sct_199-Satt191	102.1	-143.25	5.5	6.6
MIM	Sat_279-EacaMctt123	0.1	Sctt009-Satt541	0.1	-64.25	2.1	1.1
DRC	EaacMctc063-EacaMctt419	18.6	Sat_419-Satt148	72.4	64.64	2.3	2.6
	Sat_265-Eaagmcat279	5.9	Sat_286-Satt229	154.9	109.11	4.3	6.1
	EaacMcac086-Satt187	23.1	Sat_286-Satt229	154.9	-81.87	3.1	2.5
	EaacMctc063-EacaMctt419	18.6	Satt306-Satt697	91.4	-70.22	2.6	1.1
	Satt012-Satt612	80.8	Sct_195-Satt159	4.4	-62.69	2.1	3.5
							27.7

Epistatic interactions detected by MCIM using combined dataset of two locations (BREC and DRC), MIM at BREC, and MIM at DRC. [a]approximate position of the interacting loci. [b]estimated additive by additive effect ± standard deviation when corresponds, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase total isoflavones expression while a negative sign indicate that recombinant allele combinations increase total isoflavone values. [c]heritability of the additive by additive interaction effect (for MCIM) or percentage of the phenotypic variation explain for the corresponding epistatic interaction (MIM). Below each column is the sum of all interactions for the whole epistatic model, in percentage.

REFERENCES

Beggs CJ, Stolzer-Jehle A, Wellmann E (1985) Isoflavonoid formation as an indicator of UV stress in bean (Phaseolus vulgaris L.) leaves: the significance of photorepair in assessing potential damage by increased solar UV-B radiation. Plant Physiol. **79(3)**:630-634.

Bennett JO, Yu O, Heatherly LG, Krishnan HB (2004) Accumulation of genistein and daidzein, soybean isoflavones implicated in promoting human health, is significantly elevated by irrigation. Agric. Food Chem. 52:7574-7579.

Caldwell CR., Britz SJ, Mirecki RM (2005) Effect of temperature, elevated carbon dioxide, and drought during seed development on the isoflavone content of dwarf soybean [Glycine max (L.) Merrill] grown in controlled environments. J. Agric. Food Chem. 53:1125-1129.

Chew V (1977) Comparisons among treatment means in an analysis of variance. USDA ARS/H/6.

Chiari L, Piovesan ND, Naoe LK, Jose IC, Viana JMS, Moreira MA, Barros EG (2004) Genetic parameters relating isoflavone and protein content in soybean seeds. Euphytica 138:55-60.

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol. Biol. **53**:733-743.

Doerge RW (2001) Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews **3**:43-52.

Eldridge A, Kwolek W (1983) Soybean isoflavones: Effect of the environment and variety on composition. J. Agric. Food Chem. **31**:394-396.

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait in line crosses using flanking markers. Heredity **69**:315-324.

Hill J, Becker HC, Tigerstedt PMA (1998) Quantitative and ecological aspects of plant breeding. Chapman and Hall, London.

Hoeck JA, Fehr WR, Murphy PA, Welke GA (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci. **40**:48-51.

Juenger TE, Sen S, Stowe KA, Simms EL (2005) Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. Genetica **123**:87-105.

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics **152**:1203-1216.

Kassem MA, Meksem K, Iqbal MJ, Njiti VN, Banz WJ, Winters TA, Wood A, Lightfoot DA (2004) Definition of soybean genomic regions that control seed phytoestrogen amounts. J Bio & Biotech 2004:1:52-60.

Kassem MA, Shultz J, Meksem K, Cho Y, Wood AJ, Iqbal MJ, Lightfoot DA (2006) An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits. Theor. Appl. Genet. **113**:1015-1026.

Kim JA, Hong SB, Jung WS, Yu CY, Ma KH, Gwag JG, Chung IM (2007) Comparison of isoflavones composition in seed, embryo, cotyledon and seed coat of cooked-with-rice and vegetable soybean (*Glycine max* L.) varieties. Food Chemistry **102**:738-744.

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLPs linkage maps. Genetics **121**:185-199.

Lawson CGR, Rolfe BG, Djordjevic MA (1996) Rhizobium inoculation induces condition-dependent changes in the flavonoid composition of root exudates from Trifolium subterraneum. Australian Journal of Plant Physiology. **23**:93-101.

Lozovaya VV, Lygin AV, Ulanov AV, Nelson RL, Dayde J, and Widhohm JM (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci. **45**:1934-1940.

McKhann HI, Paiva NL, Dixon RA, Hirsch AM (1998) Expression of genes for enzymes of the flavonoid biosynthetic pathway in the early stages of the Rhizobium-legume symbiosis. Adv. Exp. Med. Biol. **439**:45-54.

Mao Y, Da Y (2004) Statistical power for detecting epistasis QTL effects under the F-2 design. Genet. Sel. Evol. **37:**129-150.

Mayer M (2005) A comparison of regression interval mapping and multiple interval mapping for linked QTL. Heredity **94**:599-605.

Mebrahtu T, Mohamed A, Wang CY, Andebrhan T (2004) Analysis of isoflavone contents in vegetable soybeans. Plant Foods for Human Nutrition **59**:55-61.

Meksem K, Njiti VN, Banz WJ, Iqbal MJ, Kassem MyM, Hyten DL, Yuang J, Winters TA, Lightfoot DA (2001) Genomic regions that underlie soybean seed isoflavone content. J. Biom. & Biotech. 1:1 (2001) 38-44.

Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. Am. J. Clin. Nutr. **70**: (suppl) 439-S-450-S.

Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc. Natl. Acad. USA. **104**:17909-17915.

Ososki AL, Kennelly EJ (2003) Phytoestrogens: a review of the present state of research. Phytother. Res. **17**:845-869.

Peters NK, Frost JW, Long SR (1986) A plant flavones, luteolin, induces expression of Rhizobium meliloti nodulation genes. Science **233**:977-980.

Phillips PC (2008) Epistasis, the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews **9**:855-867.

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M, Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. **45**:2454-2462.

Rochfort S, Panozzo J (2007) Phytochemicals for health, the role of pulses. J. Agric. Food Chem. **55**:7981-7994.

Stafford HA (1997) Roles of flavonoids in symbiotic and defense functions in legume roots. Bot. Rev. **63**:27-39.

Subramanian S, Hu X, Lu G, Odelland JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol. Biol. **54:**623-639.

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol. **137**:1345-1353.

Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. Plant J. **48**:261-273.

Subramanian S, Stacey G, Yu O (2007) Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci. **12**:282-285.

Tsukamoto C, Shimada S, Igita K, Kudou S, Kokubun M, Okubo K, Kitamura K (1995) Factors affecting isoflavone content in soybean seeds: Changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J. Agric. Food Chem. 43:1184-1192.

Usui T (2006) Pharmaceutical prospects of phytoestrogens. Endocrine J. **53(**1):7-20.

Variyar PS, Limaye A, Sharma A (2004) Radiation-induced enhancement of antioxidant contents of soybean (Glycine max Merrill). J. Agric. Food Chem. **52**:3385-3388.

Vyn TJ, Yin X, Bruulsema TW, Jackson CC, Rajcan I, Brouder SM (2002) Potassium Fertilization Effects on Isoflavone Concentrations in Soybean (*Glycine max* L. Merr.). J. Agric. Food Chem. **50**: 3501-3506.

Wang H, Murphy PA (1994) Isoflavone Composition of American and Japanese Soybeans in Iowa: Effects of Variety, Crop Year, and Location. J. Agric. Food Chem. **42**:1674-1677.

Whitlock MC (1995) Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. **26**:601-629.

Wu X, Blake S, Sleper DA, Shannon G, Cregan P, Nguyen HT (2009) QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor. Appl. Genet.

Yang J, Zhu J (2005) Methods for predicting superior genotypes under multiple environments based on QTL effects. Theor. Appl. Genet. **110**:1268-1274.

Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics **23**:1527-1536.

Yi N, Xu S (2002) Mapping Quantitative Trait Loci with Epistatic Effects. Genet. Res. Camp. **79**:185-198.

Yu O, McGonigle B (2005) Metabolic engineering of isoflavone biosynthesis. Advances in Agronomy. **86**:147-190.

Zhang J, Subramanian S, Zhang Y, Yu O (2007) Flavone synthases from Medicago truncatula are flavanone-2-hydroxylases and are important for nodulation. Plant Physiol **144**:741-751.

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90**:10972-10976.

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics **136**:1457-1468.

Zeng ZB, Kao CH, Basten CJ (2000) Estimating the genetic architecture of quantitative traits. Genet. Res. **74**:279-289.

CHAPTER 3

Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds.

"*The environment is everything that isn't me"*
(Albert Einstein)

ABSTRACT

Soybean (*Glycine max* [L] Merr.) seed isoflavones have long been considered a desirable trait to target in selection programs for their involvement on fitness for humans and the plant *per se*. However, attempts to modify seed isoflavone content have not always produced the expected results because their genetic basis is polygenic and complex. Undoubtedly, the extreme variability that seed isoflavones display over environments has obscured our understanding of the genetics involved. In this study, a mapping population of RILs with three replicates was analyzed in two locations and in two years. We found a total of thirty five main-effect genomic regions controlling genistein, daidzein, glycitein and total isoflavones accumulation in seeds. The use of distinct environments permitted detection of a great number of environment-modulated and minor-effect QTL. Our findings suggest that isoflavone seed concentration is controlled by a complex network of multiple minor-effect loci interconnected by a dense epistatic map of interactions. We hypothesize that the magnitude and significance of the main and epistatic effects of the nodes in the network will vary depending on the genetic background and environmental conditions. In an attempt to unravel the genetic architecture underlying the traits studied, we searched on a genome-wide scale for

genomic regions homologous to the most important identified isoflavone biosynthetic genes. We identified putative candidate genes for several of the main-effect and epistatic QTL and for QTL reported by other groups.

INTRODUCTION

Considerable evidence has implicated isoflavones in the fitness of both humans and plants. A search of the literature reveals thousands of articles and subsequent reviews describing effects on human health associated with isoflavone consumption and possible molecular mechanisms of action (reviews: Rochfort and Panozzo 2007; Zhang and Yu, 2008). Within the plant itself, isoflavones play a critical role in defense against fungal pathogens (Subramanian et al, 2004, 2005), and they also are required for the establishment and perdurability of nodules in rhizhobium-plant symbiotic associations (Subramanian et al, 2006, 2007). Consequently, there is an increasing interest in altering the isoflavone content of soybean commercial varieties, which requires an understanding of the genetics governing their synthesis and accumulation. However, unmasking the genetics underlying isoflavone accumulation in seeds is challenging because: i) isoflavones display a broad range of variability over environments due to the many factors that affect their synthesis and accumulation (Eldridge and Kwolek, 1983; Wang and Murphy 1994); ii) many QTL with small individual effects contribute in an additive manner (Kassem et al, 2004, 2006; Primomo et al, 2005; Zeng et al, 2009; Gutierrez-Gonzalez et al. 2009); iii) epistatic interactions are responsible for a great proportion of the observed phenotypic variance (Gutierrez-Gonzalez et al. 2009); and iv) soybean features a complex genome which has undergone several whole genome duplication

events. As a result, tissue-differential expression or loss-of-expression or function of some of the resulting paralogs may have occurred.

Soybean seed isoflavone levels are subjected to a great oscillation because many biotic and abiotic factors influence their synthesis and accumulation (Tsukamoto et al. 1995; Stafford 1997; Dhaubhadel et al. 2003; Bennett et al. 2004; Lozovaya et al. 2005; Zhang and Yu, 2008). For example, in a four-location field trial a cultivar was found to fluctuate from 460 to 1950 $\mu g\ g^{-1}$ in its isoflavone levels in seeds. Even within the same environment and year, an almost 3-fold variation was reported for a single cultivar (Eldridge and Kwolek, 1983). Nevertheless, in spite of the environmental interactions, the control of isoflavone content in seeds is largely genetic (Hoeck et al. 2000; Mebrahtu et al. 2004; Gutierrez-Gonzalez et al, 2009; Murphy et al. 2009), and numerous minor-effect QTL have been found to determine soybean isoflavone amounts (Kassem et al, 2004, 2006; Primomo et al, 2005). Unfortunately, many of them were identified solely based upon a particular environment, location and/or year. QTL that are stable over multiple environments are more useful in breeding programs because they might contribute to a consistent increased phenotype under changing conditions.

A set of enzymes within the phenylpropanoid pathway are responsible for the biosynthesis of the three known soybean isoflavones: genistein, daidzein, and glycitein (Fig. 3.1). Phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxilase (C4H), and 4coumarate:CoA ligase (4CL) are the first enzymes in the pathway. They convert the amino acid phenylalanine into p-Coumaroyl-CoA. In the next step, the critical enzyme chalcone synthase (CHS) acts either alone or together with chalcone reductase (CHR) to form isoliquiritigenin and naringenin chalcone, respectively. Both are substrates of

chalcone isomerase (CHI), which converts them into liquiritigenin and naringenin. These two compounds are the precursors of the soybean isoflavones daidzein and genistein, formed after the action of the key enzyme isoflavone synthase (IFS). The enzyme flavanone 3-hydroxylase (F3H) competes with IFS in the utilization of naringenin, in a branch of the pathway leading to the formation of flavonols, condensed tannins and anthocyanins. Daidzein can also act as a substrate, and after a series of reactions catalyzed by the enzymes isoflavone hydroxylase (I2'H) and isoflavone reductase (IFR), among others, leads to the synthesis of glyceollins. The biochemical steps leading to the third soybean isoflavone, glycitein, are not entirely known, although isoliquiritigenin is likely a precursor (Yu et al. 2005).

Cellular systems are intrinsically complex and their components often interact with one another in such an elaborate way that they may never be entirely understood. At the gene level, those interactions are known as epistasis. Under the quantitative genetics point of view, an epistatic interaction is revealed when a particular combination of alleles at distinct loci produces a different (non-additive) phenotype than when they are found apart. Thus, it is the particular combination of alleles that results in a change in the phenotype, i.e. the same allele can produce different phenotypes in a different genetic background. Epistatic effects have often been disregarded in mapping studies and left as a background distortion (Carlborg and Haley, 2004). However, since additional knowledge has been garnered over the last several years, epistasis has passed from being considered the exception to the norm for many traits, and inter-gene interactions have been proven to account for a great proportion of the variability of complex traits, including isoflavones (Doerge 2001; Philips 2008; Gutierrez-Gonzalez et al. 2009). This 'extra' source of

variation has been suggested to play an important role in biological processes such as heterosis, fitness and adaptation to different environments, and subsequently to evolution and speciation in natural populations (Wright 1931; Doebley et al. 1995; Whitlock 1995; Weinreich et al, 2005; Malmberg and Mauricio, 2005; Melchinger et al. 2007). Despite its importance, detecting epistasis genome-wide is not straightforward in experimental populations, and is certainly more difficult than for main-effect QTL. Statistical power to detect pairwise epistatic interactions is lower than for main QTL because tests of significance must be conducted for two intervals rather than just one, and consequently a higher critical threshold per test must be applied to overcome the problem of multiple tests (Melchinger et al, 2007). This can be translated into small-effect interactions that would remain undetected unless a larger number of individuals are considered.

Interestingly, epistasis has been found to correlate with genomic complexity and the number of chromosomes (Sanjuan and Elena, 2006; Wilfert L. 2008). The genome of soybean is known to be quite complex, owing to two presumed recent whole genome duplications (recently reviewed by Shoemaker et al. 2008). As a result, a single-copy gene in Arabidopsis can typically be expected to be present as two or four homologous genes in soybean (Shoemaker et al. 2006). Gene duplication is frequently associated with either tissue-specific differential expression (a process termed subfunctionalization), the acquisition of a new mode of action (neofunctionalization), or loss of expression or function of one or more copies (pseudogene formation) (Doyle et al. 2008). Naturally, phenylpropanoid genes are not oblivious to this complexity and attempts to genetically modify the pathway have sometimes led to unexpected results. For example, over-

expression of key enzymes such PAL, CHS, or IFS, either independently or combined, have failed to enhance isoflavone accumulation (Koes et al. 1994; Zernova et al 2009).

Over 50 loci have been reported in the few QTL mapping studies conducted on isoflavones (Kassem et al. 2004, 2006; Primomo et al, 2005; Zeng et al. 2009). However, a great majority has not been confirmed by using different parental lines, locations, or years. Attempts to validating the QTL have failed in part due to the nature of isoflavone QTL, with predominant minor main-effect, for which uncovering the experimental design needs to be optimized with, for instance, large number of individuals, enough and evenly distributed markers, and an adequate mapping methodology. We previously used the Essex × PI437654 cross to study the genetic control of seed isoflavone content (Gutierrez-Gonzalez et al. 2009). Herein, we first added data from one additional year and two additional locations with three replications each to perform a comprehensive 2×2×3-factor QTL analysis aimed at discovering not only the most stable main and epistatic QTL over environments, but also the QTL environment-modulated. Second, we accomplished an exhaustive search for candidate genes with homology to important genes for isoflavone synthesis. Genes were placed onto the genetic map together with the QTL identified in an attempt to find candidate genes and to begin to decipher the genetic network underling soybean isoflavone seed content.

MATERIALS AND METHODS

Plant Material and Growing Conditions

In a previous study we used a cross between Essex (low-isoflavone content) and the plant introduction PI 437654 (high-isoflavone content) to map isoflavone-content QTL (Gutierrez-Gonzalez et al. 2009). The same 205 F_7-derived recombinant inbred lines (RILs) mapping population was planted in the following year at the same two locations: University of Missouri Bradford Research and Extension Center (BREC, 36º 58' N) and the University of Missouri Delta Research Center (DRC, 36º 44' N), with three replications per location. Plants were maintained under irrigated conditions until physiological maturity. The four different environments (2 years × 2 locations) were subsequently used to perform an ample QTL mapping analysis, for which seed samples were harvested from a pool of at least three plants.

Isoflavone Extraction and Quantification

Genistein, daidzein, and glycitein amounts were determined as detailed in Gutierrez-Gonzalez et al, 2009. Briefly, ~ 20 seeds were ground to a fine powder and extracted with 7 mL of 80% methanol at 55 °C for 2 h, vortexing every 30 min. The supernatant was cleaned using Fisherbrand 0.45 μm 25 mm nylon syringe filters (Fisher Scientific, Pittsburgh, PA). 10 μl of the filtered extraction was used for reverse-phase HPLC on an Agilent 1100 HPLC system (Santa Clara, CA). Separation and elution were performed by an 18 min linear gradient starting with 20% methanol / 80% 10mM ammonium acetate (v/v) (pH 5.6) and finishing with 100% methanol at 1 ml/min. A RP-C18 Lunar C2 column was used (Phenomenex, La Jolla, CA). Metabolites were detected

by photodiode array. Identification and quantification of each isoflavone component were based on available standards (Indofine Chemical Co., Somerville, NJ).

Statistical analysis, Linkage Map and QTL Analysis

Statistical analysis was performed using the SAS STAT 9.1 program (SAS Institute Inc., Cary, NC). For the ANOVA, the pooled linear model contained the effect of environment, replication (environment), genotype, and environment × genotype Effects were tested using PROC GLM. Heritability in the broad-sense over environments (H^2) was calculated according to Hill et al. (1998) for which the variance components were determined by the PROC Mixed. The linkage map was previously described in Wu et al. 2009, and contained a total of 276 markers (SSR and AFLP) distributed on 26 linkage groups. The mixed-model based composite interval mapping implemented in QTLNetwork v2.0 (Institute of Bioinformatics, Zhejiang University, Hangzhou, China) was used for the QTL mapping analysis and run with two-years input data and three replications per location all combined in a 2×3×2 scheme. Critical F-value was assessed by permutation test using 1000 permutations. QTL effects were estimated using Markov chain Monte Carlo method. Genome scan was performed using 10 cM window size and 1 cM walk speed. Candidate interval selection, epistatic effects, and putative QTL detection were calculated with an experimental-wise Type I error of α=0.05, α=0.001, α=0.001, respectively.

Candidate gene identification

A list of National Center for Biotechnology Information (NCBI) entries with domains or features known to be involved in isoflavone synthesis was downloaded

(http://www.ncbi.nlm.nih.gov/) and assembled (Supplementary Table 5). Nucleotide sequence or predicted protein sequence for these genes were used in BLAST and tBLASTn searches, respectively, against the soybean whole genome sequence version Glyma1.0 (http://www.phytozome.net/soybean). An expected value threshold of 0.1 was used for searches, and all entries were manually evaluated. Regions which featured only a small match (<150 bps) were excluded. Putative pseudogenes were also included if within or in the proximity of a QTL. The nearest known SSR or SNP markers adjacent to the identified gene model were identified using the generic genome browser on www.soybase.org. In cases of a match, sequence was evaluated by recursive BLAST searches against NCBI (http://www.ncbi.nlm.nih.gov/) using the Viridiplantae nucleotide and protein databases to assign Glyma1 gene models to known genes. In the majority of cases we were unable to assign a specific previously identified gene; these entries have been labeled "putative" The complete list of identified gene models is present in Supplementary Table 5.

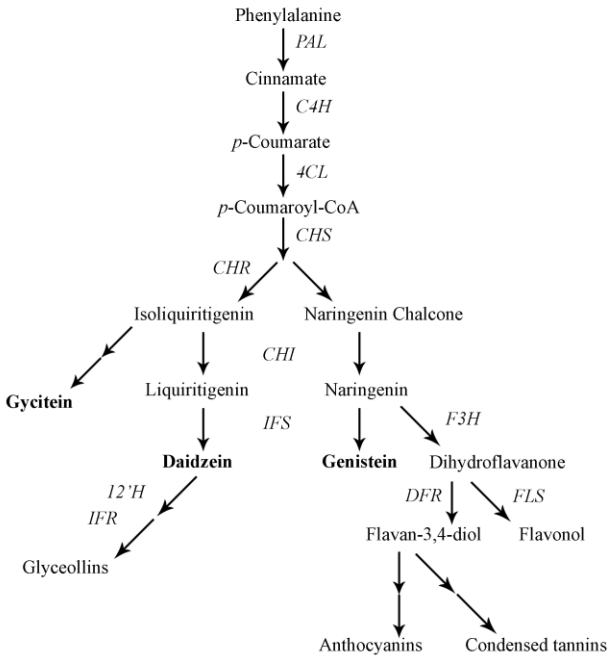

Fig. 3.1. Schematic representation of the phenylpropanoid pathway (adapted from Yu and McGonigle, 2005). Phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxilase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), chalcone reductase (CHR), chalcone isomerase (CHI), isoflavone synthase (IFS), flavanone 3-hydroxylase (F3H), dihydroflavonol reductase (DFR), flavonol synthase (FLS), isoflavone hydroxylase (12'H), and isoflavone reductase (IFR).

RESULTS

Genetic and Phenotypic Variation within Mapping Population

Analysis of variance (ANOVA) and heritabilities in the broad sense (H^2) were calculated over the combined four environments as they are good indicators of the origin of the variation within segregating mapping populations (Table 3.1). Results confirmed that the accumulation of isoflavones in soybean seeds is highly influenced by genetic (G), environmental (E), and G × E interaction effects ($P<0.0001$). Furthermore, the high heritability values found suggested that, despite of the environmental influence, most of

the phenotypic variation is determined by genes. In addition, because H^2 estimated on RILs encompasses only additive and additive-by-additive epistatic genetic variances, the involved genes could be acting alone or interacting. Frequency distribution and normal-distribution parameters of isoflavone seed content between locations and years clearly indicated quantitative inheritance of these traits (Table 3.1, Fig 3.2 and 3.3). Although parental lines did not greatly differ in genistein, daidzein and total isoflavones in some environments (Fig 3.2 and 3.3), considerable transgressive segregation was found, which indicated that both parents bear positive-effect alleles for isoflavone synthesis, and ultimately suggested that an elevated number of QTL might be segregating and likely to be detected in the mapping analysis.

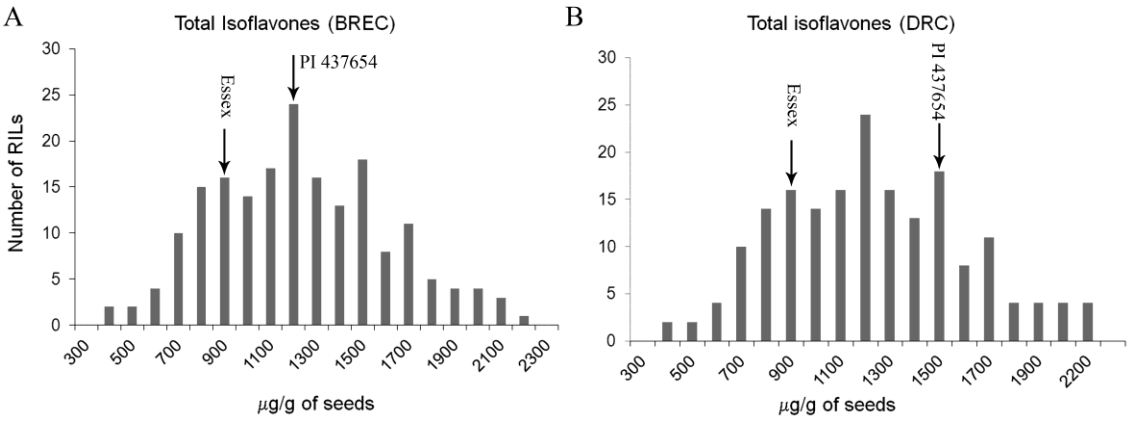

Fig. 3.2. Distribution of averaged total isoflavones in three replications of Essex × PI 437654 RILs growing in two field locations, BREC (A) and DRC (B) in year 2007. See figure 3 for individual isoflavones (for year 2006 data see Chapter 2). Arrows indicate the position of the two parental lines. Horizontal axis shows each particular isoflavone seed content in μg/g of seeds.

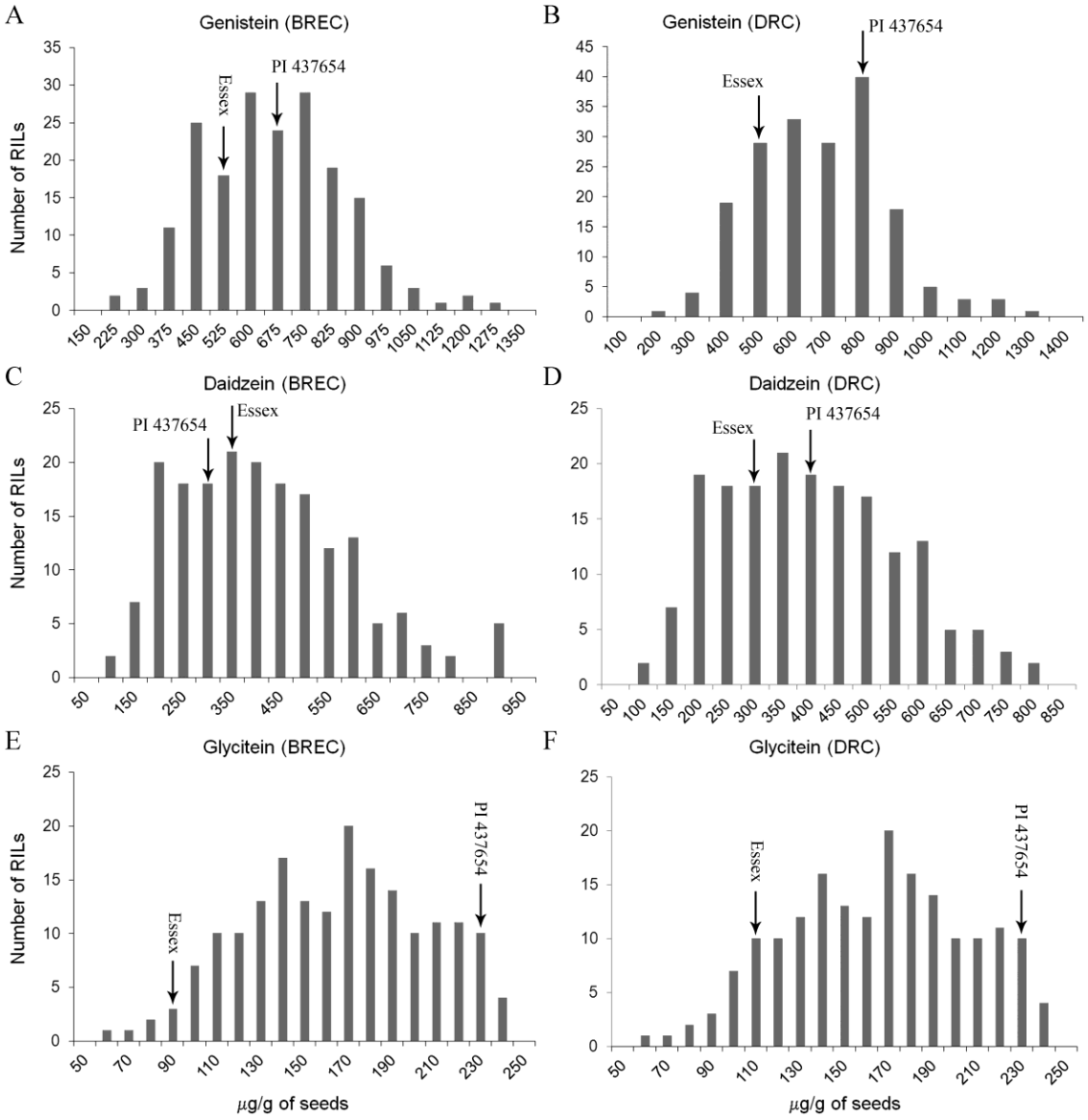

Fig. 3.3. Distribution of average genistein (A, B), daidzein (C, D), and glycitein (E, F), in three replications of Essex × PI 437654 RILs growing in two field locations, BREC (A,C, E) and DRC (B, D, F) in year 2007. Arrows indicate the position of the two parental lines. Horizontal axis shows each particular isoflavone seed content in μg/g of seeds.

QTL Mapping shows that numerous minor-effect loci control isoflavone amounts

Developing RILs may be time-consuming but allows for testing the same genetic background under different conditions. Seed isoflavone content is highly dependable on the environment (Table 3.1). Hence, it is not surprising that very distinct sets of QTL data have been reported when the same mapping populations were independently analyzed under different conditions (Primomo et al. 2005; Gutierrez-Gonzalez et al. 2009). In this study, data from four different environments were used to perform an integrated analysis by a mixed linear model. Running data from different locations together has been proven to be a powerful tool for discovering minor-effect QTL in complex traits (Gutierrez-Gonzalez et al., 2009), and might also allow the identification of the most stable QTL over the pooled locations. As a result, ten additive main genomic regions were found to influence genistein accumulation in seeds (Table 3.2 and Fig 3.4). A major QTL was found in chromosome 5 (Gm05, following the new nomenclature for soybean chromosomes) named *qGEN5*, with an estimated heritability of the additive effect, $h^2(a)$, of 5.5%. Other identified important regions were in Gm02, Gm15, Gm13, Gm19 and Gm07, which were named *qGEN2*, *qGEN15*, *qGEN13*, *qGEN19*, and *qGEN7*, respectively. All were previously reported (Gutierrez-Gonzalez et al. 2009). Importantly, two newly identified loci were found in Gm12 and Gm20, *qGEN12* and *qGEN20*. Two other regions explained genistein accumulation although only in one year at BREC: *qGEN4* and *qGEN2_2*. The loci *qGEN5*, *qGEN2*, and *qGEN19* also displayed an additive × environment (A × E) interaction effect, reflecting differences in accumulation over locations (Table 3.3.1). For daidzein, nine different regions were exerting influence on its accumulation (Table 3.2 and Fig 3.4). Two key loci were found in Gm05 and Gm12,

qDAI5 (h^2(a)=8.3%) and *qDAI12* (h^2(a)=4.3%), respectively. Other influencing loci were mapped in Gm08, Gm04, Gm01, Gm02, and Gm16: *qDAI8*, *qDAI4*, *qDAI1*, *qDAI2*, *qDAI16*, respectively, and two in Gm17, *qDAI7* and *qDAI7_2*. Apart from their additive effect, *qDAI2*, *qDAI12*, and *qDAI17_2* also showed a significant A × E interaction effect (Table 3.3.1). Glycitein accumulation was found to be influenced by a total of six different genomic regions (Table 3.3 and Fig 3.4). Despite of being the most variable isoflavone in soybean seeds (Table 3.1) several minor-effect QTL were found. Among them, *qGLY15* and *qGLY6* explained a greater percentage of the variation than any other QTL. Newly found QTLs were detected in Gm05, Gm06, Gm02, and Gm09 (*qGLY5*, *qGLY6*, *qGLY2*, and *qGLY9*). Reflecting the variability of the trait, five out of six loci had an A × E interaction effect (Table 3.2). In the latter instance, total isoflavone content was mapped as the sum of the three soy isoflavones. As a result, ten genomic regions were considered as influencing isoflavone content in seeds. The most influencing locus was again placed in Gm05 (*qTOT5*). This was previously reported, as well as they were *qTOT2*, *qTOT15*, *qTOT12*, *qTOT19*, and *qTOT7*. Newly discovered regions were found in Gm04 (*qTOT4*), Gm02 (*qTOT2_2*), and Gm16 (*qTOT16*). When considering each location independently, another locus was detected in Gm08 at BREC (*qTOT8*). Five QTLs also bared a component of A × E interaction effect (Table 3.2).

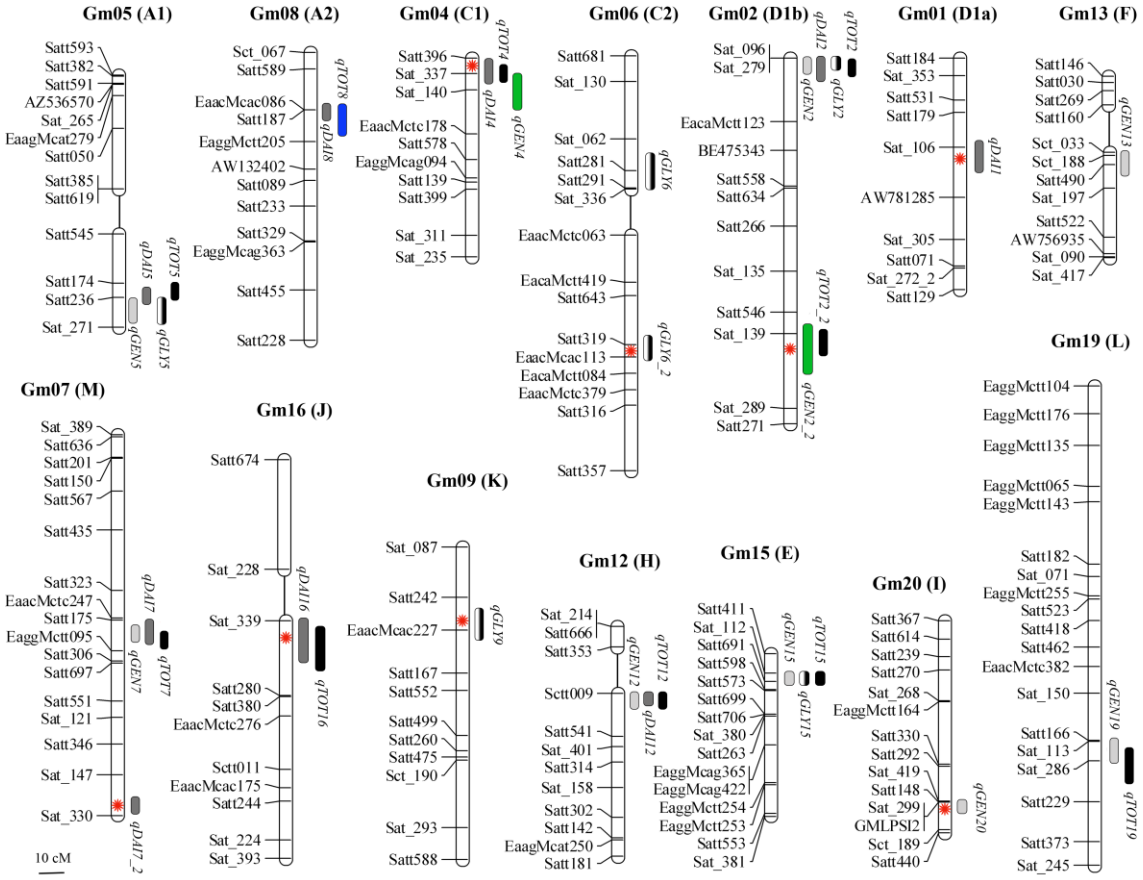

Fig. 3.4. Linkage group map summarizing QTL locations detected for distinct isoflavone species. Light grey ovals indicate loci associated with genistein, dark grey ovals indicate loci associated with daidzein, black ovals indicate loci associated with total isoflavone, and hatched ovals indicate loci associated with glycitein. Blue and green color ovals indicate particular loci discovered only in one location, blue for total isoflavone content at BREC in year 2006 and green at BREC in 2007. Flanking and other key markers used for linkage analysis are depicted at the left side of the linkage group. The name of the QTL, shown aside each oval, is a composite of the influenced trait: genistein (*gen*), daidzein (*dai*), glycitein (*gly*) and total of isoflavones (*tot*), followed by the chromosome number. Red asterisks show newly discovered QTLs while no asterisks indicate confirmed QTL.

Epistasis is decisive in determining seed isoflavone amounts

Accumulated evidence suggests that gene-gene interactions may be accountable for a great part of the phenotypic variation observed in complex traits. Particularly, additive × additive (AA) epistatic interactions have been proven to be responsible for a great part of isoflavone seed amounts in certain environments, even more than main QTL (Gutierrez-Gonzalez et al. 2009). Despite its interest, mapping epistasis in complex traits is a daunting task because of the great number of pair-wise combinations implicated and their small individual effects. Similarly to what occurred for additive main-effect QTL, performing a combined analysis over different environments allows not only recovery of a great number of small-effect AA digenic epistasis, but also the more stable-over-environment, and the interactions with an A × E component. A total of fourteen, twenty one, seven, and eight AA epistatic interactions were found to influence genistein, daidzein, glycitein and total isoflavones accumulation, respectively (Tables 3.4 and 3.5). In addition to the AA epistatic main-effect, some of the interactions also possessed an AA × E interaction effect (Table 3.8 and 3.9). The importance of epistasis was most revealed for daidzein, for which the percentage of the variation explained for by the sum of epistatic interactions was 22.7%. Importantly, the maximum contribution by any particular interaction was only 3.6%. All chromosomes had at least one epistatic connection preferentially with a distinct chromosome. Interactions between two loci within the same chromosome were rare. Overall, our results suggested that epistasis is a major determinant of phenotypic variance for isoflavone seed contents, and that many small individual interactions contribute the total effect.

Isoflavone synthesizing genes as candidate genes for additive QTL

The recent release of the whole genome sequence Glyma1.01 assembly (http://www.phytozome.net/soybean) for Williams 82 provides a powerful tool with which to interrogate QTL data. Previously reported genes for isoflavone biosynthesis (Gutierrez-Gonzalez, unpublished; Zhang and Yu, 2009) were used in BLAST searches against the whole genome sequence to identify homologous regions in the genome with assigned or putative functions (Table 3.5). All twenty soybean chromosomes had regions sharing a high percentage of homology with genes of known function in the phenylpropanoid pathway (Fig 3.5). In an effort to add more information onto the network map, we also included other reliable QTL reported earlier by our and other groups (Kassem et al. 2004, 2006; Primomo et al. 2005; Gutierrez-Gonzalez et al. 2009). In total, out of the twenty soybean chromosomes that form the soybean genome, nineteen were found bearing at least one additive QTL accounting for seed isoflavone accumulation, reflecting the wide-spread distribution of isoflavone influencing loci. In addition, to make the genetic map as informative as possible, all homologous sequences found during the blast search were also placed on the map at their approximate positions, including the ones with only putative function or potential pseudogenes. Candidate genes were considered those falling in the interval of confidence (IC) of a main-effect QTL, assuming the limits of the mapping resolution.

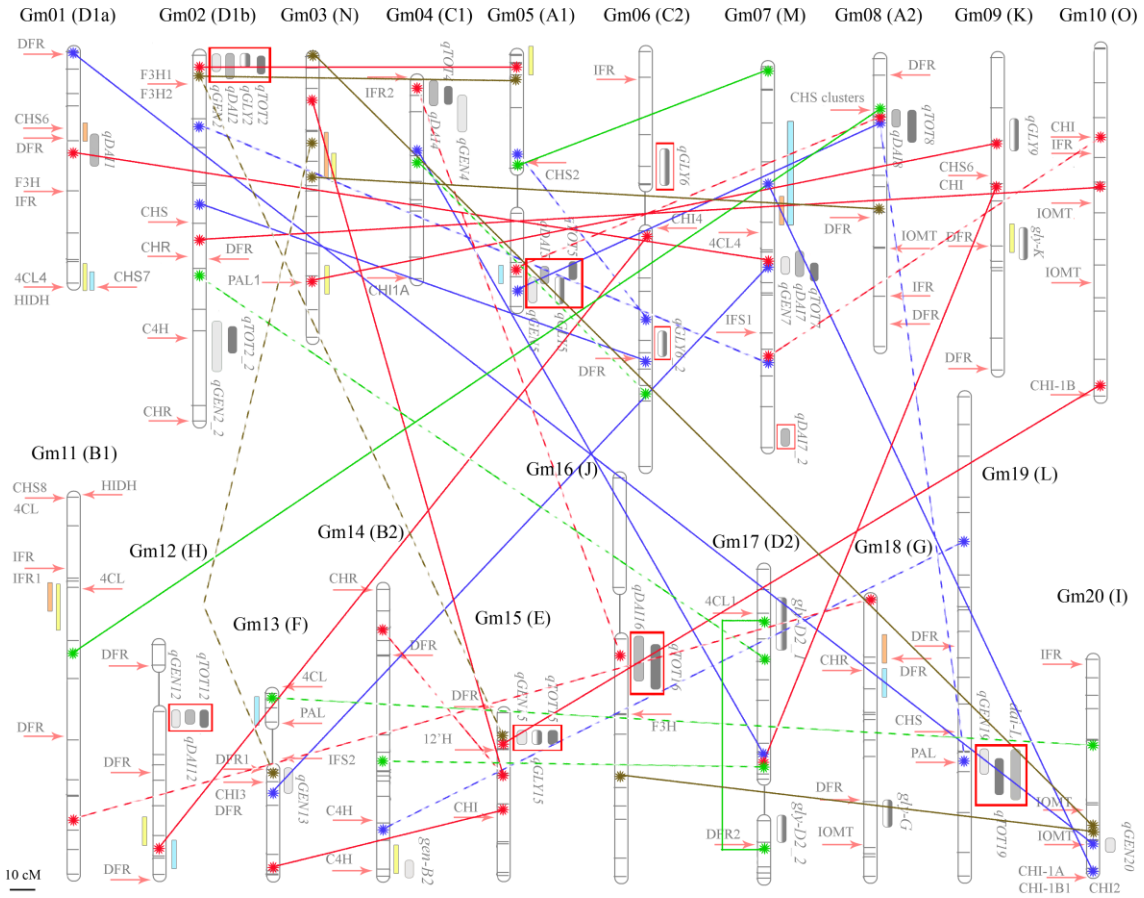

Fig. 3.5. Network of main-effect and interacting QTL. Homologous genes are located on the chromosomes by a red arrow and their abbreviated name. Light grey ovals indicate loci associated with genistein, dark grey ovals indicate loci associated with daidzein, black ovals indicate loci associated with total isoflavone, and hatched ovals indicate loci associated with glycitein. The name of the QTL, shown aside each oval, is a composite of the influenced trait: genistein (*gen*), daidzein (*dai*), glycitein (*gly*) and total of isoflavones (*tot*), followed by the chromosome number. Red squares surround loci with effect-by-environment interaction. In an effort to make the map more informative QTL reported for ours and other groups: (Primomo et al 2005) blue squares, (Kassem et al 2004) yellow and (Kassem et al 2006) light orange. Lines indicate en epistatic interaction between the interconnected QTL: blue for genistein, red for daidzein, green for glycitein, and brown for total isoflavone content. Dotted lines reflect epistasis with effect-by-environment interaction component.

Accordingly, several matches were found between QTL and biosynthetic genes. In Gm01, *qDAI1* and another QTL reported for glycitein synthesis (Kassem et al. 2006) overlapped a copy of chalcone synthase (*CHS6*) and a copy of dihidroflavonol reductase (*DFR*). It has been suggested that *CHS6* is present as a single copy in Gm09 (Matsumura et al. 2005; Tuteja and Vodkin 2008). However, BLAST searches revealed a second matching locus in Gm01 (Glyma01g22880.1), which features only few SNPs when compared to the NCBI entry for *CHS6* (Figs. 3.6-3.9). Another region at the end of the chromosome was previously reported to account for glycitein seed content (Kassem et al. 2004; Primomo et al. 2005). This locus embraced three phenylpropanoid genes (*CHS7*, *4CL4*, and *HIDH*), suggesting that the three could act together to enhance isoflavone contents. If a single genomic region harbors QTL for genistein, daidzein, glycitein and total isoflavones, it is certainly a good candidate to be considered for marker assisted selection. Chances are that the underlying gene could be an enzyme acting early in the pathway or a *trans*-acting factor, which impacts expression or activity of one or more of these genes. One such region is found at the beginning of Gm02 corresponding to two *F3H* genes (*F3H1* and *F3H2*). In the same chromosome, two other QTL were found to explain genistein and total isoflavones (*qTOT_2* and *qGEN2*). A copy of *C4H* is centered in the IC, which presumably would account for the QTL. Another *C4H* copy on Gm14 is found within the interval of the QTL *gen-B2*, and also of two other QTL for genistein and daidzein previously reported (Kassem et al. 2004). In Gm03 another QTL, in this case for glycitein, reported by the same group (2004 and 2006) coincides with a copy of other important gene, *PAL1*. At the beginning of Gm04 three QTL (*qGEN4*, *qDAI4*, and *qTOT4*) were in the same region in which an isoflavone reductase copy (*IFR2*) was

mapped. The IC of locus *qGLY6_2* in Gm06 also appears to be covering a *DFR* gene. Several genomic regions in Gm07 account for the synthesis of distinct isoflavones (*qGEN7*, *qDAI7*, *qTOT7*, and *qDAI7_2*). However, despite the observation that critical isoflavone genes (*IFS1* and *4CL4*) are present in this chromosome, those QTL remain without a clear candidate, and so do two additional regions reported by other groups. An extensive area of about 50 cM was shown to influence genistein, daidzein, glycitein and total isoflavone accumulation (Primomo et al. 2005). This region was apparently narrowed down to ~15 cM by another group (Kassem et al. 2006), because the latter is included in the former and both accounted for the same isoflavone compounds. The *4CL4* gene is less than 5 cM apart from this region and about 10 cM from *qGEN7*, *qDAI7*, and *qTOT7*. Numerous QTL reported for this chromosome suggests the presence of a large number of polymorphisms associated with isoflavone synthesis.

In a previous publication a QTL for glycitein was found in Gm09 (Gutierrez-Gonzalez et al. 2009), which matched almost exactly a locus reported by Kassem et al. (2004) for daidzein. A *DFR* gene is within their associated IC, although DFR is not directly involved in daidzein biosynthesis. In Gm11, a region identified to affect glycitein content in seeds (Kassem et al. 2004, 2006) was found to overlap a *4CL* homologous. A chalcone isomerase gene (*CHI3*) and two dihidroflavonone reductase homologues (*DFR* and *DFR1*) were located in the region identified by *qGEN13*, in Gm13. An isoflavone synthase gene (*IFS2*) in the vicinity of *qGEN13* could also be a candidate for the QTL. In this chromosome another QTL was reported for glycitein (Primomo et al. 2005), with a *PAL* and *4CL* copies lying within the IC and nearby the QTL (<5 cM), respectively. One of the two *C4H* homologous copies in Gm14 is within the genomic area delimited by

gen-B2 (Gutierrez-Gonzalez et al. 2009) and by two other QTL for genistein and daidzein (Kassem et al. 2004).

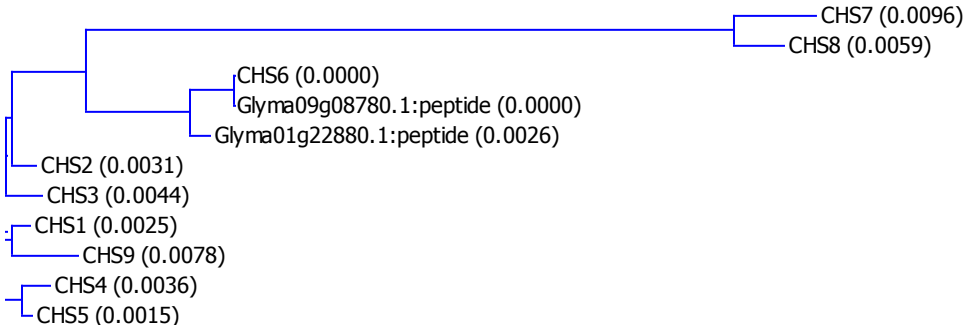

Fig. 3.6. Neighbor-joining phyllogenetic tree generated using AlignX (Invitrogen) using NCBI soybean chalcone synthase protein entries (NCBI entries: CHS1 ABB30178; CHS2 CAA46590; CHS3 X53958; CHS4 CAA36317; CHS5 AAB01004; CHS6 AAA33951; CHS7 AAA33950; CHS8 AAO67373; CHS9 ABQ63059) and the two putative CHS6 predicted proteins present in Glyma1.01 (Glyma01g22880.1 and Glyma09g08780.1).

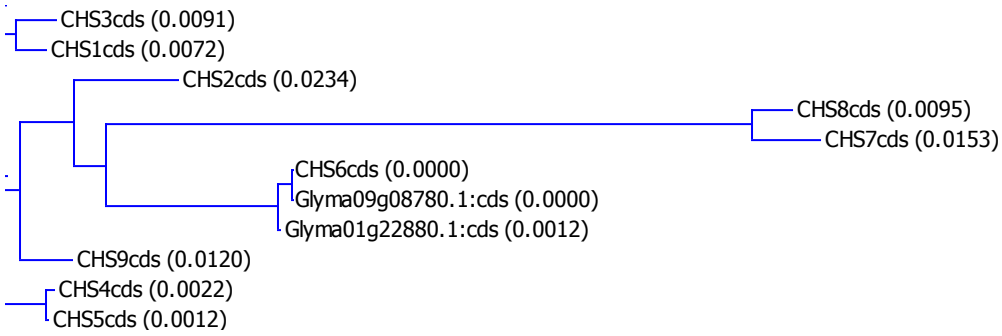

Fig. 3.7. Neighbor-joining phyllogenetic tree generated using AlignX (Invitrogen) using NCBI soybean chalcone synthase coding sequence entries coding sequence (NCBI entries: CHS1 DQ239918; CHS2 X65636; CHs3 X53958; CHS4 X52097; CHS5 L07647; CHS6 L03352; CHS7 M98871; CHS8 AY237728; CHS9 EF623853) and two putative CHS6 predicted coding sequences present in Glyma1.01 (Glyma01g22880.1 and Glyma09g08780.1).

Fig. 3.8. Protein alignment of NCBI soybean chalcone synthase VI protein sequence and the two putative CHS6 predicted predicted proteins present in Glyma1.01 (Glyma01g22880.1 and Glyma09g08780.1).

```
                                  1                                                  50
L03352      CHS6cds      (1)  ATGGTGAGTGTTGAAGAGATTCGTAAGGCACAACGTGCAGAAGGCCCTGC
Glyma01g22880.1:cds      (1)  ATGGTGAGTGTTGAAGAGATTCGTAAGGCACAACGTGCAGAAGGCCCTGC
Glyma09g08780.1:cds      (1)  ATGGTGAGTGTTGAAGAGATTCGTAAGGCACAACGTGCAGAAGGCCCTGC
                                 51                                                 100
L03352      CHS6cds     (51)  CACCGTGATGGCTATTGGCACGGCCACTCCTCCCAACTGCGTGGATCAGA
Glyma01g22880.1:cds     (51)  CACCGTGATGGCTATTGGCACGGCCACTCCTCCCAACTGCGTGGATCAGA
Glyma09g08780.1:cds     (51)  CACCGTGATGGCTATTGGCACGGCCACTCCTCCCAACTGCGTGGATCAGA
                                101                                                 150
L03352      CHS6cds    (101)  GTACCTATCCTGACTATTATTTCCGCATCACCAACAGTGACCACATGAAC
Glyma01g22880.1:cds    (101)  GTACCTATCCTGACTATTATTTCCGCATCACCAACAGTGACCACATGACC
Glyma09g08780.1:cds    (101)  GTACCTATCCTGACTATTATTTCCGCATCACCAACAGTGACCACATGAAC
                                151                                                 200
L03352      CHS6cds    (151)  GAGCTCAAAGAAAAGTTCAAGCGCATGTGTGATAAGTCAATGATTAAGAA
Glyma01g22880.1:cds    (151)  GAGCTCAAAGAAAAGTTCAAGCGCATGTGTGATAAGTCAATGATTAAGAA
Glyma09g08780.1:cds    (151)  GAGCTCAAAGAAAAGTTCAAGCGCATGTGTGATAAGTCAATGATTAAGAA
                                201                                                 250
L03352      CHS6cds    (201)  ACGATACATGTACTTAAATGAAGAGATCCTGAAGGAGAATCCAAGTGTTT
Glyma01g22880.1:cds    (201)  ACGATACATGTACTTAAATGAAGAGATCCTGAAGGAGAATCCGAGTGTTT
Glyma09g08780.1:cds    (201)  ACGATACATGTACTTAAATGAAGAGATCCTGAAGGAGAATCCAAGTGTTT
                                251                                                 300
L03352      CHS6cds    (251)  GTGCCTATATGGAACCTTCATTGGATGCAAGGCAAGACATGGTGGTTGTG
Glyma01g22880.1:cds    (251)  GTGCCTATATGGCACCTTCATTGGATGCAAGGCAAGACATGGTGGTTGTG
Glyma09g08780.1:cds    (251)  GTGCCTATATGGAACCTTCATTGGATGCAAGGCAAGACATGGTGGTTGTG
                                301                                                 350
L03352      CHS6cds    (301)  GAGGTACCAAAGTTGGGAAAAGAGGCTGCAACAAAGGCAATCAAGGAATG
Glyma01g22880.1:cds    (301)  GAGGTACCAAAGTTGGGAAAAGAGGCTGCAACAAAGGCAATCAAGGAATG
Glyma09g08780.1:cds    (301)  GAGGTACCAAAGTTGGGAAAAGAGGCTGCAACAAAGGCAATCAAGGAATG
                                351                                                 400
L03352      CHS6cds    (351)  GGGTCAACCCAAGTCCAAGATTACTCATCTCATCTTCTGCACCACTAGTG
Glyma01g22880.1:cds    (351)  GGGTCAACCCAAGTCCAAGATTACTCATCTCATCTTCTGCACCACTAGTG
Glyma09g08780.1:cds    (351)  GGGTCAACCCAAGTCCAAGATTACTCATCTCATCTTCTGCACCACTAGTG
                                401                                                 450
L03352      CHS6cds    (401)  GTGTGGACATGCCTGGTGCTGATTATCAGCTCACAAAACTGCTAGGACTT
Glyma01g22880.1:cds    (401)  GTGTGGACATGCCTGGTGCTGATTATCAGCTCACAAAACTGCTAGGACTT
Glyma09g08780.1:cds    (401)  GTGTGGACATGCCTGGTGCTGATTATCAGCTCACAAAACTGCTAGGACTT
                                451                                                 500
L03352      CHS6cds    (451)  CGTCCCTCCGTCAAGCGTTACATGATGTACCAACAAGGCTGCTTTGCTGG
Glyma01g22880.1:cds    (451)  CGTCCCTCCGTCAAGCGTTACATGATGTACCAACAAGGCTGCTTTGCTGG
Glyma09g08780.1:cds    (451)  CGTCCCTCCGTCAAGCGTTACATGATGTACCAACAAGGCTGCTTTGCTGG
                                501                                                 550
L03352      CHS6cds    (501)  TGGCACGGTGCTTCGTCTGGCGAAAGACTTGGCCGAAAACAACACGGGTG
Glyma01g22880.1:cds    (501)  TGGCACGGTGCTTCGTCTGGCGAAAGACTTGGCCGAAAACAACACGGGTG
Glyma09g08780.1:cds    (501)  TGGCACGGTGCTTCGTCTGGCGAAAGACTTGGCCGAAAACAACACGGGTG
                                551                                                 600
L03352      CHS6cds    (551)  CTCGTGTGCTCGTCGTGTGTTCAGAGATCACAGCAGTCACATTTCGCGGC
Glyma01g22880.1:cds    (551)  CTCGTGTGCTCGTCGTGTGTTCAGAGATCACAGCAGTCACATTTCGCGGC
Glyma09g08780.1:cds    (551)  CTCGTGTGCTCGTCGTGTGTTCAGAGATCACAGCAGTCACATTTCGCGGC
                                601                                                 650
L03352      CHS6cds    (601)  CCGAGTGACACCCATCTTGATAGCCTTGTTGGGCAAGCCTTGTTTGGAGA
Glyma01g22880.1:cds    (601)  CCGAGTGACACCCATCTTGATAGCCTTGTTGGGCAAGCCTTGTTTGGAGA
Glyma09g08780.1:cds    (601)  CCGAGTGACACCCATCTTGATAGCCTTGTTGGGCAAGCCTTGTTTGGAGA
                                651                                                 700
L03352      CHS6cds    (651)  TGGTGCAGCTGCTGTCATTGTTGGATCAGACCCTTTGCCTGCTGAAAAGC
Glyma01g22880.1:cds    (651)  TGGTGCAGCTGCTGTCATTCTTGGATCAGACCCTTTGCCTGCTGAAAAGC
Glyma09g08780.1:cds    (651)  TGGTGCAGCTGCTGTCATTGTTGGATCAGACCCTTTGCCTGCTGAAAAGC
                                701                                                 750
L03352      CHS6cds    (701)  CTTTGTTTGAGCTTGTGTGGACTGCACAAACAATCCTGCCAGACAGTGAA
Glyma01g22880.1:cds    (701)  CTTTGTTTGAGCTTGTGTGGACTGCACAAACAATCCTTCCAGACAGTGAA
Glyma09g08780.1:cds    (701)  CTTTGTTTGAGCTTGTGTGGACTGCACAAACAATCCTGCCAGACAGTGAA
```

```
                                751                                     800
L03352      CHS6cds    (751) GGGGCTATTGATGGCCACCTTCGCGAAGTAGGACTCACTTTCCATCTCCT
Glyma01g22880.1:cds    (751) GGGGCTATTGATGGCCACCTTCGCGAAGTAGGACTCACTTTCCATCTCCT
Glyma09g08780.1:cds    (751) GGGGCTATTGATGGCCACCTTCGCGAAGTAGGACTCACTTTCCATCTCCT
                                801                                     850
L03352      CHS6cds    (801) CAAGGATGTTCCTGGACTCATCTCGAAGAACATCCAAAAGGCCTTGGTTG
Glyma01g22880.1:cds    (801) CAAGGATGTTCCTGGACTCATCTCGAAGAACATCCAAAAGGCCTTGGTTG
Glyma09g08780.1:cds    (801) CAAGGATGTTCCTGGACTCATCTCGAAGAACATCCAAAAGGCCTTGGTTG
                                851                                     900
L03352      CHS6cds    (851) AAGCCTTCCAACCCTTGGGAATTGATGATTACAACTCTATCTTTTGGATT
Glyma01g22880.1:cds    (851) AAGCCTTCCAACCCTTGGGAATTGATGATTACAACTCTATCTTTTGGATT
Glyma09g08780.1:cds    (851) AAGCCTTCCAACCCTTGGGAATTGATGATTACAACTCTATCTTTTGGATT
                                901                                     950
L03352      CHS6cds    (901) GCACACCCTGGTGGACCAGCAATATTGGACCAAGTTGAGGCTAAGTTAGG
Glyma01g22880.1:cds    (901) GCACACCCTGGTGGACCAGCAATATTGGACCAAGTTGAGGCTAAGTTAGG
Glyma09g08780.1:cds    (901) GCACACCCTGGTGGACCAGCAATATTGGACCAAGTTGAGGCTAAGTTAGG
                                951                                    1000
L03352      CHS6cds    (951) CTTGAAACCTGAAAAAATGGAAGCTACTAGACATGTGCTCAGCGAGTATG
Glyma01g22880.1:cds    (951) CTTGAAACCTGAAAAAATGGAAGCTACTAGACATGTGCTCAGCGAGTATG
Glyma09g08780.1:cds    (951) CTTGAAACCTGAAAAAATGGAAGCTACTAGACATGTGCTCAGCGAGTATG
                               1001                                    1050
L03352      CHS6cds   (1001) GTAACATGTCAAGTGCATGCGTGTTGTTCATCTTGGATCAAATGAGGAAG
Glyma01g22880.1:cds   (1001) GTAACATGTCAAGTGCATGCGTGTTGTTCATCTTGGATCAAATGAGGAAG
Glyma09g08780.1:cds   (1001) GTAACATGTCAAGTGCATGCGTGTTGTTCATCTTGGATCAAATGAGGAAG
                               1051                                    1100
L03352      CHS6cds   (1051) AAGTCAATAGAAAATGGACTTGGCACCACTGGTGAAGGGCTTGAATGGGG
Glyma01g22880.1:cds   (1051) AAGTCAATAGAAAATGGACTTGGCACCACTGGTGAAGGGCTTGAATGGGG
Glyma09g08780.1:cds   (1051) AAGTCAATAGAAAATGGACTTGGCACCACTGGTGAAGGGCTTGAATGGGG
                               1101                                    1150
L03352      CHS6cds   (1101) TGTGCTATTTGGTTTTGGCCCTGGACTCACTGTTGAGACTGTTGTGCTTC
Glyma01g22880.1:cds   (1101) TGTGCTATTTGGTTTTGGCCCTGGACTCACTGTTGAGACTGTTGTGCTTC
Glyma09g08780.1:cds   (1101) TGTGCTATTTGGTTTTGGCCCTGGACTCACTGTTGAGACTGTTGTGCTTC
                               1151        1167
L03352      CHS6cds   (1151) GCAGTGTCACAGTCTAA
Glyma01g22880.1:cds   (1151) GCAGTGTCACAGTCTAA
Glyma09g08780.1:cds   (1151) GCAGTGTCACAGTCTAA
```

Fig. 3.9. Alignment of NCBI soybean chalcone synthase VI coding region sequence and the two putative CHS6 predicted coding regions present in Glyma1.01 (Glyma01g22880.1 and Glyma09g08780.1).

The only known copy of the isoflavone hydroxylase (*I2'H*) gene is in the region delimited by the IC of the QTL *qGEN15*, *qGLY15*, and *qTOT15*, in Gm15. Two glycitein loci were previously reported by our group in Gm17 (*gly-D2_1* and *gly-D2_2*), had a copy of *4CL1* and *DFR2* as candidate genes, respectively. Although the *CHR* homolog present on Gm18 does not lie in any of the QTL detected herein, it is actually included in QTL regions for glycitein content (Primomo et al. 2005). A *DFR* homolog is in the area demarcated by the IC of two QTL for glycitein and total isoflavones (Kassem et al. 2006), although it is likely that these two represent a single QTL. Another QTL for glycitein in Gm18 (*gly-G*) has a copy of *DFR* within its limits. A phenylalanine ammonia-lyase copy (*PAL*) located on Gm19 is clearly a good candidate for *qGEN19*, *qTOT19*, and *dai-L*. Finally, in Gm20, a copy of the *IOMT* gene is within the region demarcated by *qGEN20*. In addition, copies of *CHI-1A*, *CHI-1B1*, and *CHI2* are present in close proximity (<10 cM).

DISCUSSION

Genetic Control of Seed Isoflavone Content

The phenotype of complex traits is the result of diverse genetic and environmental factors, many of which have been found to interact with one another. As a result, isoflavone content in soybean seeds is highly variable among lines, locations, and within cultivars. The origin of such variability was found in genetic (G), environmental (E), and G × E interaction factors. This grade of unpredictability has long hampered the use of molecular breeding technologies, such as marker-assisted selection (MAS), for the

development of increased or decreased isoflavone lines. Different MAS approaches have been successfully attempted targeting traits governed by many minor-effect QTL (Bernardo, 2008). However, these approaches rely on finding a close marker-trait association relatively independent from the environment. Despite the environmental interference, heritability in the broad-sense was found high for all traits, which in a population of RILs indicates that the observed phenotypic variability is largely under the genetic control of additive loci, either by itself or by AA epistasis. Supporting the broad range of variation observed for isoflavones, several loci also have an E and a QTL × E interaction effect (A × E or AA × E) by themselves.

Combining data of several environments allowed detecting thirty-five main-effects QTL. Nearly 92% of them individually account for less than 5% of the phenotypic variability, suggesting that isoflavone seed concentrations are governed by numerous minor-effect QTL. Many of the main-effect QTL were also interacting, implying that they might have different absolute effects in different genetic backgrounds. However, most of the interactions occurred between loci that bear no additive effect. Our results suggest that epistatic network of interactions for isoflavones is largely changed when plants grow in different environments and may be one of the main causes of their phenotypic variability over locations and years. Corroborating this observation, approximately one-third of the epistatic interactions detected in this analysis also showed interaction with the environment (QTL × E). The same proportion of QTL × E interactions was found for different traits in a study in maize (Melchinger et al. 1998). In a recent study, it has been shown that an epistatic interaction can have a positive or negative effect upon plant fitness depending on the environment by regulating the

salicylic acid stress signaling pathway (Alcazar et al. 2009). We hypothesize that the whole network of main-effect and epistatic QTL changes to confront and adapt to external stimuli. This is in agreement with the role of isoflavones in defense mechanisms against damaging agents such as pathogens, UV radiation and cold stress (Beggs et al. 1985; Tsukamoto et al. 1995; Stafford 1997; Variyar et al. 2004; Subramanian et al. 2005; Naoumkina at al 2007). However, the specific reason why isoflavones are directed towards accumulation in seeds remains elusive.

Some observations suggest that we might currently perceive only the tip of the iceberg and that the genetic network controlling isoflavone synthesis is likely to be even more intricate than previously thought. First, the threshold applied to QTL detection must be very strict to deal with the multi-test issue and keep the rate of false positives low. For such complex traits likely imply that many minor-effect QTL would remain undiscovered, especially in sample mapping populations. This issue is even more critical for epistatic QTL because two intervals must be tested. For instance, taking into account both main and epistatic QTL, the model was only able to explain 37.9% (additive:25.8 + epistatic:12.1), 49.1% (25.4+23.7), 16.7% (10.2+6.5), and 40% (27.5+12.5) for genistein, daidzein, glycitein and total isoflavones, respectively. The high percentage of remaining unassigned variance is likely due to minor-effect loci and epistasis that did not reach the threshold level of detection, further suggesting that numerous minor-effect loci are involved in isoflavone synthesis. Underestimating the number of QTL causes overestimation of the genetic effects of the ones identified because of what is known as the Beavis effect (Beavis, 1994; Xu S, 2003). Thus, special care must be taken when considering the reliability of the parameter estimates. Second, only polymorphic loci will

segregate in the progeny and consequently be detected. Important but monomorphic loci may remain undiscovered. This is especially crucial for epistatic QTL as both interacting loci must be polymorphic, as otherwise they will not create phenotypic variance. Third, a population of RILs may underestimate the total epistasis if dominant interactions exist (Kearsey et al. 2003). Effectively, due to the structure of the mapping population, it is not possible to map QTL bearing dominant (D) main or epistatic effect (DD, AD or DA). However, due to the smaller number of genotypic classes, a RIL population increases the statistical power to detect the remaining AA component, which is the more useful for MAS because it is heritable (Goodnight 1988). Moreover, the self-pollinating nature of soybean, with usually less than 1% cross-pollination (Ahrent and Caviness, 1994), does not suggest these effects to be important in natural populations. Fourth, third- or higher-order epistatic interactions are not reflected on the mapping analysis but could exist and be an important component for complex traits such as isoflavones (Holland 2007; Rowe et al. 2008). As a last observation, despite the marker-dense linkage map and the threshold values considered for QTL detection, the presence of spurious QTL cannot be completely discarded due to the limited number of individuals sampled. Nevertheless, large sample size in experimental populations is not only economically unfeasible but also does not guarantee detecting all possible minor-effect QTL (Otto et al. 2000).

Complex networks of interacting genes govern isoflavone content in soybean seeds

The soybean genome is believed to have undergone at least two independent duplications from a diploid ancestor to render the actual polyploid (Shoemaker et al. 2006). This degree of duplication is also reflected in the phenylpropanoid genes and increases the difficulty in performing gene-function association analysis because

polyploidization may bring about gain-of-function, loss-of-function or neo-functionalization of certain copies (Doyle et al. 2008). It is also common to find tissue-specificity for some gene copies. For example, *CHS* has nine paralogues distributed along seven chromosomes, however, *CHS4*, *CHS5* and *CHS6* were not found in seeds at detectable levels (Gutierrez-Gonzalez et al. unpublished).

A separate discussion of the *CHS* genes present on Gm08 is merited. According to BAC sequencing, 12 *CHS*-related genes are thought to be present on Gm08, organized in two clustered regions (Tuteja and Vokdin, 2008). One such region is composed of two perfect inverted repeated regions each containing *CHS1*, *CHS3, CHS4* genes. This region is commonly referred to as the *I* locus, and although the specific silencing RNA is unclear, appears to act through siRNA-mediated gene silencing to inhibit seed coat pigmentation by specific degradation of seed coat specific *CHS7* and *CHS8* transcripts (Todd and Vodkin, 1996, Kasai et al., 2004, Tuteja et al., 2004). The *CHS6*, *CHS7* and *CHS8* genes are not apparently present in clusters on Gm08 but rather in single copy regions on other chromosomes (Tuteja and Vodkin 2008). In addition, at least four additional *CHS* genes (*CHS5, CHS3, CHS1*, and *CHS9*) are thought to be present in another *CHS* cluster on Gm08, for a total of 12 *CHS* genes on Gm08 (Tuteja and Vodkin, 2008). The current Glyma 1.01 assembly of Gm08 contains only 7 *CHS*-corresponding regions (6 *CHS* gene models, and an additional matching region which does not have an annotated gene model, Table 3.5) as determined by BLAST searches using *CHS* sequence as queries (http://www.phytozome.net/soybean). Evaluation of sequence corresponding to regions surrounding the *CHS* cluster on the Gm08 pseudo-molecule revealed four nearby regions of ambiguous sequence (e.g "nnnnnn"). It seems likely that a miss-assembly has

occurred in this region and the number of *CHS* genes present on Gm08 remains to be resolved. It is not currently clear to what extent other similar regions with gene repeats may have impacted on the assembly of the whole genome shotgun sequence.

Furthermore, as many as twenty regions were found sharing homology to the *DFR* gene. Although several of them may be pseudogenes, genes inactivated by the accumulation of deleterious mutations, it may be quite difficult to ascertain in which tissues, if any, these genes are expressed if they share a high percentage of homology. In addition, we currently have only a single sequenced genome (Williams 82) for reference. It is highly likely that other cultivars will have differing numbers and dispositions of phenylpropanoid genes. In particular, large scale rearrangements and deletions have been noted for the *CHS* gene clusters present on Gm08 (Tuteja et al. 2004).

The network of genes and interactions appears to have several interconnected neuralgic centers. The *CHS* gene cluster in Gm08 emerges as key node controlling the synthesis and accumulation of all individual isoflavones. The cluster rests within the IC of two important QTL for daidzein and total isoflavone content. In addition, it has connections with three other influential QTL and candidate genes: in Gm07(*4CL4*), Gm05, and Gm19(*PAL*). Importantly, all three isoflavones have at least one epistatic line converging in the *CHS* cluster. One of those pair wise interactions connected the cluster with the focal locus at the end of Gm05, and explained both genistein and daidzein, which is a rare phenomenon and further validated the interaction. The locus itself accounted for genistein, daidzein, glycitein, and total isoflavones, and it is by far the most principal of all QTL reported herein in terms of explained percentage of variance and additive value. For the QTL of largest effect on Gm05 (*qGEN5*, *qDAI5*, *qGLY5* and

qTOT5), no known biosynthetic genes were found to be located within this region. Perhaps a heretofore unknown isoflavone biosynthetic gene is located within this region. Alternately, a *trans*-acting factor affecting the expression or activity of isoflavone biosynthetic genes may be present (transcription factor, ubiquitin-related protein, etc). Consistent with this hypothesis, the epistatic interactions that this locus has with another central locus might suggest that a *trans*-regulatory mechanism is implicated and controls the expression of the *CHS* genes present in the Gm08 cluster. Whether this region contains such a factor remains for future work. Reinforcing its central role, the *CHS* cluster is also epistaticaly connected with other two regions both accounting for genistein, daidzein and total isoflavones. The first interaction is with a locus in Gm19/(*PAL*) and it is implicated in genistein accumulation. The second is involved in glycitein synthesis with a locus in Gm11, which does not have a candidate gene identified but a locus in the proximity has been implicated in glycitein production (Kassem et al. 2004, 2006). There is a third interaction with the locus in Gm07 for genistein, daidzein and the sum of all isoflavones. This strategic node is also connected with two other important areas in the genome, as a subnet of interrelated loci: The first epistatic interaction is shown to explain daidzein content and it is with the region in Gm01/(*CHS6*, *DFR*) of *qDAI1* and a QTL reported for glycitein by Kassem et al (2006). The second is an interaction for genistein with Gm13/(*IFS2*, *CHI3*, *DFR1*, *DFR*), which itself also explains genistein seed concentrations in an additive main-effect manner. Three loci located in Gm02/(*F3H1*, *F3H2*)-Gm05-Gm15/(*I2'H*) formed another key three-node axe for isoflavone synthesis. A focal locus on Gm02 was responsible for the accumulation of all individual isoflavones. It also featured two epistatic interactions accounting for

daidzein and total isoflavones, respectively, with a same locus in Gm05, which was reported to be responsible for daidzein accumulation (Kassem et al. 2004), but lacking of a candidate gene. The locus in Gm02 was also linked by means of an epistatic interaction for total isoflavones to a QTL in Gm15 for genistein, glycitein, and total seed isoflavone content, which is also connected with Gm10(*CHI-1B*). Finally, two other important additive nodes, in Gm04 (*qGEN4*, *qDAI4*, *qTOT4*) and Gm16 (*qDAI16*, *qTOT16*), were connected by a daidzein epistatic union.

Downstream and substrate-competing enzymes are less tractable to an intuitive interpretation, and likely require metabolite quantification to more precisely assess flux before considering their role as candidate genes. It is also difficult to establish a relationship with enzymes which govern reactions occurring far removed, or in another biosynthetic branch, from the QTL. This may be the case of for example the QTL region in Gm02 explaining the accumulation of all isoflavones (*qGEN2*, *qDAI2*, *qGLY2*, and *qTOT2*) and overlapping a copy of *F3H*. Other examples are: *IFR2* for *qDAI4*, *qTOT4*, and *qGEN4* in Gm04; *DFR2* for *gly-D2_2* in Gm17; *DFR* and *IOMT* for *gly-G* in Gm18; *12'H* for *qGEN15*, *qGLY15*, and *qTOT15* in Gm15; and *DRF* for *gly-k* in Gm09. The later QTL also overlaps a region reported by Kassem et al. (2004) for daidzein synthesis. It seems likely that another unknown factor aside from the candidate gene present in the region is responsible.

Having placed the phenylpropanoid genes on the genetic map also offered us the possibility to formulate hypotheses about the genetic basis underlying epistasis. Surprisingly, a great number of the epistatic QTL have isoflavone synthesizing genes located in their IC (Fig 3.5), suggesting that a great number of the interacting loci might

have a tractable genetic basis. Epistasis might be a critical factor for fitness-related traits in some plant species (Mei et al. 2003; Malmberg et al. 2005). This is in perfect agreement with the role that isoflavones have in the fitness of soybean plants. However, epistatic isoflavone QTL are subjected to dramatic adjustments, even more that their additive counterparts, when different environments are considered (Primomo et al. 2005; Gutierrez-Gonzalez et al. 2009). Despite these difficulties, researchers have successfully assigned genes to epistatic QTL that also had additive effect (McMullen et al. 2001; Kroymann et al. 2005; Sweigart et al. 2006; Alcazar et al. 2009). Finding the genes underlying epistatic QTL with no main-effect is more challenging but in some cases they have been revealed (Lark et al. 1995; Ehrenreich et al. 2007). Trying to assess the complete network of isoflavone epistasis appears to be a colossal task because the number, environmentally-influenced, and small-effect of the interactions. Moreover, only genes on the phenylpropanoid pathway were placed in the linkage map, and although epistasis is more likely to occur between genes on the same or related pathways (Segre et al. 2005), undoubtedly interactions with other enzymes or transcription factors exist. Under these premises, one should be very cautious when intending to assign candidate genes to epistatic QTL. However, considering the genetic network as a whole might help not only in validating the individual components but also in deciphering genes underlying epistasis. For instance, if an epistatic locus also bears a main-effect, it may warrant further study. Moreover, if both epistatic and main-effect QTL account for the same trait it seems more likely not to be a spurious QTL. This extreme, repeatable observed in our analysis, further validated many of the epistatic interactions, as opposed to being artifacts due to a limited sample size. For example, in Gm17 two independent main-effect QTL

accounted for glycitein concentrations. Both had candidate genes in their IC, and both were found to interact with each other by an epistatic interaction, which accounted for glycitein accumulation itself.

Overall, our results suggest that isoflavone accumulation in soybean seeds is controlled by a complex environment-adaptable network of interacting nodes. We could also place robust candidate genes for several main-effect loci. Nevertheless, it remains for future research to determine the nature of the proposed allelic differences between candidate genes in the lines we examined, and the manner in which these differences correlate with impacts on isoflavone content of seeds. Validation of the entire network of interactions, however, is likely to remain a monumental task due to the numerous small-effect QTL involved and their environmental unpredictability.

TABLES

Table 3.1. ANOVA, random effects, and heritability over environments of the mapping population.

				Effects					
Trait	Mean ± SD	Range	CV(%)[d]	Genetic	Env	G x E	H^2	Skewness	Kurtosis
Genistein	730.4 ± 288.3	82–1980	39.47	<0.0001	<0.0001	<0.0001	0.87	0.409	-0.078
Daidzein	417.0 ± 210.1	33–1477	50.37	<0.0001	<0.0001	<0.0001	0.91	0.798	0.990
Glycitein	126.8 ± 61.8	0–342	48.75	<0.0001	<0.0001	<0.0001	0.69	0.460	-0.580
Total Isoflav	1272.7 ± 485.0	186–3008	38.11	<0.0001	<0.0001	<0.0001	0.89	0.350	-0.186

[a] Mean ± Standard deviation (SD) and range units in µg of isoflavones per gram of seeds.
[b] P values of random effects over environments: genetic, environmental, and g x e interaction.
[c] Heritability of the traits in the broad sense over environments.
d Phenotypic CV combined over locations and years.

Table 3.2. Additive QTLs for genistein and daidzein accumulation in soybean seeds of Essex × PI 437654.

Name[a]	Interval	IC[b]	A±SE[c]	P-Value	h^2(a)
Genistein					
*qGEN5**	SATT236-SAT_271	27.4-34.4	82.1±5.2	0.00000	5.5
*qGEN2**	SAT_279-EACAMCTT123	0.0-6.0	54±4.3	0.00000	2.0
*qGEN15**	SAT_112-SATT691	8.0-11.0	58.9±4.6	0.00000	2.3
*qGEN13**	SATT490-SAT_197	3.4-11.0	25.3±4.4	0.00000	1.3
qGEN12	SCTT009-SATT541	0.0-6.0	41.9±4.3	0.00000	1.2
qGEN20	GMLPSI2-SCT_189	72.5-74.5	-32.5±4.1	0.00000	0.4
*qGEN19**	SAT_113-SAT_286	141.4-149.8	-47.5±4.5	0.00000	1.9
*qGEN7**	SATT175-EAGGMCTT095	76.1-83.1	42.9±4.9	0.00000	2.3
qGEN4	SAT_337-SAT_140	6.2-17.6	-25.5±7.7	0.00089	3.3
qGEN2_2	SAT_139-SAT_289	107.6-126.2	-49.6±9.4	0.00000	4.3
Daidzein					
*qDAI5**	SATT174-SATT236	21.5-28.4	50.3±3.2	0.00000	8.3
*qDAI8**	SATT187-EAGGMCTT205	23.0-25.1	-27.9±2.9	0.00000	2.2
qDAI4	SATT396-SAT_337	0.0-6.0	-27.5±3.2	0.00000	1.2
qDAI1	SAT_106-AW781285	34.7-42.6	20.9±3.1	0.00000	1.2
*qDAI2**	SAT_279-EACAMCTT123	0.0-8.0	27.5±3.2	0.00000	1.4
*qDAI12**	SCTT009-SATT541	0.0-3.0	63.6±3.1	0.00000	4.3
qDAI16	SAT_339-SATT280	0.0-15.0	-58.8±3.9	0.00000	2.3
*qDAI7**	SATT175-EAGGMCTT095	75.1-83.1	55.7±3.5	0.00000	2.4
qDAI7_2	SAT_147-SAT_330	145.7-151.7	-38.5±3.1	0.00000	1.0

[a]name given to a particular QTL, *GEN* and *DAI* for genistein and daidzein content, respectively, followed by the chromosome number and a number when more than one in the same chromosome. Asterisks highlight QTL reported in our previous study (Gutierrez-Gonzalez et al, 2009). [b]Interval of confidence in centiMorgans with respect to the first marker in the LG. [c]main additive effect in mg/g plus/minus standard error. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL. P-values represent the significance of each effect.

Table 3.3. Additive QTLs for glycitein and total isoflavone accumulation in soybean seeds of Essex × PI 437654.

Name[a]	Interval	IC[b]	A±SE[c]	P-Value	h^2(a)
Glycitein					
qGLY5	SATT236-SAT_271	26.4-35.4	7.1±1.1	0.00000	1.4
*qGLY6**	SATT281-SATT291	40.0-52.7	-7.1±0.9	0.00000	1.4
qGLY6_2	SATT319-EAACMCAC113	41.2-48.8	9.2±0.9	0.00000	1.0
qGLY2	SAT_279-EACAMCTT123	0.0-5.0	5.9±0.8	0.00000	0.9
*qGLY15**	SAT_112-SATT691	8.0-11.0	10.7±1	0.00000	2.1
qGLY9	SATT242-EAACMCAC227	25.1-36.3	6.4±0.9	0.00000	0.9
Total					
*qTOT5**	SATT174-SATT236	20.5-25.4	161.7±8.3	0.00000	7.0
qTOT4	SAT_337-SAT_140	3.0-9.2	-104.1±7.8	0.00000	1.1
*qTOT2**	SAT_279-EACAMCTT123	0.0-7.0	87.4±8.1	0.00000	2.1
qTOT2_2	SAT_139-SAT_289	107.6-118.2	-63.2±7.7	0.00000	1.9
*qTOT15**	SAT_112-SATT691	8.0-11.0	52.9±8.5	0.00000	1.7
*qTOT12**	SCTT009-SATT541	0.0-4.0	95.2±7.8	0.00000	3.4
qTOT16	SAT_339-SATT280	3.0-19.0	-125.8±11.2	0.00000	1.3
*qTOT19**	SAT_286-SATT229	144.8-159.9	-69.9±8.8	0.00000	1.4
*qTOT7**	SATT175-EAGGMCTT095	76.1-85.1	115.8±9	0.00000	3.1
*qTOT8**	SATT187-EAGGMCTT205	23.0-29.1	-55.5±14.0	0.00009	3.0

[a]name given to a particular QTL, *GLY* and *TOT* for glycitein and total isoflavone content, respectively, followed by the chromosome number and a number when more than one in the same chromosome. Asterisks highlight QTL reported in our previous study (Gutierrez-Gonzalez et al, 2009). [b]Interval of confidence in centiMorgans with respect to the first marker in the LG. [c]main additive effect in mg/g plus/minus standard error. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL. P-values represent the significance of each effect.

Table 3.4. Additive by additive epistatic interactions for genistein and daidzein.

interval_i[a]	interval_j[a]	AA±SE[b]	P-Value	h^2(aa)[c]
Genistein				
SATT236-SAT_271	SATT187-EAGGMCTT205	24.9±5.2	0.000002	0.6
SATT187-EAGGMCTT205	SAT_113-SAT_286	-15.3±4.5	0.000626	0.8
SATT490-SAT_197	SATT175-EAGGMCTT095	-31.5±5.2	0.000000	1.8
EAAGMCAT279-SATT050	SATT484-AQ851479	23.3±4.5	0.000000	0.5
SATT050-SATT385	SATT643-SATT319	48.7±6.9	0.000000	0.4
SATT187-EAGGMCTT205	SATT270-SAT_268	-23±4.3	0.000000	0.6
SATT233-SATT329	SATT484-AQ851479	32.2±4.6	0.000000	0.8
AW620774-SATT534	EAGGMCTT143-SATT182	-24.1±5.7	0.000024	0.4
EAACMCTC178-SATT578	SAT_222-EACAMCTT122	36.1±4.6	0.000000	1.4
SATT184-SAT_353	GMLPSI2-SCT_189	-27.7±4.1	0.000000	0.9
EACAMCTT123-BE475343	SATT346-SAT_147	-14.8±4.8	0.002152	0.3
SATT634-SATT266	EAACMCAC175-SATT244	23.8±4.8	0.000001	1.1
SATT634-SATT266	EAACMCTC276-SCTT011	20.9±6.1	0.000661	0.1
SCT_189-SATT440	SATT435-SATT323	-33.6±4.2	0.000000	1.1
Daidzein				
SATT174-SATT236	SATT187-EAGGMCTT205	20.8±3.2	0.000000	0.6
SATT187-EAGGMCTT205	SATT175-EAGGMCTT095	16.7±3.5	0.000002	0.9
SATT396-SAT_337	SAT_339-SATT280	15.9±4.2	0.000201	0.6
SAT_106-AW781285	SATT175-EAGGMCTT095	-18.7±3.7	0.000001	1.1
SATT591-AZ536570	SAT_279-EACAMCTT123	14.9±3.2	0.000003	0.7
SAT_272-SAT_411	SATT674-SAT_228	-23.3±3.1	0.000000	1.2
SATT359-EACAMCTT067	SATT309-RHG1_INDEL	18.4±3.5	0.000000	1.1
SAT_342-EAACMCTC384	SAT_380-SATT263	17.2±3.5	0.000001	1.0
SAT_062-SATT281	EAACMCAC175-SATT244	23±3.3	0.000000	0.7
EAACMCTC063-EACAMCTT41	SATT302-SATT142	49.5±3.8	0.000000	3.6
EAACMCTC063-EACAMCTT41	SATT522-AW756935	14.4±3.4	0.000033	0.3
SATT266-SAT_135	SAT_282-SAT_242	38.9±4	0.000000	0.9
SAT_333-SCT_192	SAT_284-SATT372	39.9±3.2	0.000000	1.3
EACAMCTT122-SATT082	SATT167-SATT552	39.1±3.3	0.000000	2.8
SATT691-SATT598	SAT_190-SCAA001	-23.1±4.3	0.000000	0.6
SAT_380-SATT263	EAGGMCTT098-SATT009	29±3.3	0.000000	1.9
SAT_380-SATT263	SATT490-SAT_197	28±3.2	0.000000	0.4
EAGGMCAG365-EAGGMCAG42	SATT522-AW756935	-26.1±3	0.000000	1.4
SATT242-EAACMCAC227	SATT339-SATT257	-11.6±3.3	0.000585	0.2
EAGGMCTT176-EAGGMCT135	SATT683-SAT_275	21.4±4.2	0.000000	0.4
SAT_121-SATT346	SATT259-SATT188	-31.9±4	0.000000	1.0

[a]Marker interval within each epistatic interaction occurs. [b]estimated additive by additive effect ± standard error, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase total isoflavones expression while a negative sign indicate that recombinant allele combinations increase total isoflavone values. [c]heritability of the additive by additive interaction effect (%). P-values represent the significance of each effect.

Table 3.5. Additive by additive epistatic interactions for glycitein and total isoflavones.

interval_i[a]	interval_j[a]	AA±SE[b]	P-Value	h^2(aa)[c]
Glycitein				
SATT385-SATT619	SAT_389-SATT636	7.2±0.9	0.000000	1.2
EAACMCAC086-SATT187	SAT_247-SATT519	-4.4±1.1	0.000028	0.3
SATT304-SATT416	EACAMCTT122-SATT082	2.4±1	0.014723	0.3
EAACMCTC178-SATT578	EAACMCTC379-SATT316	4.6±0.9	0.000001	0.6
SAT_135-SATT546	SAT_284-SATT372	8.6±1.1	0.000000	1.2
SATT458-SAT_284	SAT_086-SATT256	6.5±1	0.000000	1.0
SATT269-SATT160	EAGGMCTT164-SATT330	-2.7±0.9	0.005685	0.2
Total isoflavones				
SAT_279-EACAMCTT123	SAT_112-SATT691	-60.8±9.1	0.000000	1.6
SAT_265-EAAGMCAT279	SAT_279-EACAMCTT123	54.3±9.9	0.000000	1.0
SATT089-SATT233	SAT_275-SATT387	109.8±10.2	0.000000	2.1
SATT455-SATT228	SATT304-SATT416	77.2±9.1	0.000000	1.2
SATT281-SATT291	AW756935-SAT_090	62.5±7.9	0.000000	2.1
SCT_033-SCT_188	SATT683-SAT_275	53.2±8.7	0.000000	0.6
SATT292-SAT_419	EAACMCTC276-SCTT011	66.4±11.2	0.000000	1.1
SCT_189-SATT440	SCT_195-SATT159	80.1±7.8	0.000000	2.4

[a]Marker interval within each epistatic interaction occurs. [b]estimated additive by additive effect ± standard error, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase total isoflavones expression while a negative sign indicate that recombinant allele combinations increase total isoflavone values. [c]heritability of the additive by additive interaction effect (%). P-values represent the significance of each effect.

Table 3.6. Additive by environment interaction effect for genestein and daidzein.

Name[a]	Interval[b]	AE1±SE	P-Value	AE2±SE	P-Value	AE3±SE	P-Value	AE4±SE	P-Value	h^2(ae)
Genistein										
qGEN5	SATT236-SAT_271	-27.9±10	0.005461	24.6±9.9	0.013538	-28.2±9.9	0.004260	31.9±9.8	0.001175	0.7
qGEN2	SAT_279-EACAMCTT123					-15.6±7.9	0.046630	25.2±7.8	0.001260	0.3
qGEN15	SAT_112-SATT691									
qGEN13	SATT490-SAT_197									
qGEN12	SCTT009-SATT541									
qGEN20	GMLPSI2-SCT_189									
qGEN19	SAT_113-SAT_286	33.6±8.6	0.000101	-18.2±8.5	0.032716					0.3
qGEN7	SATT175-EAGGMCTT095									
qGEN4	SAT_337-SAT_140									
qGEN2_2	SAT_139-SAT_289									
Daidzein										
qDAI5	SATT174-SATT236									
qDAI8	SATT187-EAGGMCTT205									
qDAI4	SATT396-SAT_337									
qDAI1	SAT_106-AW781285									
qDAI2	SAT_279-EACAMCTT123							17.6±5.7	0.002088	0.3
qDAI12	SCTT009-SATT541			12.6±5.5	0.022917					0.2
qDAI16	SAT_339-SATT280					17.4±7.0	0.013073	-14.6±7.0	0.036985	0.3
qDAI7	SATT175-EAGGMCTT095									
qDAI7_2	SAT_147-SAT_330					10.6±5.0	0.034565			0.2

[a]name given to a particular QTL, *gen* and *dai* for genistein and daidzein content, respectively, followed by the chromosome number and a number when more than one in the same chromosome. [b]Interval of confidence in centiMorgans with respect to the first marker in the LG. AE1, AE2, AE3, and AE4 are the environment × interaction effect in µg/g plus/minus standard error for BREC_06 (AE1), DRC_06 (AE2),BREC_07 (AE3), and DRC_07 (AE4). $h(ae)^2$ is the heritability of the additive by environment interaction effects. P-values represent the significance of each effect.

Table 3.7. Additive by environment interaction effect for glycitein and total isoflavones.

name[a]	interval[b]	AE1±SE	P-Value	AE2±SE	P-Value	AE3±SE	P-Value	AE4±SE	P-Value	h^2(ae)
Glycitein										
qGLY5	SATT236-SAT_271			8.6±2.1	0.00007	-11.6±2.1	0.00000			0.0109
qGLY6	SATT281-SATT291	4.3±1.6	0.008795							0.0035
qGLY6_2	SATT319-EAACMCAC113			4.8±1.8	0.00613			-4.3±1.7	0.01466	0.0027
qGLY2	SAT_279-EACAMCTT123	6.8±1.7	0.000118			-8.8±1.7	0.00000			0.0068
qGLY15	SAT_112-SATT691			3.2±1.6	0.04452					0.0014
qGLY9	SATT242-EAACMCAC227									
Total										
qTOT5	SATT174-SATT236			34.5±15.2	0.02296	-39.2±15.1	0.00953			0.0043
qTOT4	SAT_337-SAT_140									
qTOT2	SAT_279-EACAMCTT123					-35.6±15.0	0.01740	46.2±14.9	0.00197	0.0043
qTOT2_2	SAT_139-SAT_289									
qTOT15	SAT_112-SATT691									
qTOT12	SCTT009-SATT541			30.0±13.6	0.02762					0.0028
qTOT16	SAT_339-SATT280					40.1±19.5	0.03993	-40.1±19.4	0.03881	0.0027
qTOT19	SAT_286-SATT229	29.4±14.8	0.04716							0.0021
qTOT7	SATT175-EAGGMCTT095									
qTOT8	SATT187-EAGGMCTT205									

[a]name given to a particular QTL, gen and dai for genistein and daidzein content, respectively, followed by the chromosome number and a number when more than one in the same chromosome. [b]Interval of confidence in centiMorgans with respect to the first marker in the LG. AE1, AE2, AE3, and AE4 are the environment × interaction effect in μg/g plus/minus standard error for BREC_06 (AE1), DRC_06 (AE2),BREC_07 (AE3), and DRC_07 (AE4). h(ae)2 is the heritability of the additive by environment interaction effects. P-values represent the significance of each effect.

Table 3.8. Additive-additive by environment interaction effect for each epistatic interaction for genistein and daidzein.

interval_i[a]	interval_j[a]	AAE1±SE[b]	P-Value	AAE2±SE[b]	P-Value	AAE3±SE[b]	P-Value	AAE4±SE[b]	P-Value	h^2(aae)[c]
Genistein										
SATT236-SAT_271	SATT187-EAGGMCTT205									
SATT187-EAGGMCTT205	SAT_113-SAT_286	-18.2±8	0.022991							0.3
SATT490-SAT_197	SATT175-EAGGMCTT095									
EAAGMCAT279-SATT050	SATT484-AQ851479					21.3±8.2	0.009923			0.1
SATT050-SATT385	SATT643-SATT319									
SATT187-EAGGMCTT205	SATT270-SAT_268									
SATT233-SATT329	SATT484-AQ851479									
AW620774-SATT534	EAGGMCTT143-SATT182			-26±10.6	0.014249	33.7±10.5	0.001427			0.4
EAACMCTC178-SATT578	SAT_222-EACAMCTT122									
SATT184-SAT_353	GMLPSI2-SCT_189									
EACAMCTT123-BE475343	SATT346-SAT_147	26.6±9.3	0.004397			-30.5±9.2	0.000944	19.4±9.1	0.033904	0.6
SATT634-SATT266	EAACMCAC175-SATT244									
SATT634-SATT266	EAACMCTC276-SCTT011									
SCT_189-SATT440	SATT435-SATT323									
Daidzein										
SATT174-SATT236	SATT187-EAGGMCTT205	12.4±5.8	0.034559			-13.5±5.8	0.019341	-5.3±5.7		0.2
SATT187-EAGGMCTT205	SATT175-EAGGMCTT095									
SATT396-SAT_337	SAT_339-SATT280									
SAT_106-AW781285	SATT175-EAGGMCTT095									
SATT591-AZ536570	SAT_279-EACAMCTT123									
SAT_272-SAT_411	SATT674-SAT_228									
SATT359-EACAMCTT067	SATT309-RHG1_INDEL					12.6±6.2	0.040912			0.1
SAT_342-EAACMCTC384	SAT_380-SATT263	-12.5±6.3	0.048070					12.7±6.2	0.041402	0.2
SAT_062-SATT281	EAACMCAC175-SATT244							-11.2±5.7	0.048627	0.2
EAACMCTC063-EACAMCTT41	SATT302-SATT142									
EAACMCTC063-EACAMCTT41	SATT522-AW756935									
SATT266-SAT_135	SAT_282-SAT_242									
SAT_333-SCT_192	SAT_284-SATT372									
EACAMCTT122-SATT082	SATT167-SATT552									
SATT691-SATT598	SAT_190-SCAA001									
SAT_380-SATT263	EAGGMCTT098-SATT009									
SAT_380-SATT263	SATT490-SAT_197									
EAGGMCAG365-EAGGMCAG42	SATT522-AW756935									
SATT242-EAACMCAC227	SATT339-SATT257									
EAGGMCTT176-EAGGMCT135	SATT683-SAT_275									
SAT_121-SATT346	SATT259-SATT188							23.9±7.4	0.001270	0.3

[a]Marker interval within each epistatic interaction occurs. [b]AAE1, AAE2, AAE3, and AAE4 are the additive-additive by environment interaction effect in μg/g plus/minus standard error for BREC_06 (AE1), DRC_06 (AE2),BREC_07 (AE3), and DRC_07 (AE4). h(aae)2 is the heritability of the additive-additive by environment interaction effects. P-values represent the significance of each effect. [c]heritability of the additive-additive by environment interaction effect (%). P-values represent the significance of each effect.

Table 3.9. Additive-additive by environment interaction effect for each epistatic interaction for glycitein and total isoflavones.

interval_i	interval_j	AAE1±SE	P-Value	AAE2±SE	P-Value	AAE3±SE	P-Value	AAE4±SE	P-Value	h^2(aae)
Glycitein										
SATT385-SATT619	SAT_389-SATT636									
EAACMCAC086-SATT187	SAT_247-SATT519									
SATT304-SATT416	EACAMCTT122-SATT082	5.6±2	0.004862	-8.1±1.9	0.000031					0.6
EAACMCTC178-SATT578	EAACMCTC379-SATT316							-3.2±1.5	0.030736	0.2
SAT_135-SATT546	SAT_284-SATT372			-4.6±1.9	0.015952			4.2±1.9	0.026453	0.3
SATT458-SAT_284	SAT_086-SATT256									
SATT269-SATT160	EAGGMCTT164-SATT330	8.5±1.9	0.000020			-7.6±1.9	0.000069	-4.3±1.9	0.022641	0.8
Total isoflavones										
SAT_279-EACAMCTT123	SAT_112-SATT691									
SAT_265-EAAGMCAT279	SAT_279-EACAMCTT123									
SATT089-SATT233	SAT_275-SATT387							-33.7±16.9	0.046146	0.2
SATT455-SATT228	SATT304-SATT416									
SATT281-SATT291	AW756935-SAT_090									
SCT_033-SCT_188	SATT683-SAT_275									
SATT292-SAT_419	EAACMCTC276-SCTT011	-46.7±19.7	0.017936							0.3
SCT_189-SATT440	SCT_195-SATT159									

[a]Marker interval within each epistatic interaction occurs. [b]AAE1, AAE2, AAE3, and AAE4 are the additive-additive by environment interaction effect in µg/g plus/minus standard error for BREC_06 (AE1), DRC_06 (AE2),BREC_07 (AE3), and DRC_07 (AE4). h(aae)2 is the heritability of the additive-additive by environment interaction effects. P-values represent the significance of each effect. [c]heritability of the additive-additive by environment interaction effect (%). P-values represent the significance of each effect.

REFERENCES

Ahrent DK and Caviness CE (1994) Natural cross-pollination of twelve soybean cultivars in Arkansas. Crop Sci. **34**:376-378.

Alcazar R, Garcia AV, Parker JE and Reymond M (2009) Incremental steps toward incompatibility revealed by Arabidopsis epistatic interactions modulating salicylic acid pathway activation. PROC NAT ACAD SCI **106**:334-339.

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In proceedings of the 49th annual corn and sorghum industry research conference Washington DC.

Beggs CJ, Stolzer-Jehle A, Wellmann E (1985) Isoflavonoid formation as an indicator of UV stress in bean (Phaseolus vulgaris L.) leaves: the significance of photorepair in assessing potential damage by increased solar UV-B radiation. Plant Physiol. **79(3)**:630-634.

Bennett JO, Yu O, Heatherly LG, Krishnan HB (2004) Accumulation of genistein and daidzein, soybean isoflavones implicated in promoting human health, is significantly elevated by irrigation. Agric. Food Chem. 52:7574-7579.

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. **48**:1649-1664.

Carlborg O and Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nature reviews **5**:618-625.

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol. Biol. **53**:733-743.

Doebley J, Stec A, and Gustus C (1995) Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. Genetics **141**:333-346.

Doerge RW (2001) Mapping and analysis of quantitative trait loci in experimental populations. Nature Reviews **3**:43-52.

Doyle JJ, Flagel LE, Paterson HA, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary Genetics of Genome Merger and Doubling in Plants Annual Review of Genetics **42**:443-461

Eldridge A, Kwolek W (1983) Soybean isoflavones: Effect of the environment and variety on composition. J. Agric. Food Chem. **31**:394-396.

Ehrenreich IM, Stafford PA, and Purugganan MD (2007) The genetic architecture of shoot branching in Arabidopsis thaliana: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics **176**:1223-1236.

Goodnight CJ (1988) Epistasis and the effect of founder events on the additive genetic variance. Evolution **42**:441-454.

Gutierrez-Gonzalez JJ, Wu X, Zhang J, Lee JD, Ellersieck M, Shannon JG, Yu O, Nguyen HT, Sleper DA (2009) Genetic control of soybean seed isoflavone content: Importance of statistical model and epistasis in complex traits. Theor. Appl. Genet. *In press*.

Hill J, Becker HC, Tigerstedt PMA (1998) Quantitative and ecological aspects of plant breeding. Chapman and Hall, London.

Hoeck JA, Fehr WR, Murphy PA, Welke GA (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci. **40**:48-51.

Holland JB (2007) Genetic architecture of complex traits in plants. Current Opinion in Plant Biology **10**:156-161.

Juenger TE, Sen S, Stowe KA, Simms EL (2005) Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. Genetica **123**:87-105.

Kasai A, Watarai M, Yumoto S, Akada S, Ishikawa R, Harada T, Niizeki M, and Senda M (2004) Influence of PTGS on Chalcone Synthase Gene Family in Yellow Soybean Seed Coat. Breed. Sci. **54**:355-360.

Kassem MA, Meksem K, Iqbal MJ, Njiti VN, Banz WJ, Winters TA, Wood A, Lightfoot DA (2004) Definition of soybean genomic regions that control seed phytoestrogen amounts. J Bio & Biotech 2004:1:52-60.

Kassem MA, Shultz J, Meksem K, Cho Y, Wood AJ, Iqbal MJ, Lightfoot DA (2006) An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits. Theor. Appl. Genet. **113**:1015-1026.

Kearsey MJ, Pooni HS, and Syed NH (2003) Genetics of quantitative traits in Arabidopsis thaliana. Heredity **91**:45-464.

Koes RE, Quattrocchio F, and King AA (1994) The flavonoid biosynthetic pathway in plants-Function and evolution. Bioessays **16**:123-132.

Kroymann J and Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. Nature **435**:95-98.

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLPs linkage maps. Genetics **121**:185-199.

Lark KG, Chase K, Adler F, Mansur LM, and Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. PROC NAT ACAD SCI **92**:4656-4660.

Lozovaya VV, Lygin AV, Ulanov AV, Nelson RL, Dayde J, and Widhohm JM (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci. **45**:1934-1940.

Malmberg RL and Mauricio R (2005) QTL-based evidence for the role of epistasis in evolution. Genet. Res. Camb. **86**:89-95.

Malmberg RL, Held S, Waits A, and Mauricio R (2005) Epistasis for fitness-related quantitative traits in Arabidopsis thaliana grown in the field and in the greenhouse. Genetics **171**:2013-2027.

Mao Y, Da Y (2004) Statistical power for detecting epistasis QTL effects under the F-2 design. Genet. Sel. Evol. **37:**129-150.

Matsumura H, Watanabe S, Harada K, Senda M, Akada S, Kawasaki S, Dubouzet EG, Minaka N, and Takahashi R (2005) Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean. Theor. Appl. Genet. **110**:1203-1209.

McMullen MD, Snook M, Lee EA, Byrne PF, Kross H, Musket TA, Houchins K, and Coe EH Jr (2001) The biological basis of epistasis between quantitative trait loci for flavones and 3-deoxyanthocyanin synthesis in maize (*Zea mays* L.). Genome **44**:667-676.

Mebrahtu T, Mohamed A, Wang CY, Andebrhan T (2004) Analysis of isoflavone contents in vegetable soybeans. Plant Foods for Human Nutrition **59**:55-61.

Mei HW, Li ZK, Shu QY, Guo LB, Wang YP, Yu XQ, Ying CS, Luo LJ (2003) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet **107**:89-101.

Meksem K, Njiti VN, Banz WJ, Iqbal MJ, Kassem MyM, Hyten DL, Yuang J, Winters TA, Lightfoot DA (2001) Genomic regions that underlie soybean seed isoflavone content. J. Biom. & Biotech. 1:1 (2001) 38-44.

Melchinger AE, Utz HF, Schon CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics **149**:383-403.

Melchinger AE, Piepho HP, Utz HF, Muminovic J, Wegenast TW, Torjek O, Altmann T and Kusterer B (2007) Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. Genetics **177**:1827-1837.

Murphy SE, Lee EA, Woodrow L, Seguin P, Kumar J, Rajcan I, and Ablett GR (2009) Genotype × Environment interaction and stability for isoflavone content in soybean. Crop Sci. **49**:1313-1321.

Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc. Natl. Acad. USA. **104**:17909-17915.

Otto SP and Jones CD (2000) Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. Genetics **156**:2093-2107.

Phillips PC (2008) Epistasis, the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews **9**:855-867.

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M, Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. **45**:2454-2462.

Rochfort S, Panozzo J (2007) Phytochemicals for health, the role of pulses. J. Agric. Food Chem. **55**:7981-7994.

Rowe HC, Hansen BG, Halkier BA, and Kliebenstein DJ (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell **20**:1199-1216.

Sanjuan R, Elena SF (2006) Epistasis correlates to genomic complexity. PROC NAT ACAD SCI **103**:14402-14405.

Segre D, DeLuna A, Church GM, and Kishony R (2005) Modular epistasis in yeast metabolism. Nat. Genet. **37**:77-83.

Shoemaker RC, Schlueter J and Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Current Opinion in Plant Biology **9**:104-109.

Shoemaker, RC, Sculueter JA and Jackson SA (2008) Soybean Genome Structure and Organization. Book chapter in Genetics and Genomics of Soybean G. Stacey (ed.), Springer Science+Business Media, LLC

Stafford HA (1997) Roles of flavonoids in symbiotic and defense functions in legume roots. Bot. Rev. **63**:27-39.

Subramanian S, Hu X, Lu G, Odelland JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol. Biol. **54:**623-639.

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol. **137**:1345-1353.

Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. Plant J. **48**:261-273.

Subramanian S, Stacey G, Yu O (2007) Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci. **12**:282-285.

Sweigart AL, Fishman L, and Willis JH (2006) A simple genetic incompatibility causes hybrid male sterility in mimulus. Genetics **172**:2465-2479.

Todd JJ and Vodkin LO (1996) Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell **8**:687-699.

Tsukamoto C, Shimada S, Igita K, Kudou S, Kokubun M, Okubo K, Kitamura K (1995) Factors affecting isoflavone content in soybean seeds: Changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J. Agric. Food Chem. 43:1184-1192.

Tuteja JH, Clough SJ, Chan WC, and Vodkin LO (2004) Tissue-specific gene silencing mediated by a natural occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell **16**:819-835.

Tuteja J.H. and Vodkin L.O. (2008) Structural features of the endogenous CHS silencing and target loci in the soybean genome. The Plant Genome. Crop Sci. 48(S1) S49-S68.

Variyar PS, Limaye A, Sharma A (2004) Radiation-induced enhancement of antioxidant contents of soybean (Glycine max Merrill). J. Agric. Food Chem. **52**:3385-3388.

Wang H, Murphy PA (1994) Isoflavone Composition of American and Japanese Soybeans in Iowa: Effects of Variety, Crop Year, and Location. J. Agric. Food Chem. **42**:1674-1677.

Weinreich DM, Watson RA, and Chao L (2005) Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution **59**:1165-1174.

Whitlock MC (1995) Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. **26**:601-629.

Wilfert L and Schmid-Hempel P (2008) The genetic architecture of susceptibility to parasites. BMC Evolutionary Biology **8**:187-1994.

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97-159.

Wu X, Blake S, Sleper DA, Shannon G, Cregan P, Nguyen HT (2008) QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor. Appl. Genet. 118:1093-1105.

Xu S (2003) Theoretical basis of the Beavis effect. Genetics **165**:2259-2268.

Yi N, Xu S (2002) Mapping Quantitative Trait Loci with Epistatic Effects. Genet. Res. Camp. **79**:185-198.

Yu O, McGonigle B (2005) Metabolic engineering of isoflavone biosynthesis. Advances in Agronomy. **86**:147-190.

Zeng G, Li D, Han Y, Teng W, Wang J, Qiu L, and Li W (2009) Identification of QTL underlying isoflavone contents in soybean seeds among multiple environments. Theor Appl Genet 118:1455-1463.

Zhang J, and Yu O (2009) Metabolic engineering of isoflavone biosynthesis in seeds. *In* Modification of seed composition to promote health and nutrition. Hari Krishnan (ed). Agronomy Monograph Series, pp151-177.

Zernova OV, Lygin AV, Widholm JM, and Lozovaya VV (2009) Modification of isoflavones in soybean seeds via expression of multiple phenolic biosynthetic genes. Plant Physiol. Biochem. Doi:101016/j.plaphy.2009.05.

CHAPTER 4

Confirmation of a major locus underlying individual and total isoflavone amounts in soybean seeds.

"*Imagination is more important than knowledge"*
(Albert Einstein)

INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) isoflavones may confer plant resistance to pathogens and distinct abiotic stresses (Beggs et al, 1985; Tsukamoto et al, 1995; Dakora and Philips, 1996; Moy et al, 2004; Subramanian et al, 2005). They have also garnered much attention because of their potential use for fabrication of pharmaceuticals to prevent important human diseases, such as cancer (Usui 2006; Rochfort and Panozzo, 2007). Thus, plant cultivars with high isoflavone seed levels may have biological and human-health-related advantages. Also, low-isoflavone lines may be more adequate for certain human subgroups, for instance infants and pregnant women (for review see Zhang and Yu, 2009). Consequently, there might be a potential market for developing varying isoflavone-content cultivars. Marker assisted selection (MAS) is broadly used for plant geneticists because it can greatly accelerate the process of obtaining new improved varieties, as compared to phenotypic selection. Nevertheless, the use of MAS requires a prior knowledge of reliable marker-trait associations relatively well-preserved over environments.

Although a linkage map is not strictly required for MAS, having a dense marker genetic map greatly facilitates finding strong marker-gene links because it allows for implementation of QTL mapping approaches other than ANOVA tests. Linkage-map based mapping algorithms, such as composite interval mapping (Zeng 1993, 1994) and multiple interval mapping (Kao et al. 1999; Zeng et al. 2000) have been proven to be more accurate in determining QTL locations and effects than simple marker-trait associations (Kao et al. 1999). Preferred markers for soybean researchers included single sequence repeats (SSR), restriction fragment length polymorphism (RFLP), and amplification fragment length polymorphisms (AFLP). Recently, single nucleotide polymorphisms (SNPs) are increasingly becoming available for the soybean genome, which opens enormous opportunities to create high-density maps because of their abundance. Genotyping of segregating mapping populations is labor-consuming and consequently it has been a limiting factor in QTL studies. However, with the advent of high-throughput techniques performing comprehensive genome-wide marker studies at a reasonable cost it is less of an issue. One of which, involving multiplexing SNP assays using the Illumina GoldenGate system, has proven to be effective even in discriminating paralogous genes in the complex soybean duplicated genome (Hyten et al, 2008).

Crop improvement is undeniably attached to genetic variability. Regrettably, soybean genetic variability is less compared to many other common cultivated species such as maize, rice, barley, and sugarcane, which limits the plant material within which to screen for desirable genes. The use of plant species that have been introduced to areas outside their natural ranges (Plant introductions, PI) allows searching for new sources of alleles that have been lost in the process of domestication and establishing commercial

varieties, especially in traits not considered important until recently. However, when the need arises, one can use this rich source of genetic variability for newly discovered important traits. PI 437654, with origin in China, has very high isoflavone content, sometimes up to 4-fold more in comparison to some commercial varieties (Gutierrez-Gonzalez et al., 2009), which makes it very suitable for isoflavone studies. In addition it is resistant to the soybean cyst nematode (SCN) races 1, 3, 4, and 5, and partially resistant to race 2 (Anand et al. 1984, 1988; Arelli et. 1997). QTL mapping and marker-trait associations are powerful tools for geneticists and plant breeders. It is estimated that more than 1200 quantitative trait loci have been reported for twelve major crop species (Bernardo 2008). However, very few of them have been confirmed in a different genetic background or environment. Herein, we have constructed a new mapping population and conducted a QTL study with the objective of: i) serving as a confirmation population of our previous QTL study on Essex × PI 437654, to check for consisting QTL. Sharing the high-isoflavone parent likely increases the chance of detecting common QTL. ii) because of the large difference in isoflavone content between the parental lines, being able to identify new important QTL that may be used in breeding programs.

MATERIALS AND METHODS

Plant Material and Growing Conditions

A recombinant inbred line (RIL) mapping population comprised of 188 F_{7-8} derived progeny was developed from a cross between the isoflavone-contrasting varieties Magellan (low-content) and PI437654 (high-content). The population was planted in

irrigated fields in two-row plots with three replications at two locations: University of Missouri Bradford Research and Extension Center (BREC, 36º 58' N) and the University of Missouri Delta Research Center (DRC, 36º 44' N). Seed samples were harvested from a pool of at least three plants and quantified for individual isoflavone content. Effects were tested using PROC GLM in SAS STAT 9.1 (SAS Institute Inc., Cary, NC). The pooled linear model contained the effect of environment, replication (environment), genotype, and environment × genotype. Variance components to calculate heritability estimates were determined using PROC Mixed.

Isoflavone Extraction and Quantification

Approximately 2.5 grams (~ 20 seeds) of soybean seeds were ground to a fine powder using a commercial coffee grinder. The powder was extracted with 7 mL of 80% methanol at 55 °C for 2 h, vortexing every 30 min. After centrifugation (6076 RCF (×g), 5 min), the supernatant was filtered using Fisherbrand 0.45 μm 25 mm nylon syringe filters (Fisher Scientific, Pittsburgh, PA). Samples were analyzed by reverse-phase HPLC on an Agilent 1100 high-performance liquid chromatography (HPLC) system (Santa Clara, CA). Separation and elution were accomplished using an 18 min linear gradient initiated with 20% methanol / 80% 10mM ammonium acetate (v/v) (pH 5.6) and completed with 100% methanol at a flow rate of 1 mL/min. A RP-C18 Lunar C2 column was used (Phenomix, La Jolla, CA). Detection of metabolites was achieved by photodiode array. Identification and quantification of each isoflavone component were based on available standards (Indofine Chemical Co., Somerville, NJ). Measurements are given as micrograms of isoflavone per gram of seeds plus/minus standard deviation, when corresponds (μg/g ± SD).

Single nucleotide polymorphism (SNP) analysis

The universal soybean linkage panel 1.0 (the USLP 1.0) containing 1,536 SNP loci that were mapped onto the integrated molecular genetic linkage map as previously described (Hyten et al. 2008) was utilized to genotype the RIL mapping population using the Illumina GoldenGate assay (Fan et al. 2006). For the assay, a total of 5 ul of 50 ng/ul of RNase-treated genomic DNA sample was activated by biotinylation, followed by the oligonucleotide/target annealing step, in which the SNP-specific oligonucleotide annealed to the activated DNA by ramping the temperature from 70^{o}C to 30^{o}C over 2 h. Three oligonucleotide sequences, two allele-specific oligos (ASO) and one locus-specific oligo (LSO) designed for each SNP, contain regions of genomic complementary and universal PCR primer sites. The LSO also contained a unique IllumiCode sequence complementary to a particular bead type. Following the assay oligonucleotide hybridization, wash steps to remove excess and mishybridized oligos, and adding a master mix of extension and ligation, the assay of oligonucleotide extension and ligation were implemented at 45^{o}C for 15 min. Basically, DNA polymerase was used to extend the ASO and fill the gaps between the ASO and LSO; and a DNA ligase was employed to seal the nick between the extended ASO and LSO to form PCR template that can be amplified with three universal PCR primers. These universal primers were 5'labeled with Cy3, Cy5, and biotin. The PCR amplification was then performed following a thermal cycler program: 10 min at 37^{o}C followed by 3 min at 95^{o}C; 34 cyclers of denaturation at 95^{o}C for 35 sec, annealing at 56^{o}C for 35 sec, and extension at 72^{o}C for 2 min. A 10-min extension at 72^{o}C followed the last cycler.

The resulting double-stranded PCR products were immobilized onto paramagnetic particles followed by steps of washing and denaturing. The released single-stranded DNAs were then hybridized to their complementary bead type through their unique IllumiCode sequence of the Sentrix array matrix (SAM) under a temperature gradient program for at least 12 h. The hybridized SAM was rinsed and dried for 20 min in dark. The array imaging was performed using the Illumina BeadStation (Illumina, San Diego, CA).

The allele calling for each SNP locus was conducted with the BeadStudio 3.0 software (Illumina, San Diego, CA) based on the intensities detected from the two channels, Cy3 and Cy5, for the two respective alleles of each SNP. The clusters of homozygote and heterozygous genotypes for each SNP were manually checked and identified for polymorphisms between the two parental lines. The polymorphic SNP loci were then employed for the genetic linkage analysis and QTL mapping.

Linkage Map and QTL Analysis

Linkage map contained a total of 741 SNPs markers which were distributed on 27 linkage groups (LG) covering 2008.9 cM, with an average distance between markers of 2.7 cM. Statistical analysis was performed using the SAS STAT 9.1 program, and heritability in the broad-sense over environments was calculated according to Hill et al. (1998). QTLNetwork v2.0 (Institute of Bioinformatics, Zhejiang University, Hangzhou, China) was chosen to perform a mixed-model based composite interval mapping for the QTL mapping analysis and run with the two locations input data and three replications per location all combined in a 2×3 scheme. In addition, each environment was run

separately. Critical F-value was assessed by permutation test using 1000 permutations. QTL effects were estimated using Markov chain Monte Carlo method. Candidate interval selection, epistatic effects, and putative QTL detection were calculated with an experimental-wise Type I error of α=0.1, α=0.001, α=0.001, respectively. Significance level for candidate intervals was relaxed to allow more genomic intervals to be included for further testing. Genome scan was performed using 10 cM window size and 1 cM walk speed.

RESULTS AND DISCUSSION

Genetic and Phenotypic Variation within Mapping Population

Parental lines were chosen because of their great difference in isoflavone content. Despite the low-isoflavone content of Magellan seeds, transgressive segregation was observed (Fig. 4.1), which implies that this parent might also bear positive-effect alleles. An analysis of variance (ANOVA) was conducted over environments (Table 4.1), indicating that accumulation of the distinct isoflavones in soybean seeds is largely governed by genetic (G), environmental (E), and G × E interaction effects ($P<0.0001$).

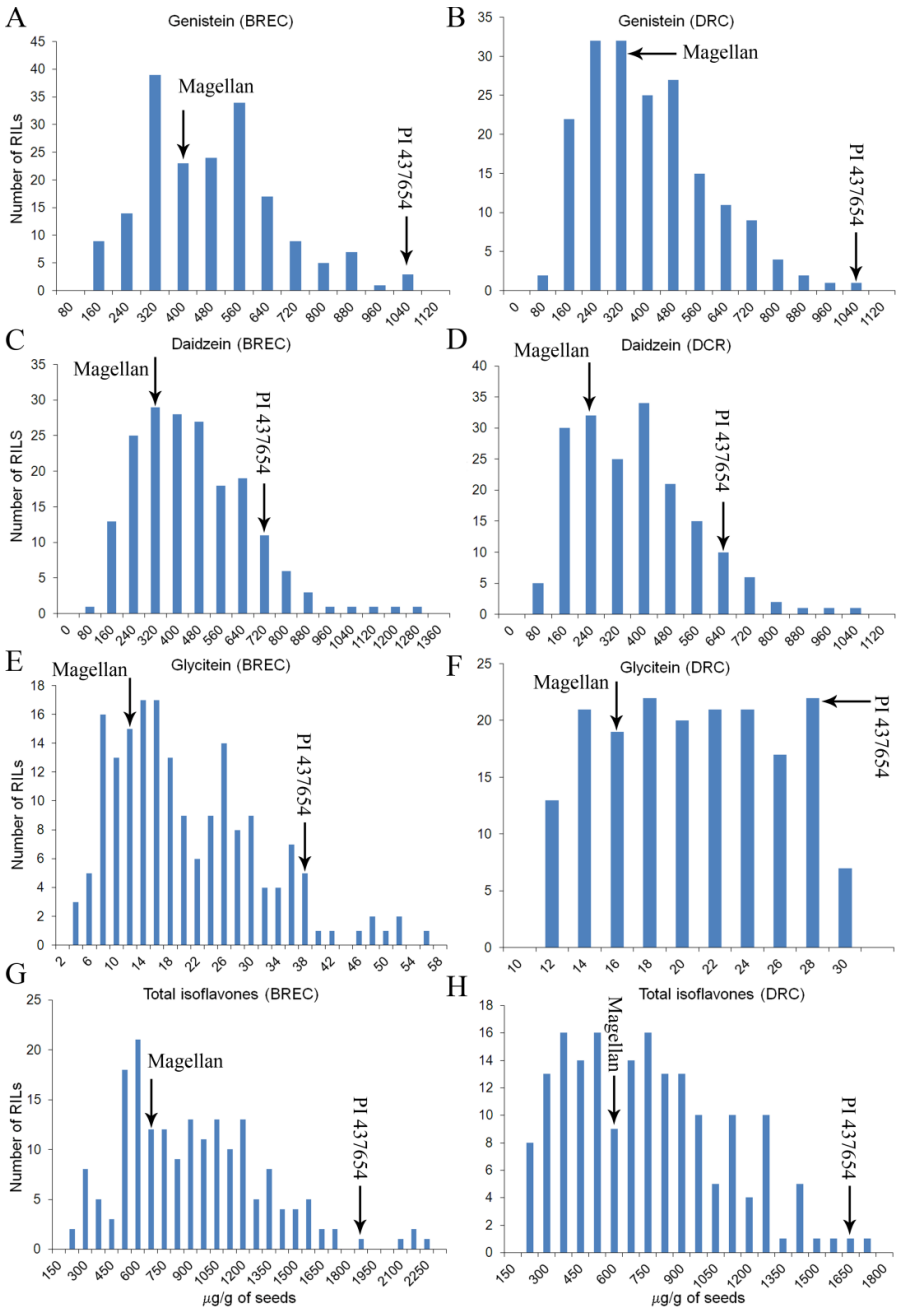

Fig. 4.1. Distribution of average genistein, daidzein, glycitein, and total isoflavones in three replications of Magellan × PI 437654 RILs growing in two field locations, BREC and DRC in year 2007. Arrows indicate the position of the two parental lines. Horizontal axis shows each particular isoflavone seed content in μg/g of seeds.

The influence of genes and environment on isoflavone synthesis has been extensively reported (Eldridge et al. 1983; Wang and Murphy 1994; Tsukamoto et al. 1995; Hoeck et al. 2000; Mebrahtu et al. 2004; Caldwell et al. 2005; and Lozovaya et al. 2005), and there are so many external affecting aspects that even subtle changes at the micro-environmental level could lead to significant differences in isoflavone accumulation. The distribution of frequencies (Fig. 4.1), and the normality-associated parameters (skewness and kurtosis) support the notion of these traits being inherited in a quantitative manner, with many genes exerting small individual effects. However, glycitein appears to separate from normality. This could be due to glycitein being present in small amounts in seeds, and it has also been shown to be the more variable of all isoflavones. Alternatively, it could also suggest that few glycitein-synthesizing genes are segregating in the progeny, or the presence of a major QTL. Results strongly support the first hypothesis as only three loci were found to affect glycitein amounts. Heritability in the broad sense (H^2) exhibited elevated numbers (>90%) for genistein, daidzein and total isoflavones (Table 4.1), indicating that despite of the influence of the environment, the recorded variation is on average largely under genetic control.

QTL main-effects and interactions

A total of twenty-seven genomic regions were identified having an effect on genistein, daidzein, glycitein, and total isoflavone seed content (Tables 4.2 and 4.3). The high-density marker map allowed not only narrowing down the confidence intervals in which the QTL were placed but also detecting a great number of minor-effect loci. A major QTL was found in Gm05 (*Glycine max* chromosome 5), former linkage group (LG) A1, and accounted for all three plus total isoflavone accumulation in seeds (*qGEN5*,

qDAI5, *qGLY5*, and *qTOT5*). This was also reported by our and other groups as a main isoflavone-affecting region Primomo et al, 2005; Gutierrez-Gonzalez et al, 2009). To our knowledge this is the first time that the QTL is reported to affect glycitein concentrations. Only this only locus was able to explain nearly one-third of genistein, daidzein and total isoflavones phenotypic variation, between 27.8 and 33.9 %, and it is the isoflavone QTL with the greatest effects ever reported. The proportion explained was comparatively smaller but still the largest for glycitein, with 7.6% of the variance. These observations and the additive-effect values strongly suggest that Gm05 bears an important gene for isoflavone synthesis, likely either an early enzyme in the phenylpropanoid pathway or a *trans*-acting factor. Fine mapping of the region would certainly narrow down the confidence interval and help in the identification of possible candidate genes. Another locus was also found influencing all four traits studied (*qGEN6*, *qDAI6*, *qGLY6*, and *qTOT6*). This was located in Gm06 (LG C2) and accounted for ~5%, on average, of the phenotypic variance of genistein, daidzein, and total isoflavone amounts. Several other minor-effect QTL were identified for individual and total isoflavones, indicating that isoflavone accumulation in soybean seeds is largely governed by loci with small individual additive effects. Nevertheless, surprisingly only three genomic regions were associated with glycitein content in seeds. Likely this is related to the departure of glycitein form normality, especially in one location (Fig. 4.1), which is also reflected in the parameters skewness and kurtosis (Table 4.1).

Several epistatic interactions were found to be significant for individual and total isoflavone concentrations (Table 4.4). However, no single epistatic QTL were detected for glycitein, the minor and more variable soybean isoflavone. Epistatic interactions have

been shown to be a key genetic factor in determining isoflavone final seed contents. Although epistatic effects are frequently larger than for main QTL, their use in breeding programs is nearly impracticable. The main obstacle is the variability experienced by epistatic QTL when changing growing conditions.

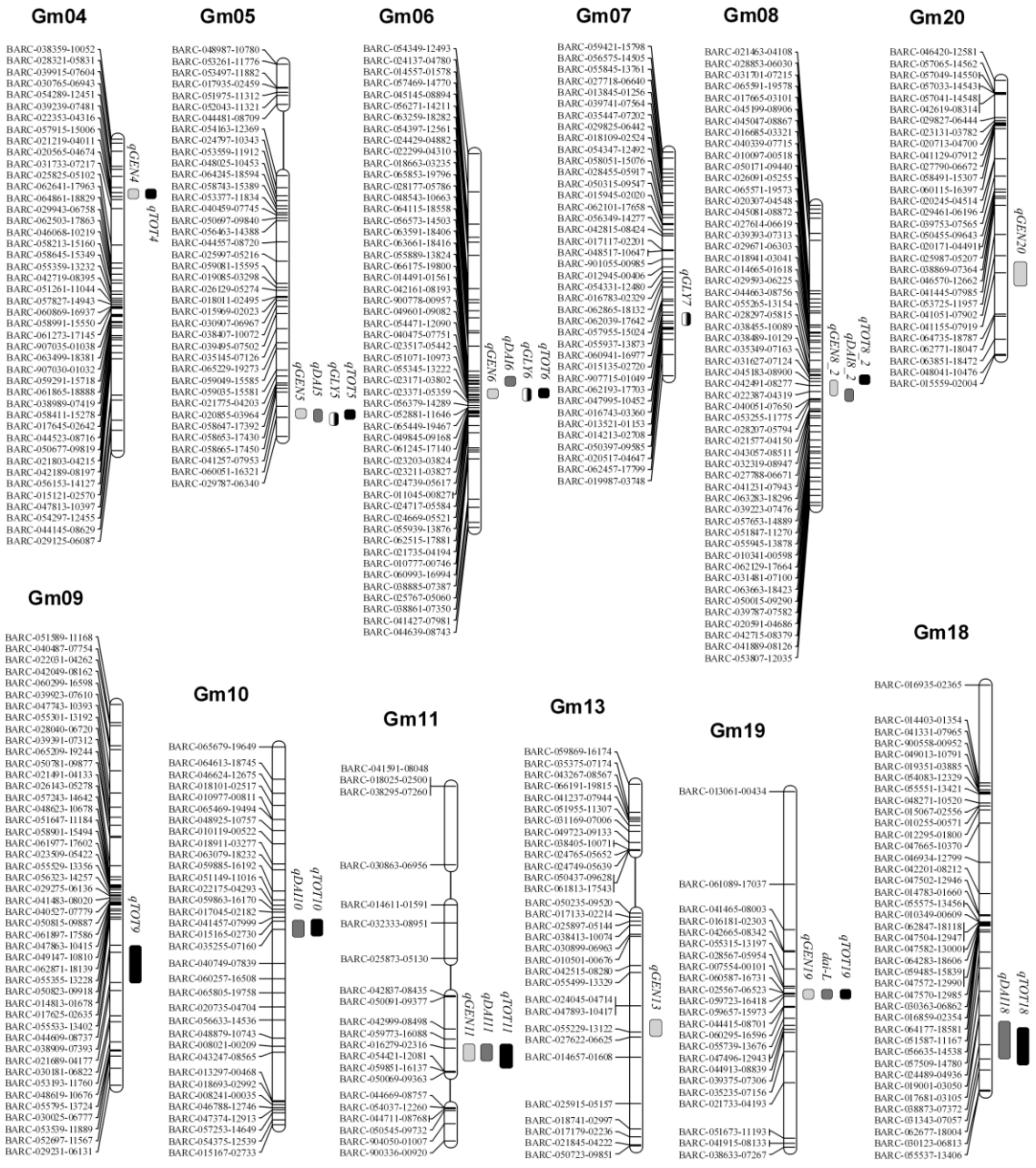

Fig. 4.2. Linkage map constructed for Magellan × PI 437654 population. SNP markers are shown at the left side of each linkage group. Isoflavone QTL are depicted at the right side. Light grey ovals indicate loci associated with genistein, dark grey ovals indicate loci associated with daidzein, black ovals indicate loci associated with total isoflavone, and hatched ovals indicate loci associated with glycitein. The name of the QTL, shown nearby each oval, is a composite of the influenced trait: genistein (*gen*), daidzein (*dai*), glycitein (*gly*) and total of isoflavones (*tot*), followed by the chromosome number. Red asterisks indicate confirmation QTL.

Table 4.1. ANOVA, random effects, and heritability over environments of the Magellan x PI 437654 mapping population.

				Effects					
Trait	Mean ± SD	Range	CV(%)	Genetic	Env	G x E	H^2	Skewness	Kurtosis
Genistein	408.8 ± 202.1	58–1297	49.4	<0.0001	<0.0001	<0.0001	0.90	0.707	0.370
Daidzein	384.8 ± 212.2	49–2013	55.1	<0.0001	<0.0001	<0.0001	0.95	1.300	4.486
Glycitein	20.3 ± 13.3	0–277	65.5	<0.0001	<0.0001	<0.0001	-	9.972	171.320
Total Isoflav	813.7 ± 413.4	124–3174	50.8	<0.0001	<0.0001	<0.0001	0.95	0.940	1.708

[a] Mean ± Standard deviation (SD) and range units in µg of isoflavones per gram of seeds.
[b] P values of random effects over environments: genetic, environmental, and g x e interaction.
[c] Heritability of the traits in the broad sense over environments.
[d] Phenotypic CV combined over locations

Table 4.2. Additive QTL for genistein and daidzein accumulation in soybean seeds

QTL[a]	interval[b]	range	A±SE[c]	P-Value	h^2(a)
Genistein					
qGEN4	BARC-057915_15006-BARC-021219_04011	16.1-17.6	-23.8±4	0.000000	1.3
qGEN5	BARC-041257_07953-BARC-060051_16321	73.1-74.8	133.7±4.1	0.000000	33.9
qGEN6	BARC-040475_07751-BARC-023517_05442	73.4-74.7	32.7±4	0.000000	3.1
qGEN8_2	BARC-022387_04319-BARC-040051_07650	54.5-58.5	-29.2±4	0.000000	2.4
qGEN11	BARC-016279_02316-BARC-054421_12081	15.6-19.2	24.9±4	0.000000	2.2
qGEN13	BARC-055229_13122-BARC-027622_06625	32.6-38.3	-34.3±4	0.000000	2.6
qGEN19	BARC-059723_16418-BARC-059657_15973	62.4-63.2	30.6±4	0.000000	1.1
qGEN20	BARC-041155_07919-BARC-064735_18787	56.3-63.5	-25.6±4	0.000000	1.1
Daidzein					
qDAI5	BARC-041257_07953-BARC-060051_16321	73.1-75.8	125.6±4.4	0.000000	31.4
qDAI6	BARC-066175_19800-BARC-014491_01561	69.0-71.3	52.6±4.2	0.000000	6.6
qDAI8_2	BARC-022387_04319-BARC-040051_07650	56.7-60.2	-27.7±4.3	0.000000	1.9
qDAI10	BARC-035255_07160-BARC-040749_07839	54.8-58.3	-24.3±4.2	0.000000	1.7
qDAI11	BARC-016279_02316-BARC-054421_12081	15.6-20.2	31.3±4.2	0.000000	2.0
qDAI18	BARC-017681_03105-BARC-038873_07372	104.1-115.3	-13.7±4.4	0.002101	1.6
dai-L	BARC-059723_16418-BARC-059657_15973	62.4-63.2	35.7±4.2	0.000000	1.2

[a]name given to a particular QTL, *GEN* and *DAI* for genistein and daidzein content, respectively. [b]Interval of confidence in centiMorgans with respect to the first marker in the LG. [c]main additive effect in mg/g plus/minus standard error. Mean effect of substituting both Magellan alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL. P-values represent the significance of each effect.

Table 4.3. Additive QTL for glycitein and total isoflavone accumulation in soybean seeds

QTL[a]	interval[b]	range	A±SE[c]	P-Value	h^2(a)
Glycitein					
qGLY5	BARC-060051_16321-BARC-029787_06340	73.1-77.8	3.8±0.4	0.000000	7.6
qGLY6	BARC-054471_12090-BARC-040475_07751	72.3-76.2	1.1±0.4	0.004506	0.7
qGLY7	BARC-062193_17703-BARC-047995_10452	50.4-53.0	-1.3±0.4	0.000622	1.0
Total isof.					
qTOT4	BARC-057915_15006-BARC-021219_04011	16.1-17.6	-65.8±7.9	0.000000	1.3
qTOT5	BARC-041257_07953-BARC-060051_16321	73.1-74.8	227.1±8.3	0.000000	27.8
qTOT6	BARC-040475_07751-BARC-023517_05442	73.4-74.7	80.4±8.1	0.000000	4.9
qTOT8_2	BARC-031627_07124-BARC-045183_08900	52.6-55.5	-50.5±7.9	0.000000	1.5
qTOT9	BARC-044609_08737-BARC-038909_07393	74.3-85.6	-60.6±8.6	0.000000	1.9
qTOT10	BARC-035255_07160-BARC-040749_07839	53.8-58.3	-50.6±7.9	0.000000	0.9
qTOT11	BARC-016279_02316-BARC-054421_12081	15.6-22.2	57.1±8.2	0.000000	1.6
qTOT18	BARC-038873_07372-BARC-031343_07057	106.1-117.3	-76.1±8.1	0.000000	2.5
qTOT19	BARC-059723_16418-BARC-059657_15973	62.4-63.2	40.5±8	0.000000	0.8

[a]name given to a particular QTL, GLY and *TOT* for glycitein and total isoflavone content, respectively. [b]Interval of confidence in centiMorgans with respect to the first marker in the LG. [c]main additive effect in mg/g plus/minus standard error. Mean effect of substituting both Magellan alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL. P-values represent the significance of each effect.

Table 4.4. Additive by additive epistatic interactions for genistein, daidzein, and total isoflavones.

interval_i[a]	interval_j[a]	AA±SE[b]	P-Value	h^2(aa)[c]
Genistein				
BARC-041257_07953-BARC-060051_16321	BARC-055229_13122-BARC-027622_06625	-24.9±4.2	0.000000	1.2
BARC-050661_09809-BARC-050325_09554	BARC-010777_00746-BARC-060993_16994	6.6±3.9	0.097149	0.2
BARC-031263_07027-BARC-026065_05240	BARC-062647_17964-BARC-052789_11619	35.9±4.6	0.000000	1.8
BARC-045053_08869-BARC-055637_13558	BARC-023509_05422-BARC-055529_13356	-30.4±4	0.000000	1.9
BARC-029825_06442-BARC-018109_02524	BARC-030171_06819-BARC-018923_03037	-30.2±6.7	0.000008	1.0
BARC-059869_16174-BARC-035375_07174	BARC-031363_07063-BARC-040761_07846	10.3±4.5	0.022084	1.1
BARC-059869_16174-BARC-035375_07174	BARC-012703_00380-BARC-017933_02457	11.2±4.5	0.012535	0.3
BARC-016891_02361-BARC-038405_10071	BARC-012703_00380-BARC-017933_02457	11.1±4	0.005493	0.2
BARC-011591_00299-BARC-048855_10738	BARC-044415_08701-BARC-060295_16596	-29.3±4.3	0.000000	1.7
Daidzein				
BARC-041257_07953-BARC-060051_16321	BARC-066175_19800-BARC-014491_01561	24.3±4.4	0.000000	1.2
BARC-041257_07953-BARC-060051_16321	BARC-022387_04319-BARC-040051_07650	-16.5±4.5	0.000271	0.9
BARC-044557_08720-BARC-025997_05216	BARC-013637_01186-BARC-050543_09730	28.4±4.7	0.000000	1.4
BARC-065229_19273-BARC-059049_15585	BARC-064301_18614-BARC-016831_02340	14.7±5.1	0.004209	0.2
BARC-053261_11776-BARC-053497_11882	BARC-062101_17658-BARC-056349_14277	19.5±4.6	0.000025	1.1
BARC-053261_11776-BARC-053497_11882	BARC-013845_01256-BARC-039741_07564	15.5±4.6	0.000925	0.6
BARC-054471_12090-BARC-040475_07751	BARC-041649_08056-BARC-024045_04714	-22.6±4.3	0.000000	0.8
BARC-022031_04262-BARC-042049_08162	BARC-044669_08757-BARC-054037_12260	-21.8±4.5	0.000001	0.9
BARC-065229_19273-BARC-059049_15585	BARC-031627_07124-BARC-045183_08900	-23.7±8.1	0.003403	0.6
Total isoflavones				
BARC-035255_07160-BARC-040749_07839	BARC-016279_02316-BARC-054421_12081	-41.4±8.3	0.000001	1.0
BARC-017645_02642-BARC-044523_08716	BARC-029825_06442-BARC-018109_02524	56.5±12	0.000003	0.9
BARC-050781_09877-BARC-021491_04133	BARC-050437_09628-BARC-061813_17543	-47.4±8.2	0.000000	1.3
BARC-008241_00035-BARC-046788_12746	BARC-020713_04700-BARC-041129_07912	48.5±8	0.000000	1.6

[a]Marker interval within each epistatic interaction occurs. [b]estimated additive by additive effect ± standard error, a positive sign for epistatic effects indicate that parental allele combinations at the two loci involved in epistasis increase total isoflavones expression while a negative sign indicate that recombinant allele combinations increase total isoflavone values. [c]heritability of the additive by additive interaction effect (%). P-values represent the significance of each effect.

REFERENCES

Anand SC, Gallo KM, Baker IA, Hartwig EE (1988) Soybean plant introductions with resistance to races 4 or 5 of soybean cyst nematode. Crop Sci. **28**:563-564.

Anand SC and Gallo KM (1984) Identification of additional soybean germplasm with resistance to race 3 of the soybean cyst nematode. Pl Dis. **68**:593-595.

Arelli APR , Wilcox JA, Meyers O Jr, Gibson PT (1997) Soybean Germplasm Resistant to Races 1 and 2 of Heterodera glycines. Crop Sci. **37**:1367-1369.

Beggs CJ, Stolzer-Jehle A, Wellmann E (1985) Isoflavonoid formation as an indicator of UV stress in bean (Phaseolus vulgaris L.) leaves: the significance of photorepair in assessing potential damage by increased solar UV-B radiation. Plant Physiol. **79(3)**:630-634.

Caldwell CR., Britz SJ, Mirecki RM (2005) Effect of temperature, elevated carbon dioxide, and drought during seed development on the isoflavone content of dwarf soybean [Glycine max (L.) Merrill] grown in controlled environments. J. Agric. Food Chem. 53:1125-1129.

Dakora FD, and Philips DA (1996) Diverse functions of isoflavonoids in legumes transcend anti-microbial definitions of phytoalexins. Physiol. Mol. Plant Pathol. 49:1-20.

Eldridge A, Kwolek W (1983) Soybean isoflavones: Effect of the environment and variety on composition. J. Agric. Food Chem. **31**:394-396.

Fan J, Gunderson K, Bibikova M, Yeakley J, Chen J, Garcia EW, Lebruska L, Laurent M, Shen R, Barker D (2006) Illumina universal bead arrays. Methods Enzymol 410:57-73

Hoeck JA, Fehr WR, Murphy PA, Welke GA (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci. **40**:48-51.

Hyten D, Song Q, Choi I, Yoon M, Specht J, Matukumalli L, Nelson R, Shoemaker R, Young N, Cregan P (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945-952

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics **152**:1203-1216.

Lozovaya VV, Lygin AV, Ulanov AV, Nelson RL, Dayde J, and Widhohm JM (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci. **45**:1934-1940.

Mebrahtu T, Mohamed A, Wang CY, Andebrhan T (2004) Analysis of isoflavone contents in vegetable soybeans. Plant Foods for Human Nutrition **59**:55-61.

Moy P, Qutob D, Chapman BP, Atkinson I, and Gijzen M (2004) Patterns of gene expression upon infection of soybean plants by Phytophthora sojae. Mol. Plant Microbe Interact. 17:1051-1062.

Rochfort S, Panozzo J (2007) Phytochemicals for health, the role of pulses. J. Agric. Food Chem. **55**:7981-7994.

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol. **137**:1345-1353.

Tsukamoto C, Shimada S, Igita K, Kudou S, Kokubun M, Okubo K, Kitamura K (1995) Factors affecting isoflavone content in soybean seeds: Changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J. Agric. Food Chem. 43:1184-1192.

Usui T (2006) Pharmaceutical prospects of phytoestrogens. Endocrine J. **53**(1):7-20.

Wang H, Murphy PA (1994) Isoflavone Composition of American and Japanese Soybeans in Iowa: Effects of Variety, Crop Year, and Location. J. Agric. Food Chem. **42**:1674-1677.

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90**:10972-10976.

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics **136**:1457-1468.

Zeng ZB, Kao CH, Basten CJ (2000) Estimating the genetic architecture of quantitative traits. Genet. Res. **74**:279-289.

Zhang J, and Yu O (2009) Metabolic engineering of isoflavone biosynthesis in seeds. *In* Modification of seed composition to promote health and nutrition. Hari Krishnan (ed). Agronomy Monograph Series, pp151-177.

CHAPTER 5

Differential expression of isoflavone biosynthetic genes in soybean during water deficits

"*If the facts don't fit the theory, change the facts"*
(Albert Einstein)

ABSTRACT

Numerous environmental factors influence isoflavone accumulation and have long hampered their genetic dissection. Temperature and water regimes are two of the most significant abiotic factors. However, while the effects of temperature have been broadly studied, not much is known about how water scarcity might affect isoflavone content in seeds. Studies have shown that accumulation of isoflavones is promoted by well-water conditions, but the molecular basis remains elusive. In the present work, several intensities of water stress were evaluated at various critical stages of soybean (*Glycine max* (L.) Merr.) seed development, in both field and controlled environments. Results suggested that only long-term progressive drought spanning most of seed developmental stages significantly decreased isoflavone content in seeds. The reduction is proportional to the intensity of the stress and appears to occur in a genotype-dependent manner. However, regardless of water regime, isoflavone compounds were mainly accumulated in the later seed developmental stages. Transcripts of the most important genes for isoflavone biosynthesis were also quantified from samples collected at key seed developmental stages under well-watered and long-term water-deficit conditions. Expression of *CHS7, CHS8,* and *IFS2* correlated with isoflavone accumulation under

well-watered conditions. Interestingly, we found the two isoflavone synthase genes in soybean (*IFS1* and *IFS2*) showing different patterns of expression. The abundance of *IFS1* transcripts was maintained at a constant rate, whereas *IFS2* was down-regulated and highly correlated with isoflavone accumulation under both water-deficit and well-watered conditions, suggesting *IFS2* as a main contributor to isoflavone diminution under drought.

INTRODUCTION

Isoflavonoids are a group of secondary metabolites present mostly in leguminous plants and particularly abundant in soybean. There is an increasing interest in these compounds because they have long been associated with important preventive and therapeutic medicinal properties (reviews Messina, 1999; Rochfort and Panozzo, 2007). Nevertheless, a complete understanding of their synthesis and accumulation has been impeded because they are highly influenced by genetic and environmental factors which are not always well understood (Wang and Murphy, 1994; Hoeck et al. 2000; Mebrahtu et al. 2004; Gutierrez-Gonzalez et al. 2009). Illustrating their complexity, a great majority of the discovered isoflavone-associated QTL bear minor-effects, and epistasis, or interactions between effects of alleles from two or more genetic loci, can account for as much as 50% of the variation for a particular isoflavone in a given environment (Gutierrez-Gonzalez et al. 2009). Influencing environmental factors consist of both biotic, such as wounding, nodulation and pathogen attack, and abiotic elements: temperature, water regime, UV light, soil nutrient content, and carbon dioxide (Dixon and

Paiva, 1995; Lozovaya et al. 2005; Subramanian et al. 2006, and 2007; Naoumkina et al. 2007).

Genistein, daidzein and glycitein, the known soybean isoflavones, are synthesized by a branch of the phenylpropanoid pathway. This extended metabolic route is also involved in the synthesis of other important plant compounds such as tannins, lignins, lignans, anthocyanins, flavones, flavonols, and the soybean phytoalexins glyceollins, which are pterocarpans that possess antimicrobial activities (Yu and McGonigle, 2005). The precursor in the pathway is the amino acid phenylalanine, which in the initiating step is removed of its amine group to render cinnamic acid by the enzyme phenylalanine ammonia lyase (*PAL*). The first critical enzyme for flavonoid synthesis is chalcone synthase (*CHS*), which exists as a multi-gene family in soybean, although not all copies are expressed in seeds at detectable levels (Dhaubhadel et al. 2007). Other important enzymes in the pathway for isoflavone synthesis are chalcone isomerase (*CHI*), which converts chalcones to flavanones, and chalcone reductase (*CHR*), required for daidzein and glycitein formation. However, the enzyme that specifically differentiates isoflavone-producing plant species from those with no isoflavone content is isoflavone synthase (*IFS*), an endoplasmic reticulum integral cytochrome P450 monooxygenase, which catalyzes a 2,3 aryl ring migration of flavanones to their corresponding isoflavones (Akashi et al. 1999; Steele et al. 1999; Jung et al. 2000). In the soybean genome, IFS is present in two copies, *IFS1* and *IFS2*, that differ by a few amino acids. Both convert naringenin and liquiritigenin to genistein and daidzein, respectively. Despite their homology, *IFS1* and *IFS2* are differentially regulated at the transcriptional level. For instance, although both copies contribute to the level of isoflavones in seeds (Cheng et al.

2008), the expression of *IFS2* has been shown to increase during late stages of seed development while *IFS1* transcripts are kept more constant (Dhaubhadel et al. 2003, 2007). Moreover, only *IFS2* was shown to be induced in soybean hypocotyls and transgenic roots in response to pathogen attack (Dhaubhadel et al. 2003; Subramanian el al., 2004).

Isoflavones are found in roots, seedlings and flowers, being especially abundant in seeds and leaves. Within seeds, distinct tissues have the capability to produce isoflavones, and expression of important genes in the pathway has been found in pods, embryo, and seed coat (Dhaubhadel et al. 2003). Interestingly, activation of the pathway in seeds has shown to be tissue-dependent. For instance, both hypocotyls and cotyledons have the ability to synthesize isoflavones. However, isoflavone amounts in cotyledons appear to be more influenced by the environment than that of hypocotyls (Berger et al. 2008). Tissue specificity of *IFS* genes is also observed in other plant organs. While *IFS1* is expressed in many tissues, *IFS2* transcripts appear to be absent in, for instance, leaves and stems, or detected only if induced in response to stresses (Dhaubhadel et al. 2003).

Soybean plant growth and development are traditionally divided according to six vegetative (V) and eight reproductive (R) stages. R1 and R2 are considered flowering stages, while R3 and R4 encompass pod formation. Seed development occurs in stages R5 through R8. Studies have shown that final seed composition, including isoflavones, is heavily influenced by environmental stresses during R5-R7 (Gayler and Sykes, 1985; Brevedan and Egli, 2003; Lozovaya et al, 2005). Seed development in the R5 stage is characterized by a rapid increase in weight and nutrient accumulation, which continues until R6. Seeds in R6 stage fill the pod cavity but are still immature. Typical of the R7

stage, the seed coat begins to turn from green to either tan or yellow, depending on the cultivar. At this point, accumulation of dried weight ceases and soybean has reached its physiological maturity (McWilliams, 1999). Accumulation of isoflavones in soybean seeds takes place during the later stages of seed maturation (Dhaubhadel et al. 2007), and research suggests that their levels are greatly influenced by water availability during this period. Nevertheless, little is known about the timing and magnitude of the water deprivation required to exert a significant effect, and it is yet to be determined at which stage of seed development drought might cause more variation. Bennett et al. (2004) found that total isoflavone content increased by as much as 2.5 fold in irrigated vs. non-irrigated soybean crops during a two-year field study. Lozovaya et al. (2005) studied the effect of temperature and soil moisture status during seed development under controlled conditions and concluded that high soil moisture increased daidzein, genistein, and total isoflavones, but not glycitein when the stress was imposed at the beginning of the R6 stage. Expression patterns in well-watered conditions have also shown that increased expression of *PAL*, *CHS7*, *CHS8*, and *IFS2* temporally coincided with the onset of embryonic isoflavonoid accumulation. In addition, transcriptional abundance of *CHS7*, *CHS8*, and *IFS2* highly correlated with final isoflavone content (Dhaubhadel et al. 2007).

Drought is an increasing problem in many regions of the world and hence making the study of plant-water relations of extreme importance. Understanding the mechanisms regulating seed isoflavone accumulation under water-limiting conditions would be of vital importance for development of high-isoflavone-content soybean cultivars. We have conducted a series of experiments targeting critical phases of seed formation (R5-R8) to assess how the extent of water deficit might affect isoflavone accumulation in seeds. An

extended profiling comparing expression levels of essential genes and isoflavone content in watered vs. water-limiting conditions was performed. It was found that, while the expression of *IFS1* and other important isoflavone genes remained unchanged, *IFS2* expression was down-regulated in the latest stages of seed filling under water stress. Importantly, this change correlated with isoflavone seed content. These results suggested that *IFS2* is a main contributor to decreasing isoflavone concentrations in soybean seeds under drought conditions.

MATERIALS AND METHODS

Field drought treatment

For the field study, ten isoflavone-contrasting $F_{2:7}$ recombinant inbred lines from a cross between Essex (low-isoflavone content) and plant introduction PI 437654 (high-isoflavone), were planted in two-row plots under irrigated and non-irrigated conditions following a split-plot design with three replications at the University of Missouri Bradford Research and Extension Center (BREC, 38º 9’ N) and at the University of Missouri Delta Research Center (DRC, 36º 44’ N). Drought was imposed at the onset of R4 stage by a cessation of watering until physiological maturity. Stress intensity was quantified by means of soil moisture content (SMC), using a HH2 Moisture Meter and WET sensor (Delta-T Devices Ltd., Cambridge, UK), every five days from the beginning of R4 stage. Readings were taken at 25 cm deep in three randomly chosen places per plot. Three independent measurements were taken per reading point, and averaged. Plants were allowed to grow until physiological maturity and dry seeds were harvested and pooled. From each line, three independent samples were used for isoflavone

quantification as described below. Records of temperatures and precipitation during the testing period can be found at http://aes.missouri.edu/bradford/weather/ and http://aes.missouri.edu/delta/weather/.

Severe drought stress treatment in greenhouse

For the severe drought stress study at R5 stage (40 days after pollination, (DAP)), seeds of genotypes Williams 82, Magellan and PI 437654 were planted in 19-liter pots in a randomized complete block design (RCBD) with four replications of one well-watered and two stressed plants in controlled greenhouse conditions at the University of Missouri. Pots were weighed and adjusted to the same substrate amount by weight. Growth conditions were set at 30ºC/20ºC day/night temperatures, 16/8 hr day/night cycles, 800 µmol m^{-2} s^{-1} light intensity, and 60% relative air humidity.

Plants were allowed to grow under well-watered conditions until the onset of R5 stage at which time the stress treatment was imposed by completely withholding water. During drought stress, SMC was measured as described. In addition, measurements of leaf water potential (LWP), using Model 610 Pressure Chamber instrument (PMS Instrument Company, Albany, OR), and leaf relative water content (LRWC) following procedures described by Barrs and Weatherley (1962), were taken. When plants reached a LWP of -2.0 MPa, soil moisture levels were reestablished to well-watered values by re-watering, and maintained until plants reached maturity. Mature dried seeds were harvested and their isoflavone levels quantified.

Severe stress treatment in growth chamber

Severe drought stress was also imposed at R6 stage (60 DAP) in a separated growth chamber setting. Williams 82 seeds were grown in a growth chamber with three replicates under well-watered conditions until the beginning of R6 stage, at which time drought stress was imposed by completely withholding water. Seed samples were collected before and after the stress period, and RNA was extracted to perform qRT-PCR. Growth conditions were set to 27ºC/18ºC day/night temperatures and 14/10 hr day/night photoperiod, and 65% relative air humidity. After the stress period, plants were maintained in irrigated (control) conditions until maturity. Mature dried seeds were harvested and quantified for isoflavone levels.

Long-term drought experiment

To assess the level of stress required to induce varying levels of isoflavone accumulation, a long-period stress protocol has been adapted from the procedures and methods described by Brevedan et al. (2003). Williams 82 seeds were germinated in a growth chamber and received the same amount of water until growth stage R5 (30 DAP). For drought treatment, water was withheld until SMC reached the desired levels. Four different soil moisture levels at 25, 20, and 15% SMC, and at soil saturation capacity (well-watered conditions as a control) were tested, with three biological replicates for each moisture level. To maintain the drought treatment, stressed pots were weighed individually and again 24 h later. The difference between the two measurements was considered as the amount of water loss for each plant per day and, therefore, the amount of water to be applied to each particular pot to keep the SMC at the stressed levels

mentioned above until physiological maturity. Well-watered plants received enough water to maintain SMC between 40 to 45%. Once the low soil moisture conditions were imposed, drought parameters were measured (SMC and LWP) every five days. SMC was tracked by means of PR2 profile probe and the HH2 moisture meter (Delta-T Devices Ltd., Cambridge, UK). Measurements were taken for each pot at 5 and 30 cm depths, repeated three times in succession, and averaged. Growth chamber conditions were set to 27ºC/18ºC day/night temperatures, 14/10 hr day/night photoperiod, and 65% humidity. Mature dried seeds of each SMC level were collected and quantified for isoflavone contents.

Once the threshold level of stress to produce changes was established at 15% SMC (Fig. 5.7), four replicates of Williams 82 and RCAT-Angora were grown in a separate experiment under the aforementioned conditions. Mature and immature seed samples were taken at five time points and used for qRT-PCR and isoflavone quantification as explained below.

RNA isolation and cDNA synthesis

For gene expression analysis, Williams 82 seeds were collected from the long-term stress treatment at the R5e (early R5, 30 DAP), R5L (late R5, 45 DAP), R6 (60 DAP), R7 (70 DAP), and mature dried seeds (DS) stages. In addition, for transcript analysis of severe stress at R6, seed samples were also taken before and after the stress period. Transcripts of selected genes from the phenylpropanoid pathway were measured and quantified by quantitative real-time PCR (qRT-PCR) using gene-specific primers (Table 5.1). Total RNA was isolated from plants with TRIZOL reagent according to the

manufacture supplied protocol (Invitrogen, Carlsbad, CA). RNA concentration and quality were measured using NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE), prior to DNase digestion. For each sample, 18 ug of total RNA was digested in a volume of 50 µl with Turbo DNA-free DNase I (Ambion/Applied Biosystems, Austin, TX) to remove genomic DNA contamination according to the instructions provided by the manufacturer. After DNase I digestion, RNA concentration was determined again with the NanoDrop spectrophotometer, and first-strand cDNA synthesis was performed using 1 µg of DNAse-treated total RNA in 20 µl reaction using iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA), following the manufacturer provided instructions.

Quantitative RT-PCR

Gene-specific primers were designed using Primer Express software Version 2.0.0 (Applied Biosystems, Foster City, CA). Primer specificity was confirmed by blasting each primer sequence against soybean genome sequences at Phytozome (http://www.phytozome.net/) using the BLASTN algorithm. Quantitative RT-PCR reactions were performed in 384-well plates using the ABI 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA). SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA) was used for real-time DNA quantification. 0.4 µM primer concentration and 2 µl of prepared cDNA were used in a final volume of 10 µl per well. PCR reactions were set up as follows: 50ºC for 2 min, 95 ºC for 10 min, and 40 amplification cycles of 95ºC for 15 sec and 60ºC 1 min. Data were analyzed with SDS 2.2.1 software package (Applied Biosystems, Foster City, CA). A cyclophilin (CYP2, TC224926), a eukaryotic elongation factor 1-alpha (ELFα, TC203954), an F-box family

protein (CD397253), and an actin gene (ACT2/7, TC204150), were selected as housekeeping genes for comparison of gene expression levels in long-term and severe stress (Aldrich et al. unpublished). Gene expression levels were calculated following the guidelines of geNorm (Vandesompele et al. 2002). The Ct values were normalized to the geometrically averaged expression levels of those reference genes.

Isoflavone Extraction and Quantification

For dry seeds, approximately 2 g of soybean seeds (~ 15 seeds) were ground to a fine powder using a General Electric seed grinder (model 5XBG008, New York City, NY). For fresh tissue samples, 1~2 g of seeds were ground in liquid nitrogen with mortars and pestles. In both cases, the powder was used to extract isoflavones with 7 mL of 80% methanol at 55 °C for 2 h, vortexing every 30 min. After centrifugation (6076 RCF (×g), 5 min), the supernatant was filtered using Fisherbrand 0.45 μm 25 mm nylon syringe filters (Fisher Scientific, Pittsburgh, PA). Samples were analyzed by reverse-phase HPLC on an Agilent 1100 high-performance liquid chromatography (HPLC) system (Santa Clara, CA). Separation and elution were accomplished using an 18 min linear gradient initiated with 20% methanol / 80% 10mM ammonium acetate (v/v) (pH 5.6) and completed with 100% methanol at a flow rate of 1 mL min^{-1}. A RP-C18 Lunar C2 column was used (Phenomix, La Jolla, CA). Detection of metabolites was achieved by photodiode array. Identification and quantification of each isoflavone component were based on comparison of the retention times of the respective standards (Indofine Chemical Co., Somerville, NJ). Measurements are given as micrograms of isoflavone per gram of seeds plus/minus standard error (μg/g ± SE), when applicable.

RESULTS AND DISCUSSION

Small changes in soil moisture content had no effect on isoflavone concentrations in field trials

Evidence suggests that isoflavone synthesis and accumulation in seeds is affected by water regime. However, not much is known about the magnitude, time course or length of the stress required. We aimed to assess how water limitations influence isoflavone content in seeds by first conducting a field study. Ten soybean isoflavone-contrasting recombinant inbred lines with three replications were tested in two field locations under irrigated and non-irrigated conditions. Mature dried seeds were quantified for individual and total isoflavones and no significant differences (Student t-test, $P>0.05$) were found in any of the lines at either location (Figs. 5.1A and 5.1B). Stress intensity was assessed by periodical readings of soil moisture content (SMC) during last phases of embryogenesis, from the onset of R4 stage until physiological maturity (Fig. 5.1C). Irrigated plots were watered accordingly to keep SMC above 30%, while non-irrigated plants were not supplied with water, other than that from natural precipitation. Up to a 2.5-fold decrease has been reported in seed isoflavones in soybean cultivars growing in non-irrigated vs. irrigated fields (Bennett et al. 2004). Nevertheless, in our two-location experiment we found no statistical difference between irrigated and non-irrigated plots in any individual isoflavones. This discrepancy could be attributable to the uncontrollable natural precipitation. Several rain storms were recorded in both locations during this period, and it is reflected in the graph by the rising peaks of SMC at particular points. Although moisture levels measured on irrigated plots were always higher than plots that did not receive any extra water, this difference may not be significant (5-10% SMC) in

critical phases, from late R5 and during the entire R6 and R7 stages. Results from this field experiment suggest that small differences in SMC may not be enough to produce significant changes in isoflavone content, and that more severe drought stress may be required to exhibit changes in isoflavone content.

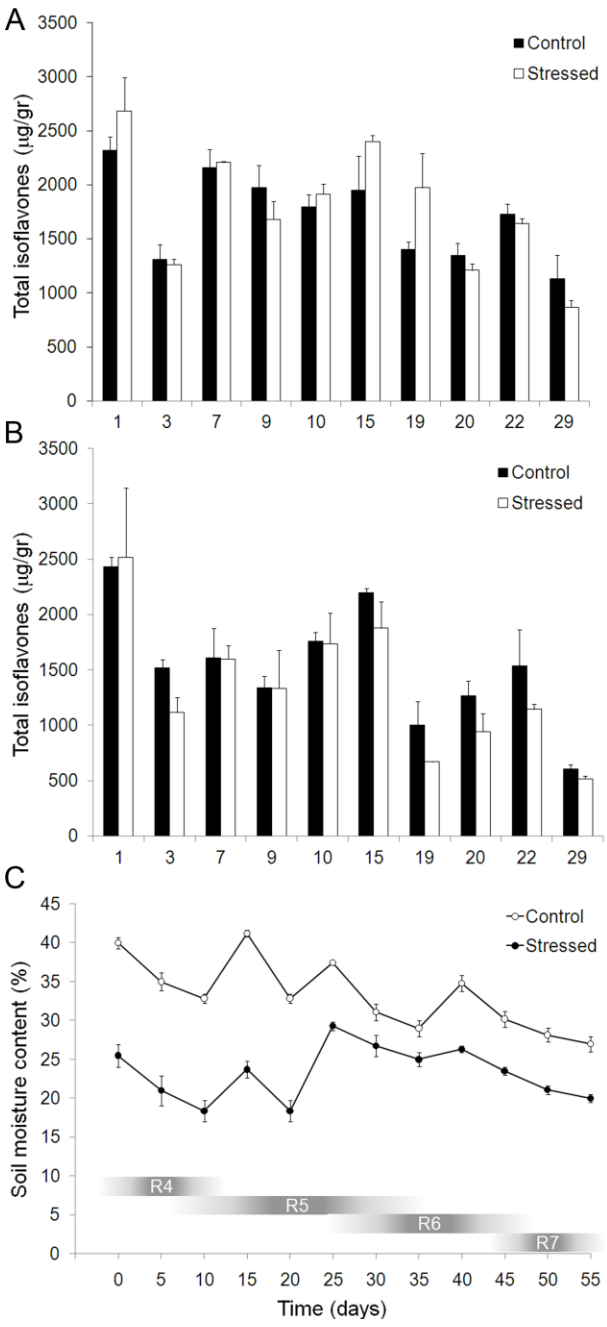

Fig. 5.1 *Isoflavone levels were not altered by small changes in soil moisture content during seed maturation*. Ten Recombinant Inbred Lines developed from a cross between Essex and PI437654 were grown under irrigated (control) and non-irrigated (stressed) conditions at Bradford Research and Extension Center (**a**) and Delta Research Center (**b**). Columns represent total isoflavones of mature dried seeds. Level of stress was monitored by taking soil moisture content measurements every 5 days. SMC at BREC was shown (**c**). Data are the means of measurements from three independent biological replicates. Bars represent the SE

Short periods of severe stress during seed filling did not affect isoflavone accumulation

Because mild stress conditions in the field appeared to contribute insignificantly on the variation of isoflavone content in seeds, we then studied the effect of more severe stresses. An inherent drawback to field assays is the exposure to uncontrollable weather conditions. To minimize these random external factors, experiments were implemented in controlled environments, where more drastic water-limiting conditions can be reached by completely withholding water. However, severe stress can only be maintained for short periods of time, after which plants must be re-watered to enable proper seed development. Two independent experiments were set up to target the middle of R5 and R6 stages, respectively, as they are crucial for accumulation of secondary metabolites (Pedersen, 2004; McWilliams et al. 1999). Seeds of 40 and 60 days after pollination (DAP) were selected as representatives of each respective stage.

In the first experiment, Williams 82, Magellan, and PI 437654 lines were grown in greenhouse under controlled conditions. Stress was imposed at R5 (40 DAP), and maintained until plants reached -2 MPa leaf water potential (LWP). At that point plants had between 60-65% leaf relative water content (LRWC) and 5-10% SMC (Figs. 5.2A, 5.2B, and 5.2C). After the stress period, plants were re-watered and maintained in well-watered conditions until physiological maturity. Whereas the variation of seed size, measured as 100-seed weight, was statistically significant in all three soybean cultivars ($P<0.001$, data not shown), there was no significant difference ($P>0.05$) in total isoflavone content of mature dried seeds between irrigated and non-irrigated pots in either genotype (Fig. 5.2D). There was also no variation ($P>0.05$) when the aggregate

composition of individual isoflavones, suggesting severe water stress during R5 stage alone had no significant impact on isoflavone levels.

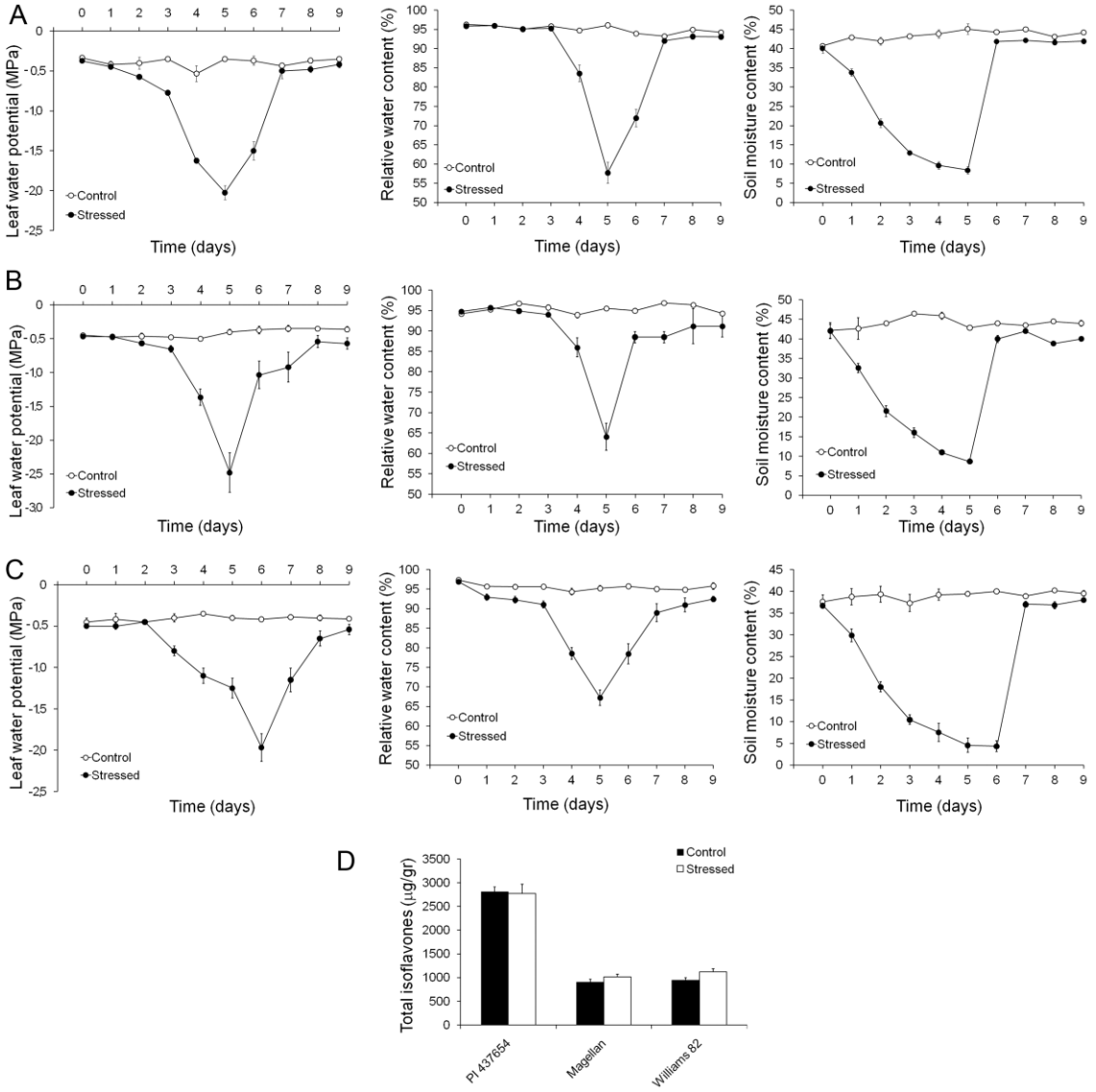

Fig. 5.2 *Severe drought stress at R5 stage (40 DAP) produce no change in isoflavone concentrations*. Plants of three genotypes: Magellan (**a**), Williams 82 (**b**) and PI 437654 (**c**) were grown in greenhouses in well-watered conditions until stress treatment. At R5 stage (40 DAP) they were withhold water until they reached -2 MPa leaf water potential after which they were re-watered and grown in well-watered conditions (control). Water stress was followed by daily measurements of leaf water potential (MPa), relative water content (%), and soil moisture content (%). Total isoflavones were quantified from mature dried seeds plants (**d**). Data are the means of measurements from three independent biological replicates. Bars represent the SE.

In the second experiment targeting the R6 stage, the stress was imposed to Williams 82 plants at 60 DAP by completely withholding water until plants reached -2 MPa LWP, after which time plants were re-watered and kept well-watered until maturity. Similarly, no differences (P>0.05) were found in either total isoflavones or individual compounds in mature dried seeds (Fig. 5.3A).

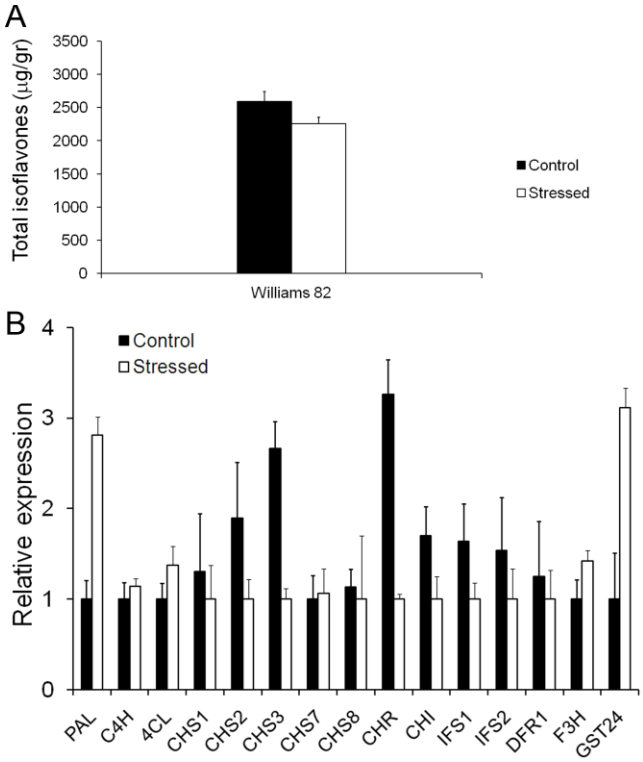

Fig. 5.3 (**a**) *Severe drought stress at R6 (60 DAP) produce no changes in isoflavone concentrations*. Plants of Williams 82 were grown in growth chambers in well-water conditions until stress treatment. At R6 stage, stress was imposed by complete water withholding and kept until plants reached -2.0 MPa. after which they were allowed to grow in well-watered conditions. Total isoflavones were quantified using mature dried seeds. (**b**) *Gene expression of some isoflavone-related genes altered significantly during water stress*. Transcripts were quantified by real-time PCR using seed samples taken immediately after the stress period, before re-watering and compared to well-watered plants (controls). Data are the means of measurements from four independent biological replicates. Bars represent the SE.

Taking both studies together, these results suggested that short periods of fast and severe dehydration in critical phases of seed filling had no effect on isoflavone concentrations, even though these treatments were able to decrease seed size significantly. For example, we found 15.2 % diminution in the weight of 100 seeds of Williams 82 grown in water-stressed plants compared to non-stressed ones. Short periods of water stress during R6-R8 stages have been shown to reduce seed size but not the number of seeds (McWilliams et al. 1999). Our results seem to indicate that after a period of short time-lasting stress, seeds of re-watered plants had time to recover physiologically to reach the well-watered levels of seed compounds.

To gain more knowledge on what might occur at the molecular lever during severe stress periods, a transcript profiling was performed after the stress period at R6 stage, prior to re-watering (Fig. 5.3B). Primers were designed to examine the most important genes for isoflavone biosynthesis (Table 5.1). Most genes, including *IFS1*, *IFS2*, *CHS7*, and *CHS8*, showed no significant differences in expression between well-watered and water-stressed plants. However, *PAL1* and the stress-induced control *GST24* were up-regulated by 3-fold in severe stress conditions. On the other hand, *CHS3*, and *CHR* were down-regulated. The fact that only the first enzyme in the pathway, *PAL1*, but no other important gene for isoflavone synthesis displayed increased expression might suggest that the phenylpropanoid pathway is directing the metabolic flux to compounds other than isoflavones. Downstream products have been found only to accumulate to high amounts under stress situations (Yu and McGonigle, 2005). Oppositely, *CHR* was down-regulated by 3-fold. *CHR* constitutes a branch point in the phenylpropanoid pathway and it is necessary for biosynthesis of glyceollins and for isoflavones daidzein and glycitein,

which also might suggest that under severe stress the flux is directed towards naringenin side. Naringenin in turn is a precursor for many other compounds such as tannins, anthocyanins, flavonols and flavones, which have been shown to be induced under a variety of abiotic stresses (Dixon and Paiva, 1995). For instance, anthocyanins are important for protection against abiotic-induced oxidative damage and in conferring certain osmotic advantages (Chalker-Scott, 1999; Winkel-Shirley, 2002; Wahid and Ghazanfar, 2006). In addition, in its role as protective against oxidative damage, glutathione S-transferase (*GST24*) levels frequently increased after numerous environmental stimuli, including dehydration stress (McGonigle et al. 2000). The fact that most of the transcripts remained unaltered after a brief period of severe stress does not support the hypothesis of a strong stress-induced physiological change. Alternately, a second hypothesis appears to be more plausible. Plants reached a LWP between –2.0 to –2.5 MPa in 5-6 days, depending on the genotype. However, they were stressed for only the last 2 to 3 days, at least the aerial part, which might be a short period, and metabolic pathways leading to isoflavone synthesis in seeds may need more time to adjust and to produce substantial changes. Nevertheless, further studies are required, remaining to future work to perform a comprehensive analysis of the phenylpropanoid metabolite composition in different seed stages to fully ascertain the metabolic flux and the evolution of the pathway during severe stress periods.

Slow-progressing water stress spanning seed maturation period decreased isoflavone seed content in growth chamber

Since short periods of complete water withholding during critical seed developmental stages failed to produce any effect on isoflavone accumulation, we then applied procedures to analyze long-term drought stress spanning the entire seed-filling period and thus resembling the gradual progression of drought in field conditions. This long-term stress protocol was design to impose drought stress at the critical period, when isoflavone accumulation reaches the highest levels, in the latest stages of seed formation (Dhaubhadel et al. 2007). To determine the severity of the water-withholding required to significantly affect isoflavone synthesis, we first tested three different stress intensities (Fig. 5.4). Stressed treatments were initiated at the beginning of R5 stage (30 DAP) by suppressing any addition of water until soil in pots reached 25, 20 and 15% SMC, respectively (see Materials and Methods). Each pot soil moisture level was maintained thereafter by addition of the proper amount of water until physiological maturity and dried seeds were harvested. The experiment was designed to replicate drought conditions experienced in the field in which water stress may span several weeks and thus more suitable for studies involving seed development. Interestingly, the results showed that the intensity of water deficit affected total isoflavone accumulation: as the severity of the stress increased fewer isoflavones were formed. No differences were found when plants grew at 25 and 20% SMC when compared to the well-watered controls (Fig. 5.4); however, mature dried seeds had significantly ($P<0.05$) less amount of isoflavones when grown in 15% SMC pots. Thus, out of the three stress intensities, the largest difference in isoflavone content occurred when plants were maintained in approximately 15% SMC,

and thereby that level was used in subsequent long-term drought experiments. Results strongly suggest that water deprivation spanning the entire seed filling decrease isoflavone concentrations in an intensity-dependent manner.

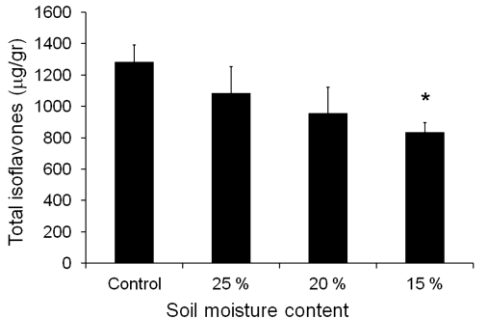

Fig. 5.4 *Long-term drought affects isoflavone accumulation in seeds*. Williams 82 was used to assess the stress required to significantly diminish seed isoflavone content. Three soil moisture content (SMC) were tested versus well-watered pots (control). Only when SMC was maintained at 15% or below, a significant difference in isoflavone concentrations was noticed. Data are the means of measurements from four independent biological replicates. Bars represent the SE. Significance level: * $P<0.05$

Isoflavone seed profiling in well-watered and long-term drought stress conditions

After assessing the level of stress needed to induce differences in isoflavone synthesis, we then tested cultivars RCAT-Angora and Williams 82 in a growth chamber maintaining 15% SMC from the onset of R5 until completion of seed development. To maintain the stress level, pot SMC was controlled by periodical readings (Fig. 5.5A). As measurements in soil may not reflect the actual stress in plants, leaf dehydration state was also monitored by means of LWP (Fig. 5.5B). Seed samples representatives of stages R5 early (R5e), R5 late (R5L), R6, R7, and dried mature seeds (DS) were collected at 30, 45, 60 and 70 DAP, respectively, and quantified for individual isoflavones. Significant differences in total isoflavones between well-watered and drought-stressed plants were only found in the later developmental stages, R7 and DS, in both genotypes: for RCAT-Angora (R7 and DS, $P<0.05$), and Williams 82 (R7, $P<0.001$; DS, $P<0.01$) (Fig. 5C). When comparing individual isoflavones, significant differences were found for genistein in both RCAT-Angora (R7, $P<0.05$ and DS, $P<0.01$), and in Williams 82 (R7, $P<0.001$ and DS, $P<0.05$); daidzein Williams 82 (R7, $P<0.001$ and DS, $P<0.01$); and glycitein in Williams 82 (R7, $P<0.05$ and DS, $P<0.05$). However, no differences were found in RCAT-Angora for daidzein and glycitein. Traces could be found as early as 30 DAP (R5e stage); nevertheless, isoflavones were largely accumulated in the late stages of seed maturation regardless of water regime. Isoflavones synthesized in the hypocotyls are likely to be the primary source of the small amounts of isoflavones detected in early phases (R5) because isoflavone accumulation in cotyledons begins later, at the R6 stage (Berger et al. 2008). Together, results suggest that total isoflavones decrease their levels when seed development occurs under water-limiting conditions. However, there might

also be a genetic component involved because cultivar differences in percentage of reduction were observed. Effectively, Williams 82 plants experienced a reduction in total isoflavone content by 30.5% when stressed by long-term drought while RCAT-Angora only to 14%. Varietal percentage variation was also reported by Bennett et al. (2004). Nelson et al. (2002) noticed that drought stress lowered isoflavone levels by more than 50% for some cultivars but only 5% for others. Similar responses were observed for increases in temperature. Thus, although there was always a reduction in seed isoflavones under drought stress conditions, it appears to occur in a genotype-dependent manner.

Seed isoflavone biosynthetic genes transcript profiling

Differences in isoflavone contents between well-water and long-term water-stressed conditions suggested that expression of genes involved in isoflavone biosynthesis might be influenced by water deficit. To test this hypothesis, samples of Williams 82 developing at 15% SMC or non-stress conditions were collected at the same five time points than for the isoflavone seed profiling (R5e, R5L, R6, R7, and DS) to conduct qRT-PCR with primers targeting the most important genes for isoflavone biosynthesis (Table 5.1). The phenylpropanoid pathway genes for qRT-PCR analysis were selected based on previous literature and summarized in our recent review article (Zhang and Yu, 2008). We pursued three objectives: (i) development of a complete profile of isoflavone transcript accumulation along with seed developmental phases under both well-watered and water-stressed conditions; (ii) investigation of possible correlations of expression of phenylpropanoid pathway genes with isoflavone accumulation; and (iii) study the relationship between isoflavone biosynthesis and expression of phenylpropanoid genes under drought stress. Results showed that key genes

had different patterns of expression, depending on water regime. In well-watered conditions, the expression patterns of these genes were grouped into five categories. The first group encompasses genes that maintained a steady level of expression along all four stages: *CHI1A*, *CHS2*, *CHS3* and *C4H*. The second group was formed by genes that were expressed at low levels at the beginning of seed development, peaked in intermediate stages and decreased at the latest steps: *PAL1*, *CHS1*, and *CHR*. The third group included genes that had higher expression at early stages and then decreased towards the end of seed maturation: *F3H*, *DFR1*, and *4CL*. The fourth cluster includes genes that progressively increased their expression reaching maximum levels in the latest stages: *CHS7*, *CHS8* and *IFS2*. Transcripts of *IFS1* showed a unique expression pattern and as a result, it was considered a fifth group. They were induced at a low rate at the beginning of seed filling and from late R5 increased and kept steadily constant. A few transcripts, *CHS4*, *CHS5*, *CHS6*, *IFR* and *CHI1B1* had very little or no expression in the conditions tested, and hence they were not included. Results from the dried mature seed stage were also discarded for clustering as it represented a special phase in which tissue was basically dehydrating, and most likely transcription was not occurring similar to other tissues.

Plants subjected to long-time progressive drought stress showed no change in the expression pattern of some genes when compared with well-watered plants. Belonging to this group were: *4CL*, *CHS1*, *CHS2*, *CHS3*, *CHR*, *CHI1A* and *IFS1*. Nevertheless, other important genes in the pathway had differences in regulation at particular stage(s). Genes clearly up-regulated in well-watered conditions at some points of seed maturation were: *DFR1* and *F3H* at 30 DAP; *DFR1* at 45 DAP; *CHS7*, *CHS8*, and *IFS2* at 65 DAP. On the

other hand, drought up-regulated genes were: *C4H* at 45 DAP; and finally, *PAL1*, *CHS7*, *CHS8*, and *GST24* at 55 DAP. Certain amounts of *CHS7*, *CHS8*, *IFS1*, and *IFS2* transcripts were also found in matured dried seeds, suggesting that isoflavones accumulate in seeds even during the desiccation process, and that the rapid accumulation at final stages is not only because of the concentration of compounds due to water loss.

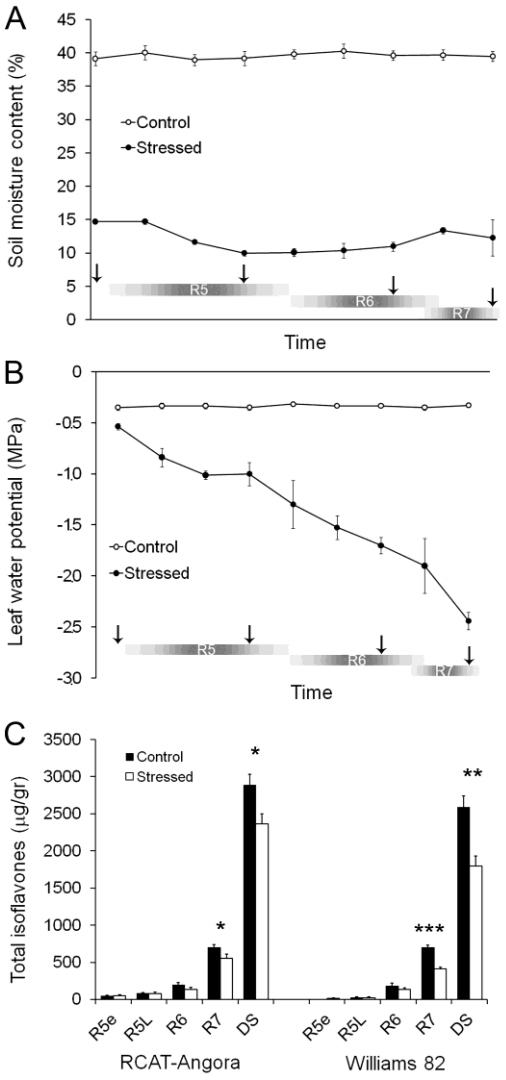

Fig. 5.5 *Long-term drought spanning seed filling reduces isoflavone accumulation*. (**a**) Two cultivars, RCAT-Angora and Williams 82, were grown in growth chambers in well-watered (control) and in 15% SMC pots (stressed). Water stress was initiated at the onset of R5 stage (30 DAP) and maintained until physiological maturity. Samples collected at R5e (30 DAP), R5L (45), R6 (60), R7 (70), and dried seed (DS). Soil moisture content (**b**) and leaf water potential (**c**) were measured every 5 days. Arrows point at the time where samples were collected. Data are the means of measurements from four independent biological replicates. Bars represent the SE. Significance level: * $P<0.05$, ** $P<0.01$, and *** $P<0.001$

In well-watered conditions, *CHS7*, *CHS8* and *IFS2* correlated precisely with isoflavone content. Dhaubhadel et al (2007) also reported these three genes to play an important role in isoflavonoid synthesis in soybean seeds developed in normal moisture levels. Importantly, only *IFS2* correlated with isoflavone content under water-stressed conditions (Fig. 5.6). The other two genes, *CHS7* and *CHS8*, were expressed at a higher rate and reached a maximum at R6 stage (60 DAP) when plants were grown under long-term water deficit. In non-stressed plants, both peaked at later stages, likely reflecting a slowdown in metabolism. *F3H* and *DFR1*, the enzymes downstream of genistein synthesis, are mainly expressed in an early phase of seed formation, and might indicate that at this time the pathway directs their intermediate metabolites to the production of condensed tannins and anthocyanins. This occurs with the practical absence of isoflavones at this stage of seed formation. The fact that both were up regulated in well-watered conditions seemed not to exert any effect in depleting isoflavone content because of the early stage in which it occurred.

In water-deficit plants, *IFS2* transcripts were also accumulated progressively to reach the maximum at the latest stage, but at a much lesser rate than in non-stressed plants. This coincides with the lower isoflavone amounts observed in drought conditions. Importantly, significant differences were found only in the late stages (R7 and DS) in both total isoflavone content and *IFS2* transcript accumulation. Differential regulation of *IFS2* under stressed conditions has also been confirmed by previous reports. Dhaubhadel et al. (2003) showed that *IFS2* is induced in response to pathogen attack, but *IFS1* expression was kept constant.

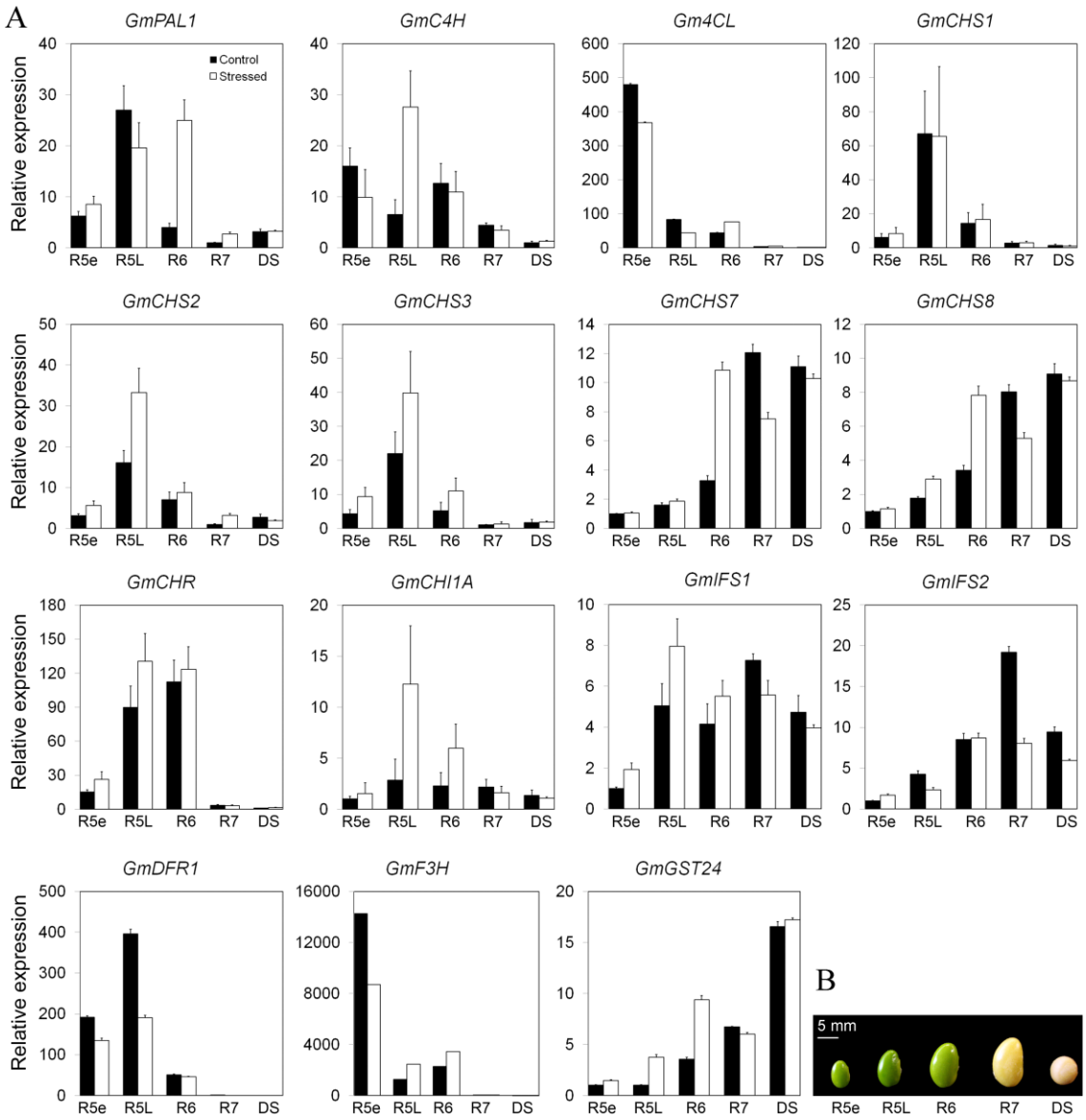

Fig. 5.6 *Gene expression of many isoflavone-related genes altered significantly during water stress.* Samples were collected from Williams 82 plants growing under long-term progressive stress from beginning of R5 stage until mature dried seeds. (A) Primers were designed to target selected genes for isoflavone synthesis. Samples were taken at R5 early (R5e, 30 DAP), R5 late (R5L, 45 DAP), R6 (60 DAP), R7 (70 DAP), and dried seeds (DS). Data are the means of measurements from four independent biological replicates. Bars represent the SE. (B) Comparison between the sizes and colors of the different stages.

Our results indicated that, although both *IFS1* and *IFS2* contributed to isoflavone levels, *IFS2* is down-regulated and it is coincident with isoflavone reduced accumulation under long-term water-stressed conditions. Moreover, reduction in seed isoflavones due to water stress appears to be genotype-dependent. One can hypothesize that, if *IFS2* accounts for this decreasing, as our results suggested, some differential transcriptional regulation or change in enzyme activity among genotypes may occur. This is supported by the highly polymorphic nucleotide sequence of *IFS1* and *IFS2* in the 3'-UTR, 5'-UTR, but also in the coding region. Importantly, those polymorphisms were associated with isoflavone concentrations in soybean seeds (Cheng et al. 2008).

Overall, because of their implications in human health, knowing how water deficit affects isoflavone content in seeds would be crucial to develop soybean cultivars with increased isoflavone levels under drought stress. Regardless of water regime, isoflavone compounds are mainly accumulated in the later seed developmental stages. Transcripts of *CHS7*, *CHS8*, and *IFS2* correlated with isoflavone accumulation under well-watered conditions. Results also indicate *IFS2* as being differentially regulated by water regime, while *IFS1* expression is kept more constant. Our data also suggested that differences in accumulation of isoflavones under irrigated and non-irrigated situations was related to the activity of the *IFS2* enzyme, and that catabolic reactions leading to isoflavone degradation appeared not to be relevant.

Fig. 5.7 *Long-term water-deficit stress produced intense defoliation*. Plants growing in the growth chamber under progressive long-term water withholding (left) reached maturity earlier than well-watered plants (right). Picture was taken in R7 stage (70 DAP). Stressed plants had also less number of pods per plant. Water-deficit plants were maintained at 15% SMC from R5 until physiological maturity.

TABLES

Table 5.1 Primer pairs used in quantitative RT-PCR

Name	Sequence ID	Enzyme	Fordward Primer	Reverse Primer
GmPAL1	X52953	Phenylalanine ammonia-lyase 1	*5'-AGCAACACAACCAGGATGTCAA-3'*	*5'-CAATTGCTTGGCAAAGTGCA-3'*
GmC4H	X92437	Cinnamic acid 4-hydroxylase	*5'-AGGCGAGATCAACGAAGACAAC-3'*	*5'-GTTCACAAGCTCAGCAATGCC-3'*
Gm4CL	X69955.	4-coumarate:CoA ligase	5'-AGGCAATGTACGTGGACAAGCT-3'	5'-TCCGAGAGGACAGAGAAGTGGA-3'
GmCHS1	X54644	Chalcone synthase 1	*5'-AAGCGCATGTGTGATAAGTCGA-3'*	*5'-TTGCATCCAACGAAGGTGC-3'*
GmCHS2	X65636.	Chalcone synthase 2	*5'-TATGGCACCTTCATTGGATGC-3'*	*5'-GCTGGTGGTGCAAAAAATGAG-3'*
GmCHS3	X53958.	Chalcone synthase 3	*5'-GAGATCCGTAATGCACAACGTG-3'*	*5'-CTTTGAGCTCGGTCATGTGCT-3'*
GmCHS4	X52097.	Chalcone synthase 4	*5'-CCTTCCAAGCCACTTTGCA-3'*	*5'-CTGGAGCAAAGGATGAAAGTGA-3'*
GmCHS5	L07647	Chalcone synthase 5	*5'-CACTTTGCCACATTCATTCC-3'*	*5'-TGTGAATGAACTAATGAAGCTATAGC-3'*
GmCHS6	L03352	Chalcone synthase 6	*5'-ACCAACAGTGACCACATGAACG-3'*	*5'-GGCACAAACACTTGGATTCTCC-3'*
GmCHS7	M98871	Chalcone synthase 7	*5'-AACCCACCAAACCGTGTTGAT-3'*	*5'-CTTGTCACACATGCGCTGAAAT-3'*
GmCHS8	AY237728	Chalcone synthase 8	*5'-ATGGAGCTGCTGCTGTCATTG-3'*	*5'-CCTCACGAAGGTGTCCATCAA-3'*
GmCHR	X55730	Chalcone reductase	*5'-CAAAGCCATTGGAGTCAGCAA-3'*	*5'-CCATGCAAGGTTCATCTCCACT-3'*
GmCHI1A	AY595413	Chalcone isomerase 1A Type II	*5'-GGCGCTGAATACTCAAAGAAGG-3'*	*5'-AGAGGCACCAGGTGCAAAATT-3'*
GmCHI1B1	AY595414	Chalcone isomerase 1B1 Type II	*5'-AGCTGAATTGCTCGACTCCCT-3'*	*5'-CAGATTGCATATGTGCCACACA-3'*
GmIFS1	AF195798	Isoflavone synthase 1	*5'-AGAATTCCGTCCCGAGAGGTT-3'*	*5'-TGCCATTCCTGAAGTAGCCAA-3'*
GmIFS2	AF195799	Isoflavone synthase 2	*5'-AATGTGCCCTGGAGTCAATCTG-3'*	*5'-GGCGTCACCACCCTTCAATAT-3'*
GmIFR	AJ003245	Isoflavone reductase	*5'-AGATGGAAATGTGAAAGGAGCG-3'*	*5'-TGTGCACGGCTTTGTTCAAG-3'*
GmF3H	AY595420.	Flavanone 3-hydroxylase	*5'-TTACCTGGCCCAGGAGAAAAC-3'*	*5'-ATTCCGGCAAGAGAAATCACTG-3'*
GmDFR1	AF167556	Dihydroflavonol-4-reductase	*5'-TTGTTGTCGGTCCCTTTCTGA-3'*	*5'-GTGGACGAATTGACCTTGCTTT-3'*
GmDFR2	EF187612.	Dihydroflavonol-4-reductase 2	*5'-CCATGGATTTTGACTCCAAGGA-3'*	*5'-CTTCGGACAGTTTTGGCCTTC-3'*
MtbGluc2	TC94967	b-glucosidase2	*5'-TGGCACAGCATCTTCTGCATAT-3'*	*5'-TACGAGTCAACCGCAACGTCT-3'*
MtbGluc3	TC107558	b-glucosidase3	*5'-AATGAACCATGGAGTGTGAGCA-3'*	*5'-TGAATCACCACCTGTGCAATTC-3'*
GmGST24	AF243379.	Glutathione S-transferase	*5'-TGACTCCAGATCCCAAGGTGAT-3'*	*5'-TGAAGAAATCCCCAGCCAAA-3'*
CYP2	TC224926	Cyclophilin	*5'-CGGGACCAGTGTGCTTCTTCA-3'*	*5'-CCCCTCCACTACAAAGGCTCG-3'*
ELFα	TC203954	Eukaryotic elongation factor 1-alpha	*5'-GACCTTCTTCGTTTCTCGCA-3'*	*5'-CGAACCTCTCAATCACACGC-3'*
F-box	CD397253	F-box protein family	*5'-AGATAGGGAAATGGTGCAGGT-3'*	*5'-CTAATGGCAATTGCAGCTCTC-3'*
ACT2/7	TC204150	Actin	*5'-CTTCCCTCAGCACCTTCCAA-3'*	*5'-GGTCCAGCTTTCACACTCCAT-3'*

REFERENCES

Akashi T, Aoki T, Ayabe S (1999) Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. Plant Physiol **121**:821-828.

Bennett JO, Yu O, Heatherly LG et al (2004) Accumulation of genistein and daidzein, soybean isoflavones implicated in promoting human health, is significantly elevated by irrigation. J Agric Food Chem **52**:7574-7579.

Berger M, Rasolohery CA, Cazalis R et al (2008) Isoflavone accumulation kinetics in soybean seed cotyledons and hypocotyls: distinct pathways and genetic controls. Crop Sci **48**:700-708.

Brevedan RE, Egli DB (2003) Short periods of water stress during seed filling, leaf senescence, and yield of Soybean. Crop Sci **43**:2083-2088.

Chalker-Scott L (1999) Environmental significance of anthocyanins in plant stress responses. Photochem Photobiol **70**:1-9.

Cheng H, Yu O, Yu D (2008) Polymorphisms of IFS1 and IFS2 gene are associated with isoflavone concentrations in soybean seeds. Plant Sci. **175**:505-512.

Dhaubhadel S, McGarvey BD, Williams R et al (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol. Bio. **53**: 733-743.

Dhaubhadel S, Gijzen M, Moy P, et al (2007) Transcriptome analysis reveals a critical role of CHS7 and CHS8 genes for isoflavonoid synthesis in soybean seeds. Plant Physiol **143**:326-338.

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. Plant Cell **7**:1085-1097.

Gayler KR, Sykes GE (1985) Effects of nutritional stress on the storage proteins of soybeans. Plant Physiol. **78**(3):582-585.

Gutierrez-Gonzalez JJ, Wu X, Zhang J et al (2009) Genetic control of soybean seed isoflavone content: Importance of statistical model and epistasis in complex traits. Theor Appl Genet. *In press.*

Hoeck JA, Fehr WR, Murphy P.A et al (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci **40**:48-51.

Jung W, Yu O, Lau SC et al (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nature Biotech **18**:208-212.

Lozovaya VV, Lygin AV, Ulanov AV et al (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci **45**:1934-1940.

McGonigle B, Keeler SJ, Lau SC et al (2000) A genomic approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize. Plant Physiol **124**:1105-1120.

McWilliams DA, Berglund DR, Endres GJ (1999) Soybean growth and management. North Dakota State University.

Mebrahtu T, Mohamed A, Wang CY et al (2004) Analysis of isoflavone contents in vegetable soybeans. Plant Foods for Human Nutrition **59**:55-61.

Messina MJ (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. Am J Clin Nutr **70**: (suppl) 439-S-450-S.

Naoumkina M, Farag MA, Sumner LW et al (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc Natl Acad USA. **104**:17909-17915.

Nelson R, Lygin A, Lozovaya V et al (2002) Genetic and environmental control of soybean seed isoflavone levels and composition. Proceedings of the 9th biennial Conf. of Cel. and Mol. Bio. of Soybean.

Pedersen P (2004) Soybean growth and development. Iowa State University.

Rochfort S, Panozzo J (2007) Phytochemicals for health, the role of pulses. J Agric Food Chem **55**:7981-7994.

Steele CL, Gijzen M, Qutob D et al (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. Arch Biochem Biophys 367, 146-150.

Subramanian S, Hu X, Lu G et al (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol Biol **54:**623-639.

Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. Plant J **48**:261-273

Subramanian S, Stacey G, Yu O (2007) Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci **12**:282-285.

Vandesompele J, De Preter K, Pattyn F et al (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology **3:**research0034.1-0034.11.

Wahid A, Ghazanfar A (2006) Possible involvement of some secondary metabolites in salt tolerance of sugarcane. J Plant Physiol. **163**:723-730.

Wang H, Murphy PA (1994) Isoflavone Composition of American and Japanese Soybeans in Iowa: Effects of Variety, Crop Year, and Location. J Agric Food Chem **42**:1674-1677.

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol **5**:218-223.

Yu O, McGonigle B (2005) Metabolic engineering of isoflavone biosynthesis. Advances in Agronomy. **86**:147-190.

Zhang J, Yu O (2008) Metabolic engineering of isoflavone biosynthesis in seeds. *In* Modification of seed composition to promote health and nutrition. Hari Krishnan (ed). Agronomy Monograph Series, pp 151-177.

CHAPTER 6

Conclusions, practical recommendations and further studies

Isoflavone concentration in soybean seeds has been shown to have a wide range. In our genetic material up to a 4 fold difference for averaged total isoflavones could be found among Essex × PI437654 recombinant inbred lines, and over 13 fold in the Magellan × PI437654 population (Tables 6.1-6.4). These values reflect the average of all locations, years and replications available for each cross, and thus they can be considered as a reliable and consistent measure of each line (see the CD for data of individual isoflavones). The same cultivar can also vary in its isoflavone levels. However, despite the fact that some variation was also observed, even in genetically identical individuals and within the same environment, the largest variation was consistently observed when changing environments, either to a different location or years. Effectively, within the same location and year the variation between replications does not usually exceed 10 to 20 % for total isoflavones, suggesting that an important genetic component may be involved. However, the variation noticed in different ambient conditions also indicates that the underlying genetics could be modulated or affected by the environment. This hypothesis was later confirmed in our multi-environment QTL mapping studies, in which a complex network of environmentally affected interacting loci was proposed as a model to explain seed isoflavone concentrations.

Having such variability makes it of paramount importance to choose carefully the experimental design. In our field mapping studies, the segregating population of

recombinant inbred lines (RILs) was planted in a two-row scheme with three replications per environment. To perform isoflavone quantification, seeds from a pool of at least three different plants per genotype were mixed and ground. Pooling seeds from different plants is important to minimize the effect of the variation among plants. The approximate number of seeds used in each extraction was roughly twenty (1.5 to 2.0 grams). This increased number of seeds used with respect to other isoflavone research groups also helps in attenuating the oscillation among seeds. To assess statistically the correctness of our methods or whether the variability of the traits could have an effect on the results, the coefficients of variation (CV) for the three replications were calculated and had average values ranging 0.1 to 0.2 for all traits, locations and years, which are perfectly acceptable. In addition, outliers and RILs with highly distinct numbers for the three measurements taken (three replications) were eliminated before running the software for the QTL mapping analysis. Precise phenotypic quantification is crucial for QTL mapping so special care must be taken to assign reliable isoflavone levels to RILs. All statistical parameters measured and experimental designs devised for the present thesis work appeared to perform well and be balanced between marginal gain and cost. For experiments in controlled environments (greenhouses and growth chambers), three replications were also used for studies not involving stresses. However, four replications were used, and are recommended, for experiments involving water-stresses because more variability was observed not only in the drought parameters taken but also in the isoflavone content of seeds. Importantly, growth chamber experiments are recommended over greenhouse ones because environmental conditions can be maintained in a more desirable and constant range. For example, keeping temperature constant is especially

crucial as it has been shown to exert great influence in the isoflavone concentrations in seeds. However, plants placed closer to the cooling panel in the greenhouse were considerably cooler than the ones located at the opposite side of the room. The difference was more drastic if benches were densely covered by pots because the air flow had more obstacles to go through, and differences could reach up to several degrees Celsius.

Soybean isoflavones share many enzymes in their synthesis throughout the phenylpropanoid pathway. Hence, it is not surprising that individual compounds are highly correlated. For example, total daidzein and total genistein are strongly correlated with total isoflavone content. In addition, genistein is highly correlated with daidzein (>0.9). These results are in agreement with other researches and suggest that the relative amounts of seed isoflavones cannot be dramatically altered. These findings were further reinforced due to the fact that most of the QTL found to be involved in genistein synthesis accounted also for daidzein content. It is also common to find that a single locus explains the accumulation of several isoflavones. In addition, correlations with other important compounds have been studied. For instance, according to several studies, isoflavones appear not to be correlated with protein content. However, Craig et al. (2005) reported a positive correlation in five out of seventeen cultivars analyzed, which indicates that correlation with protein might be genotype-dependent. There also seems to be a moderate negative correlation between concentration of oil and isoflavones, although it could also be influenced by the particular cultivar because a strong positive correlation has also been found for some genotypes. Altogether, these findings suggested that increased isoflavone content lines could be achieved without depriving other key

nutritional compounds; but modifying the relative amounts of individual isoflavones appears not to be possible, at least drastically.

The genetic basis that governs accumulation of isoflavones in soybean seeds is extremely complex. Many individual small-effect QTL contribute in an additive manner to the final concentrations. In addition, interactions between some of these QTL or other non-additive loci (epistasis) have been shown to account for a large percentage of the phenotypic variance for the traits. To complicate the puzzle even more, both additive loci and epistatic interactions are often influenced by the environment. Taking everything together, it is implied in our QTL mapping studies that an intricate network of interacting loci governs the synthesis and accumulation of the distinct isoflavone compounds in seeds of soybeans. In addition, this complex network is modulated by environmental conditions and reflects the intricacy of the genetics underlying isoflavone concentrations. Attempts have been made to increase the number of positive minor-effect QTLs through marker assisted selection. However, the task is time-consuming and only few QTLs can be targeted at a time. This, together with the fact that epistatic interactions are sometimes more important than the additive QTLs and unpredictable in the sense that they depend on the genetic background, does not guarantee that introducing several minor effect QTLs would have the expected outcome.

Interestingly, despite the fact that significantly influencing QTLs have minor individual effects, there is still some room for genetically modifying isoflavone content of seeds. The major QTL detected in chromosome Gm05 (former linkage group A1) opens up the possibility for developing increased isoflavone cultivars through marker assisted selection. This is by far the most important locus affecting isoflavones ever

reported and accounts for about 30 % of the variance found in genistein, daidzein, and total isoflavone content; and more than 7 % for glycitein. Currently, we are making progresses in identifying the gene behind this important QTL, which could also be used for genetic engineering of soybean plants. Several indications suggest that a *trans*-acting factor could be responsible for the QTL. If our hypothesis is finally confirmed, it would be the first TF reported specifically being involved in isoflavone synthesis.

Isoflavone content in seeds is affected by water regime in a cultivar dependent manner. The stage of seed development and length of water deprivation are factors that influence the extent of the process. Our results suggested that short periods of strong water deficit in critical phases of seed development do not exert any effect if the stress is followed by a period of irrigation. Accordingly, only drought conditions spanning the entire or most of the seed filling period would provoke changes in isoflavone contents. Later phases are probably the most critical because isoflavone synthesis peaks at the end of seed filling. Results suggest that periods of short droughts that may occur during the growing season will not affect isoflavone levels, but a more prolonged period without water will have a negative impact.

Table 6.1 High total isoflavone Essex x xPI437654 RILs. Averaged data from 2 locations in 2 years.

Sorted by total isoflavones

RIL	Genistein	Daidzein	Glycitein	Total
1	1017	1138	172	2327
11	1274	755	166	2195
121	1034	944	204	2169
208	1302	684	172	2158
155	1025	843	160	2027
126	1060	772	203	2010
161	1131	673	150	1953
15	1075	737	140	1951
229	1030	753	139	1922
187	999	734	157	1890
PI437654	1306	469	95	1869

Consistently high isoflavone lines

Table 6.2 Low total isoflavone Essex x PI437654 RILs. Averaged data from 2 locations in 2 years.

Sorted by total isoflavones

RIL	Genistein	Daidzein	Glycitein	Total
54	430	180	123	734
37	427	191	89	707
169	402	193	99	695
142	436	172	85	692
110	405	190	86	682
44	410	187	79	676
223	401	192	76	670
29	351	204	60	615
215	392	140	44	576
82	336	126	78	540

Consistently low isoflavone lines

Table 6.3 Consistently high total isoflavone content Magellan x PI437654 RILs. Averaged data of the 2 locations.

RIL	Genistein	Daidzein	Glycitein	Total
181	993	1049	34	2076
176	946	1068	34	2047
PI437	1140	703	38	1880
243	848	867	35	1750
23	841	873	34	1748
83	771	729	32	1531
53	805	678	38	1521
164	780	691	27	1498
21	628	813	38	1479
57	802	636	33	1471
155	694	725	30	1449

Table 6.4 Consistently low total isoflavone content Magellan x PI437654 RILs. Averaged data of the 2 locations.

RIL	Genistein	Daidzein	Glycitein	Total
190	138	123	12	273
226	129	124	14	267
49	150	97	14	261
117	144	105	9	259
151	148	94	11	254
213	131	93	15	239
115	117	109	9	236
96	106	88	20	215
152	86	83	9	177
172	80	65	10	155

VITA

Juan Jose Gutierrez-Gonzalez was born in Madrid, Spain in 1971. He attended The Sacred Family Private School until he entered The Polytechnic University of Madrid, where he graduated with a bachelor degree in Agronomical Engineering in 1997, with a thesis work in Designing and engineering a farm of cocks of Leon. He joined the Genetics Area Program of the Plant Sciences department at the University of Missouri in fall 2004 to pursue a PhD, goal reached in fall 2009.