GENETIC FACTORS AND BIOCHEMICAL PATHWAYS REGULATING POST-HARVEST AFLATOXIN CONTAMINATION IN CULTIVATED PEANUT

by

WALID AHMED ABDELMONEIM KORANI

(Under the Direction of Peggy Ozias-Akins)

ABSTRACT

Aflatoxin accumulates in peanut seeds in the field (pre-harvest) or during storage (post-harvest) as a result of *Aspergillus spp.* infection. It is one of the most challenging diseases for peanut as toxin biosynthesis is dependent on environment, which makes identification of resistant genotypes difficult. Therefore, this study was carried out to discover genetic factors and biochemical pathways that underlie resistance to postharvest aflatoxin contamination of cultivated peanut.

Since accumulated evidence supports the importance of the lipoxygenase (LOX) gene superfamily in plant defense against many diseases, LOX genes of cultivated peanut were identified. In addition, functional classification, evolutionary analysis, and in-depth expression analysis were carried out. Moreover, the expression responses to different diseases including aflatoxin contamination were estimated.

To identify resistant genotypes, a protocol of tracking and assaying the *A. flavus* infection and the subsequent aflatoxin accumulation was developed using a GFP-expressing strain. In addition, a phenotyping tool was designed, designated SICIA (Seed Infection Coverage and Intensity Analyzer) using a Matlab script. The application of this protocol combined with

different statistical models enabled the identification of a highly reliable resource for resistance in cultivated peanut. Moreover, this approach allowed differentiation between the genotype response to *A. flavus* infection and the ability to accumulate aflatoxins.

Genetic mapping of putative resistance genes requires DNA sequence variation. Since cultivated peanut is tetraploid, extracting true SNPs directly from next generation sequencing data using currently available filtering tools is challenging. Therefore, a machine learning tool for refining SNP calling from sequence data of polyploids was designed, designated SNP-ML.

RNA-seq analysis was carried out for *A. flavus*-infected resistant and susceptible peanut genotypes to determine the factors associated with the resistance response for aflatoxin accumulation in cultivated peanut. An R package was designed to conduct KEGG enrichment analysis for polyploids, designated keggseq. The application of this package revealed the importance of alpha-linolenic acid and protein processing in the endoplasmic reticulum in the resistance response. The analysis also included application of different tools for differential expression analysis of time course experiments, expression clustering, GO enrichment analysis, *de novo* assembly, annotation and co-expression network analysis.

INDEX WORDS: peanut, aflatoxin, *Aspergillus*, lipoxygenase, gene family, GFP, SICIA, SNP, machine learning, SNP-ML, SNP-MLer, RNA-seq, differentially expressed genes, keggseq package.

GENETIC FACTORS AND BIOCHEMICAL PATHWAYS REGULATING POST-HARVEST AFLATOXIN CONTAMINATION IN CULTIVATED PEANUT

by

WALID AHMED ABDELMONEIM KORANI

B.S., Cairo University, Egypt, 2001

M.S., Ain Shams University, Egypt, 2007

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

GENETIC FACTORS AND BIOCHEMICAL PATHWAYS REGULATING POST-HARVEST AFLATOXIN CONTAMINATION IN CULTIVATED PEANUT

by

WALID AHMED ABDELMONEIM KORANI

Major Professor:	Peggy Ozias-Akins
Committee:	Corley Holbrook
	Craig Kvien
	Scott Jackson
	Dayton Wilde

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2017

DEDICATION

I dedicate this work to whom my heartfelt thanks; to my parents for everything, to my wife Fatma for her love and support, to my sisters, professors and friends for all the support they lovingly offered along the period of my post-graduation.

ACKNOWLEDGEMENTS

I wish to express my sincere thanks, deepest gratitude and appreciation to my supervisor Peggy Ozias-Akins for suggesting problem solutions, supervision, continued assistance and Ye Chu for her support, laboratory training and continued help. Grateful appreciation and thanks to my committee members for guiding me through the progress; Peggy Ozias-Akins, Corley Holbrook, Craig Kvien; Scott Jackson; Dayton Wilde. Grateful appreciation is also extended to all my colleagues whom helped me during the work; Evelyn Saul, Joann Conner, Josh Clevenger, Stephanie Botton, Kathleen Marasigan, Jing Chen and Haiyan Liu. I thank also Abdelshafy Ragab, Egyptian Atomic Energy Authority and Ministry of Higher Education of Egypt for their support.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Peanut, or groundnut, (*Arachis hypog*aea) is one of the most important oilseed crops. Almost 50% of the peanut seed is oil. In addition, seeds are rich in protein (26%). Peanuts have a wide range of uses, *e.g.*, peanut butter, cooking oil, candies, biscuits, peanut flour and coffee, and are used in a wide range of products, *e.g.*, laxatives, atopic dermatitis treatment, shampoos, lotions, creams and soaps, stains, inks and paper (Roth 1991; Floyd *et al.,* 2000; Paller *et al.,* 2003; Mattes 2005; Shastry *et al.,* 2009; Seo *et al.,* 2013), George Washington Carver developed almost 300 peanut products (Burchard 2005). Moreover, peanut was the first source of biofuel (Knothe 2001) and it has a high biofuel production; one peanut acre produces 90 gallons of biofuel (Brown 2006). However, it is used rarely as a bioenergy crop since it is an expensive resource as compared to other crops such as flax, sunflower, or soybean.

Cultivated peanut is an allo-tetraploid (2n = 4x = 40) that was formed from spontaneous doubling of a cross between two diploid species, *i.e.*, *A. duranensis* and *A. ipaensis* (Seijo *et al.,* 2004). The two sub-genomes are very similar (Bertioli *et al.,* 2015), which increases the difficulty of finding genetic factors underlying economic traits or disease resistance.

Peanut production faces numerous diseases such as leaf spots, nematode infection and aflatoxin contamination. However, the latter is the most challenging disease since it is very sensitive to the environmental conditions as the climate directly influences the fungal community structure and quantity of aflatoxin produced by aflatoxin-producing fungi, and indirectly affects fungal interaction with a plant during development or due to wounding from insects (Cotty and

Jaime-Garcia, 2007). Aflatoxin is an acutely toxic, carcinogenic and immune-suppressive class of mycotoxins that affects domesticated animals and humans (Scheidegger and Payne, 2003) and is produced as a secondary metabolite (Diener *et al.,* 1987) upon *Aspergillus spp.* (*A. flavus and A. parasiticus*) infection. *A. flavus* is an ascomycetous fungus that can infect humans, plants, animals and insects (Klich 2007). The fungus may infect peanut and accumulate aflatoxin in the field (pre-harvest aflatoxin contamination) or during storage (post-harvest aflatoxin contamination). Drought-tolerant genotypes, or irrigation combined with good management practices may reduce pre-harvest aflatoxin contamination since drought conditions are highly correlated with aflatoxin accumulation (Holbrook *et al.,* 2000a; Nigam *et al.,* 2009). However, improving breeding lines for resistance to post-harvest aflatoxin conamination is required especially in developing countries as they lack good storage conditions and aflatoxin testing is irregular. Therefore, finding genetic factors controlling *A. flavus* infection and aflatoxin contamination for harvested peanut is important.

Lipoxygenase importance for plant defense of *Aspergillus spp.* infection and other diseases

Lipoxygenases (LOX) are a gene superfamily encoding dioxygenases that catalyze the addition of molecular oxygen atoms to polyunsaturated fatty acids. Accumulated evidence stated the importance of LOXs in plant defense (Porta and Rocha-sosa, 2002). Some LOXs were observed as positive effectors in resistance while others were found to contribute to susceptibility (Kumari *et al.* 2012; Muller *et al.,* 2014).

LOXs play an important role in *Aspergillus spp.* infection and aflatoxin contamination of peanut. However, the exact role is not clear since the response of LOXs to these fungi is controversial. The first recognized peanut LOX, *PnLOX1*, had a positive regulation with the

infection by *A. parasiticus* (Burow *et al.,* 2004). However, two other subsequently described LOXs, *PnLOX2* and *PnLOX3*, showed negative regulation upon infection by *A. flavus* (Tsitsigiannis *et al.,* 2005). On the other hand, there is evidence that *PnLOX2* and *PnLOX3* are expressed in different patterns when resistant and susceptible varieties of peanut are compared under normal and infection conditions. Kumari *et al.* (2012) found that *PnLOX2* was expressed in a resistant genotype with and without infection and an uninfected susceptible genotype whereas *PnLOX3* only was expressed in the infected resistant genotype. Muller *et al.* (2014) observed that both *PnLOX2* and *PnLOX3* are up-regulated in a susceptible genotype and down-regulated in a resistant genotype 27 hours post-infection by *A. flavus*. In addition, they found that *PnLOX4* and *PnLOX5* are up-regulated in the resistant genotype and down-regulated in the susceptible 20 and 5 hours post-infection, respectively. However, the expression pattern reversed 27 and 48 hours post-infection, respectively.

Although aflatoxins produced in soybean seeds infected by *A. flavus* is much lower than for peanut (Bean *et al.,* 1972), volatile aldehydes that are formed by the lipoxygenase pathway in soybean were found to inhibit *A. flavus* growth and the subsequent aflatoxin contamination (Doehlert *et al.,* 1993; Boue *et al.,* 2005). On the other hand, lipoxygenase-derived linoleic acid derivatives, 9S-HPODE and 13S-HPODE, stimulate morphological differentiation of *A. flavus* and *A. parasiticus* (Calvo *et al.,* 1999).

Additionally, considerable evidence in the literature supports the involvement of LOX in *Aspergillus spp.* interaction with many other plants, *e.g.*, maize (Gao *et al.,* 2009; Huang *et al.,* 2013), cottonseeds (Zeringue 1996) and almond (Mita *et al.,* 2007). Furthermore, the role of lipoxygenases is not limited to interaction between the plant and the fungus. Lipoxygenase

products can affect aflatoxin biosynthesis inside the fungus; 9S-HPODE promotes aflatoxin production whereas 13S-HPODE and 9S-HPOTE inhibit it (Burow *et al.,* 1997).

Root-knot and cyst nematodes are destructive pathogens that have huge negative economic effects for many crops. LOXs were found to play a role in interaction of plants with nematodes. *LOX3* and *LOX4* of *Arabidopsis*, which share 97% amino acid sequence similarity in the substrate-binding pocket (Caldelari *et al.,* 2011), act differently against nematode infection since a mutant lacking *LOX3* is more resistant than wild type and a mutant lacking *LOX4* is more susceptible (Ozalvo *et al.,* 2014). In maize, the expression of *ZmLOX3* increases after inoculation with root-knot nematode. In addition, the disruption of *ZmLOX3* suppresses plant growth and increases the reproduction of root-knot nematode (Gao *et al.,* 2008).

Lipoxygenases have an important role in soybean resistance to *Phakopsora pachyrhizi*, which causes soybean rust (Choi *et al.,* 2008). LOX induction increases in pearl millet seedlings after the infection by *Sclerospora graminicola*, the pathogen that causes downy mildew disease (Babitha *et al.,* 2004; Babitha *et al.,* 2006). In comparison between resistant and susceptible pearl millet cultivars to downy mildew pathogen, higher lipoxygenase activities were found in the resistant genotype under infection conditions (Shivakumar *et al.,* 2003). In rice, LOX activity rapidly increases in leaves after inoculation with an incompatible race of *Magnaporthe grisea*, the rice blast fungus (Ohta *et al.,* 1991). Additionally, *OsLOX1* increases rapidly in response to attack by brown plant hopper. Moreover, a transgenic rice line with a lower activity of *OsLOX1* was found to be less tolerant (Wang *et al.,* 2008). On the other hand, the volatiles produced by *LOX-3* in rice grains during storage are attractive for storage insects, and the cultivars lacking *LOX-3* are more resistant to storage insects (Tang *et al.,* 2009).

Differential induction of LOX activity in leaves of the tomato cv Moneymaker was found in response to inoculations with the non-host pathogen *Pseudomonas syringae* pv syringae, which induces a hypersensitive resistance response, and *P. syringae* pv tomato, which is a pathogen of tomato causing the bacterial speck disease (Koch *et al.,* 1992).

Overexpression of *TomLoxD* leads to elevated wound-induced jasmonic acid biosynthesis, increases expression of wound-responsive genes and enhances resistance to insects and necrotrophic pathogens (Yan *et al.,* 2013). Induced Systemic Resistance (ISR) triggered by *Pseudomonas putida* BTP1 in tomato is associated with a higher level of *TomLoxD* and *TomLoxF* transcription (Mariutto *et al.,* 2011). *CsLOX1* products play an important role in regulating cell death related to flower senescence and the jasmonic acid related defensive reaction of tea (*Camellia sinensis*) plant to phloem-feeders (Liu and Han, 2010).

In addition to LOX function in pathogenicity, LOX also may be involved in symbiosis of plants with other organisms *e.g., Rhizobium tropici* symbiosis with common bean (Porta and Rocha-sosa, 2000).

Lipoxygenase importance for plant development and quality traits

Many LOX paralogs have been identified within plant genomes, hence it is a large multigene family. Different LOXs are expressed differentially from tissue to tissue, which reflects the importance of LOXs in plant development. During leaf development of soybean, the neutral and most of the acidic lipoxygenase isozymes are present in greatest abundance in the youngest leaves and decline in amount as leaves age (Saravitz and Siedow, 1995). In addition, lipoxygenase activity increases during root initiation (Junghans *et al.,* 2004). *Arabidopsis LOX1* is expressed specifically during early germination (Melan *et al.,* 1994). *LOXg* expression is found in the endocarp and the mesocarp of pea pods and absent from the pod exocarp and ovules

(Rodríguez-Concepción and Beltrán 1995). In tomato, *tomloxA* is expressed in germinating seeds as well as in ripening fruits and reaches its peak during the breaker stage. However, *tomloxB* is highly accumulated in ripe fruit (Ferrie *et al.,* 1994). In potato, jasmonic acid, which is one of the oxylipin products of lipoxygenase activity, plays a critical role in microtuber formation (Matsuki *et al.,* 1992). In addition, the suppression of potato *LOX1* causes a reduction in tuber yield, decreases the average tuber size, and disrupts tuber formation (Kolomiets *et al.,* 2001).

LOX activity may positively or negatively affect plant traits. LOX activity is helpful in delaying bran deterioration of rice, *e.g.* high *LOX-1* and *LOX-2* activity is essential for red rice bran storage, while high *LOX-3* activity is essential for white rice bran storage (Zhang *et al.,* 2009). On the other hand, the activity of soybean LOXs on linolenic acid results in seed quality deterioration (Lima *et al.,* 2010). Additionally, oxylipins that are produced by LOXs may cause rancidity of seed oils of maize (Rodriguez-Saona *et al.,* 1995). Moreover, inhibition or reduction of LOX activity can decrease oil rancidity in rice seeds (Malekian *et al.,* 2000), and retard the development of oxidative rancidity and extend the shelf life of walnuts and almonds (Buranasompob *et al.,* 2007).

Hydroperoxide products that are formed from linoleic and linolenic acid due to LOX activities are metabolized with enzymatic and non-enzymatic pathways to volatile products, which may alter flavor (Gardner 1988). The rice *LOX-3* is involved in the production of volatiles that are responsible for stale flavor during storage. Development of this undesirable flavor is delayed in cultivars lacking *LOX3* (Shirasawa *et al.,* 2008). In tomato, C5 and C6 volatile compounds are among the most important contributors to consumer attraction to fresh tomatoes. Co-suppression or antisense inhibition of *TomLoxC* leads to a large reduction in C5 and C6 volatiles (Shen *et al.,* 2014) and major decrease in the flavor volatiles in both fruit and leaf (Chen

et al., 2004). Tomato peel, which has significantly higher concentrations of volatiles than flesh, has a significantly higher LOX activity (Ties and Barringer, 2012). *MdLOX1a* and *MdLOX5e* contribute in fruit aroma production of apple (Vogt *at al.,* 2013).

Mycotoxin formation in oil crops

Filamentous fungi produce a wide variety of economically important secondary metabolites that are known as extrolites. An extrolite is any outwardly directed chemical compound that is excreted or accumulated in the cell wall of living organisms (Frisvad and Samson, 2004). Many of these extrolite compounds are beneficial, such as antibiotics, food grade pigments, enzymes, vitamins, lipids, and various pharmaceuticals. However, others, such as mycotoxins, have deleterious effects (Adrio and Demain, 2003). Mycotoxins are some of the most toxic natural substances known and have been estimated to contaminate up to 25 % of the world's food production (Bennett and Klich, 2003).

There are several classes of mycotoxins, based on structural and chemical properties; the most important one is polyketides that includes sterigmatocystin and aflatoxin (Payne and Brown, 1998). Aflatoxin is a family of toxic and carcinogenic metabolites that causes a severe impact on human health, and a great loss and a high management cost for agricultural crops (Robens and Cardwell, 2003; Carbone *et al.,* 2007). There are four major classes of aflatoxins, depending on the presence of the characteristic polyketide dihydro- (B1 and G1) or tetrahydro- (B2 and G2) bisfuran rings (Ehrlich *et al.,* 2004).

Aflatoxin-producing fungi primarily belong to Aspergilli, which contain *A. flavus* and *A. parasiticus*, and are responsible for aflatoxin contamination of oil-rich crops such as corn, peanuts, cottonseed, and tree nuts (Horn 2003). *Aspergillus spp.* is the dominant infecting fungus of maize, peanut and soybean under poor storage conditions (Bhattacharya and Raha, 2002).

The biosynthesis of aflatoxins involves over 20 enzymatic reactions in a complex polyketide pathway that converts acetate and malonate to the intermediates Sterigmatocystin (ST) and O-Methylsterigmatocystin (OMST), the respective penultimate and ultimate precursors of aflatoxins. Although these precursors are chemically and structurally very similar, their accumulation differs at the species level for *Aspergilli*. Notable examples are *A. nidulans* that synthesizes only ST, *A. flavus* that makes predominantly aflatoxins, and *A. parasiticus* that generally produces either aflatoxins or OMST (Carbone *et al.,* 2007).

Aspergillus spp. infection and aflatoxin contamination (screening and resistance) for peanut

As screening peanut genotypes for resistance to aflatoxin contamination is the most critical step in the breeding process, different approaches have been developed for pre- and post-harvest screening (Mixon and Rogers, 1973; Holbrook *et al.,* 1994; Anderson *et al.,* 1996; Young and Cousin, 2001, Xue *et al.,* 2004a). Based on these methods, several peanut genotypes were reported to be resistant for pre-harvest aflatoxin contamination, *e.g.* J-11, Lampang (Kisyombe *et al.,* 1985), and post-harvest aflatoxin contamination *e.g.*, PI 337394F (Mixon and Rogers, 1973), J-11 (Kisyombe *et al.,* 1985), PI337409 (Mixon and Rogers, 1973; Kisyombe *et al.,* 1985), TG19, TG49, TG18A and TG18 (Harish *et al.,* 2005). However, very rarely does a cultivar consistently show resistance per se since aflatoxin contamination is very variable even for the same cultivar under different conditions (Blankenship *et al.,* 1984).

Green Fluorescent Protein (GFP) can be utilized to track *A. flavus* infection of peanut. It was used widely in fungal biology and for studying the fungus/plant interaction, *e.g.*, it has been used in monitoring food colonization by *A. flavus* (Du *et al.,* 1999), studying the oomycete pathogen *Phytophthora parasitica* interaction with tobacco (Bottin *et al.,* 1999), estimating the expression of an endopolygalacturonase gene of *Colletotrichum lindemuthianum* during bean

infection (Dumas *et al.,* 1999), tracking *Trichoderma harzianum* growth and activity in soil (Bae and Knudson, 2000) and studying *A. flavus* interaction with cottonseed (Rajasekaran *et al.,* 2008).

GFP is a 27-kDa protein that absorbs light at maxima of 395 and 475 and emits it at a maximum of 508 nm (Lorang *et al.,* 2001). A cDNA expressing GFP was initially cloned from the jellyfish *Aequorea victoria* in 1992 (Prasher 1992). A short time afterwards, it was successfully expressed in bacteria and *Caenorhabditis elegans* (Lorang *et al.,* 2001). Applications of GFP rapidly increased after this breakthough. GFP has many advantages as a reporter protein since it only requires UV light and oxygen for visualization (no cofactor or substrates are needed), it can be tracked *in vivo*, it allows non-destructive sampling, and it is stable against proteases, a wide range of pH and relatively high temperature (Lorang *et al.,* 2001).

Abiotic stress is an important factor controlling aflatoxin accumulation. Drought stress is the most important environmental factor affecting aflatoxin contamination of peanut as it enhances *Aspergillus* infection and increases aflatoxin accumulation (Guo *et al.,* 2006; Shan *et al.,* 2011) and the development of drought-tolerant peanut cultivars could reduce aflatoxin contamination (Holbrook *et al.,* 2000a; Guo *et al.,* 2006). On the other hand, drought tolerance does not always lead to lower aflatoxin contamination (Hamidou *et al.,* 2014). However, drought conditions that have no significant apparent effects on peanut traits, *e.g.*, pegs, pods and fibrous roots, may increase the severity of *A. parasiticus* infection (Kisyombe *et al.,* 1985). Although, susceptible peanut cultivars have higher levels of seed infection than the resistant cultivars under water stress conditions, resistant genotypes may become very susceptible under extreme water deficit conditions (Waliyar *et al.,* 2003). Moreover, aflatoxin contamination of peanut can be

related to the occurrence of soil moisture stress during pod-filling when soil temperatures are near optimal for *A. flavus* (Craufurd *et al.,* 2006).

Peanut genotypes that are selected as resistant based on *in vitro* seed colonization by *A. flavus* may contain high levels of aflatoxin when subjected to an extended period of heat and drought stress (Blankenship *et al.,* 1984). Therefore, heat stress plays an important role in aflatoxin accumulation. A longer heat stress period and higher incidence of *A. flavus* contamination can result in higher aflatoxin accumulation (Sanders *et al.,* 1985).
In addition, seed structure and chemical composition greatly influences the amount of aflatoxin formed. Resistance to fungal colonization and aflatoxin contamination was found to be associated with seed coat integrity in the peanut resistant genotypes PI 337394, PI 337409, and J11 genotypes (Asis *et al.,* 2005). Wax contents of some resistant genotypes are significantly higher than susceptible cultivars (Liang *et al.,* 2003).

No peanut cultivars have complete resistance to aflatoxin production. However, significant cultivar differences do exist (Mehan and McDonald, 1984, Xue *et al.,* 2004a). In general, peanut diploid genotypes are more resistant to seed inoculation by *A. flavus* and the subsequent aflatoxin contamination than cultivated peanuts (Xue, *et al.,* 2004b). The potential exists to associate components of resistance with molecular markers if reliable phenotyping of segregating materials can be combined with polymorphic, genome-wide molecular markers.

Peanut molecular markers and polymorphism analysis

Molecular markers generally have proven to be very useful for crop improvements and studies of crop evolution in many species (Mohan *et al.,* 1997). In addition, they help to enhance selection efficiency, reduce the intensive work of indirect selection and speed up the breeding process (Liqin *et al.,* 2004). However, low levels of polymorphism in cultivated peanut to

abundant levels in wild *Arachis* have been reported (Halward *et al.,* 1991). Therefore, abundant and highly reliable markers are still needed for cultivated peanut breeding.

Different markers were used to study polymorphism among peanuts, *i.e.*, DNA Amplification Fingerprinting (DAF) (He and Prakash, 1997), Randomly Amplified Polymorphic DNA (RAPD) (Halward *et al.,* 1992; Dwivedi *et al.,* 2001; Creste *et al.,* 2005; Mondal *et al.,* 2005; Mondal *et al.,* 2008), Amplified Fragment Length Polymorphism (AFLP) (He and Prakash, 1997; Jiang *et al.,* 2007), Inter-Simple Sequence Repeat ISSR (Mondal *et al.,* 2008; Baloch *et al.,* 2010) and Sequence Related Amplified polymorphism (SRAP) (Baloch *et al.,* 2010). However, such types of markers are dominant and most of them have low reproducibility.

Simple Sequence Repeats (SSR) delivered more interest as they are co-dominant, stable markers. In addition, the variation found in peanut using microsatellites is higher than that discovered using other markers such as RAPD and AFLP (Gimenes *et al.,* 2007). Therefore, SSRs were used widely for identifying polymorphisms among peanut genotypes, creating genetic maps and finding Quantitative Trait Loci (QTL) for different traits (Hopkins *et al.,* 1999; Raina *et al.,* 2001; Guohao *et al.,* 2003; Moretzsohn *et al.,* 2004; Jayashree *et al.,* 2005; Luo *et al.,* 2005; Jiang *et al.,* 2007; Proite *et al.,* 2007; Cuc *et al.,* 2008; Mondal *et al.,* 2008; Foncéka *et al.,* 2009; Liang *et al.,* 2009; Selvaraj *et al.,* 2009; Hong *et al.,* 2010; Jiang *et al.,* 2010; Mandoulakani *et al.,* 2010; Li *et al.,* 2011; SangIK *et al.,* 2011; Zhao *et al.,* 2012; Ren *et al.,* 2014; Huang *et al.,* 2015; Zhao *et al.,* 2016; Huang *et al.,* 2017; Wilson *et al.,* 2017).

Single Nucleotide Polymorphism (SNP) is the most abundant genome-wide source of variation among species. Therefore, it has become the most attractive marker type for scientists especially after the emergence of Next-Generation Sequencing (NGS) technology. However, extracting reliable SNPs from NGS data of polyploid species is challenging. The two progenitors

of cultivated peanut have very similar sequences (Bertioli *et al.,* 2016), which increases the difficulty of differentiating SNPs within versus between sub-genomes of cultivated peanut. Therefore, the true SNP discovery rate in tetraploid peanut using NGS data is very low (Zhou *et al.,* 2014; Khera *et al.,* 2013; Peng *et al.,* 2016). However, different filtration approaches were proposed which allows increasing the discovery rate. Recently, Sliding Window Extraction of Explicit Polymorphisms (SWEEP) was developed (Clevenger and Ozias-Akins, 2015) which produces 40% accuracy. To further improve SNP-calling accuracy, we investigated the utility of machine learning.

Machine learning applications in biology

Machine learning is a set of algorithms that facilitate pattern recognition, classification and prediction based on models derived from existing data (Tarca *et al.,* 2007). Machine learning algorithms are divided into two main categories, *i.e.,* supervised and unsupervised. Supervised algorithms predict the classes of data points based on the classes supplemented in the training dataset (classification). On the other hand, unsupervised algorithms separate the input data points into distinct groups (clustering).

The first machine learning approach was created by Rosenblatt (1958) since he built a hypothetical nervous system model called a perceptron, which is considered the first artificial neural network model, to study the organization of cognitive systems. However, such types of models did not receive the widespread attention of biologists until 1982 when they were used to extract sequences of *E.coli* translation initiation sites from a library of over 78,000 mRNA sequences (Stormo *et al.,* 1982). Afterwards, a large number of machine learning algorithms were developed, *e.g.*, decision trees, super vector machine, naive Bayes, K-nearest neighbors, K-means, random forest, dimensionality reduction, gradient boosting and models that were based

on linear or logistic regression (Tarca *et al.,* 2007). Now, there are many computational tools for machine learning modeling such as WEKA, Scikit learn package of python, machine learning toolbox of MATLAB.

Since the era of 'omics' began, intensive work has been carried out for creating *in-silico* methods for sequence structural and functional annotation to extract information from the growing sequence databases. Homology-based methods were successful in many cases since they provide a rapid, efficient, and concise analysis for gene structure and function. However, the effectiveness of such methods drops dramatically when the sequence similarity is too low and becomes useless in cases of sequences with no similarity with any known genes. In addition, sequences sharing similarities sometimes do not have the same function (Hirsh and Fraser, 2001). In such cases, the importance of *ab initio* methods becomes apparent. These methods comprise approaches that extract features directly from sequences such as splice donors/acceptors, transcription start/end and coding regions to annotate gene/protein structure or function. Machine learning was an effective tool included in different applications of *ab initio* methods to annotate gene or protein structure or function as listed below.

Neural network has been used for a variety of purposes in DNA or protein annotation, *e.g.*, locating protein-coding regions in DNA sequences (Uberbacher and Mural, 1991), DNA flow cytometry histogram analysis (Ravdin *et al.,* 1993), structured vs random DNA or RNA region discrimination (Alvager *et al.,* 1997), image analysis of DNA sequencing slab gels (Li *et al.,* 2000), rare event detection in genomes (Choe *et al.,* 2000), clustering of DNA microarray data (Sawa and Ohno-Machado, 2003), studying stability of DNA/DNA duplexes (Liu *et al.,* 2005), regulatory DNA element discovery (Firpi *et al.,* 2010), non-coding DNA function

prediction (Quang and Xie 2016) and studying DNA-protein binding site interaction (Zeng *et al.,* 2016; Dutta *et al.,* 2016).

Support vector machine was used to apply classification of single function enzymes using only protein structure information (Dobson and Doig, 2005), sequence features (Mohammed and Guda, 2015), and combined attributes such as sequence, structure and chemical properties (Borgwardt *et al.,* 2005). Additionally, other trainers were used successfully for the same purposes, *e.g.*, neural network using sequence information (Osman *et al.,* 2010). However, not all enzymes have unique reactions and these methods introduce errors when multi-functional enzymes are included (Amidi *et al.,* 2017). Therefore, multi-label classifiers were also created using different machine learning approaches (Zou *et al.,* 2013; Wang *et al.,* 2014; Amidi *et al.,* 2017). A combined model of neural network and NAÏVE Bayes classifier was used to create a protein structure predictor for protein contact maps (He *et al.,* 2017).

Genetic factors underlying aflatoxin formation and resistance

The whole genome sequence of *A. flavus* was released in 2005, which was built from 2761 scaffolds (4.5Mbp-200bp). The genome has a size of 40 Mbp and contains 13,478 predicted genes distributed on 8 chromosomes (https://www.aspergillusflavus.org/genomics/). Aflatoxin synthesis is controlled by a large gene cluster near the telomere of chromosome 3 (Amaike and Keller, 2011). Although, the whole genome of *A. parasiticus* has not been sequenced so far, the aflatoxin cluster was sequenced and observed to be very similar to that of *A. flavus* (Yu *et al.,* 2004). Bhatnagar *et al.* (2003) described 21 enzymatic steps required for aflatoxin formation, regulated by two genes, aflR and aflJ, which are involved in transcriptional activation of most of the structural genes. All genes were located in a region of 70 Kbp. Subsequently, additional genes were identified within this region; Yu *et al.* (2004) described 25

genes, Ehrlich *et al.* (2005) reported 28 genes and Georgianna and Payne (2009) mapped 30 genes. The cluster region is conserved with other species of Aspergilli such as *A. nominus*, *A. pseudotamarii* and *A. bombycis*, and even with species of other genera such as *Emericella astellata* and *Dothistroma spp.* However, only *A. flavus* and *A. parasiticus* produce aflatoxin or the related sterigmatocystin in agricultural commodities (Amaike and Keller, 2011). The aflatoxin pathway is affected by different environmental and nutritional factors such as temperature, pH, carbon and nitrogen source, stress factors, lipids, and salts since these factors may affect the globally acting transcription factor aflR (Bhatnagar *et al.,* 2003). This variability confounds study of the plant response to *A. flavus* or *A. parasiticus* infection.

Given that aflatoxin accumulation in peanut is sensitive to environmental conditions, which causes large variation in aflatoxin contents among infected seeds within genotypes, finding high-confidence genetic factors controlling resistance is very challenging. As mentioned before, lipoxygenases were studied for a long time as candidates for aflatoxin resistance. Some other proteins have been reported to affect *A. flavus* resistance such as β-1,3-glucanases, chitinases, pathogenesis-related proteins 10 and 10.1, ribosome inactivating proteins (RIPs), and zeamatin (Fountain *et al.,* 2014). In addition, WRKY transcription factors are implicated in resistance (Fountain *et al.,* 2015a). Furthermore, the drought stress-responding compounds such as reactive oxygen species (ROS) are highly associated with aflatoxin production (Jayashree and Subramanyam, 2000; Reverberi *et al.,* 2012; Fountain *et al.,* 2015b) and antioxidant enzymes are highly co-expressed with fungal growth under infection conditions (Fountain *et al.,* 2016).

This study was designed to identify the genetic factors behind the *A. flavus*/peanut interaction and affecting post-harvest aflatoxin formation. Initially, tetraploid peanut lipoxygenases were studied in detail since they were reported as important genes in plant

defense. Then, a screening approach for post-harvest aflatoxin resistance was proposed and used to identify a resistant peanut genotype. Furthermore, a SNP calling tool was created to refine the SNP calling of polyploids and to study polymorphism between the resistant genotype and the susceptible one. Finally, an RNA-seq experiment was carried out to identify the genes and pathways involved in resistance to aflatoxin contamination.

References

Adrio JL, Demain, AL (2003) Fungal biotechnology. International Microbiology **6:** 191-199

Alvager T, Graham, G, Hutchison, D, Westgard, J (1997) Neural network method to analyze data compression in DNA and RNA sequences. J Chem Inf Comput Sci **37:** 335-337.

Amaike S, Keller NP (2011) *Aspergillus flavus*. Annu Rev Phytopathol **49:** 107-133

Amidi S, Amidi, A, Vlachakis, D, Paragios, N, Zacharaki, EI (2017) Automatic single- and multi-label enzymatic function prediction by machine learning. Peer J **5:** e3095

Anderson WF, Holbrook, CC, Wilson DM (1996) Development of greenhouse screening for resistance to *Aspergillus parasiticus* infection and preharvest aflatoxin contamination in peanut. Mycopathologia **135:** 115-118

Asis R, Barrionuevo, DL, Giorda, LM, Nores, ML, Aldao, MA (2005) Aflatoxin production in six peanut (*Arachis hypogaea* L) genotypes infected with *Aspergillus flavus* and *Aspergillus parasiticus*, isolated from peanut production areas of Cordoba, Argentina. Journal Agricultural Food Chemistry **53:** 9274-80

Babitha MP, Prakash, HS, Shetty, HS (2004) Purification and properties of lipoxygenase induced in downy mildew resistant pearl millet seedlings due to infection with *Sclerospora graminicola.* Plant Science **166:** 31-39

Babitha MP, Prakash, HS, Shetty, HS (2006) Induction of lipoxygenase in downy mildew resistant seedlings of pearl millet in response to inoculation with *Sclerospora graminicola*. International Journal of Agriculture and Biology **8:** 560-564

Bae YS, Knudson GR (2000) Cotransformation of *Trichoderma harzianum* with beta-glucuronidase and green fluorescent protein genes provides a useful tool for monitoring fungal growth and activity in natural soils. Appl Environ Microbiol **66:** 810

Baloch FS, Kurt C, Arioglu, Hand Ozkan H (2010) Assaying of diversity among soybean (*Glycin max* (L) Merr) and peanut (*Arachis hypogaea* L) genotypes and DNA level. Turkish Journal of Agriculture and Forestry **34:** 285-301

Bean GA, Schillinger, JA, Klarman, WL (1972) Occurrence of aflatoxins and aflatoxins-producing strains of *Aspergillus spp.* In Soybean. Applied Microbiology **24**: 737-439

Bennett JW, Klich, M (2003) Mycotoxins. Clinical Microbiology Reviews **16:** 497-516

Bertioli D, Cannon, SB, Froenicke, L, Huang, G, Farmer, AD, *et al.* (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet **48:**438-446

Bhatnagar D, Ehrlich, KC, Cleveland, TE (2003) Molecular genetic analysis and regulation of aflatoxin biosynthesis. Appl Microbiol Biotechnol **61:** 83-93

Bhattacharya K, Raha, S (2002) Deteriorative changes of maize, groundnut and soybean seeds by fungi in storage. Mycopathologia **155:** 135-141

Blankenship PD, Cole, RJ, Sanders, TH, Hill, RA (1984) Effect of geocarposphere temperature on pre-harvest colonization of drought-stressed peanuts by *Aspergillus flavus* and subsequent aflatoxin contamination. Mycopathologia **85:** 69-74

Borgwardt KM, Ong CS, Schönauer S, Vishwanathan SVN, Smola AJ, *et al.* (2005) Protein function prediction via graph kernels. Bioinformatics **21:** i47-i56

Bottin A, Larche L, Villalba F, Gaulin E, Esquerre-Tugaye M T, *et al.* (1999) Green fluorescent protein (GFP) as gene expression reporter and vital marker for studying development and microbe-plant interaction in the tobacco pathogen *Phytophthora parasitica* var. nicotianae. FEMS Microbiol Lett **176:** 51–56.

Boue SM, Shih, BY, Carter-Wientjes, CH, Cleveland, TE (2005) Effect of soybean lipoxygenases on volatile generation and inhibition of *Aspergillus flavus* mycelial growth. Journal of Agricultural and Food Chemistry **53**: 4778−4783

Brown LR (2006) Plan B 2.0, Beyond the oil peak (Chapter 2), 21-40. W.W. Norton & Company, Inc.

Buranasompob A, Tang, J, Powers, JR, Reyes, J, Clark, S, *et al.* (2007) Lipoxygenase activity in walnuts and almonds. LWT- Food Science and Technology **40:** 893-899

Burchard PD (2005) George Washington Carver: For his time and ours, Special history study. National Park Service, United States Department of the Interior

Burow GB, Gardner, HW, Keller, NP (2000) A peanut seed lipoxygenase responsive to *Aspergillus* colonization. Plant Molecular Biology **42:** 689-701

Burow GB, Nesbitt, TC, Dunlap, J, Keller, NP (1997) Seed lipoxygenase products modulate *Aspergillus* mycotoxin biosynthesis. Molecular Plant-Microbe Interactions **10:** 380-387

Caldelari D, Wang, G, Farmer, E, Dong, X (2011) *Arabidopsis lox3 lox4* double mutants are male sterile and defective in global proliferative arrest. Plant Molecular Biology **75**: 25-33

Calvo AM, Hinze, LL, Gardner, HW, Keller, NP (1999) Sporogenic effect of polyunsaturated fatty acids on development of *Aspergillus spp.* Applied and Environmental Microbiology **65**: 3668-3673

Carbone I, Ramirez-Prado, JH, Jakobek, JL, Horn, B W (2007) Gene duplication, modularity and adaptation in the evolution of the aflatoxin gene cluster. BMC Evolutionary Biology **7:** 1-12

Chen G, Hackett, R, Walker, D, Andy, T, Zhefeng, L, *el al.* (2004) Identification of a specific isoform of tomato lipoxygenase (*TomloxC*) lnvolved in the generation of fatty acid-derived flavor compounds. Plant Physiology **136:** 2641-2651

Choe W, Ersoy, OK, Bina, M (2000) Neural network schemes for detecting rare events in human genomic DNA. Bioinformatics **16:** 1062-1072

Choi JJ, Matthews, BF, Frederick, RD, Alkharouf, NW, Schneider, KT (2008) Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. Functional and Integrative Genomics **8:** 341-359

Clevenger JP, Ozias-Akins, P (2015) SWEEP: A tool for filtering high-quality SNPs in polyploid crops. G3 (Bethesda) **5**:1797-803

Cotty PJ, Jaime-Garcia R (2007) Influences of climate on aflatoxin producing fungi and aflatoxin contamination. International Journal of Food Microbiology **119:** 109–115

Craufurd PQ, Prasad, PVV, Waliyar, F, Taheri, A (2006) Drought, pod yield, pre-harvest *Aspergillus* infection and aflatoxin contamination on peanut in Niger. Field Crop Research **98:** 20-29

Creste S, Tsai, S, Valls, J, Gimenes, M, Lopes, C (2005) Genetic characterization of Brazilian annual *Arachis* species from sections *Arachis* and *Heteranthae* using RAPD markers. Genetic Resources and Crop Evolution **52:** 1079-1086

Cuc LM, Mace, ES, Crouch, JH, Quang, VD, Long, TD, *et al.* (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biology **8:** 1-11

Diener UL, Cole, RJ, Sanders, TH, Payne, GA, Lee, LS, *et al.* (1987). Epidemiology of Aflatoxin formation by *Aspergillus flavus*. Annual Review of Phytopathology **25:** 249-270

Dobson PD, Doig AJ (2005) Predicting enzyme class from protein structure without alignments. Journal of Molecular Biology 345: 187-199

Doehlert DC, Wicklow, DT, Gardner, HW (1993) Evidence implicating the lipoxygenases pathway in providing resistance to soybeans against *Aspergillus flavus.* Phytopathology **83**: 1473-1477

Du W, Huang Z, Flaherty JE, Wells K, Payne GA (1999) Green fluorescent protein as a reporter to monitor gene expression and food colonization by *Aspergillus flavus*. Appl Environ Microbiol **65:** 834–836.

Dumas B, Centis S, Sarrazin N, Esquerre-Tugaye MT (1999) Use of green fluorescent protein to detect expression of an endopolygalacturonase gene of *Colletotrichum lindemuthianum* during bean infection. Appl Environ Microbiol **65:** 1769–1771.

Dutta S, Madan, S, Parikh, H, Sundar, D (2016) An ensemble micro neural network approach for elucidating interactions between zinc finger proteins and their target DNA. BMC Genomics **17:** 1033

Dwivedi SL, Gurtu, S, Chandra, S, Yuejin, W, Nigam, SN (2001) Assessment of genetic diversity among selected groundnut germplasm I: RAPD analysis. Plant Breeding **120:** 345-349

Ehrlich KC, Chang, PK, Yu, J, Cotty, PJ (2004) Aflatoxin biosynthesis cluster gene cypA is required for G aflatoxin formation. Applied Environmental Microbiology **70:** 6518-6524

Ehrlich KC, Yu, J, Cotty, PJ (2005) Aflatoxin biosynthesis gene clusters and flanking regions. Journal of Applied Microbiology **99:** 518-527

Ferrie BJ, Beaudoin, NN, Burkhart, WW, Bowsher, CG, Rothstein, SJ (1994) The cloning of two tomato lipoxygenase genes and their differential expression during fruit ripening. Plant Physiology **106:** 109-118

Firpi HA, Ucar, D, Tan, K (2010) Discover regulatory DNA elements using chromatin signatures and artificial neural network. Bioinformatics **26:** 1579-1586

Floyd W, Harper, D, Thompson, N (2000) Paper coating lubricant. Google Patents US 6162771 A

Foncéka D, Hodo-Abalo, T, Rivallan, R, Faye, I, Sall, MN, *et al.* (2009) Genetic mapping of wild introgressions into cultivated peanut: a way toward enlaging the geneic basis of a recent allotetraploid. BMC Plant Biology **9:** 1-13

Fountain JC, Bajaj, P, Nayak, SN, Yang, L, Pandey, MK, *et al.* (2016) Responses of *Aspergillus flavus* to oxidative stress are related to fungal development regulator, antioxidant enzyme, and secondary metabolite biosynthetic gene expression. Frontiers in Microbiology **7:** Article 2048

Fountain JC, Khera, P, Yang, L, Nayak, SN, Scully, BT, *et al.* (2015b) Resistance to *Aspergillus flavus* in maize and peanut: Molecular biology, breeding, environmental stress, and future perspectives. The Crop Journal **3:** 229-237

Fountain JC, Raruang, Y, Luo, M, Brown, RL, Guo, B, *et al.* (2015a) Potential roles of WRKY transcription factors in regulating host defense responses during *Aspergillus flavus* infection of immature maize kernels. Physiological and Molecular Plant Pathology **89:** 31-40

Fountain JC, Scully, BT, Ni, X, Kemerait, RC, Lee, RD, *et al.* (2014) Environmental influences on maize-*Aspergillus flavus* interactions and aflatoxin production. Frontiers in Microbiology **5:** Article 40

Frisvad JC and Samson RA (2004) Polyphasic taxonomy of *Penicillium* subgenus *Penicillium*, A guide to identification of food and airborne *terverticillate Penicillia* and their mycotoxins. Studies in Mycology **49:** 1-173

Gao X, Feussner, I, Betran, J, Kolomiets, MV, Keller, NP, *et al.* (2009) Inactivation of the lipoxygenase *ZmLOX3* Increases susceptibility of maize to *Aspergillus spp.* Molecular Plant-Microbe Interactions **22**: 222-231

Gao X, Feussner, I, Tumlinson, J, Kolomiets, M, Starr, J, *et al.* (2008) Maize 9-Lipoxygenase *ZmLOX3* controls development, root-specific expression of defense genes, and resistance to root-knot nematodes. Molecular Plant-Microbe Interactions **21:** 98-109

Gardner HW (1988) Lipoxygenase pathway in cereals. Advances in Cereal Science and Technology, chapter8, 161-215

Georgianna DR, Payne, GA (2009) Genetic regulation of aflatoxin biosynthesis: From gene to genome. Fungal Genetics and Biology **46:** 113-125

Gimenes MA, Hoshino, AA, Barbosa, AVG, Palmieri, DA, Lopes, CR (2007) Characterization and transferability of microsatellite markers of the cultivated peanut (*Arachis hypogaea*). BMC Plant Biology **7:** 1-17

Guo BZ, Xu, G, Cao, YG, Holbrook, CC, Lynch, RE (2006) Identification and characterization of phospholipase D and its association with drought susceptibilities in peanut (*Arachis hypogaea*). Planta **223:** 512-550

Guohao H, Meng, R, Newman, M, Gao, G, Pittman, RN, *et al.* (2003) Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L). BMC plant Biology **3:** 1-6

Halward T, Stalker, H, LaRue, E, Kochert, G (1991) Genetic variation detectable with molecular markers among un-adapted germplasm resources of cultivated peanut and related wild species. Genome **34:** 1013-1020

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

Hamidou FH, Rathore, A, Waliyar, F, Vadez, V (2014) Although drought intensity increases aflatoxin contamination, drought tolerance does not lead to less aflatoxin contamination. Field Crops Research **156:** 103-110

Harish BN, Gowda, MVC, Kusuma, VP (2005) Confectionery groundnuts resistant to seed colonization by *Aspergillus flavus.* International Arachis Newsletter **25:** 10-12

He B, Moruza, SM, Wang, Y, Shen, H, Zhang, Y (2017) NeBcon: Protein contact map prediction using neural network training coupled with naïve Bayes classifiers. Structural Bioinformatics btx164

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

Hirsh AE, Fraser, HB (2001) Protein dispensability and rate of evolution. Nature **411:** 1046-9

Holbrook CC, Kvien, CK, Rucker, KS, Wilson, DM, Hook, JE, *et al.* (2000a) Preharvest aflatoxin contamination in drought-tolerant and drought-intolerant peanut genotypes. Peanut Science **27:** 45-48

Holbrook CC, Matheron, ME, Wilson, DM, Anderson, WF, Will, ME, *et al.* (1994) Development of a large-scale field system for screening peanut for resistance to preharvest aflatoxin contamination. Peanut Science **21:** 20-22

Hong Y, Chen, X, Liang, X, Liu, H, Zhou, G, *et al.* (2010) A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L) genome. BMC Plant Biology **10:** 1-13

Hopkins MS, Casa, AM, Wang, T, Mitchell, SE, Dean, RE, *et al.* (1999) Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Science **39:** 1243-1247

Horn BW (2003) Ecology and population biology of aflatoxigenic fungi in soil. Journal of Toxicology, toxin reviews **22:** 351-379

Huang L, He, H, Chen, W, Ren, X, Chen, Y, *et al.* (2015) Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea* L.). Theoretical and Applied Genetics **128:** 1103-1115

Huang L, Ren, X, Wu, B, Li, X, Chen, W, *et al.* (2017) Development and deployment of a high-density linkage map identified quantitative trait loci for plant height in peanut (*Arachis hypogaea* L.). Scientific Reports **6:** 39478

Huang P, Franco, O, Yan, Y, Keller, N, Kolomiets, M (2013) Maize lipoxygenase *LOX2* regulates pathogenesis of mycotoxin-producing *Aspergillus flavus.* APS-MSA Joint meeting, August 10-14, Austin, Texas

Jayashree B, Ferguson, M, Ilut, D, Doyle, J, Crouch, JH (2005) Analysis of genomic sequences form peanut (*Arachis hypogaea*). Electronic Journal of Biotechnology **8:** 226-237

Jayashree T, Subramanyam, C (2000) Oxidative stress as a prerequisite for aflatoxin production by *Aspergillus parasiticus*. Free Radic Biol Med **29:** 981-985

Jiang H, Liao, B, Ren, X, Lei, Y, Mace, E, *et al.* (2007) Comparative assessment of genetic diversity of peanut (*Arachis hypogaea* L) Genotypes with Various Levels of Resistance to Bacterial Wilt Through SSR and AFLP Analyses. Journal of Genetics and Genomics **34:** 544-554

Jiang H, Ren, X, Zhang, X, Huang, J, Lei, Y, *et al.* (2010) Comparison of genetic diversity based on SSR markers between peanut mini core collections from China and ICRISAT. Acta Agronomica Sinica **36:** 1084-1091

Junghans TG, Oliveira, MA, Moreira, MA (2004) Lipoxygenase activities during development of root and nodule of soybean. Pesquisa Agropecuária Brasileira **39:** 625-630

Khera P, Upadhyaya, HD, Pandey, MK, Roorkiwal, M, Sriswathi, M, *et al.* (2013) SNP-based genetic diversity in the reference set of peanut (*Arachis spp*) by developing and applying cost-effective KASPar genotyping assays. The Plant Genome **6:** 1-11

Kisyombe CT, Beute, MK, Payne, GA (1985) Field evaluation of peanut genotypes for resistance to infection by *Aspergillus parasiticus.* Peanut Science **12:** 12-17

Klich MA (2007) *Aspergillus flavus*: the major producer of aflatoxin. Molecular Plant Pathology **8:** 713-22

Knothe GH (2001) Historical perspectives on vegetable oil-based diesel fuels. Inform **12:** 1103-1107

Koch E, Brigitt MM, Hans-Georg, E, Slusarenko, A (1992) A lipoxygenase from leaves of tomato (*Lycopersicon esculentum* Mill) is induced in response to plant pathogenic *Pseudomonads.* Plant Physiology **99:** 571-576

Kolomiets MV, Hannapel, DJ, Chen, H, Tymeson, M, Gladon, RJ (2001) Lipoxygenase is involved in the control of potato tuber development. The Plant Cell **13:** 613- 626

Kumari AM, Sucharitha, A, Devi, PUM (2012) Impact of 13-lipoxygenase on induction of resistance in peanut against aflatoxigenic *Aspergillus* Continental. Journal of Biological Science **5**: 4-13

Li Y, Chen, CY, Knapp, SJ, Culbreath, AK, Holbrook, CC, *et al.* (2011) Characterization of simple sequence repeat (SSR) markers and genetic relationships within cultivated peanut (*Arachis hypogaea* L). Peanut Science **38:** 1-10

Li Y, Schwager, C, Ansorge, W (2000) Using a neural network for lane-tracking of DNA sequencing slab gels. J Biochem Biophys Methods **45:** 65-74

Liang X, Chen, X, Hong, Y, Liu, H, Zhou, G, *et al.* (2009) Utility of EST-derived SSR in cultivate peanut (*Arachis hypogaea* L) and *Arachis* wild species. BMC Plant Biology **9:** 1-9

Liang XQ, Zhou, GY, Pan, RZ (2003) Wax and cuticle of peanut seed coat in relation to infection and aflatoxin production by *Aspergillus flavus.* Jouranl of Tropical and Subtropical Botany **11:** 11-14

Lima WA, Borém, AA, Dias, DS, Moreira, MA, Dias, LS (2010) Lipoxygenase and physiological quality of soybean seeds during storage. Seed Science and Technology **38:** 767-771

Liqin D, Fang, PZ, Changqing, C (2004) Of conventional breeding with molecular breeding research and application. Cereal Crops **24:** 280-281

Liu S, Han, B (2010) Differential expression pattern of an acidic 9/13-lipoxygenase in flower opening and senescence and in leaf response to phloem feeders in the tea plant. BMC Plant Biology **10:** 228

Liu X, Ma, L, Cheng, C, Wang, Y, Miyajima, H, *et al.* (2005) A neural network for predicting the stability of DNA/DNA duplexes. Nucleosides Nucleotides Nucleic Acids **24:** 199-209

Lorang JM, Tuorl, RP, Martinez, JP, Sawyer, TL, Redman, RS, *et al.* (2001) Green fluorescent protein is lighting up fungal biology. Applied and Environmental Microbiology **67:** 1987-1994

Luo M, Liang, XQ, Dand, P, Holbrook, CC, Bausher, MG, *et al.* (2005) Microarray-based screening of differentially expressed genes in peanut in response to *Aspergillus parasiticus* infection and drought stress. Plant Science **169:** 695-703

Malekian F, Rao, RM, Prinyawiwatkul, W, Marshall, WE, Windhauer, M, *et al.* (2000) Lipase and lipoxygenase activity, functionality and nutrient losses in rice bran during storage, Louisiana Agricultural Experiment station, LSU agricultural center, Baton Rouge, LA 70894-5055, Bulletin No. 870

Mandoulakani BA, Alami, A, Esfahani, M (2010) Association analysis for morphological traits in peanut (*Arachis hypogaea*) using microsatellite markers. Iranian Journal of Crop Science **12:** 510-519

Mariutto M, Duby, F, Adam, A, Bureau, C, Fauconnier, M, *et al.* (2011) The elicitation of a systemic resistance by *Pseudomonas putida* BTP1 in tomato involves the stimulation of two lipoxygenase isoforms. BMC Plant Biology **11:** 29

Matsuki T, Tazaki H, Fujimori T, Hogetsu, T (1992) The influences of jasmonic acid methyl ester on microtubules in potato cells and formation of potato tubers. Bioscience, Biotechnology, and Biochemistry **56:** 1329–1330

Mattes R (2005) Soup and satiety. Physilogy and Behavior **83:** 739-747

Mehan VK, McDonald, D (1984) Research on the aflatoxin problem in groundnut at ICRISAT. Plant and Soil **79:** 255-260

Melan MA, Enriquez, ELD, Peterman, KP (1994) The *LOX1* gene of *Arabidopsis* is temporally and spatially regulated in germinating seedlings. Plant Physiology **105:** 385-393

Mita G, Iannacone, R, Casey, R, Santino, A, Epifani, F, *et al.* (2007) 9-Lipoxygenase metabolism is involved in the almond/*Aspergillus* carbonarius interaction. Journal of Experimental Botany **58:** 1803-1811

Mixon AC, Roger, KM (1973) Peanut accessions resistant to seed infection by *Aspergillus flavus.* Journal of Agronomy **65:** 560-562

Mohammed A, Guda C (2015) Application of a hierarchical enzyme classification method reveals the role of gut microbiome in human metabolism. BMC Genomics **16:** S16

Mohan M, Nair, S, Bhagwat, A, Krshna, TG, Yano, M, *et al.* (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding **3:** 87-103

Mondal S, Ghosh, S, Badigannavar, AM (2005) RAPD polymorphism among groundnut genotypes differing in disease reaction to late leaf spot and rust. International *Arachis* Newsletter **25:** 27-30

Mondal S, Sutar, SR, Badigannavar, AM (2008) Comparison of RAPD and ISSR marker profile of cultivated peanut genotypes susceptible or resistant to foliar diseases. Journal of Food and Agriculture Environment **6:** 181-187

Moretzsohn MC, Hopkins, MS, Mitchell, SE, Kresovich, S, Valls, JFM, *et al.* (2004) Genetic diversity of peanut (*Arachis hypogaea* L) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biology **4:** 1-10

Muller V, Ame, MV, Carrari, V, Gieco, J, Asis R (2014) Lipoxygenase activation in peanut seed cultivars resistant and susceptible to *Aspergillus parasiticus* colonization. Phytopathology **104**: 1340-8

Nigam SN, Waliyar, F, Aruna, R, Reddy, SV, Kumar, LP, *et al.* (2009) Breeding peanut for resistance to aflatoxin contamination at ICRISAT. Peanut Science **36:** 42-49.

Ohta H, Kan, S, You-Liang, P, Furusawa, I, Jiko, S, *et al.* (1991) A lipoxygenase pathway is activated in rice after infection with the rice blast fungus *Magnaporthe grisea.* Plant Physiology **97:** 94-98

Osman MH, Liong, C, Hashim, I (2010) Hybrid learning algorithm in neural network system for enzyme classification. International Journal of Advances in Soft Computing and its Applications **2:** 209-220

Ozalvo RR, Cabrera, JJ, Escobar, CC, Christensen, SA, Borrego, EJ, *et al.* (2014) Two closely related members of *Arabidopsis* 13-lipoxygenases (13-LOXs), *LOX3* and *LOX4*, reveal distinct functions in response to plant-parasitic nematode infection. Molecular Plant Pathology **15:** 319-332

Paller AS, Nimmagadda, S, Schachner, L, Mallory, SB, Kahn, T, *et al.* (2003) Fluocinolone acetonide 0.01% in peanut oil: therapy for childhood atopic dermatitis, even in patients who are peanut sensitive. Journal of the American Academy of Dermatology **48:** 569-577

Payne GA and Brown, MP (1998) Genetics and physiology of aflatoxin biosynthesis. Annual Review of Phytopathology **36:** 329-362

Peng ZM, Gallo, BL, Tillman, D, Rowland, Wang, J (2016) Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L). Mol Genet Genomics **291**: 363-81

Porta H, Rocha-Sosa, M (2000) A *Phaseolus vulgaris* lipoxygenase gene expressed in nodules and in *Rhizobium tropici* inoculated roots. Biochimica et Biophysica Acta **1517:** 139-142

Porta H, Rocha-Sosa, M (2002) Plant lipoxygenases. physiological and molecular features. Plant Physiology **130:** 15-21

Prasher DC (1992) Primary structure of *Aequorea Victoria* green-fluorescent protein. Gene **111:** 229-233

Proite K, Leal-Bertioli, SCM, Bertioli, DJ, Moretzsohn, MC, Silva, FR, *et al.* (2007) ESTs from a wild *Arachis* species for gene discovery and marker development. BMC Plant Biology **7:** 1-10

Quang D, Xie, X: DanQ (2016) A hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. Nucleic Acids Res **44:** e107

Raina SN, Rani, B, Kojima, T, Ogihara, Y, Singh, KP, *et al.* (2001) RAPD and SSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. Genome **44:** 763-772

Rajasekaran K, Cary, JW, Cotty, PJ, Cleveland, TE (2008) Development of a GFP-expressing *Aspergillus flavus* strain to study fungal invasion, colonization, and resistance in cottonseed. Mycopathologia **165:** 89–97

Ravdin PM, Clark, GM, Hough, JJ, Owens, MA, McGuire, WL (1993) Neural network analysis of DNA flow cytometry histograms. Cytometry **14:** 74-80

Ren X, Jiang, H, Yan, Z, Chen, Y, Zhou, X, *et al.* (2014) Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. PLOS ONE **9:** e88091

Reverberi M, Punelli, M, Smith, CA, Zjalic, S, Scarpari, M, *et al.* (2012) How peroxisomes affect aflatoxin biosynthesis in *Aspergillus flavus*. PLOS ONE **7:** e48097

Robens J, Cardwell, KF (2003) The costs of mycotoxin management to the USA: management of aflatoxins in the United States. Journal of Toxicology, toxin reviews **22:** 139–152

Rodríguez-Concepción MM, Beltrán, JP (1995) Repression of the pea lipoxygenase gene *LOXg* is associated with carpel development. Plant Molecular Biology **27:** 887-899

Rodriguez-Saona LE, Barrett, DM, Selivonchick, DP (1995) Peroxidase and lipoxygenase influence on stability of polyunsaturated fatty acids in sweet corn (*Zea mays* L). Journal of Food Science **60:** 1041-1044

Rosenblatt F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev **65:** 386-408

Roth N (1991). Skin moisturizing and oil restoring lotion. Google Patents US 4986986 A

Sanders HT, Cole, RJ, Blankenship, PD, Hill, RA (1985) Relation of environmental stress duration to *Aspergillus flavus* invasion and aflatoxin production in preharvest peanuts. Peanut Science **12:** 90-93

SangIK H, SukBok, B, TaeJoung, H, MyongHee, L, KiChang, J, *et al.* (2011) The development and selection of SSR markers for identification of peanut (*Arachis hypogaea* L) varieties in Korea. Korean Journal of Breeding Science **43:** 133-138

Saravitz DM, Siedow, JN (1995) The lipoxygenase isozymes in soybean [*Glycine max* (L) Merr] leaves: changes during leaf development, after wounding, and following reproductive sink removal. Plant Physiology **107:** 535-543

Sawa T, Ohno-Machado, L (2003) A neural network-based similarity index for clustering DNA microarray data. Comput Biol Med **33:** 1-15

Scheidegger KA, Payne, GA (2003). Unlocking the secrets behind secondary metabolism: A review of *Aspergillus flavus* from pathogenicity to functional genomics. Journal of Toxicology **22:** 423-459

Seijo, J, Lavia, GI, Fernandez, A, Karpovickas, A, Ducasse, D, *et al.* (2004) Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (Legumniosae). Am J Bot **91:** 1294-1303

Selvaraj MG, Narayana, M, Schubert, AM, Ayers, JL, Baring, MR, *et al.* (2009) Identification of QTLs for pod and kernel traits in cultivated peanut by bulked segregant analysis. Electronic Journal of Biotechnology **12:** 1-10

Seo JY, Kim, SS, Kim, HJ, Liu, K, Lee, H, *et al.* (2013) Laxative effect of peanut sprout extract. Nutr Res Pract **7:** 262-266.

Shastry AV, Ben-Yoseph, EM, Walters, M, Willcocks, NA, Collins, TM, *et al.* (2009) Edible inks for ink-jet printing on edible substrates. Google Patents WO 2004003089 A1

Shan S, Zhang, T, Li, C, Yang, C, Yan, C, *et al.* (2001). Cloning and analysis of a NBS-LRR disease resistance gene candidate PnAG1 from peanut (*Arachis hypogaea* L.). Electronic Journal of Biotechnology **14:** 1424

Shen J, Tieman, D, Jones, J, Taylor, M, Schmelz, E, *et al.* (2014) A 13-lipoxygenase, *TomloxC*, is essential for synthesis of C5 flavour volatiles in tomato. Journal of Experimental Botany **65:** 419-428

Shirasawa K, Takeuchi, Y, Ebitani, T, Suzuki, Y (2008) Identification of gene for rice (*Oryza sativa*) seed lipoxygenase-3 involved in the generation of stale flavor and development of SNP markers for lipoxygenase-3 deficiency. Breeding Science **58:** 169-176

Shivakumar, PD, Geetha, HM, Shetty, HS (2003) Peroxidase activity and isozyme analysis of pearl millet seedlings and their implications in downy mildew disease resistance. Plant Science **164:** 85-93

Stormo GD, Schneider, TD, Gold, L, Ehrenfeuch, A (1982) Use of the perceptron algorithm to distinguish translation initiation sites in *E. coli*. Nucleic Acids Res **10:** 2997-3011

Tang Q, Yu, Z, Wu, Y, Zhang, Y, Liu, B (2009) Study on the relationship between lipoxygenase-3 and the characteristic of resisting storage insects of rice grain. Journal of Food, Agriculture and Environment **7:** 334-338

Tarca AL, Carey, VJ, Chen, X, Romero, R, Draghici, S (2007) Machine learning and its applications to biology. PLOS Computational Biology **3:** e116

Ties P, Barringer, S (2012) Influence of lipid content and lipoxygenase on flavor volatiles in the tomato peel and flesh. Journal of Food Science 77: C830-C837

Tsitsigiannis DI, Kunze, S, Willis, DK, Feussner, I, Keller, NP (2005) *Aspergillus* infection inhibits the expression of peanut 13S-HPODE-forming seed lipoxygenases. Molecular Plant-Microbe Interactions Journal **18**: 1081-1089

Uberbacher EC, Mural, RJ (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc Natl Acad Sci USA **88:** 11261-11265

Vogt J, Schiller, D, Ulrich, D, Schwab, W, Dunemann, F (2013) Identification of lipoxygenase (LOX) genes putatively involved in fruit flavour formation in apple (*Malus* x *domestica*). Tree Genetics and Genomes **9:** 1493-1511

Waliyar F, Traore, A, Fatondji, D, Ntare, BR (2003) Effects of irrigation interval, planting date, and cultivar on *Aspergillus flavus* and aflatoxin contamination of peanut in a sandy soil of Niger. Peanut Science **30:** 79-84

Wang R, Liu, Y, Su, N, Wan, J, Shen, W, *et al.* (2008) A novel lipoxygenase gene from developing rice seeds confers dual position specificity and responds to wounding and insect attack. Plant Molecular Biology **66:** 401-414

Wang Y, Jing R, Hua Y, Fu Y, Dai X, *et al.* (2014) Classification of multi-family enzymes by multi-label machine learning and sequence-based descriptors. Analytical Methods **6:** 6832-6840

Wilson JN, Chopra, R, Baring, MR, Selvaraj, MG, Simpson, CE, *et al.* (2017) Advanced backcross Quantitative Trait Loci (QTL) analysis of oil concentration and oil quality traits in peanut (*Arachis hypogaea* L.). Tropical Plant Biology **10:** 1-17

Xue HQ, Isleib, TG, Payne, GA, OBrain, G (2004a) Evaluation of post-harvest aflatoxin production in peanut germplasm with resistance to seed colonization and pre-harvest aflatoxin contamination. Peanut Science **31:** 124-134

Xue HQ, Isleib, TG, Stalker, HT, Payne, GA, OBrain, G (2004b) Evaluation of *Arachis* species and interspecific tetraploid lines for resistance to aflatoxin production by *Aspergillus flavus.* Peanut Science **31:** 134-141

Yan L, Zhai, Q, Wei, J, Li, S, Wang, B, *et al.* (2013) Role of tomato lipoxygenase D in wound-induced jasmonate biosynthesis and plant immunity to insect herbivores. PLOS Genetics **9:** 12

Young RK, Cousin, MA (2001) Detection of moulds producing aflatoxins in maize and peanuts by an immunoassay. International Journal of Food Microbiology **65:** 27-38

Yu, J, Bhatnagar, D, Cleveland, TE (2004) Complete sequence of aflatoxin pathway gene cluster in *Aspergillus parasiticus*. FEBS Letters **564:** 126-130

Zeng H, Edwards, MD, Liu, G, Gifford, DK (2016) Convolutional neural network architectures for predicting DNA-protein binding. Bioinformatics **32:** i121-i127

Zeringue H (1996) Possible involvement of lipoxygenase in a defense response in aflatoxigenic *Aspergillus*-cotton plant interactions. Canadian Journal of Botany **74:** 98-102

Zhang Y, Xuan, H, Zhu, X, Yang, J, He, C, *et al.* (2009) Effect of lipoxygenase activity and red seed coat on rice bran deterioration. Journal of the Science of Food and Agriculture **89:** 1904-1908

Zhao Y, Prakash, CS, He, G (2012) Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. BMC Research Notes **5:** 362

Zhou X, Dong, Y, Zhao, J, Huang, L, Ren, X, *et al.* (2016) Genomic survey sequencing for development and validation of single-locus SSR markers in peanut (*Arachis hypogaea* L.). BMC Genomics **17:** 420

Zhou XJ, Xia, Y, Ren, X, Chen, Y, Huang, L, *et al.* (2014) Construction of SNP-based genetic linkage map in cultivated peanut based on lage scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics **15:** 351

Zou Q, Chen W, Huang Y, Liu X, Jiang Y (2013) Identifying multi-functional enzyme by hierarchical multi-label classifier. Journal of Computational and Theoretical Nanoscience **10:** 1038-1043

CHAPTER 2

DUPLICATION, GENE EXPRESSION AND GENOMIC ORGANIZATION OF THE ARACHIS LIPOXYGENASE GENE FAMILY

[1]W. A. Korani, Y. Chu, and P. Ozias-Akins. To be submitted to *BMC Plant Biology*

Abstract

Lipoxygenases (LOX) are a dioxygenase gene super-family which functions to catalyze the addition of two oxygen atoms on cis, cis-1,4-pentadiene structure within fatty acids. Although, this function is simple at the molecular level, its biological impact is on plant development and response to pathogens. Lipoxygenases have a very conserved structure containing two domains, *i.e.,* PLAT and Lipoxygenase domains, which facilitates LOX gene annotation.

In soybean, lipoxygenases have been studied in detail and a large number of LOX genes, 44 genes, were identified. In peanut (*Arachis hypogaea*), only three lipoxygenases have been experimentally confirmed from seed cDNA libraries and three from root cDNA libraries. Therefore, we utilized the published reference genomes of the progenitors of cultivated peanut (*A. duranensis* and *A. ipaensis*) to carry out a comprehensive study of LOX genes in tetraploid peanut including characterization, functional classification, and large-scale expression profiling for developmental stages, nodulation and plant-pathogen interactions.

Twenty-four and 25 LOX genes were identified in *A. duranensis* and *A. ipaensis*, respectively. Among them, 20 orthologous pairs were assigned. LOX genes were distributed across most chromosomes and tended to be located within regions that have chromosomal rearrangements. Three LOX pairs are located in regions with inversions (chromosome 6 or 9), ten within regions with intra-chromosomal translocation (*e.g.,* chromosome 8) and one within an inter- chromosomal translocation region (chromosomes 6 and 9).

Eleven LOX genes were missing a PLAT domain or had a truncated Lipoxygenase domain. Therefore, functional prediction was applied to only 36 genes because of the difficulty to align genes with significantly different lengths. The genes were successfully classified into

four clusters, 13S_typeI, 13S_type_II, 9S_type_I and 9S_typeII. Differential expression profiling of 22 different developmental stages showed three prominent patterns; one group of LOX genes was highly expressed in seed tissues (different stages), another group was constitutively expressed and the third was highly expressed only in gynophore, shoot, root, pericarp and flower tissues. In addition, expression profiles showed similar patterns for orthologous genes. The expression pattern of LOX genes during plant-pathogen interaction (Aaspergilli, nematode and *Cercosporidium personatum*) revealed the response of specific LOX genes for every disease. Most affected LOX genes were down-regulated under the infection conditions compared with the respective controls.

We identified 17 active LOX gene pairs in tetraploid peanut, out of 20 total orthologous pairs, which were clustered in functional groups that reflect their cellular roles. In addition, LOX genes were grouped into discrete expression patterns across peanut developmental stages that may give an indication for their phenotypic functions. Moreover, LOX genes whose expression was altered during pre- or post-harvest aflatoxin contamination, nematode, late leaf spot infection, and nodulation were identified.

Background

Lipoxygenases (LOX) form a gene superfamily that is ubiquitously distributed in plants, fungi, algae and animals (Andreou *et al.,* 2009). In addition, LOX genes have been detected in prokaryotic organisms (Hansen *et al.,* 2013) and *Archaea* (Koval and Jarrell, 1987). Moreover, some LOX-like sequences were found in viruses (Horn *et al.,* 2015). Lipoxygenases are non-heme, iron-containing dioxygenases that recognize a cis,cis-1,4-pentadiene structure within polyunsaturated fatty acids and activate the addition of two molecular oxygen atoms to produce fatty acid hydroperoxides. The amino acid sequences of LOX genes have a conserved structure

across animals and plants; they have five conserved histidine residues in a stretch of 38 residues in addition to a sixth histidine at 160 residues downstream (Steczko *et al.,* 1992). Lipoxygenases have two distinct domains, *i.e.*, PLAT (Polycystin-1, Lipoxygenase, Alpha-Toxin) and Lipoxygenase. The PLAT domain (NCBI-CDD: pfam01477) is a small amino-terminal domain forming an eight-stranded antiparallel β-barrel. On the other hand, the Lipoxygenase domain (NCBI-CDD: pfam00305) is a long carboxy-terminal domain, which contains the active site and forms 18-22 helices in addition to one or two anti-parallel β-sheets. The conserved histidine residues and domain structure facilitate the distinction of lipoxygenases from other enzymes of living organisms.

Lipoxygenases have been classified according to the insertion position of oxygen atoms; animals have four major types, *i.e.*, 12-lipoxygenase (EC:1.13.11.31), 15-lipoxygenase (EC:1.13.11.33), 5-lipoxygenase (EC:1.13.11.34) and 8-lipoxygenase (EC:1.13.11.40) that insert dioxygen atoms at C12, C15, C5 and C8 positions of arachidonic acid, respectively. Although plants have two fatty acids containing a cis, cis-1,4-pentadiene structure, linoleic and alpha-linolenic acid, they have merely two types of lipoxygenases, 13-lipoxygenase (EC:1.13.11.12) and 9-lipoxygenase (EC:1.13.11.58) that insert dioxygen at C13 and C9 positions of these fatty acids, respectively.

Plant lipoxygenases produce cis-trans 9S- or 13S- hydroperoxy linoleic acid (9S- or 13S-HPODE) or 9S- or 13S-hydroperoxy linolenic acid (9S- or 13S-HPOTE) (Tsitsigiannis *et al.,* 2005). 13-lipoxygenase catalyzes the oxidation of linoleate acid into (9Z,11E,13S)-13-hydroperoxyoctadeca-9,11-dienoate or alpha-linolenate into (9Z,11E,13S,15Z)-13-hydroperoxyoctadeca-9,11,15-trienoate (UNIPROT: http://www.uniprot.org/). Whereas, 9-lipoxygenase catalyzes the oxidation of linoleate acid into (9S,10E,12Z)-9-hydroperoxy-10,12-

octadecadienoate (UNIPROT: http://www.uniprot.org/). These oxidative products are highly reactive as they are metabolized in various enzymatic pathways into a series of oxylipins, *e.g.,* jasmonates, epoxy hydroxy-fatty acids, hydroxyl-fatty acids, keto-fatty acid, aldehydes, alcohols, traumatin and divinylether. These compounds have important roles in plant defense mechanisms, plant development, signaling, abiotic stress and quality traits (AOCS Lipid Library: http://lipidlibrary.aocs.org/).

Peanut (*Arachis hypogaea* L.) is an important oilseed crop; it is ranked fourth in world production after soybean, cottonseed and rapeseed. Peanut kernels are very rich in protein (25-30%) and oil (45-55%) contents. Peanut oil contains 27.3-38.3 % linoleic acid, 4.7 % in high oleic lines (O'keefe *et al.,* 1993), in addition to 0.37 to 1.11 % of alpha-linolenic (Musa 2010), which are suitable substrates for lipoxygenases. Lipoxygenases have been reported to be involved in biotic stress such as *Aspergillus flavus* and *A. parasiticus* infection and the subsequent aflatoxin contamination (Tsitsigiannis *et al.,* 2005; Burow *et al.,* 2000), and nematode infection (Gao *et al.,* 2008), which are major problems that affect peanut production and quality. In addition, they may have a role in nodulation (Bueno *et al.,* 2001), an important process for peanut growth. Moreover, lipoxygenases have been reported to be involved in plant development, and seed quality and flavor (St. Angelo *et al.,* 1979; Engeseth *et al.,* 1987; Robinson *et al.,* 1995).

The first recognized peanut LOX gene, *PnLOX1*, was identified from a seed cDNA library (Burow *et al.,* 2000). It was shown to be highly induced by methyl jasmonate treatment, wounding and *A. parasiticus* infections in mature cotyledons (Burow *et al.,* 2000). Two other peanut seed LOX genes, *PnLOX2* and *PnLOX3*, were observed to be highly expressed in mature seeds (Tsitsigiannis *et al.,* 2005); however, they were repressed by *Aspergillus flavus* infection

(Tsitsigiannis *et al.,* 2005). Two additional peanut seed LOX genes were annotated from a transcriptome shotgun assembly (TSA) database, *PnLOX4* and *PnLOX5*, these two genes responded differently, compared with each other and across different peanut genotypes, as well as across multiple time points, to *A. flavus* infection (Muller *et al.,* 2014). Recently, three peanut LOX genes were identified from a root cDNA library, *i.e., PnLOX6*, *PnLOX7* and *PnLOX8* (Guo *et al.,* 2015).

In this work, we identified all possible LOX genes in tetraploid peanut utilizing the reference genome of its progenitors (*A. duranensis* and *A. ipaensis*), studied the orthology and synteny among them, and classified them into functional clusters. In addition, we profiled their expression across a wide range of developmental tissues and their response to *A. flavus*, nematode, *Cercosporidium personatum* and *Rhizobium* interaction with peanut.

Results and Discussion

LOX gene pool of cultivated peanut

Lipoxygenases have a highly conserved motif within the Lipoxygenase domain at the binding site of catalytic iron; this motif has five histidine residues in the form of H-4aa-H-4aa-H-17aa-H-8aa-H (where, H is histidine and aa is any amino acid). This structure was validated in legumes using crystallographic determination of the active site and the ligand of soybean *LOX-1* (Minor *et al.,* 1993). In addition, the full structure of plant lipoxygenase domains, which is formed from PLAT domain followed by Lipoxygenase domain was confirmed for the same enzyme of soybean by crystallography (Minor *et al.,* 1996). Moreover, we found a region of 50 residues, designated the LOX-Core having a similar amino acid frequency across 135 LOX genes that were collected from a wide range of organisms. These genes were extracted from GenBank by searching lipoxygenases that have been experimentally tested; the annotated LOXs

were excluded. These features allowed the annotation of LOX sequences from the gene pools of *A. duranensis* and *A. ipaensis* with a high level of confidence (Bertioli *et al.,* 2016). Scanning MAKER annotation of both genomes captured 24 and 25 LOX genes (out of 39313 and 44436 total genes), respectively (Table 2.1 and additional files: Tables 2.S1 and 2.S2).

The orthology analysis of these LOX genes using reciprocal best Basic Local Alignment Search Tool (BLAST) Hits (RBH) approach identified 20 orthologous pairs (Table 2.1). The genes were named by the abbreviation of the reference genome, A or B for *A. duranensis* and *A. ipaensis*, respectively, then an underscore and the number of the gene assigned according to the order of its occurrence across chromosomes. The estimated gene structure on the tetraploid genome was named by the abbreviation of *Arachis hypogaea*, Ah, followed by the source of the annotated gene between two underscores then the gene number. The source of the annotated genes were abbreviated by A, B or A/B; A for paralogous genes of *A. duranensis* without an orthologous gene within the *A. ipaensis* genome and duplicated from one LOX gene of *A. duranensis*. B is vice versa. And the orthologous pairs that had one gene originating from the *A. duranensis* genome and the other gene originating from the *A. ipaensis* genome were abbreviated by A/B. Every orthologous pair or single paralogs were considered a unique *A. hypogaea* LOX gene.

Although, RBH was used widely for identification of orthologous gene pairs across different genomes (Mattila *et al.,* 2012; Fuchsman and Rocap, 2006; Hirsh and Fraser, 2001), genes that are most similar based on BLAST best hits, may not be phylogenetically the closest (Koski and Golding, 2001). One example of this among the LOX genes we analyzed, Araip.8Z22U, had the best BLAST similarity with Aradu.2E0TL, which was annotated as phosphoenolpyruvate carboxylase. Therefore, RBH may lead to inaccurate results and needs to

be supported by other approaches for orthology study. We validated the results of RBH by creating dot-plots for all pairs and studying the flanking genes (data not shown).

To confirm the LOX gene structures, the most similar proteins were identified using NCBI-BLASTp (Additional files: Table 2.S3) using peanut predicted LOX amino acid sequences against the non-redundant protein database of NCBI. All orthologous gene pairs, Ah_A/B_ 01, 03-06, 8-11, 13-15, 17 and 18, had a PLAT domain and a complete Lipoxygenase domain, the ideal structure of LOX. However, the orthologous pair Ah_A/B_02 had two domains of each PLAT and Lipoxygenase in a structure of PLAT-Lipoxygenase-PLAT-Lipoxygenase for both A_02 and B_02, suggesting that both pairs were tandem duplications and were more likely to be four genes rather than only two. An opposite scenario was observed for Ah_A/B_07 since Aradu.U67PQ had part of a Lipoxygenase domain, and Aradu.289WG located adjacently on the chromosome in the same direction and had the second part of Lipoxygenase and a PLAT domain. Therefore, they are more likely to be the same gene that were truncated, interrupted by another segment or had a mistake in the MAKER annotation especially its orthologous gene, Araip.HGI2J, which had the two domains in the ideal structure. Similarly, the two genes Aradu.2ZL37 and Aradu.S1X34 were located adjacent to each other and in the same direction on the same chromosome. The first one had a PLAT domain and a part of a Lipoxygenase domain and the latter had the other part of the Lipxoygenase domain. Therefore, they presumably are one gene.

Based on these results, corrections were applied for genes of Ah_A/B_02 to split each gene into two genes (Ah_A/B_2A and Ah_A/B_2B) and Ah_A/B_07 to join Aradu.U67PQ and Aradu.289WG into one gene (Ad_07). In addition, Aradu.2ZL37 and Aradu.S1X34 are joined in one gene (Ah_A_26) (Table 2.1). Moreover, the analysis showed that some LOX genes had

missing segments, *e.g.,* the four genes Ah_A/B_12 and Ah_A/B_20, and Araip.Q7EYZ of Ah_A/B_19. Therefore, they are excluded from the phylogenetic analysis, as they may negatively affect multiple alignment and more likely to be pseudogenes.

A phylogenetic tree of the annotated LOX genes and the eight experimentally recognized peanut LOX genes was generated to reveal redundant genes (Additional files: Table 2.S4 and figure 2.S1). The three cDNA seed library LOX genes, *PnLOX1-3* (Tsitsigiannis *et al.,* 2005; Burow *et al.,* 2000), were identified as different forms of Ah_A/B_09. In addition, the seed LOX *PnLOX4* (Muller *et al.,* 2014) was identified as Ah_A/B_08. Additionally, the seed LOX *PnLOX5* (Muller *et al.,* 2014) and the root LOX *PnLOX7* (Guo *et al.,* 2015) were recognized as one gene, Ah_A/B_17. Root LOXs *PnLOX6* and *PnLOX8* (Minor *et al.,* 1993) matched Ah_A/B_15 and Ah_A/B_14, respectively. Interestingly, 13 new LOX genes were recognized, *i.e.*, Ah_A/B_01, Ah_A/B_02A, Ah_A/B_02B, Ah_A/B_03, Ah_A/B_04, Ah_A/B_05, Ah_A/B_06, Ah_A/B_07, Ah_A/B_10, Ah_A/B_11, Ah_A/B_13, Ah_A/B_16 and Ah_A/B_18. A similar study was carried out in legumes (Song *et al.,* 2016). However, the number of recognized LOXs in peanut was lower than those in this study.

To study if there are more LOX genes in tetraploid peanut that were not annotated from the two progenitor genomes, *de novo* assembly of RNA-seq data without genome guidance resulted in 250,802 genes (376,393 isoforms); 284,642,109 bases were assembled with a contig N50 value = 1185 and an average contig = 756.24. Among these genes, 41 genes (63 isoforms) were recognized as LOX genes. Within the peanut LOX gene pool, the shortest sequence containing a full Lipoxygenase domain was Ad_16 (2127 bp). Therefore, sequences that are shorter than 2 kbp are more likely to be pseudogenes as the functional LOXs previously reported or found in NCBI GenBank had a length greater than 2 Kbp. For that reason, we extracted

assembled genes that were captured by BLAST and had a length more than 2 kbp; 13 genes (31 isoforms) were extracted. No redundancy was found among these genes. LOX gene structure of these was confirmed by NCBI-BLASTx search. The phylogenetic tree of these genes with peanut LOX genes, described above, revealed that none of these assembled genes is a new LOX gene (Additional files: Figure 2.S2). This suggests that all LOX gene duplications took place before the hybridization of *A. duranensis* and *A. ipaensis*. In other words, the duplication of LOX genes occurred in the diploid genomes of peanut and no event occurred in the tetraploid peanut.

De novo assembly without genome guidance failed to differentiate between the two forms (homeologs) of the genes since the similarity between them is very high (> 99 %). In addition, some LOX genes were not assembled, *i.e.,* Ah_A/B_ 01, 06,17 and 20 since their expression levels were too low to be assembled. Although Ah_A/B_ 09, 10 and 13 had significant expression in some tissues, they also failed to be assembled since they were expressed in many other tissues and they had a very highly similarity with other genes, *i.e.,* Ah_A/B_ 19, 11 and 15, respectively. This increases the probability that many common k-mers will occur between them, which increases the probability of forming overlapping de Bruijn graphs (Homolog.us – Bioinformatics: http://www.homolog.us/blogs/) that complicates distinction of splicing forms of the same gene and highly similar genes.

Localization of LOX genes across chromosomes

The abundance of LOXs in peanut suggests their importance in peanut evolution especially in terms of disease resistance. Therefore, studying LOX duplications across the genome is a necessary prerequisite to elucidating duplicate gene function. Chromosomal locations and directions of LOX genes were estimated on a synthetic peanut tetraploid genome constructed by combining sequences from A-genome and B-genome progenitors, *A. duranensis*

and *A. ipaensis,* respectively (Figure 2.1). LOX genes were distributed across all chromosomes, except chromosomes A01/B01 and A05/B05, and sometimes on opposite strands. Some chromosomes had few LOX genes, *i.e.*, A04/B04 and A07/B07 (AB stands for the synthetic tetraploid genome). The majority of LOX genes were found on chromosomes A03/B03, A06/B06, A08/B08 and A09/B09, the latter had almost half of the LOX genes, nine orthologous pairs (eight full length with both domains) and two paralogs.

Twenty-six LOX genes are positioned near the ends of chromosomes and many are in rearranged chromosomal regions, *i.e.*, inverted segments, intra-chromosomal translocations and inter-chromosomal translocations. Almost half of the LOX genes are found in intra-chromosomal translocations (Figure 2.1 – green regions) as the orthologous genes were located on the same chromosome of the other parental genome however in different positions. Three gene pairs are located within inversions (Figure 2.1 – sky blue regions and additional files: Figure 2.S3). Orthologous genes at Ah_A/B_05 were surrounded by 11-S seed storage protein and potassium transporter family protein in opposite directions. The two genes of Ah_A/B_06 were surrounded by auxin response 4-like gene and DNA-directed DNA polymerase gene in opposite directions. The two pairs were located on chromosomes A06/B06. The distance between Ad_05 and Ad_06 LOX genes was 3.33 Mbp, and 3.17 Mbp between Ai_05 and Ai_06. Therefore, Ah_A/B_05 and Ah_A/B_06, were located in a large inverted region. The third inverted orthologous pair was located near of the end of chromosomes A09/B09 surrounded by spermidine synthase I and a vacuolar membrane-like protein in opposite directions. Only one orthologous pair, Ah_A/B_20, was found in an inter-chromosomal translocation as one gene was located on A09 and the second of the pair was located on B06 (Figure 2.1 – a dark blue line). However, the Lipoxygenase domains of both genes are incomplete and they are missing PLAT domains.

To validate the synthetic tetraploid genome assembly and the locations of LOX genes, PCR primers were designed to amplify the interval segments between LOX genes and the closest genes. However, most of these interval segments are too long to be amplified easily by PCR. Therefore, we tested two loci with intergenic segments predicted to be less than 3 kbp. The primers were designed based on conserved sequences across A- and B- genomes. PCR for the intergenic segments between Ad_18 and Aradu_5RZ6C and between Ai_18 and Araip_Q00X2 produces bands with a size of ~600 bp (data not shown), which are close to the predicted sizes. Similarly, amplicons of segments around Ah_A/B_04 were close to the calculated ones as the calculated segments flanking Ad_04 that were 1000 and 1863 bp had amplicons of ~1 and 2 kbp, respectively. And those calculated ones around Ai_04, which were 977 and 2389 bp had amplicons of ~1 and 2.5 kbp, respectively (Additional files: Figure 2.S4).

LOX genes are ubiquitously distributed across living organisms, from very basic forms of life (prokaryotes) up to the highly advanced organisms (human). In addition, they often are duplicated within genomes during evolution (Additional files: Table 2.S5). Our results showed that LOX genes are distributed on the genomes similarly across different eukaryotes since multiple, or sometimes many, LOXs are located on particular chromosomes and most of them tend to cluster near telomeres while other chromosomes do not have any LOXs. Soybean is the closest related crop to peanut with experimentally detailed studies for LOX genes (Additional files: Figure 2.S5). Forty-four LOX sequences, containing a Lipoxygenase domain, were retrieved from SoyBase (Soybase: http://www.soybase.org/) (Additional files: Table 2.S6). Soybean LOXs also were distributed across multiple chromosomes, nine chromosomes had 1-2 LOXs each, and five chromosomes had four to nine LOXs; most of them were clustered in small

regions near the ends of chromosomes. In addition, six chromosomes were devoid of LOX genes.

Classification of Peanut LOX genes

Andreou and Feussner (2009) classified plant, mammal and prokaryotic LOX genes into three functional groups based on amino acid sequences, *i.e.,* 13S_legumes, 9S-typeI and 13S_typeII. They used two criteria: 1) the position of added dioxygen on the fatty acid (carbon 13 or 9) and 2) the subcellular localization of the translated enzyme (typeI contains extra-plastidial enzymes and typeII contains plastidial enzymes). We constructed a phylogenetic tree using the annotated peanut LOX gene sequences along with selected plant and mammalian sequence accessions; the latter were included as an outgroup (Figure 2.2 and additional files: Table 2.S7). Unlike Andreou and Feussner (2009), we excluded prokaryotes from our analysis since the lipoxygenase structure differs from that of plants and animals (Hansen *et al.,* 2013). In addition, we included only complete LOX genes as truncated ones may distort the multiple alignment and consequently affect the integrity of the phylogenetic tree.

All functionally characterized LOX genes that clustered within the 13S_typeII group (Figure 2.2 - green color) had evidence of chloroplast localization and the production of 13-hydroxyperoxides as major products; *Arabidopsis thaliana* had three genes within this group, all with chloroplast transit peptides, jasmonate-inducibility and photosynthetic light reactivity (Bell and Mullet, 1993; Frenkel *et al.,* 2009; Kilaru *et al.,* 2011; UNIPROT: http://www.uniprot.org/). Similarly, LOXs from other species of this group had predicted transit peptides for targeting the enzymes to chloroplasts and tendency to produce 13S-hydroperoxides, *i.e.*, two potato LOXs (*LOX2:St:1* and *LOX2:St:2*), two tomato LOXs (*LOX2:Le:1* and *LOC2:Le:2*), moss *PpLOX1* and tobacco *LOX2:Nt* (Royo *et al.,* 1996; Heitz *et al.,* 1997; Chen *et al.,* 2004; Shen *et al.,* 2014;

Senger *et al.,* 2005). Since peanut LOX genes Ah_A/B_ 01, 03, 04, 07, 16 and 19 were clustered within this group, they are more likely to be chloroplastic 13S-lipoxygenases. Moreover, NCBI-BLASTp search (Additional files: Table 2.S3) confirmed the similarity of these genes to 13S_typeII since all of them had best hits with 13S-lipoxygenases with a chloroplast localization feature.

The other two groups contained type-I LOX genes since there was no evidence for the presence of chloroplast transit peptides. LOX genes of potato and *Arabidopsis* that bordered the 9S_typeI group (Figure 2.2 - blue color) are recognized to be 9S- lipoxygenases (Kilaru *et al.,* 2011; Royo *et al.,* 1996). There are three peanut genes located among this group, *i.e.,* Ah_A/B_06, Ah_A/B_17 and Ah_A/B_18. However, the latter contains a signal for chloroplast translocation since NCBI-BLASTp search gave best hits with chloroplastic 9S- lipoxygenases. On the other hand, a NCBI-BLASTp search for the other two genes gave best hits with 9S-lipoxygenases without chloroplastic signals. Therefore, Ah_A/B_18 is more likely to be a new (fourth) group of, *i.e.*, 9S_typeII (Figure 2.2 - dark blue color).

The classification of 13S_legume LOXs as 13S_typeI (Figure 2.2 - red color) is erroneous since four out of seven soybean LOX genes, which form the main structure of this group, produce only 50-60% of 13S- hydroperoxy products (Chen *et al.,* 2004; Youn *et al.,* 2006). Moreover, all peanut genes included in this group had best hits of NCBI-BLASTp search with 9S-lipoxygenases except Ah_A/B_09, which had best hits with 13S-lipoxygenases. However, this gene is equivalent to *PnLOX1*, which produces 30% 9S-hydroperoxy products (Burow *et al.,* 2000), and *PnLOX2* and *PnLOX3*, which produce 16% 9S-hydroperoxy (Tsitsigiannis *et al.,* 2005). Therefore, this group presents LOX genes that produce a combination of 9S- and 13S- hydroperoxy products and it may be roughly divided into two sub-

groups depending on the ratio of 13S- to 9S- hydroperoxy products, *i.e.,* 13S-typeI_SG1, which has a high ratio (> 0.4), and 13S-typeI_SG2, which has a moderate to low ratio (< 0.4).

Interestingly, all peanut LOX genes that belong to 13S_typeII group were located on chromosomes A02/B02, A03/B03, A06/B06 and A10/B10, except Ah_A/B_16 and Ah_A/B_17 (chromosome A09/B09). In addition, all LOX genes that belong to 13S_typeI group were located on chromosome A08/B08 and two clusters on chromosome A09/B09. Therefore, LOXs that have similar function tend to be clustered together.

The expression profiling of peanut LOX genes at developmental stages

Figure 2.3 represents the RNA-seq expression profiles of peanut LOX genes across 22 different peanut tissues, including leaves, shoots, flowers, gynophores, pericarps, and seeds (Additional files: Table 2.S8) (Clevenger *et al.,* 2016a). The profile revealed three prominent expression patterns (EP) *i.e.,* EP-I, EP-II and EP-III, in addition to unique patterns for some genes. Moreover, genes that do not have a complete LOX gene structure, *i.e.,* Ah_A/B_ 12, 20 and 21-27 (which are excluded from the phylogenetic tree – Figure 2.2), did not show significant expression in any tissue, except for Ah_A/B_20. This pair lacks the PLAT domain and part of the Lipoxygenase domain. Therefore, while expressed, function may be impaired. Although, Ah_A/B_01 and Ah_A/B_19 have complete LOX gene structures, they may be pseudogenes since they have low expression in all tissues. Interestingly, all other genes had significant expression levels across the tissues and the two orthologs for each pair had similar expression profiles. Southern and northern blot analysis of several LOX orthologous pairs using ortholog-specific probes confirmed their copy number and expression patterns (Additional files: Table 2.S9 and figures 2.S6 and 2.S7).

EP-I comprised two orthologous pairs, Ah_A/B_8 (*PnLOX4* (Muller *et al.,* 2014)) and Ah_A/B_9 (*PnLOX1-3* (Tsitsigiannis *et al.,* 2005; Burow *et al.,* 2000)) that are highly expressed in mature and immature seeds. The expression of *PnLOX1-4* changes significantly upon *in vitro* infection by *Aspergillus spp.* (Tsitsigiannis *et al.,* 2005; Burow *et al.,* 2000; Muller *et al.,* 2014). There is evidence of seed LOX gene response in *Aspergillus spp.* interaction with soybean (Doehlert *et al.,* 1993; Mellon and Cotty, 2002), maize (Gao *et al.,* 2009; Huang *et al.,* 2013), cotton (Zeringue Jr 1996) and almond (Buranasompob *et al.,* 2007). Furthermore, LOX products can affect aflatoxin biosynthesis in the fungus (Kumari *et al.,* 2011). Therefore, Ah_A/B_08 and Ah_A/B_09 are more likely to have roles in *Aspergillus spp.* interaction with peanut and aflatoxin biosynthesis inside the fungi.

Seed LOX genes can affect seed quality positively or negatively. Their activity is helpful in delaying bran deterioration in rice (Zhang *et al.,* 2009). On the other hand, oxylipins that are produced by LOX proteins cause oil rancidity in seeds of soybean (Rodriguez-Saona *et al.,* 1995), walnuts and almond (Buranasompob *et al.,* 2007). In addition, oxylipins are metabolized to volatile products, which may alter seed flavor (Gardner 1988). These products may be responsible for stale flavor in rice after storage (Shirasawa *et al.,* 2008) and fruit aroma in apple (Vogt *et al.,* 2013), yet they may contribute to C6 (Chen *et al.,* 2004) and C5 (Shen *et al.,* 2014) flavor volatiles, which are important for consumer liking of fresh tomatoes. Therefore, Ah_A/B_08 and Ah_A/B_09 may also have roles in seed quality and flavor traits of peanut.

EP-II comprised mainly Ah_A/B_ 04, 14, 16, 17 and 18. These genes were ubiquitously expressed across all tissues. On the other hand, EP_III comprised six LOX genes, *i.e.,* Ah_A/B_ 03, 05, 10, 11, 13 and 15, which are expressed in gynophore, shoot, root, pod and flower (Ah_A/B_15 is expressed also in leaves) but at different levels. LOX genes play a role in leaf

development in soybean (Saravitz and Siedow, 1995), early germination of *Arabidopsis* (Melan *et al.,* 1994), carpel development of pea (Rodriguez-Concepcion and Beltran 1995), seed germination and fruit ripening of tomato (Ferrie *et al.,* 1994) and microtuber formation in potato (Matsuki *et al.,* 1992; Kolomiets *et al.,* 2001). Therefore, the expression of genes that belong to these two groups, across a wide range of tissues and in different stages with different levels, supports the idea that they may have functions in peanut development.

Peanut LOX genes in the response to plant-pathogen interaction

Many reports support the inclusion of LOX genes in plant-pathogen interaction for different diseases, *e.g.,* soybean interaction with *Phakopsora pachyrhizi* (which causes soybean rust) (Choi *et al.,* 2008), pearl millet infection by *Sclerospora graminicola* (the pathogen that causes downy mildew disease) (Babitha *et al.,* 2004; Babitha *et al.,* 2006), the interaction between rice and *Magnaporthe grisea*, (which causes rice blast fungus) (Ohta *et al.,* 1991), the resistance to insects and necrotrophic pathogens in tomato (Yan *et al.,* 2013) and defensive reaction of tea to phloem-feeders (Liu and Han, 2010).

Expression profiles of peanut LOX genes in response to different pathogens and nodulation is presented in Figure 2.4. Ah_A/B_8 and Ah_A/B_9, which showed high expression in seed tissues, were the most responsive to pre- and post-harvest aflatoxin contamination. The genes were down-regulated in response to infection as compared with controls in more resistant (C76-16, NC3033, Tifguard) and more susceptible (A72, Florida07 and Tifrunner) genotypes to pre-harvest aflatoxin (Clevenger *et al.,* 2016b), and in both resistant (ICG 1471) and susceptible (Florida-07) genotypes to post-harvest aflatoxin (Chapter 3). These results are contrary to those reported by Burow *et al.* (2000) where up-regulation of *PnLOX1* (Ah_A/B_9) after infection by *A. parasiticus* was observed using northern blot analysis. These conflicting results may be due to

probe choice since they observed only one band in southern blot analysis, yet at least two bands would be predicted based on our observations. In addition, Tsitsigiannis *et al.* (2005) tested two forms of the same gene (Ah_A/B_09), *PnLOX2* and *PnLOX3*, using northern blot and qPCR and reported opposite results from previous work, *i.e.,* down-regulation for both after the infection of *A. flavus*. Moreover, our northern blot analysis of three peanut genotypes (GT-C20, Tifrunner and Florunner) showed down-regulation in peanut after infection by *A. flavus* (additional files: Figure 2.S8). Genes Ah_A/B_17, Ad_25 and Ai_27 were responsive also for infection in both pre- and post- harvest aflatoxin analysis. Therefore, they may play a minor role in response to infection as they had lower expression in seed tissues. The differential response between the resistant versus susceptible genotype is not clear. For instance, Ah_A/B_08 of Florida-07 (susceptible for pre- and post- harvest aflatoxin contamination) was the most down regulated as a result of infection as compared with those of the three resistant genotypes in pre-harvest aflatoxin analysis. On the other hand, this gene of ICG-1471, which is resistant to post-harvest aflatoxin contamination, was more down regulated than that of the susceptible genotypes (Florida-07) in post-harvest aflatoxin analysis.

LOX genes were reported to play important roles in *Arabidopsis* and maize responses to nematodes (Gao *et al.,* 2008; Ozalvo *et al.,* 2014). Ah_A/B_ 10, 14, 15 and 16 had high responses to nematode infection (Figure 2.4) as their expression was very high and was extremely reduced by the infection. They also had a significant high expression in the root tissue (Figure 2.3). Similar to *A. flavus* infection, LOX genes are down-regulated by nematode infection in both the resistant (Tifguard) and susceptible (Gregory) genotypes. Ah_A/B_ 06, 07 and 02 had a high response to late leaf spot infection (Figure 2.4). They also were expressed highly in the leaf tissues especially Ah_A/B_02, which was expressed only in leaf tissue (Figure

2.3). Therefore, they may be candidate genes in leaf development and have a function in the response to late leaf spot. Again, LOXs are more likely to be down-regulated by *Cercosporidium personatum* infection. Interestingly, LOXs are very specific in their response to pathogens since LOXs that responded to *A. flavus* infection are different from those that responded to nematode infection and both are different from *C. personatum* responsive LOXs. This suggests that the pathways including these gene groups do not interact, especially since these genes have tissue specific expression patterns (Figure 2.3) and knocking out a particular LOX expressed in one tissue in response to pathogen infection may not affect response to another disease organism (Martins *et al.,* 2002). Additionally, most LOX genes responsive to diseases are 13S_typeI or 13S_typeII. Therefore, 13S-hydroperoxides are important in disease response pathways either in cytoplasm or plastids.

Ah_A/B_ 03, 13, 14, 15 and 16 were the most affected genes during nodulation stages (Figure 2.4). They had significantly different expression profiles between roots and nodules (Figure 2.3), suggesting they are involved in the symbiosis between peanut and *Bradyrhizobium spp.* since this role was notified for LOX genes in some plants, *e.g.,* common bean (Porta and Rocha-Sosa, 2000).

Conclusions

Among the annotated genes of *A. duranensis* and *A. ipaensis*, 28 genes were recognized as candidate LOX genes in tetraploid peanut, 21 orthologous pairs (including the two orthologous pairs resultied from Ah_A/B_02 splitting) and 7 paralogs. Seventeen orthologous pairs had a complete LOX gene structure and were expressed in peanut tissues (not pseudogenes). Thirteen orthologous pairs out of them had three discrete expression patterns and each of the others had unique expression profiles. The expression patterns reflected the

importance of LOX genes in peanut and their putative functions in peanut development, biotic stress (Aspergilli, nematode and *C. personatum*) and symbiosis (nodulation). In addition, LOX genes were functionally classified into four clusters, one of which had two sub-clusters and another one merely contained one gene. Moreover, LOX genes tend to be located in clusters near to the chromosome ends since most of LOX genes of peanut and many other organisms had this feature. Finally, *de novo* assembly of RNA-seq data failed to discover a new class of LOX genes. Therefore, all LOX duplication events occurred before the tetraploidization of peanut.

Methods

Data resources

The genomes and MAKER annotations of diploid progenitors of cultivated tetraploid peanut (*Arachis duranensis* and *A. ipaensis*) were downloaded from PeanutBase (PeanutBase: http://peanutbase.org/). RNA-seq data of 22 different tissues of tetraploid cultivar Tifrunner were obtained from the transcriptome project of tetraploid peanut (courtesy of Peggy Ozias-Akins lab in collaboration with Brian Scheffler) (Clevenger *et al.,* 2016a). The RNA-seq data of pre-harvest, post-harvest aflatoxin, nematode and late leaf spot infections, and nodulation were collected from different projects (courtesy of Peggy Ozias-Akins lab). Published LOX gene and protein sequences were collected from NCBI, SoyBase, MaizeGDB and UniProt databases (O'keefe *et al.,* 1993; Soybase: http://www.soybase.org/; Maize GDB: http://www.maizegdb.org/).

Synteny analysis

Blast+ was used to carry out reciprocal BLAST analysis (Camacho *et al.,* 2009); two nucleotide libraries were constructed for the transcripts of *A. duranensis* and *A. ipaensis*, BLASTn was applied for the LOX gene pool of *A. duranensis* against the library of *A. ipaensis*

and vice versa. Geneious was used to create dot-plot graphs for protein and DNA sequences (Geneious version 8.1.6 http://www.geneious.com, (Kearse *et al.,* 2012)). Flanking genes of LOX genes were studied using the visualization tool of PeanutBase (PeanutBase: http://peanutbase.org/). The LOX pairs that passed the reciprocal BLAST and at least one of the dot plots or flanking genes filters were considered orthologous (Additional files: Table S10).

Clustering of LOX genes

BLASTx and BLASTp were applied to the transcripts and the annotated protein sequences of LOX genes for studying structure. Geneious was used for applying multiple alignment and creating UPGMA phylogenetic trees (Geneious version 8.1.6 http://www.geneious.com, (Kearse *et al.,* 2012)).

RNA-seq analysis

All paired-end read files were mapped to a synthetic tetraploid peanut genome, which merged *A. duranensis* and *A. ipaensis* genomes, using Tophat version 2.0.13 (Trapnell *et al.,* 2009). Cufflinks, Cufflmerge and Cuffdiff were used for applying the differential expression analysis of tissue samples (Cufflinks version 2.2.1, (Trapnell *et al.,* 2012)) and CummeRbund R package (Goff *et al.,* 2013) were used for visualizing the expression profiles in ($+1\text{Log}_{10}$) FPKM (Fragment Per Kilobase Million); ($+1\text{Log}_{10}$) TPM (Transcripts Per Kilobase Million) were calculated for biotic stress RNA-seq experiments using R 3.2.2.

Plant materials and nucleic acid extraction

Young leaf tissues of Tifrunner and GTC-20 genotypes were collected for southern blot analysis. Modified CTAB method was used for DNA extraction (Doyle and Doyle, 1987). Leaf, root, pericarp, mature and yellow-I stage seeds were collected from GTC-20 for probe preparation and northern blot analysis. RNA was extracted by homogenizing 100 mg of frozen

ground tissues (powdered) in 1 ml trizol and 200 µl of chloroform. RNA was precipitated using 250 µl of 3M sodium acetate and 250 µl of isopropanol. RNA was tested on formaldehyde agarose gel and measured by nanodorp (Thermo Scientific - Version 2000).

The Validation of LOX gene loci

PCR primers were designed for two loci of LOX and adjacent genes (Additional files: Table 2.S11) using Geneious 8.1.6 (Kearse *et al.,* 2012). DNA of *A. duranensis* and *A. ipaensis* accessions V14167 and K30076 respectively (courtesy of Peggy Ozias-Akins lab) was used for PCR.

Designing and preparing the probes

cDNA was constructed from RNA of root, leaves and seeds tissues using SuperScript® III RT (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. LOX gene specific primer pairs were designed by Vector NTI®Suite V6.0 Software (InforMax, Bethesda, MD) (Additional files: Table 2.S11). An amplicon for every LOX gene was selected after PCR and purified by QIAquick Gel Extraction Kit (QIAGEN Inc. Valencia, CA) according to the manufacturer's instructions. The cleaned amplicons were ligated to PCR-4-Topo vector (Invitrogen, Carlasbad, CA) and transformed to NEB-5-alpha competent *E. coli* cells (New England BioLabs Inc., Ipswich, MA). Isolated colonies were collected and sequenced by Sanger capillary DNA sequencing (Georgia Genomics Facility, University of Georgia, Athens, GA).

Plasmids were extracted from the enriched colonies using QIA prep spin miniprep kit (QIAGEN Inc. Valencia, CA) according to the manufacturer's instructions and PCR was conducted. Amplicons were purified by QIAquick Gel Extraction Kit (QIAGEN Inc. Valencia, CA) according to the manufacturer's instructions. Twenty-five ng of every probe was labeled by α^{32}P-dCTP using Random Primer DNA Labeling Kit (Roche, Indianapolis, IN) according to the

manufacturer's instructions. Unincorporated label was removed by Sephadex G-50 (Sigma, St. Louis, MO).

The specificity of every probe was tested against all probes by running 80 pg in an agarose gel, transferring to GeneScreen Plus® nylon membrane (PerkinElmer Inc., Waltham, Massachusetts) then applying hybridization according to Sambrook and Russell (2001) (Additional files: Figure 2.S9).

Southern blot analysis

Ten µl Tifrunner and GTC-20 DNAs were digested overnight using *Eco*RV, *Xba*I and *Sac*I restriction nucleases. The digested DNA was run on an agarose gel overnight at 30 V. The gels were transferred to GeneScreen Plus® nylon membranes (PerkinElmer Inc., Waltham, Massachusetts) overnight in 0.4N sodium hydroxide. DNA was fixed at 80° C for 2 h. the hybridization was carried out overnight at 65° C according to Sambrook and Russell (2001). The signals were visualized by the overnight exposure to X-ray film.

Northern blot analysis

Fifteen µl of RNA of the previously described tissues were applied to formaldehyde agarose gel overnight at 30 V. RNA was transferred to GeneScreen Plus® nylon membranes (PerkinElmer Inc., Waltham, Massachusetts) overnight in 20X SSC and fixed at 80° C for 2 h and hybridized with labeled probes overnight according to Sambrook and Russell (2001). The signals were visualized by the exposure to X-ray film overnight.

De novo assembly of RNA-seq data and sequence analysis for LOX genes

De novo transcript sequence assembly was carried out for concatenated files of the 22 tissues using Trinity 2.0.6 (Haas *et al.,* 2013). One hundred thirty five different LOX proteins of different organisms were collected from GenBank. Fifty residues, including active site and

surrounding segments were extracted and designated LOX-Core. Motif signatures were analyzed using WEBLOGO Version 2.8.2 (WEBLOGO: http://weblogo.berkeley.edu/) (Additional files: Figure 2.S10). BLASTx was applied for the assembled transcripts against a library of LOX-Core sequences using BLAST+ (Camacho *et al.,* 2009).

Tables

Table 2.1. Orthologous pairs and paralogs of peanut LOX genes depending on *A. duranensis* and *A. ipaensis* genomes, their locations and designated shortcut names.

Orthologous pairs / paralogous ID	Chromosome	A-genome		B-genome	
		PeanutBase ID	Short-form ID	PeanutBase ID	Short-form ID
Ah_A/B_01	02	Aradu.XZG8N	Ad_01	Araip.849ER	Ai_01
Ah_A/B_02A	03	Aradu.Q5K4W	Ad_02	Araip.MN7KE	Ai_02
Ah_A/B_02B	03				
Ah_A/B_03	03	Aradu.W07KG	Ad_03	Araip.NWR3L	Ai_03
Ah_A/B_04	03	Aradu.C88Z1	Ad_04	Araip.X1W86	Ai_04
Ah_A/B_05	06	Aradu.8D3SW	Ad_05	Araip.2KP3T	Ai_05
Ah_A/B_06	06	Aradu.G99LQ	Ad_06	Araip.W6TLM	Ai_06
Ah_A/B_07	06	Aradu.U67PQ	Ad_07a	Araip.HGI2J	Ai_07
		Aradu.289WG	Ad_07b		
Ah_A/B_08	08	Aradu.AC956	Ad_08	Araip.DH1Z0	Ai_08
Ah_A/B_09	08	Aradu.WX5KP	Ad_09	Araip.E99Y9	Ai_09
Ah_A/B_10	09	Aradu.SK1BS	Ad_10	Araip.5F6MD	Ai_10
Ah_A/B_11	09	Aradu.AE16G	Ad_11	Araip.T64GQ	Ai_11
Ah_A/B_12	09	Aradu.GJ1CE	Ad_12	Araip.7V9BH	Ai_12
Ah_A/B_13	09	Aradu.FM0YX	Ad_13	Araip.GV48H	Ai_13
Ah_A/B_14	09	Aradu.TJL9X	Ad_14	Araip.K56RN	Ai_14
Ah_A/B_15	09	Aradu.951UC	Ad_15	Araip.Q8LFT	Ai_15
Ah_A/B_16	09	Aradu.LNK8S	Ad_16	Araip.VN0A4	Ai_16
Ah_A/B_17	09	Aradu.C3RV0	Ad_17	Araip.3GK67	Ai_17
Ah_A/B_18	09	Aradu.KZX2M	Ad_18	Araip.D6PZJ	Ai_18
Ah_A/B_19	10	Aradu.AS232	Ad_19	Araip.Q7EYZ	Ai_19
Ah_A/B_20	09/06	Aradu.5E1NU	Ad_20	Araip.G6789	Ai_20
Ah_B_21	02	-	-	Araip.HUQ4W	Ai_21
Ah_B_22	04	-	-	Araip.J97JQ	Ai_22
Ah_B_23	06	-	-	Araip.P423S	Ai_23
Ah_B_24	07	-	-	Araip.8Z22U	Ai_24
Ah_A_25	08	Aradu.KXZ9V	Ad_25	-	-
Ah_A_26	09	Aradu.2ZL37	Ad_26a	-	-
		Aradu.S1X34	Ad_26b	-	-
Ah_B_27	09	-	-	Araip.LB22X	Ai_27

Figures

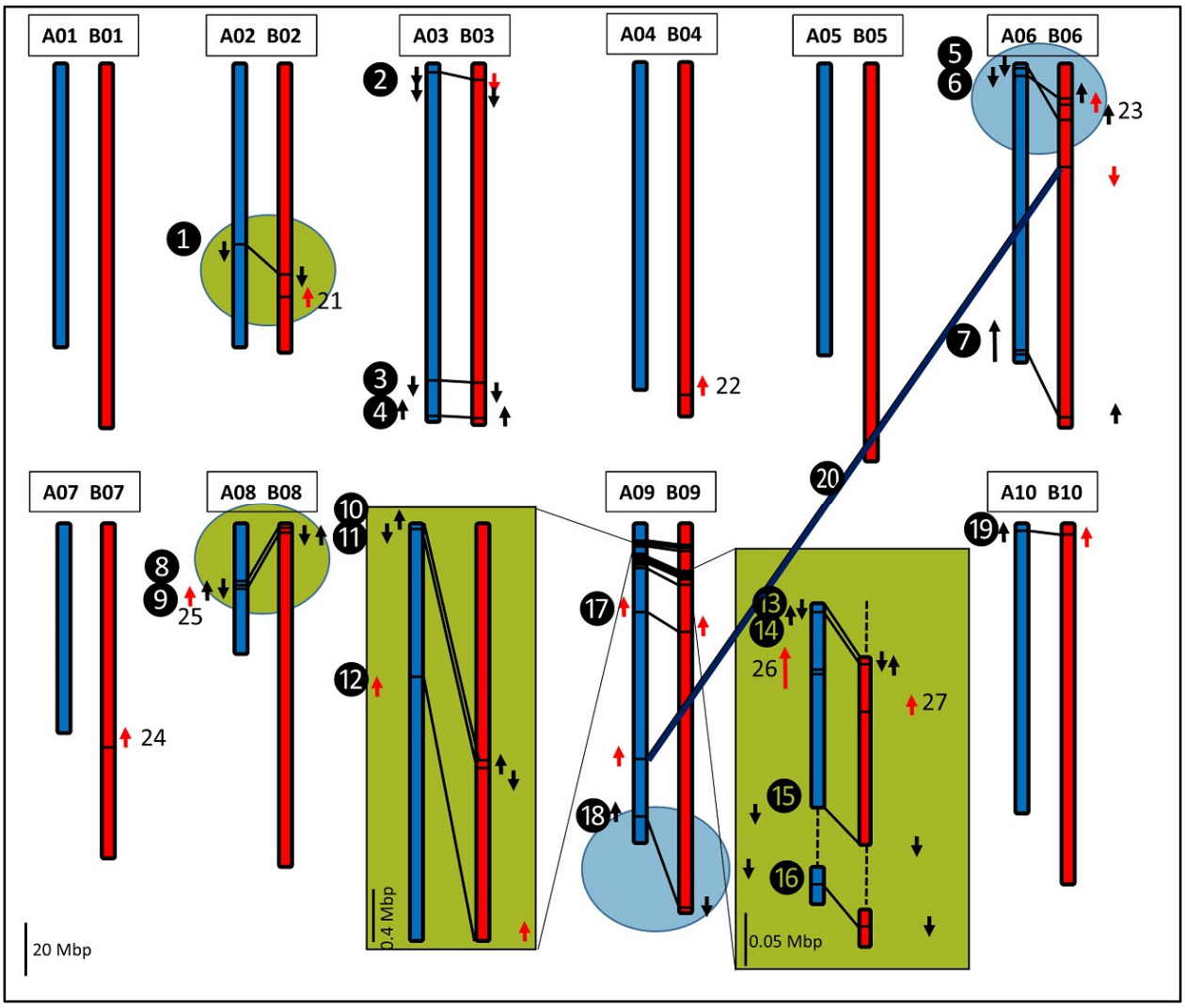


Figure 2.1: Locations and directions peanut LOX genes on a synthetic tetraploid genome based on *A. duranensis* (A-genome) and *A. ipaensis* (B-genome): Up arrow: the gene locates on the forward strand (3' to 5' direction), down arrow: the gene locates on reverse strand (5' to 3' direction), black arrow: genes with complete Lipoxygenase domain, red arrow: genes with incomplete lipoxygenase domain, numbers from 1-20 in black circles: locations of orthologous gene pairs; numbers 21-27: locations of paralogs; longer arrows of orthologous pairs 7 and 26: two incomplete genes are joined in one gene; duplicated arrow of the orthologous pair 2: genes with structures of two LOX genes (every one is a duplicated LOX gene); regions marked with

green: regions with intra-translocation; regions marked with sky blue: regions with inversion; dark blue line of the orthologous pair 20: a region with inter-translocation; lines between two genes: the two genes have a synteny; Mbp: Million base pairs, rectangles: a magnified part of the chromosome.

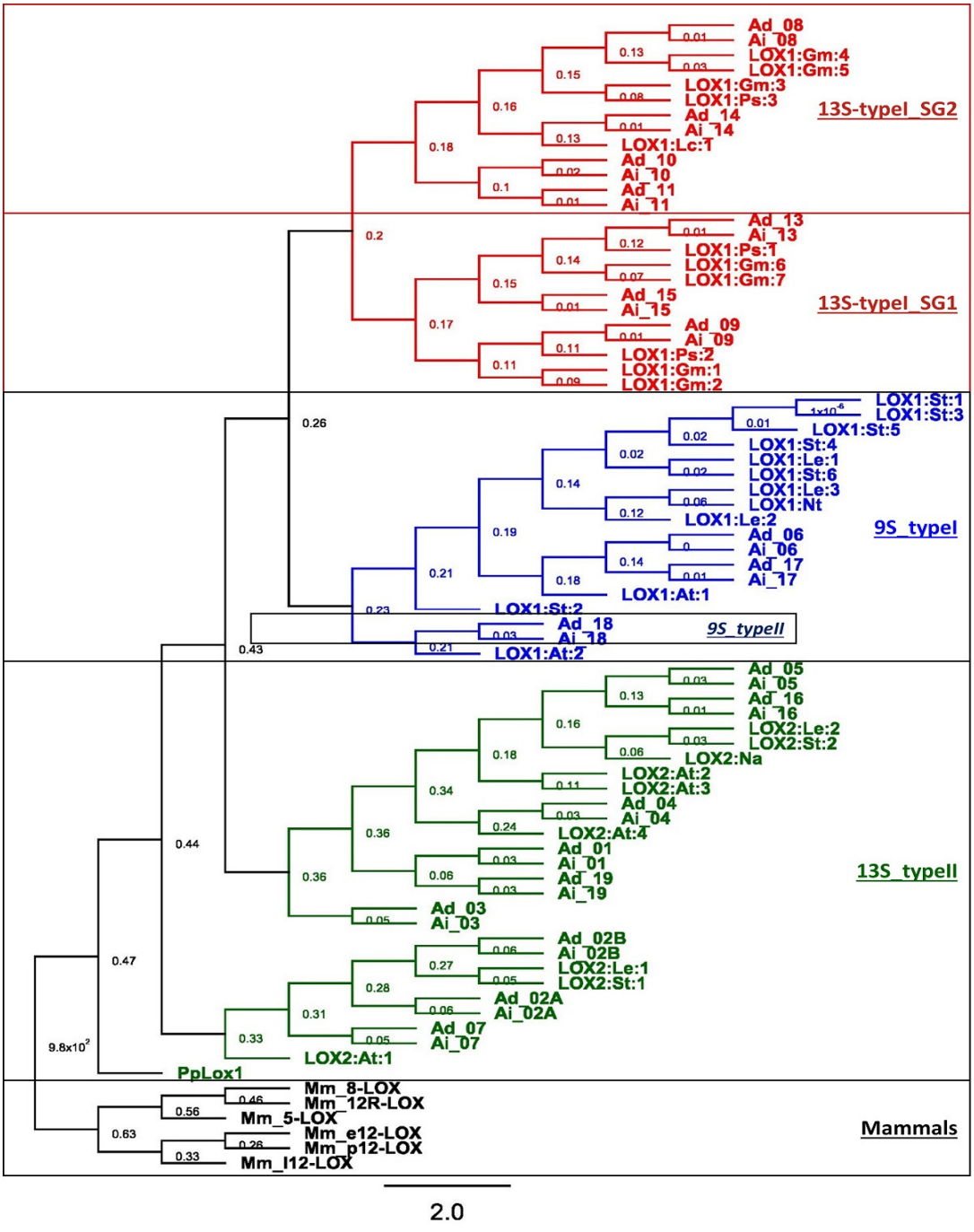


Figure 2.2: Classification of peanut LOX genes based on previously published phylogenetic tree (Andreou and Feussner, 2009): GenBank IDs of the used accessions are presented in supplemental table 2.7.

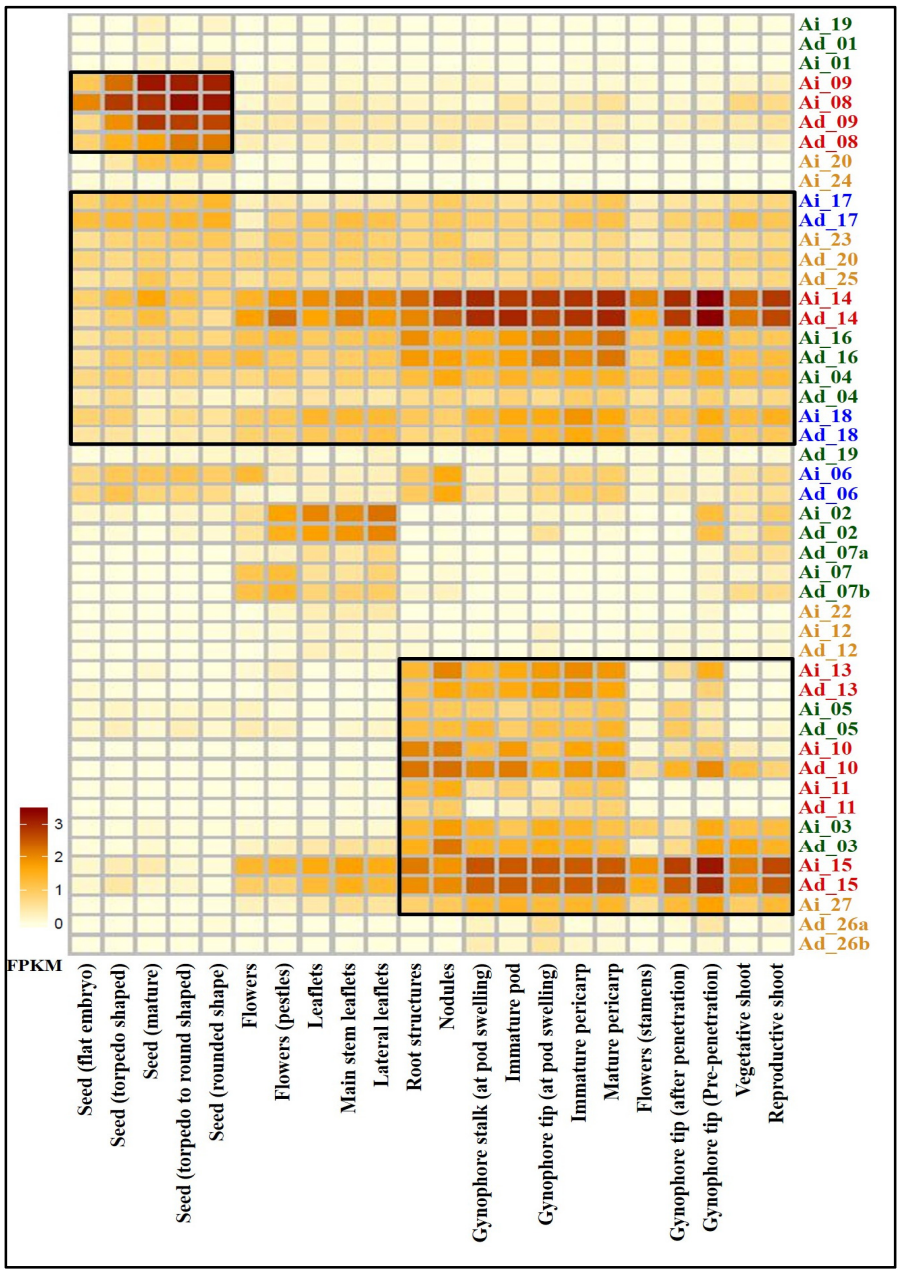


Figure 2.3: RNA-seq differential expression profile of peanut LOX genes: Upper rectangle: EP-I (Expression Pattern 1); middle rectangle: EP-II (Expression Pattern 2); lower rectangle: EP-III (Expression Pattern 3); font colors match the figure 2.2 classification; a description of the tissues is represented in additional files: Table S8.

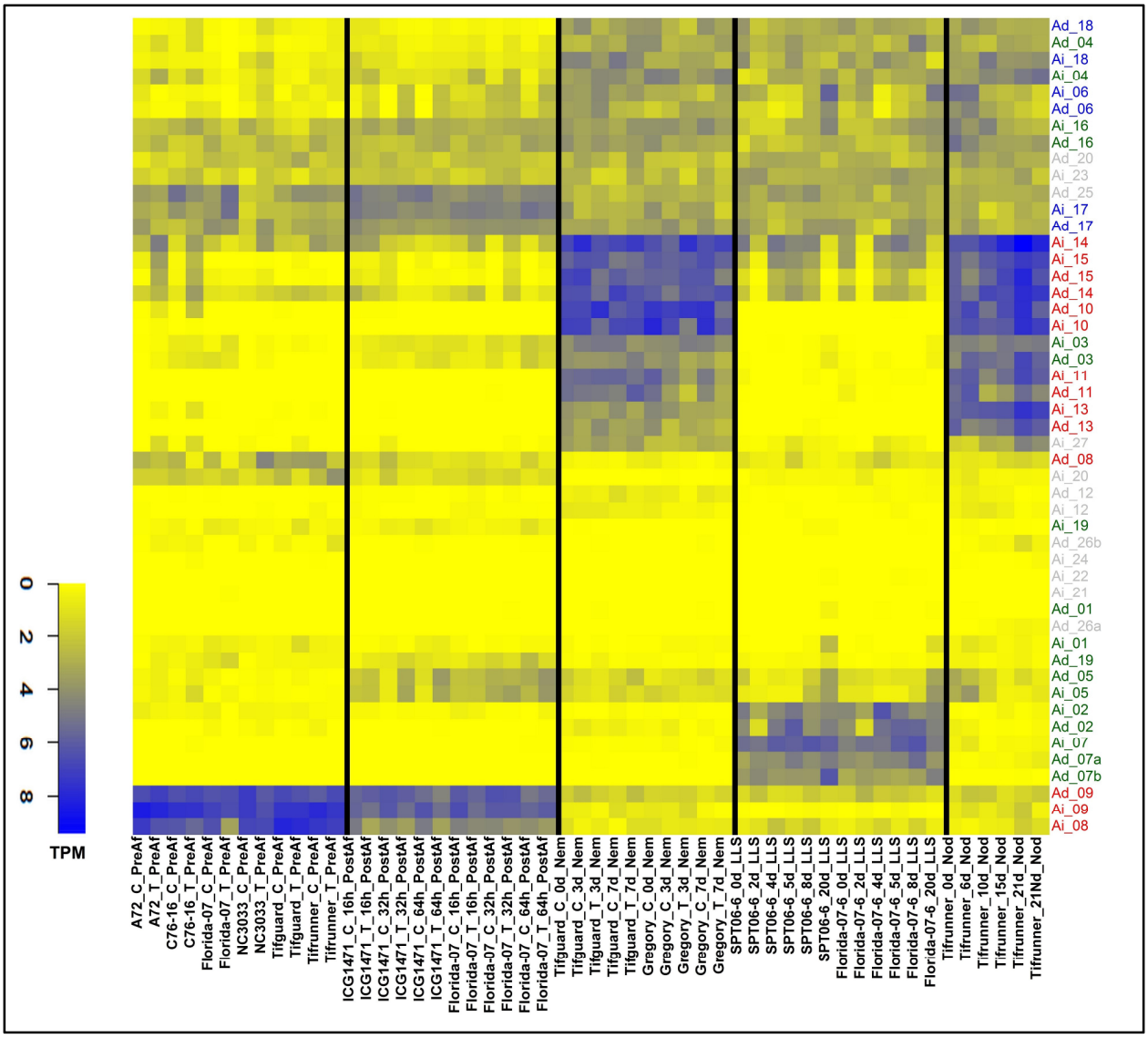


Figure 2.4: RNA-seq profile of peanut interaction with pathogens and nodulation: PreAf: Pre-harvest Aflatoxin; PostAf: Post-harvest Aflatoxin; Nem: Nematode; LLS: Late Leaf Spot; Nod: Nodulation; font colors match the figure 2.2 classification.

References

Andreou A, Brodhun, F, Feussner, I (2009) Biosynthesis of oxylipins in non-mammals. Prog Lipid Res **48:** 148-170

Andreou A, Feussner, I (2009) Lipoxygenases–structure and reaction mechanism. Phytochemistry **70:** 1504-1510

Babitha MP, Prakash, HS, Shetty, HS (2006) Induction of lipoxygenase in downy mildew resistant seedlings of pearl millet in response to inoculation with *Sclerospora graminicola*. International Journal of Agriculture and Biology **8:** 560-564

Babitha MP, Prakash, HS, Shetty, HS (2004) Purification and properties of lipoxygenase induced in downy mildew resistant pearl millet seedlings due to infection with *Sclerospora graminicola*. Plant Science **166:** 31-39

Bell E, Mullet, JE (1993) Characterization of an *Arabidopsis* lipoxygenase gene responsive to methyl jasmonate and wounding. Plant Physiol **103:** 1133-1137

Bertioli DJ, Cannon, SB, Froenicke, L, Huang, G, Farmer, AD, *et al.* (2016) The genome sequences of *Arachis duranensis* and A*rachis ipaensis*, the diploid ancestors of cultivated peanut. Nature genetics

Bueno P, Soto, MJ, Rodríguez-Rosales, MP, Sanjuan, J, Olivares, J, *et al.* (2001) Time-course of lipoxygenase, antioxidant enzyme activities and H2O2 accumulation during the early stages of *Rhizobium*–legume symbiosis. New Phytologist **152:** 91-96

Buranasompob A, Tang, J, Powers, JR, Reyes, J, Clark, S, *et al.* (2007) Lipoxygenase activity in walnuts and almonds. LWT- Food Science and Technology **40:** 893-899

Burow GB, Gardner, HW, Keller, NP (2000) A peanut seed lipoxygenase responsive to *Aspergillus* colonization. Plant Mol Biol **42:** 689-701

Camacho C, Coulouris, G, Avagyan, V, Ma, N, Papadopoulos, J, *et al.* (2009) BLAST+: architecture and applications. BMC Bioinformatics **10:** 421

Chen G, Hackett, R, Walker, D, Taylor, A, Lin, Z, *et al.* (2004) Identification of a specific isoform of tomato lipoxygenase (*TomloxC*) involved in the generation of fatty acid-derived flavor compounds. Plant Physiol **136:** 2641-2651

Choi JJ, Alkharouf, NW, Schneider, KT, Matthews, BF, Frederick, RD (2008) Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. Funct Integr Genomics **8:** 341-359

Clevenger J, Chu, Y, Scheffler, B, Ozias-Akins, P (2016a) A Developmental Transcriptome Map for Allotetraploid *Arachis hypogaea*. Frontiers in Plant Science **7:** 1446

Clevenger J, Marasigan, K, Liakos, V, Sobolev, V, Vellidis, G, *et al.* (2016b) RNA sequencing of contaminated seeds reveals the state of the seed permissive for pre-harvest aflatoxin contamination and points to a potential susceptibility factor. Toxins **8:** 317

Doehlert DC, Wicklow, DT, Gardner, HW (1993) Evidence implicating the lipoxygenases pathway in providing resistance to soybeans against *Aspergillus flavus*. Phytopathology **83:** 473-1477

Doyle JJ, Doyle, JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bul **19:** 11-15

Engeseth N, Klein, B, Warner, K (1987) Lipoxygenase lsoenzymes in soybeans: Effects on crude oil quality. Journal of Food Science **52:** 1015-1019

Ferrie BJ, Beaudoin, N, Burkhart, W, Bowsher, CG, Rothstein, SJ (1994) The cloning of two tomato lipoxygenase genes and their differential expression during fruit ripening. Plant Physiol **106:** 109-118

Frenkel M, Kulheim, C, Jankanpaa, HJ, Skogstrom, O, Dall'osto, L, *et al.* (2009) Improper excess light energy dissipation in *Arabidopsis* results in a metabolic reprogramming. BMC Plant Biol **9:** 12

Fuchsman CA, Rocap, G (2006) Whole-genome reciprocal BLAST analysis reveals that planctomycetes do not share an unusually large number of genes with *Eukarya* and *Archaea*. Appl Environ Microbiol **72:** 6841-6844

Gao X, Brodhagen, M, Isakeit, T, Brown, SH, Gobel, C, *et al.* (2009) Inactivation of the lipoxygenase *ZmLOX3* increases susceptibility of maize to *Aspergillus spp*. Mol Plant Microbe Interact **22:** 222-231

Gao X, Starr, J, Gobel, C, Engelberth, J, Feussner, I, *et al.* (2008) Maize 9-lipoxygenase *ZmLOX3* controls development, root-specific expression of defense genes, and resistance to root-knot nematodes. Mol Plant Microbe Interact **21:** 98-109

Gardner HW (1988) Lipoxygenase pathway in cereals. Advances in cereal science and technology, Chapter8. American Association of Cereal Chemists, Incorporated SI PaUl, Minnesota, USA 161-215

Goff L, Trapnell, C, Kelley, D (2013) CummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. R package version **2:**

Guo Y, Abernathy, B, Zeng, Y, Ozias-Akins, P (2015) TILLING by sequencing to identify induced mutations in stress resistance genes of peanut (*Arachis hypogaea*). BMC genomics **16:** 157

Haas BJ, Papanicolaou, A, Yassour, M, Grabherr, M, Blood, PD, *et al.* (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc **8:** 1494-1512

Hansen J, Garreta, A, Benincasa, M, Fuste, MC, Busquets, M, *et al.* (2013) Bacterial lipoxygenases, a new subfamily of enzymes? A phylogenetic approach. Appl Microbiol Biotechnol **97:** 4737-4747

Heitz T, Bergey, DR, Ryan, CA (1997) A gene encoding a chloroplast-targeted lipoxygenase in tomato leaves is transiently induced by wounding, systemin, and methyl jasmonate. Plant Physiol **114:** 1085-1093

Hirsh AE, Fraser, HB (2001) Protein dispensability and rate of evolution. Nature **411:** 1046-1049

Horn T, Adel, S, Schumann, R, Sur, S, Kakularam, KR, *et al.* (2015) Evolutionary aspects of lipoxygenases and genetic diversity of human leukotriene signaling. Progress in Lipid Research **57:** 13-39

Huang P, Franco, O, Yan, Y, Keller, N, Kolomiets, M (2013) Maize lipoxygenase *LOX2* regulates pathogenesis of mycotoxin-producing *Aspergillus flavus*. APS-MSA Joint meeting, August 10-14, Austin, Texas

Kearse M, Moir, R, Wilson, A, Stones-Havas, S, Cheung, M, *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:** 1647-1649

Kilaru A, Herrfurth, C, Keereetaweep, J, Hornung, E, Venables, BJ, *et al.* (2011) Lipoxygenase-mediated oxidation of polyunsaturated N-acylethanolamines in *Arabidopsis*. J Biol Chem **286:** 15205-15214

Kolomiets MV, Hannapel, DJ, Chen, H, Tymeson, M, Gladon, RJ (2001) Lipoxygenase is involved in the control of potato tuber development. Plant Cell **13:** 613-626

Koski LB, Golding, GB (2001) The closest BLAST hit is often not the nearest neighbor. J Mol Evol **52:** 540-542

Koval SF, Jarrell, KF (1987) Ultrastructure and biochemistry of the cell wall of *Methanococcus voltae*. Journal of Bacteriology **169:** 1298-1306

Kumari AM, Devi, PM, Sucharitha, A (2011) Differential effect of lipoxygenase on aflatoxin production by *Aspergillus spp*. International Journal of Plant Pathology, **2:** 153-164

Liu S, Han, B (2010) Differential expression pattern of an acidic 9/13-lipoxygenase in flower opening and senescence and in leaf response to phloem feeders in the tea plant. BMC Plant Biol **10:** 228

Martins CaO, Sediyama, CS, Oliveira, MGDA, Reis, MS, Rocha, VS, *et al.* (2002) Resistance to stem canker, frogeye leaf spot and powdery mildew of soybean lines lacking lipoxigenases in the seeds. Scientia Agricola **59:** 701-705

Matsuki T, Tazaki, H, Fujimori, T, Hogetsu, T (1992) The influences of jasmonic acid methyl ester on microtubules in potato cells and formation of potato tubers. Bioscience, Biotechnology, and Biochemistry **56:** 1329–1330

Mattila TM, Bechsgaard, JS, Hansen, TT, Schierup, MH, Bilde, T (2012) Orthologous genes identified by transcriptome sequencing in the spider genus *Stegodyphus*. BMC genomics **13:** 70

Melan MA, Enriquez, A, Peterman, TK (1994) The *LOX1* gene of *Arabidopsis* is temporally and spatially regulated in germinating seedlings. Plant Physiol **105:** 385-393

Mellon JK, Cotty, PJ (2002) No effect of soybean lipoxygenase on aflatoxin production in *Aspergillus flavus*-inoculated seeds. J Food Prot **65:** 1984-1987

Minor W, Steczko, J, Bolin, JT, Otwinowski, Z, Axelrod, B (1993) Crystallographic determination of the active site iron and its ligands in soybean lipoxygenase L-1. Biochemistry **32:** 6320-6323

Minor W, Steczko, J, Stec, B, Otwinowski, Z, Bolin, JT, *et al.* (1996) Crystal structure of soybean lipoxygenase L-1 at 1.4 A resolution. Biochemistry **35:** 10687-10701

Muller V, Ame, MV, Carrari, V, Gieco, J, Asis, R (2014) Lipoxygenase activation in peanut seed cultivars resistant and susceptible to *Aspergillus parasiticus* colonization. Phytopathology **104:** 1340-1348

Musa ÖM (2010) Some nutritional characteristics of kernel and oil of peanut (Arachis hypogaea L.). Journal of Oleo Science **59:** 1-5

O'keefe S, Wiley, V, Knauft, D (1993) Comparison of oxidative stability of high-and normal-oleic peanut oils. Journal of the American Oil Chemists' Society **70:** 489-492

Ohta H, Shida, K, Peng, YL, Furusawa, I, Shishiyama, J, *et al.* (1991) A lipoxygenase pathway is activated in rice after infection with the rice blast fungus *Magnaporthe grisea*. Plant Physiol **97:** 94-98

Ozalvo R, Cabrera, J, Escobar, C, Christensen, SA, Borrego, EJ, *et al.* (2014) Two closely related members of *Arabidopsis* 13-lipoxygenases (13-LOXs), *LOX3* and *LOX4*, reveal distinct functions in response to plant-parasitic nematode infection. Mol Plant Pathol **15:** 319-332

Porta H, Rocha-Sosa, M (2000) A *Phaseolus vulgaris* lipoxygenase gene expressed in nodules and in *Rhizobium* tropici inoculated roots. Biochim Biophys Acta **1517:** 139-142

Robinson DS, Wu, Z, Domoney, C, Casey, R (1995) Lipoxygenases and the quality of foods. Food Chemistry **54:** 33-43

Rodriguez-Concepcion M, Beltran, JP (1995) Repression of the pea lipoxygenase gene *LOXg* is associated with carpel development. Plant Mol Biol **27:** 887-899

Rodriguez-Saona LE, Barrett, DM, Selivonchick, DP (1995) Peroxidase and lipoxygenase influence on stability of polyunsaturated fatty acids in sweet corn (*Zea mays* L.). Journal of Food Science **6:** 1041-1044

Royo J, Vancanneyt, G, Perez, AG, Sanz, C, Stormann, K, *et al.* (1996) Characterization of three potato lipoxygenases with distinct enzymatic activities and different organ-specific and wound-regulated expression patterns. J Biol Chem **271:** 21012-21019

Sambrook J, Fritsch, E, Maniatis, T (2001) Molecular cloning: a laboratory manual, third edition, Cold Spring Laboratory. Cold Spring Harbor, NY, USA

Saravitz DM, Siedow, JN (1995) The lipoxygenase isozymes in soybean [*Glycine max* (L.) Merr.] leaves (changes during leaf development, after wounding, and following reproductive sink removal). Plant Physiol **107:** 535-543

Senger T, Wichard, T, Kunze, S, Gobel, C, Lerchl, J, *et al.* (2005) A multifunctional lipoxygenase with fatty acid hydroperoxide cleaving activity from the moss *Physcomitrella patens*. J Biol Chem **280:** 7588-7596

Shen J, Tieman, D, Jones, JB, Taylor, MG, Schmelz, E, *et al.* (2014) A 13-lipoxygenase, *TomloxC*, is essential for synthesis of C5 flavour volatiles in tomato. J Exp Bot **65:** 419-428

Shirasawa K, Takeuchi, Y, Ebitani, T, Suzuki, Y (2008) Identification of gene for rice (*Oryza sativa*) seed lipoxygenase-3 involved in the generation of stale flavor and development of SNP markers for lipoxygenase-3 deficiency. Breeding Science **58:** 169-176

Song H, Wang, P, Li, C, Han, S, Lopez-Baltazar, J, *et al.* (2016) Identification of lipoxygenase (LOX) genes from legumes and their responses in wild type and cultivated peanut upon *Aspergillus flavus* infection. Scientific reports **6:**

St. Angelo AJ, Kuck, JC, Ory, RL (1979) Role of lipoxygenase and lipid oxidation in quality of oilseeds. Journal of Agricultural and Food Chemistry **27:** 229-234

Steczko J, Donoho, GP, Clemens, JC, Dixon, JE, Axelrod, B (1992) Conserved histidine residues in soybean lipoxygenase: functional consequences of their replacement. Biochemistry **31:** 4053-4057

Trapnell C, Pachter, L, Salzberg, SL (2009) TopHat: discovering splice junctions with RNA-seq. Bioinformatics **25:** 1105-1111

Trapnell C, Roberts, A, Goff, L, Pertea, G, Kim, D, *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc **7:** 562-578

Tsitsigiannis DI, Kunze, S, Willis, DK, Feussner, I, Keller, NP (2005) *Aspergillus* infection inhibits the expression of peanut 13S-HPODE-forming seed lipoxygenases. Mol Plant Microbe Interact **18:** 1081-1089

Vogt J, Schiller, D, Ulrich, D, Schwab, W, Dunemann, F (2013) Identification of lipoxygenase (LOX) genes putatively involved in fruit flavour formation in apple (*Malus* x *domestica*). Tree Genetics and Genomes **9:** 1493-1511

Yan L, Zhai, Q, Wei, J, Li, S, Wang, B, *et al.* (2013) Role of tomato lipoxygenase D in wound-induced jasmonate biosynthesis and plant immunity to insect herbivores. PLoS Genet **9:** e1003964

Youn B, Sellhorn, GE, Mirchel, RJ, Gaffney, BJ, Grimes, HD, *et al.* (2006) Crystal structures of vegetative soybean lipoxygenase VLX-B and VLX-D, and comparisons with seed isoforms *LOX-1* and *LOX-3*. Proteins **65:** 1008-1020

Zeringue Jr H (1996) Possible involvement of lipoxygenase in a defense response in aflatoxigenic *Aspergillus*–cotton plant interactions. Canadian journal of botany **74:** 98-102

Zhang Y, Xuan, H, Zhu, X, Yang, J, He, C, *et al.* (2009) Effect of lipoxygenase activity and red seed coat on rice bran deterioration. Journal of the Science of Food and Agriculture, **89:** 1904-1908

Appendix 2.A

Supplemental Materials

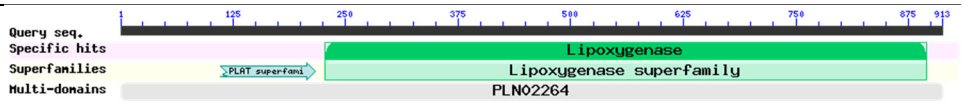
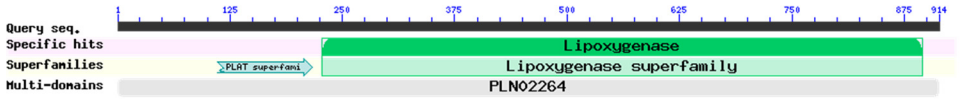
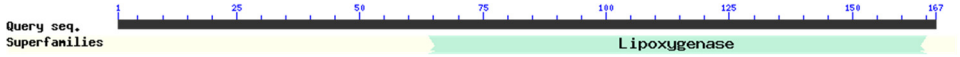
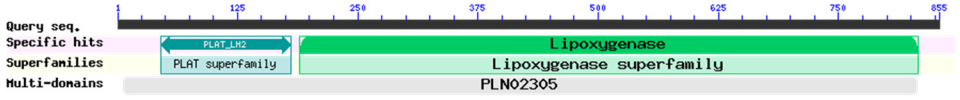
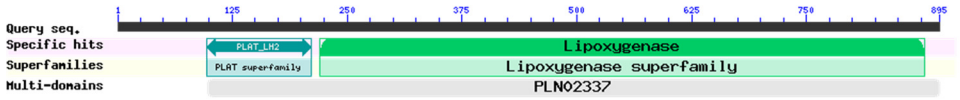
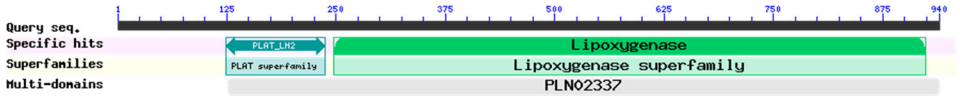
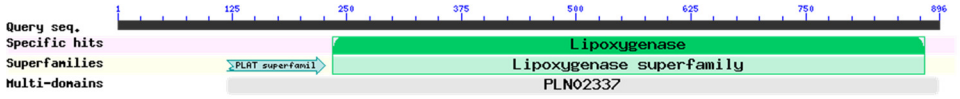
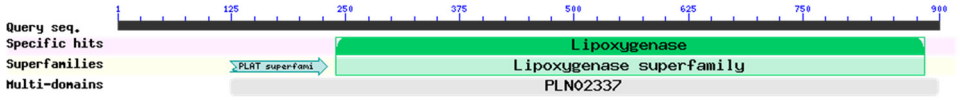
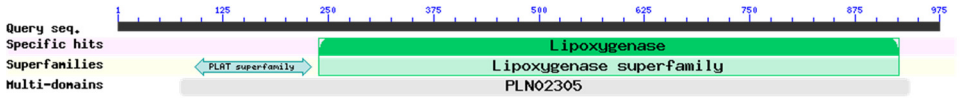
Table 2.S1: Orthologous pairs of peanut LOX genes depending on A. duranensis and A. ipaensis genomes, which have identified by reciprocal BLAST and confirmed by dot plots.

Locus	Chromosome	A-genome		B-genome	
		PeanutBase ID	Shortcut ID	PeanutBase ID	Shortcut ID
1	02	Aradu.XZG8N	Ad_01	Araip.849ER	Ai_01
2	03	Aradu.Q5K4W	Ad_02	Araip.MN7KE	Ai_02
3	03	Aradu.W07KG	Ad_03	Araip.NWR3L	Ai_03
4	03	Aradu.C88Z1	Ad_04	Araip.X1W86	Ai_04
5	06	Aradu.8D3SW	Ad_05	Araip.2KP3T	Ai_05
6	06	Aradu.G99LQ	Ad_06	Araip.W6TLM	Ai_06
7	06	Aradu.U67PQ	Ad_07	Araip.HGI2J	Ai_07
8	08	Aradu.AC956	Ad_08	Araip.DH1Z0	Ai_08
9	08	Aradu.WX5KP	Ad_09	Araip.E99Y9	Ai_09
10	09	Aradu.SK1BS	Ad_10	Araip.5F6MD	Ai_10
11	09	Aradu.AE16G	Ad_11	Araip.T64GQ	Ai_11
12	09	Aradu.GJ1CE	Ad_12	Araip.7V9BH	Ai_12
13	09	Aradu.FM0YX	Ad_13	Araip.GV48H	Ai_13
14	09	Aradu.TJL9X	Ad_14	Araip.K56RN	Ai_14
15	09	Aradu.951UC	Ad_15	Araip.Q8LFT	Ai_15
16	09	Aradu.LNK8S	Ad_16	Araip.VN0A4	Ai_16
17	09	Aradu.C3RV0	Ad_17	Araip.3GK67	Ai_17
18	09	Aradu.KZX2M	Ad_18	Araip.D6PZJ	Ai_18
19	10	Aradu.AS232	Ad_19	Araip.Q7EYZ	Ai_19
20	09/06	Aradu.5E1NU	Ad_20	Araip.G6789	Ai_20

Table 2.S2: Paralogous peanut LOX genes, which did not meet the reciprocal BLAST criteria and have a synteny with other genes that they probably duplicated from.

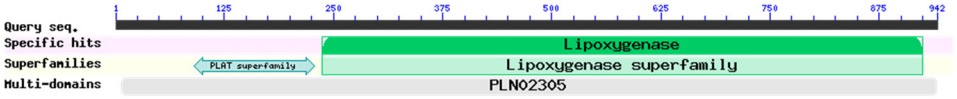
Chromosome	PeanutBase ID of paralogous LOX genes	PeanutBase ID of the best hits (paralogs)	PeanutBase ID of the best hits of the paralogs
A06	Aradu.289WG	Araip.HGI2J	Aradu.U67PQ
A08	Aradu.KXZ9V	Araip.E99Y9	Aradu.WX5KP
A09	Aradu.2ZL37	Araip.LB22X	Aradu.951UC
A09	Aradu.S1X34	Araip.Q8LFT	Aradu.951UC
B02	Araip.HUQ4W	Aradu.951UC	Araip.Q8LFT
B04	Araip.J97JQ	Aradu.Q5K4W	Araip.MN7KE
B06	Araip.P423S	Aradu.AC956	Araip.DH1Z0
B07	Araip.8Z22U	Aradu.2E0TL	-
B09	Araip.LB22X	Aradu.951UC	Araip.Q8LFT

Table 2.S3: BLASTp analysis of every peanut LOX translated gene.

ID	Protein structure and BLASTp best hits
01_A02	 Lipoxygenase 3, chloroplastic [Glycine soja] gb|KHN08522.1| Length: 833
01_B02	 Lipoxygenase 3, chloroplastic [Glycine soja] Sequence ID: gb|KHN08522.1|Length: 833
Araip.HUQ4W	 Lipoxygenase [Macrotyloma uniflorum] Sequence ID: gb|AIL90390.1|Length: 867
02_A03	 Linoleate 13S-lipoxygenase 2-1, related protein [Medicago truncatula] Sequence ID: gb|KEH30273.1|Length: 906
02A_B03	 linoleate 13S-lipoxygenase 2-1, related protein [Medicago truncatula] Sequence ID: gb|KEH30273.1|Length: 906
02A_A03	 Linoleate 13S-lipoxygenase 2-1 [Morus notabilis] Sequence ID: ref|XP_010087356.1|
02B_B03	 Linoleate 13S-lipoxygenase 2-1 [Morus notabilis] Sequence ID: ref|XP_010087356.1|Length: 912
03_A03	 Linoleate 13S-lipoxygenase 3-1, chloroplastic [Glycine soja] Sequence ID: gb|KHN07257.1|Length: 910
03_B03	 Linoleate 13S-lipoxygenase 3-1, chloroplastic [Glycine soja] Sequence ID: gb|KHN07257.1|Length: 910
04_A03	

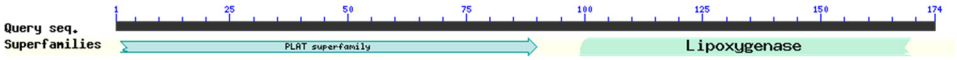
Lipoxygenase [Theobroma cacao]
Sequence ID: ref|XP_007045017.1|Length: 914

04_B03



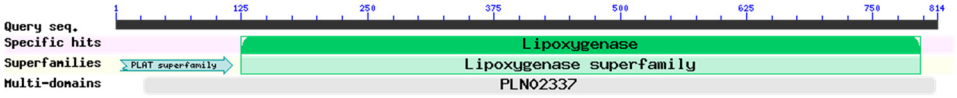
Lipoxygenase [Theobroma cacao]
Sequence ID: ref|XP_007045017.1|Length: 914

Araip.J97JQ



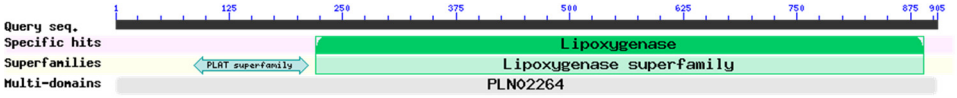
Linoleate 13S-lipoxygenase 2-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN21876.1|Length: 653

05_A06



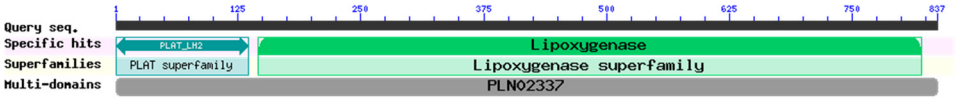
Lipoxygenase [Sesbania rostrata]
Sequence ID: emb|CAC43237.1|Length: 922

05_B06



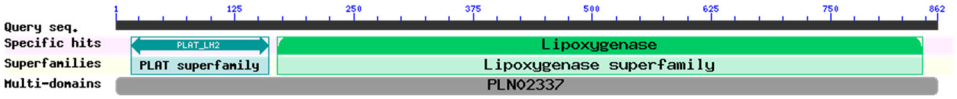
Lipoxygenase [Sesbania rostrata]
Sequence ID: emb|CAC43237.1|Length: 922

06_A06



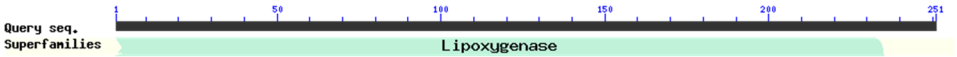
Lipoxygenase [Corylus avellana]
Sequence ID: emb|CAD10740.1|Length: 873

06_B06



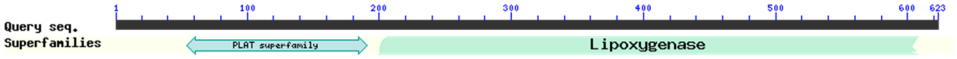
Lipoxygenase [Corylus avellana]
Sequence ID: emb|CAD10740.1|Length: 873

07_A06



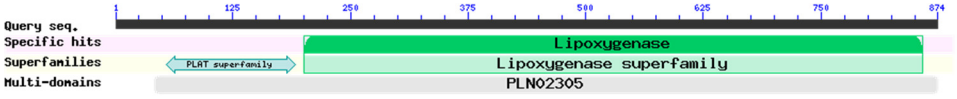
Linoleate 13S-lipoxygenase 2-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN46337.1|Length: 859

Aradu.289WG



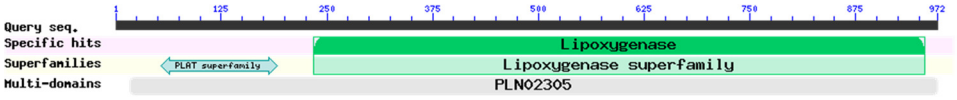
Linoleate 13S-lipoxygenase 2-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN21708.1|Length: 854

07_A06 &
Aradu.289WG



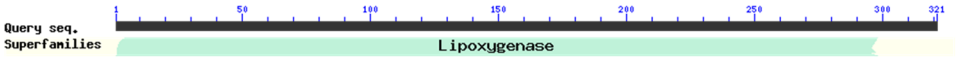
Linoleate 13S-lipoxygenase 2-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN21708.1|Length: 854

07-B06



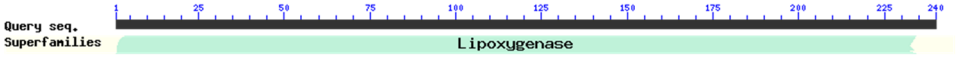
Linoleate 13S-lipoxygenase 2-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN21708.1|Length: 85

Araip.P423S



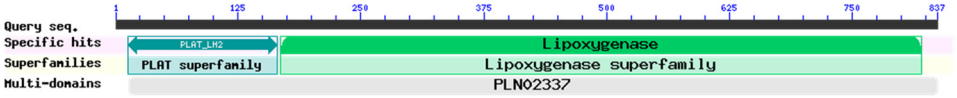
Linoleate 9S-lipoxygenase [Phaseolus vulgaris]
Sequence ID: sp|P27481.1|LOXB_PHAVULength: 741

Araip.8Z22U



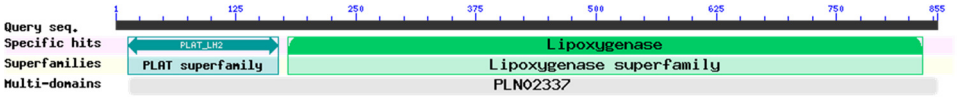
Linoleate 9S-lipoxygenase-4 [Glycine soja]
Sequence ID: gb|KHN32712.1|Length: 720

08_A08



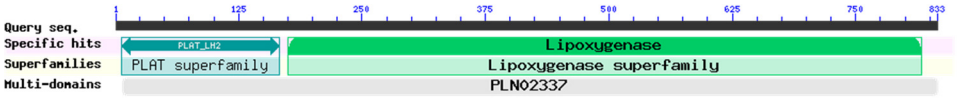
Linoleate 9S-lipoxygenase-4 [Glycine max]
Sequence ID: sp|P38417.1|LOX4_SOYBNLength: 853

08_B08



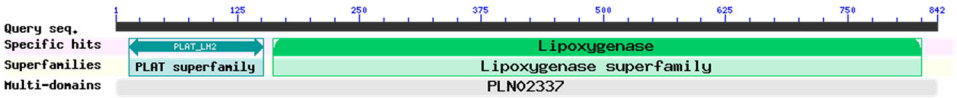
Linoleate 9S-lipoxygenase-4 [Glycine max]
Sequence ID: sp|P38417.1|LOX4_SOYBNLength: 853

09_A08



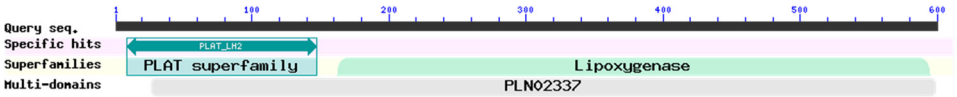
13-lipoxygenase [Arachis hypogaea]
Sequence ID: gb|AAY87056.1|Length: 863

09_B08



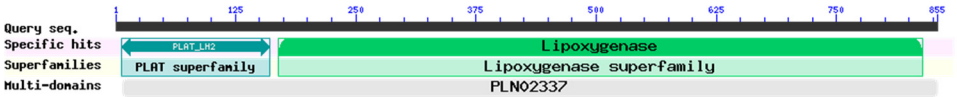
13-lipoxygenase [Arachis hypogaea]
Sequence ID: gb|AAY87057.1|Length: 863

Aradu.KXZ9V



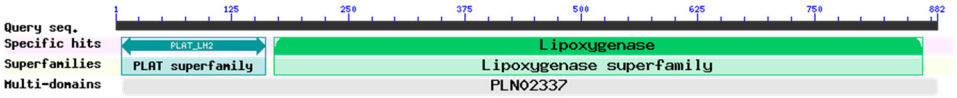
13-lipoxygenase [Arachis hypogaea]
Sequence ID: gb|AAY87056.1|Length: 863

10_A09



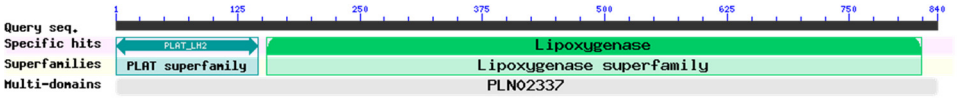
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

10_B09



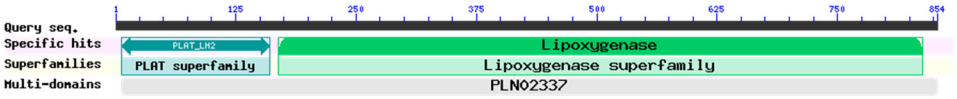
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

11_A09



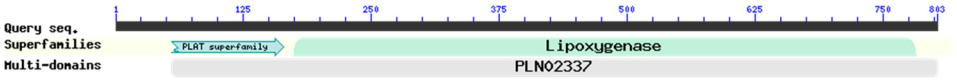
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

11_B09



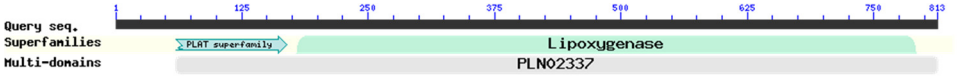
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

12_A09



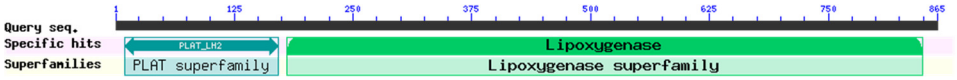
Linoleate 9S-lipoxygenase 5, chloroplastic [Glycine soja]
Sequence ID: gb|KHN32739.1|Length: 744

12_B09



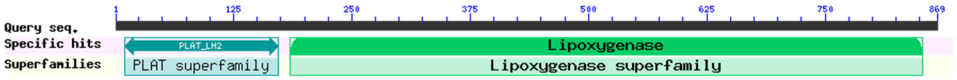
Linoleate 9S-lipoxygenase 5, chloroplastic [Glycine soja]
Sequence ID: gb|KHN32739.1|Length: 744

13_A09



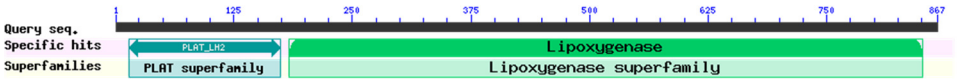
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

13_B09



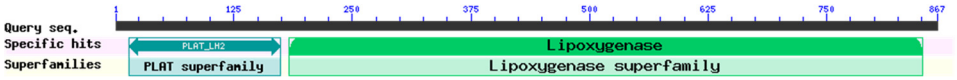
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

14_A09



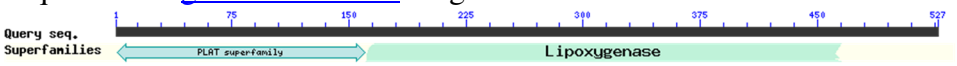
Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

14_B09



Seed linoleate 9S-lipoxygenase-3 [Glycine soja]
Sequence ID: gb|KHN32710.1|Length: 857

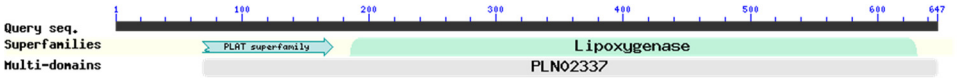
Aradu.2ZL37



Aradu.S1X34

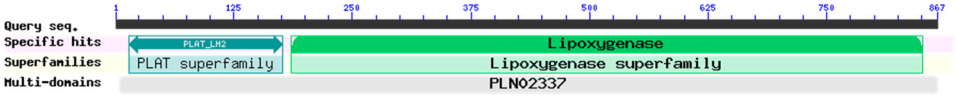
lipoxygenase-9 [Glycine max]
Sequence ID: ref|NP_001237323.1|Length: 865

Araip.LB22X



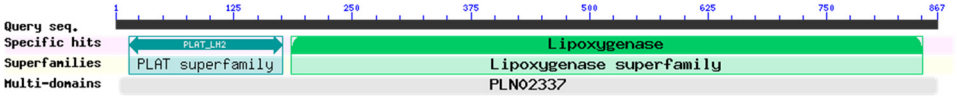
Linoleate 9S-lipoxygenase 1 [Glycine soja]
Sequence ID: gb|KHN01371.1|Length: 846

15_A09



lipoxygenase-9 [Glycine max]
Sequence ID: ref|NP_001237323.1|Length: 865

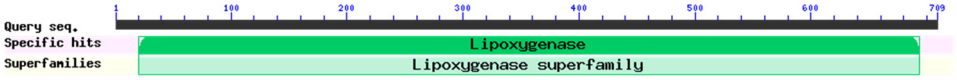
15_B09



lipoxygenase-9 [Glycine max]

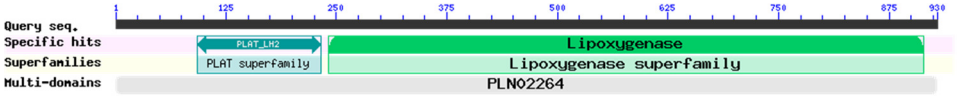
Sequence ID: ref|NP_001237323.1|Length: 865

16_A09



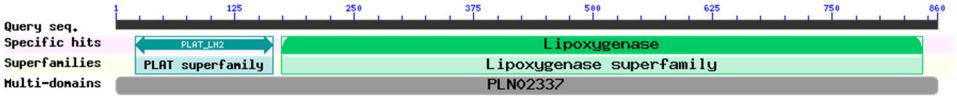
Linoleate 13S-lipoxygenase 3-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN39622.1|Length: 922

16_B09



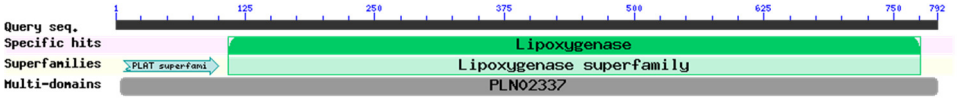
Linoleate 13S-lipoxygenase 3-1, chloroplastic [Glycine soja]
Sequence ID: gb|KHN39622.1|Length: 922

17_A09



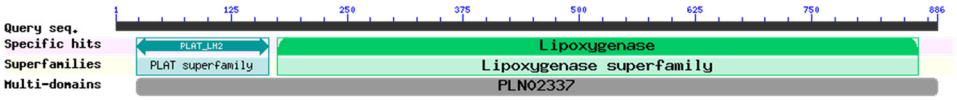
putative lipoxygenase-9 [Arachis hypogaea]
Sequence ID: gb|AJT49215.1|Length: 860

17_B09



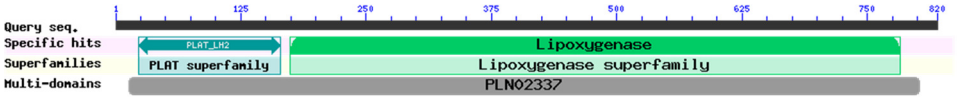
putative lipoxygenase-9 [Arachis hypogaea]
Sequence ID: gb|AJT49215.1|Length: 860

18_A09



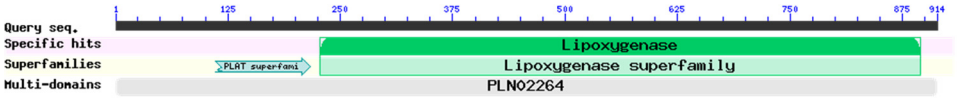
Linoleate 9S-lipoxygenase 5, chloroplastic [Glycine soja]
Sequence ID: gb|KHN02707.1|Length: 850

18_B09



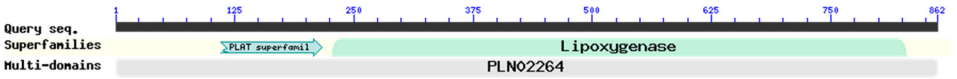
Linoleate 9S-lipoxygenase 5, chloroplastic [Glycine soja]
Sequence ID: gb|KHN02707.1|Length: 850

19_A10



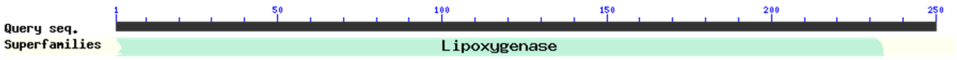
Lipoxygenase 3, chloroplastic [Glycine soja]
Sequence ID: gb|KHN08522.1|Length: 833

19_B10



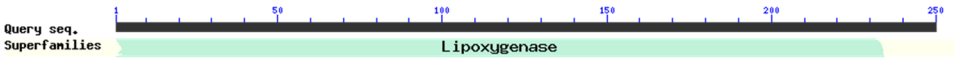
Lipoxygenase 3, chloroplastic [Glycine soja]
Sequence ID: gb|KHN08522.1|Length: 833

20_A09



13-lipoxygenase [Arachis hypogaea]
Sequence ID: gb|AAY87056.1|Length: 863

20_B06



13-lipoxygenase [Arachis hypogaea]
Sequence ID: gb|AAY87056.1|Length: 863

Table 2.S4: Published peanut LOX genes.

Name	Gene bank ID
pnLOX1	AAF60270.1
pnLOX2	AAY87056.1
pnLOX3	AAY87057.1
pnLOX4	EZ722311.1
pnLOX5	JR564445.1
pnLOX6_c05	AJT49213
pnLOX6_e10	AJT49214
pnLOX7_g02	AJT49215
pnLOX8_h07	AJT49216

Table 2.S5: LOX gene number in some plants based on NCBI search and SoyBase annotation file.

Species	No. of LOXs
Arabidopsis	6
Rice	14
Apple	36
Mei	18
Peach	16
Strawberry	14
Grape	12
Pear	23
Barely	3
Wheat	9
Medicago truncatula	32
soybean	44
peanut	8

Table 2.S6: Soybean LOX genes, which contained Lipoxygenase domain, extracted from the annotation file of SoyBase.

Feature	Chromosome	Start	End	Assembly
Glyma.03g2373	Gm03	43723483	43730370	Glyma2.0
Glyma.04g1055	Gm04	10440510	10445278	Glyma2.0
Glyma.04g1059	Gm04	10789400	10791837	Glyma2.0
Glyma.05g0986	Gm05	26207237	26209499	Glyma2.0
Glyma.07g0067	Gm07	491733	496064	Glyma2.0
Glyma.07g0069	Gm07	503752	509348	Glyma2.0
Glyma.07g0070	Gm07	509819	514700	Glyma2.0
Glyma.07g0071	Gm07	529398	533680	Glyma2.0
Glyma.07g0348	Gm07	2763836	2770113	Glyma2.0
Glyma.07g0349	Gm07	2782231	2787974	Glyma2.0
Glyma.07g0399	Gm07	3288274	3294751	Glyma2.0
Glyma.07g1778	Gm07	33819873	33820090	Glyma2.0
Glyma.07g1968	Gm07	36509005	36518982	Glyma2.0
Glyma.08g1891	Gm08	15163027	15164686	Glyma2.0
Glyma.08g1892	Gm08	15172904	15178499	Glyma2.0
Glyma.08g1893	Gm08	15186020	15191205	Glyma2.0
Glyma.08g1894	Gm08	15193038	15201472	Glyma2.0
Glyma.08g1895	Gm08	15206364	15212242	Glyma2.0
Glyma.08g1896	Gm08	15235197	15239971	Glyma2.0
Glyma.08g1897	Gm08	15242887	15243655	Glyma2.0
Glyma.08g1898	Gm08	15249052	15254212	Glyma2.0
Glyma.10g1539	Gm10	38898358	38904608	Glyma2.0
Glyma.11g1302	Gm11	9903923	9913136	Glyma2.0
Glyma.11g1303	Gm11	9925260	9936297	Glyma2.0
Glyma.12g0547	Gm12	3949453	3957386	Glyma2.0
Glyma.13g0303	Gm13	9773782	9780355	Glyma2.0
Glyma.13g0759	Gm13	17965521	17975820	Glyma2.0
Glyma.13g2390	Gm13	34926963	34931964	Glyma2.0
Glyma.13g3475	Gm13	43761727	43766023	Glyma2.0
Glyma.13g3476	Gm13	43769021	43773290	Glyma2.0
Glyma.13g3477	Gm13	43773475	43780320	Glyma2.0
Glyma.13g3478	Gm13	43797692	43803266	Glyma2.0
Glyma.14g1735	Gm14	42911150	42913226	Glyma2.0
Glyma.15g0263	Gm15	2123754	2128104	Glyma2.0
Glyma.15g0264	Gm15	2130531	2134563	Glyma2.0
Glyma.15g0265	Gm15	2142191	2147489	Glyma2.0
Glyma.15g2333	Gm15	43893441	43895221	Glyma2.0
Glyma.16g0087	Gm16	735622	742337	Glyma2.0
Glyma.16g0962	Gm16	18245957	18246472	Glyma2.0
Glyma.19g2633	Gm19	50591809	50596423	Glyma2.0
Glyma.20g0537	Gm20	12341598	12348808	Glyma2.0
Glyma.20g0540	Gm20	12411805	12423973	Glyma2.0
Glyma.20g0541	Gm20	12486755	12493110	Glyma2.0
Glyma.20g2344	Gm20	46695256	46696124	Glyma2.0

Table 2.S7: Gene bank ID of accessions that used in the LOX gene classifications.

Name	Genbank ID
LOX1:At:1	Q06327
LOX1:At:2	CAC19365
LOX1:Gm:1	AAA33986
LOX1:Gm:2	AAA33987
LOX1:Gm:3	CAA31664
LOX1:Gm:4	BAA03101
LOX1:Gm:5	AAB67732
LOX1:Gm:6	AAA96817
LOX1:Gm:7	S13381
LOX1:Lc:1	CAA50483
LOX1:Le:1	P38415
LOX1:Le:2	P38416
LOX1:Le:3	AAG21691
LOX1:Nt	S57964
LOX1:Ps:1	AAB71759
LOX1:Ps:2	CAA55318
LOX1:Ps:3	CAA55319
LOX1:St:1	S44940
LOX1:St:2	AAD09202
LOX1:St:3	P37831
LOX1:St:4	CAA64766
LOX1:St:5	CAA64765
LOX1:St:6	AAB67860
LOX2:At:1	P38418
LOX2:At:2	AAF79461
LOX2:At:3	CAC19364
LOX2:At:4	AAB65766
LOX2:Le:1	AAB65767
LOX2:Le:2	AAB65767
LOX2:Na	AAP83138
LOX2:St:1	CAA65268
LOX2:St:2	CAA65269
PpLOX1	CAE47464
Mm_5-LOX	AAC37673
Mm_8-LOX	CAA75003
Mm_12R-LOX	CAA74714
Mm_e12-LOX	AAA20659
Mm_l12-LOX	AAA20658
Mm_p12-LOX	CAA67625

Table 2.S8: The description of tissues that were used in RNA-seq differential expression analysis.

Tissue	Stage
Leaf	leaflets partially open
Leaf	leaflets partially open, from mainstem
Leaf	leaflets partially open, from laterals
Vegetative shoot	first flower, from mainstem
Reproductive shoot	first flower, from laterals
Root structures	10 d post-emergence
Nodules	25 d post-emergence
Flowers	Pedals,keel, hypanthium sepals
Flower	Fully open, morning of anthesis; pistles
Flower	Fully open, morning of anthesis; stamens
Gynophore tip	elongating peg prior to soil penetration
Gynophore tip	elongating peg after 24 h soil penetration
Gynophore tip	At pod swelling (Pattee stage 1)
Gynophore “stalk”	At pod swelling (Pattee stage 1)
Pod	Pericarp very watery, embryo very small
Pericarp	Pericarp soft, not as watery
Seed	embryo flat
Pericarp	beginning to show cracks or cottony
Seed	Torpedo shaped
Seed	Torpedo to round shaped
Seed	Round shaped
Seed	seed coat beginning to dry out

Table 2.S9: BLASTn search using probe sequences against a library of all annotated transcripts of *A. duranensis* and *A. ipaensis.*

The query sequence id	Subject	% id	alignment length	mis-matches	gap openings	q.start	q.end	s.start	s.end	e-value	bit score
P_01	Araip.DH1Z0.1	100	296	0	0	1	296	2279	2574	5.00E-167	587
P_01	Aradu.AC956.1	99.32	296	2	0	1	296	2159	2454	3.00E-162	571
P_02	Araip.E99Y9.1	100	316	0	0	1	316	2389	2704	6.00E-179	626
P_02	Aradu.WX5KP.1	99.05	316	3	0	1	316	2320	2635	9.00E-172	603
P_02	Araip.G6789.1	98.42	316	5	0	1	316	1883	2198	6.00E-167	587
P_02	Aradu.KXZ9V.1	98.42	316	5	0	1	316	3238	3553	6.00E-167	587
P_02	Aradu.5E1NU.1	97.15	316	9	0	1	316	1074	1389	2.00E-157	555
P_03a	Araip.W6TLM.1	99.69	321	1	0	1	321	2259	2579	2.00E-179	628
P_03a	Aradu.G99LQ.1	99.69	321	1	0	1	321	2284	2604	2.00E-179	628
P_03b	Araip.W6TLM.1	100	333	0	0	1	333	446	778	0	660
P_03b	Aradu.G99LQ.1	99.7	333	1	0	1	333	471	803	0	652
P_04_	Aradu.C3RV0.1	100	332	0	0	1	332	2248	2579	0	658
P_04	Araip.3GK67.1	98.19	331	6	0	1	331	1977	2307	2.00E-173	609
P_05	Araip.K56RN.1	99.69	321	1	0	1	321	1560	1880	2.00E-179	628
P_05	Aradu.TJL9X.1	99.07	321	3	0	1	321	1550	1870	1.00E-174	613
P_06	Araip.Q8LFT.1	99.66	293	1	0	1	293	2431	2723	8.00E-163	573
P_06	Aradu.951UC.1	99.32	293	2	0	1	293	2439	2731	2.00E-160	565

Table 2.S10: Evidence of orthologs and paralogs of LOX genes

Locus	Chromosome	A-genome	B-genome	Reciprocal	Full	Transcript	flanking
1	02	Aradu.XZG8N	Araip.849ER	1	1	1	0
2	03	Aradu.Q5K4W	Araip.MN7KE	1	1	1	2
3	03	Aradu.W07KG	Araip.NWR3L	1	1	1	2
4	03	Aradu.C88Z1	Araip.X1W86	1	1	1	2
5	06	Aradu.8D3SW	Araip.2KP3T	1	0	1	2
6	06	Aradu.G99LQ	Araip.W6TLM	1	0	1	2
7	06	Aradu.U67PQ	Araip.HGI2J	1	1	1	2
8	08	Aradu.AC956	Araip.DH1Z0	1	1	1	2
9	08	Aradu.WX5KP	Araip.E99Y9	1	1	1	2
10	09	Aradu.SK1BS	Araip.5F6MD	1	1	1	1
11	09	Aradu.AE16G	Araip.T64GQ	1	1	1	1
12	09	Aradu.GJ1CE	Araip.7V9BH	1	1	1	1
13	09	Aradu.FM0YX	Araip.GV48H	1	1	1	2
14	09	Aradu.TJL9X	Araip.K56RN	1	1	1	2
15	09	Aradu.951UC	Araip.Q8LFT	1	1	1	2
16	09	Aradu.LNK8S	Araip.VN0A4	1	1	1	1
17	09	Aradu.C3RV0	Araip.3GK67	1	1	1	1
18	09	Aradu.KZX2M	Araip.D6PZJ	1	0	1	2
19	10	Aradu.AS232	Araip.Q7EYZ	1	1	1	2
20	09/10	Aradu.5E1NU	Araip.G6789	1	0	1	0
21	06	Aradu.289WG	-	0	-	-	-
22	08	Aradu.KXZ9V	-	0	-	-	-
23	09	Aradu.2ZL37	-	0	-	-	-
24	09	Aradu.S1X34	-	0	-	-	-
25	02	-	Araip.HUQ4W	0	-	-	-
26	04	-	Araip.J97JQ	0	-	-	-
27	06	-	Araip.P423S	0	-	-	-
28	07	-	Araip.8Z22U	0	-	-	-
29	09	-	Araip.LB22X	0	-	-	-

Table 2.S11: PCR primer pairs of probes, assembly validation and sequencing.

Primer ID	Forward sequence	Reverse sequence
P_01	5'-TGCCTGAGAAAGGGTCTCCT-3'	5'-AGCTGGTCCAGAACGATTTT-3'
P_02	5'-TCCATATGGAGGGCTTATCC-3'	5'-CCCATCTTTGTTCTTCTCTG-3'
P_03a	5'-TCCATGCTGCAGTTAACTTT-3'	5'-TTCTCTTCTCAATCTCTGCT-3'
P_03b	5'-CCTGGTGCATTCTTAATTAG-3'	5'-ACGAGCATATTCTGAACCTT-3'
P_04	5'-CGAACCGTCCAGCCATAAGT-3'	5'-GGGCATCTTAACAGGTCCAT-3'
P_05	5'-CCTATGCTACAAGAACCGTT-3'	5'-TGTCTCTATAATGTGGGAGA-3'
P_06	TTACCGGAACTCGAAAGCGA-3'	5'-GCAGGCCCAATTCTGTTTCT-3'
Araip.D6PZJ_Araip.Q00X2	5'-CAAGGGAGGGTGGTCTCACTGGAAAA-3'	5'-GCCATGTATGAACCTTGCAGCTTCCT-3'
Aradu.5RZ6C_Aradu.KZX2M	5'-AGGAAGCTGCAAGGTTCATACATGGC-3'	5'-GGTCTCACTGGAAAGGGAATCCCCAA-3'
Aradu.GB953_Aradu.C88Z1	5'-GTTAGTAGAGGCGTATGAGCGGGAGG-3'	5'-GGTCCAGGTGTAACTGGTCGTGGAAT-3'
Aradu.C88Z1_Aradu.P0D43	5'-TCCTTCCAGAAGAAACCACGGAACCA-3'	5'-TTGCAAATTAGCCTTACCAGCAAGTT-3'
Araip.X1W86_Araip.HTH9H	5'-AGTCACTTGGCACTCTCAACTCAGCA-3'	5'-CTCTCATCACAACAATGCCCTCTGCC-3'
pSMART2IFD	5'-TCACACAGGAAACAGCTATGA-3'	5'-CCTCTTCGCTATTACGCCAGC-3'

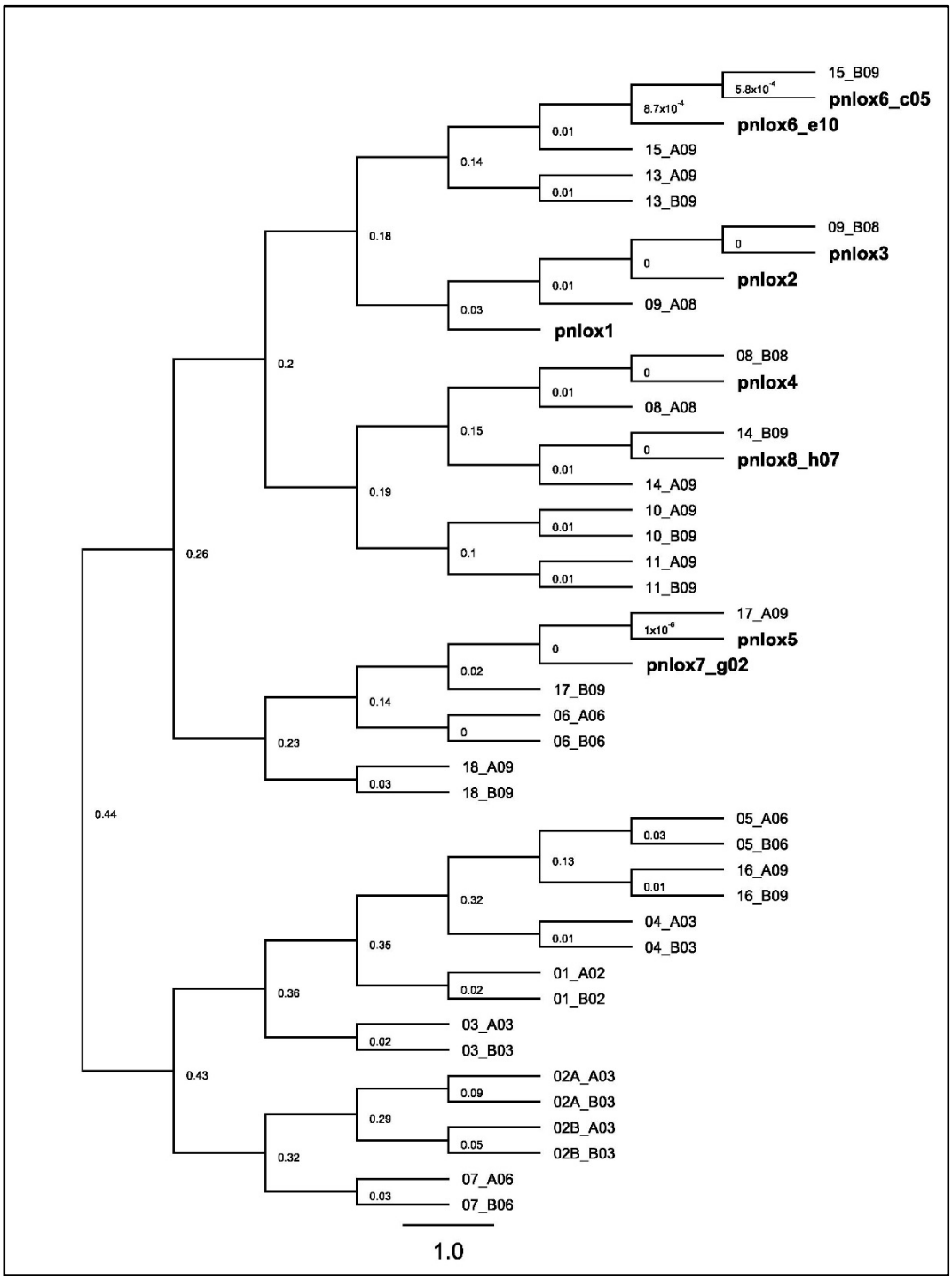


Figure 2.S1: A phylogenetic tree based on multiple alignment of annotated and published peanut LOX genes.

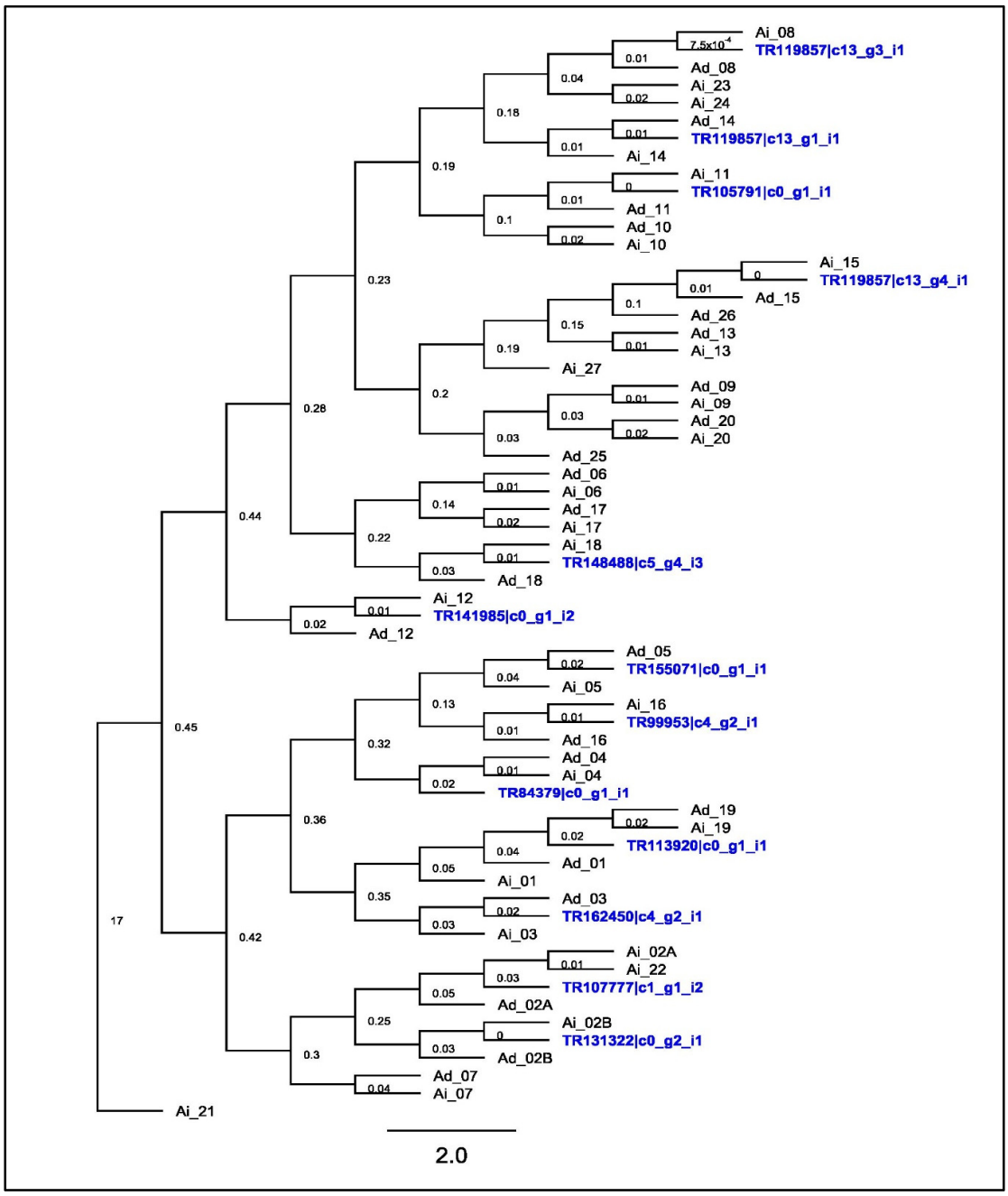
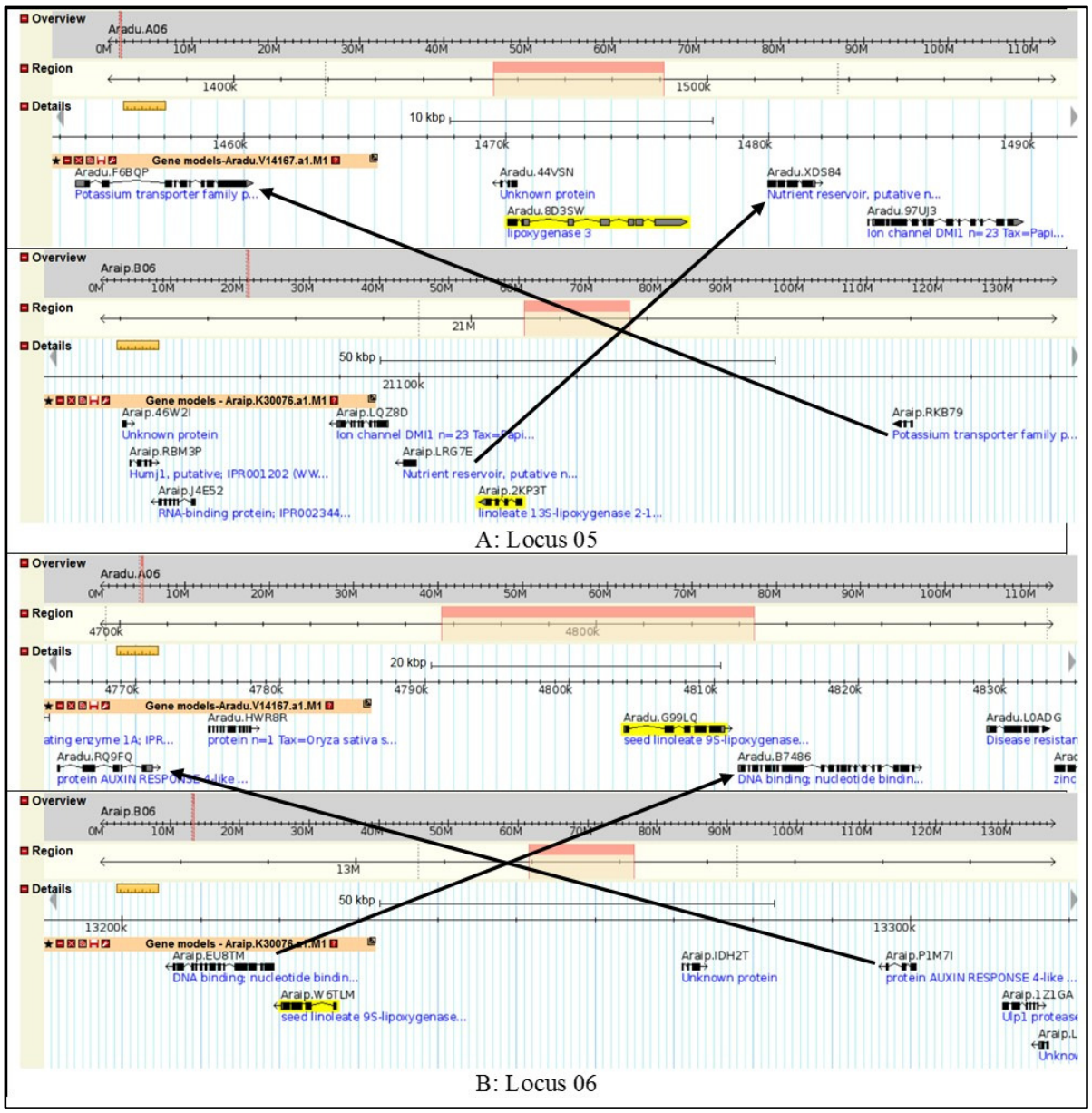


Figure 2.S2: Phylogenetic tree based on multiple alignment of 13 assembled LOX genes and peanut LOX gene pool.


Overview
Aradu.A06
Region
1400k
1500k
Details
10 kbp
1460k
1470k
1480k
1490k
Gene models-Aradu.V14167.a1.M1
Aradu.F6BQP
Potassium transporter family p...
Aradu.44VSN
Unknown protein
Aradu.XDS84
Nutrient reservoir, putative n...
Aradu.8D3SW
lipoxygenase 3
Aradu.97UJ3
Ion channel DMI1 n=23 Tax=Papi...
Overview
Araip.B06
Region
21M
Details
50 kbp
21100k
Gene models - Araip.K30076.a1.M1
Araip.46W2I
Unknown protein
Araip.LQZ8D
Ion channel DMI1 n=23 Tax=Papi...
Araip.RKB79
Potassium transporter family p...
Araip.RBM3P
Humj1, putative; IPR001202 (WW...
Araip.LRG7E
Nutrient reservoir, putative n...
Araip.J4E52
RNA-binding protein; IPR002344...
Araip.2KP3T
linoleate 13S-lipoxygenase 2-1...
A: Locus 05
Overview
Aradu.A06
Region
4700k
4800k
Details
20 kbp
4770k
4780k
4790k
4800k
4810k
4820k
4830k
Gene models-Aradu.V14167.a1.M1
Aradu.HWR8R
protein n=1 Tax=Oryza sativa s...
Aradu.G99LQ
seed linoleate 9S-lipoxygenase...
Aradu.L0ADG
Aradu.RQ9FQ
protein AUXIN RESPONSE 4-like ...
Aradu.B7486
DNA binding; nucleotide bindin...
Overview
Araip.B06
Region
13M
Details
50 kbp
13200k
13300k
Gene models - Araip.K30076.a1.M1
Araip.EU8TM
DNA binding; nucleotide bindin...
Araip.IDH2T
Unknown protein
Araip.P1M7I
protein AUXIN RESPONSE 4-like ...
Araip.W6TLM
seed linoleate 9S-lipoxygenase...
Araip.1Z1GA
Ulp1 protease
B: Locus 06

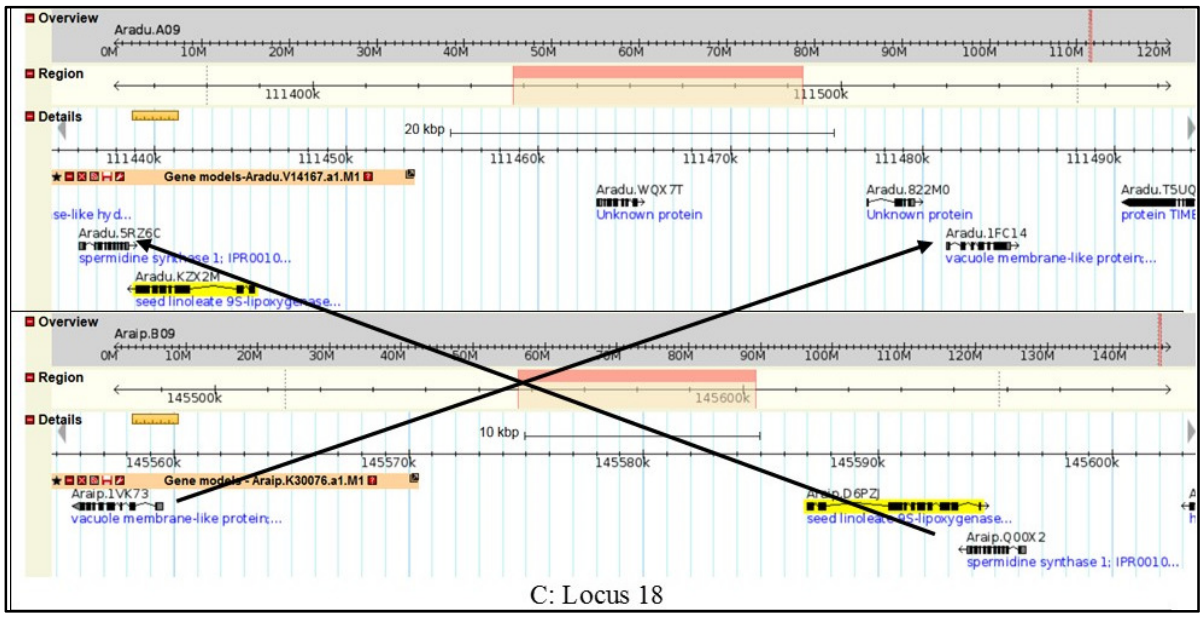


Figure 2.S3: LOX genes included within inverted regions between A-genome (upper panels) and B-genome (lower panels).

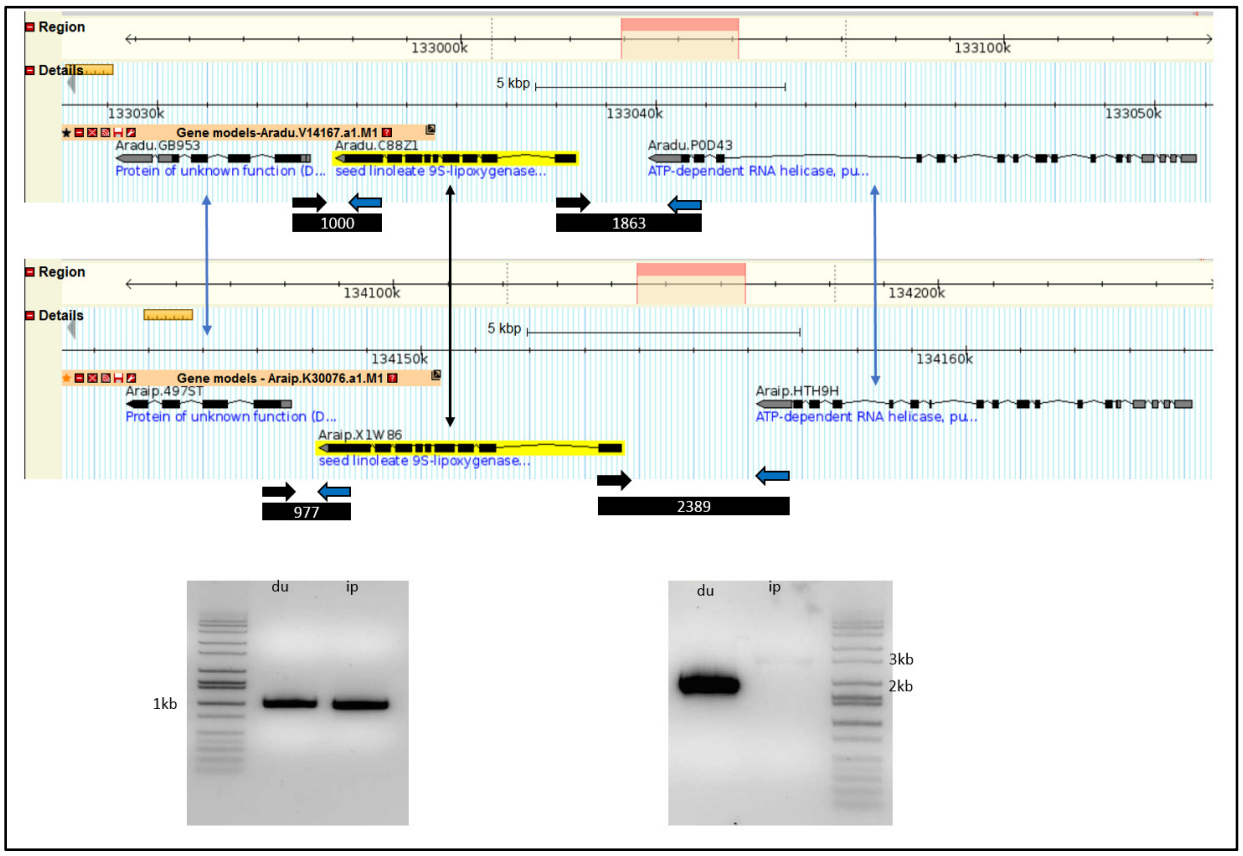


Figure 2.S4: PCR for interval regions at locus 04.

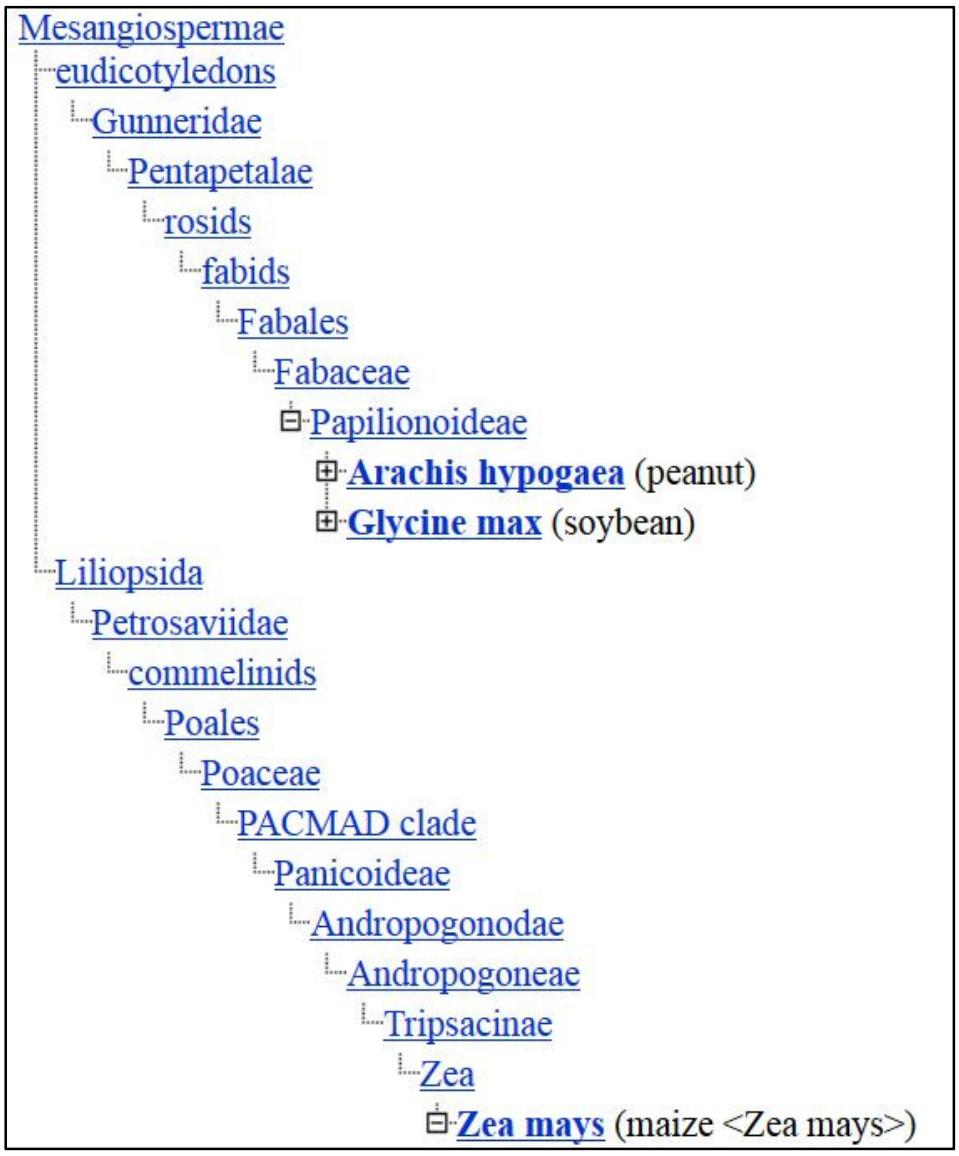


Figure 2.S5: The divergence among peanut, soybean and maize.

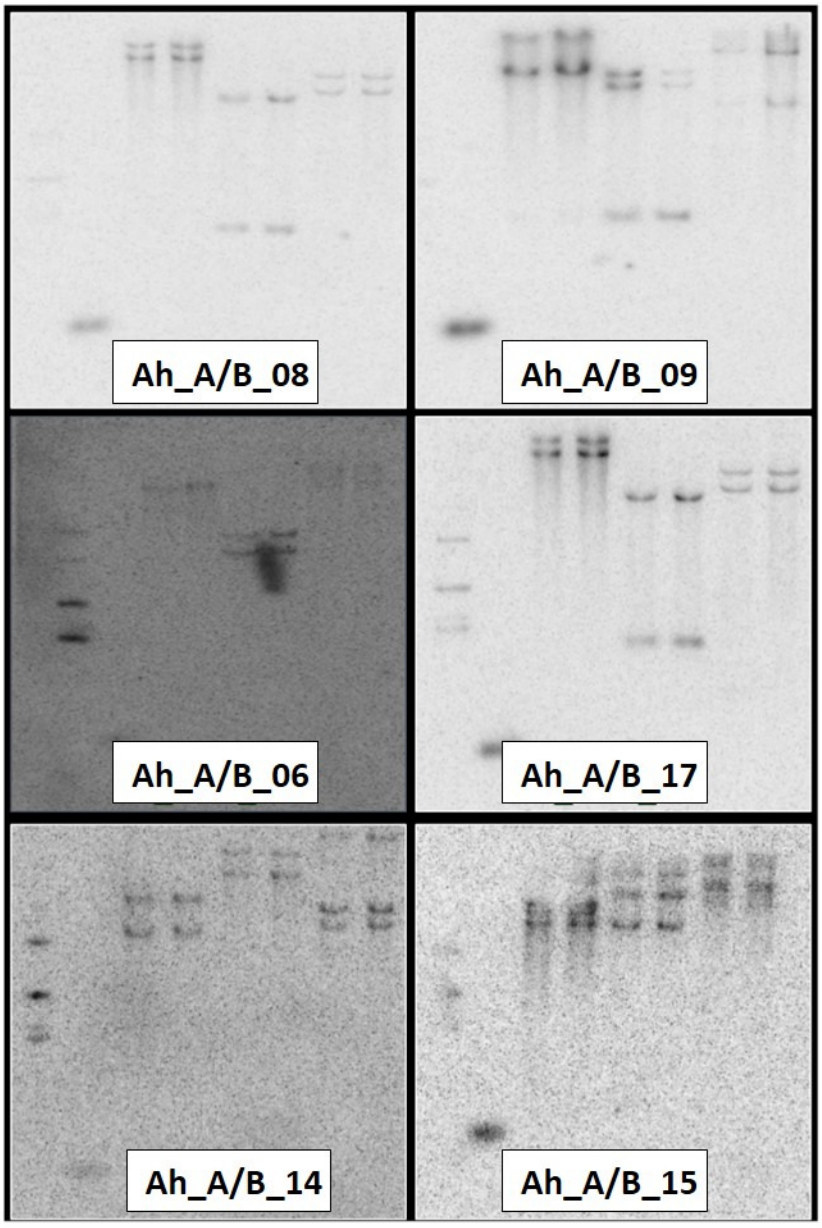


Figure 2.S6: Southern blot analysis of six peanut LOX genes: lanes from left to right: marker, positive control, GTC-*Eco*RV, TIF-*Eco*RV, GTC-*Xba*I, TIF-*Xba*I, GTC-*Sac*I, TIF-*Sac*I.

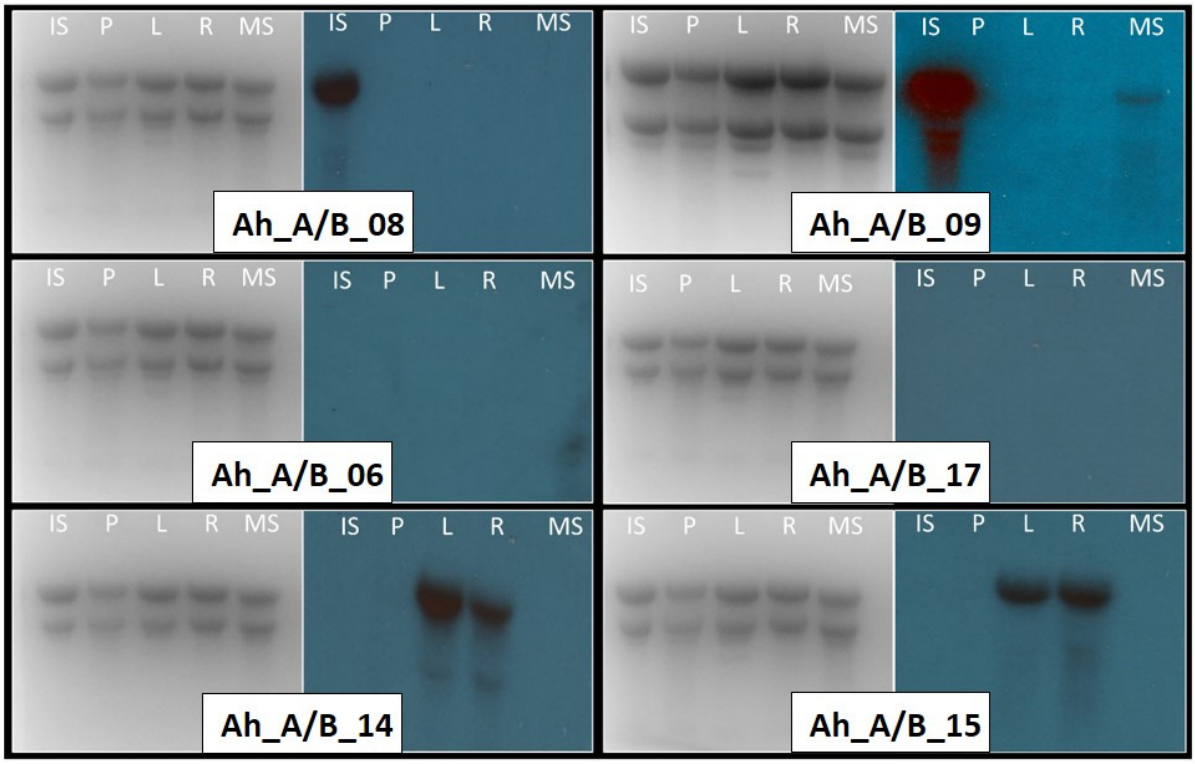


Figure 2.S7: Northern blot analysis of six peanut LOX genes: IS: Immature seed, P: Pericarp, L: Leaf, R: Root, MS: Mature seed.

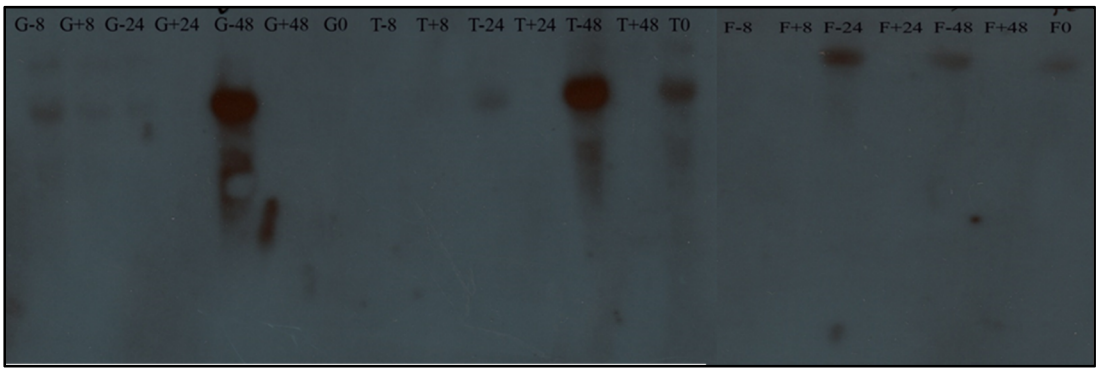


Figure S8. Northern blot analysis of Ah_A/B_9 after inoculation by *A. flavus*: 0,8,24 and 48: hours after infection, +: treatments, -: controls (mock), G: GT-C20, T: Tifrunner, F: Florunner.

Southern blot analysis (Additional files: Figure 2.S8-upper panel) of Ah_A/B_08 produced two bands for all tested restriction enzymes in both cultivars (GT-C20 and Tifrunner), which represent Ad_08 and Ai_08 alleles. However, the analysis of Ah_A/B_09 produced three bands with *Xba*I and two different intensity bands with the other two restriction enzymes (one band is roughly twice in intensify of another). This may have resulted from the two homeologs (Ad_09 and Ai_09) in addition to Ad_25, Ad_20 or Ai_20 since they have a high similarity to Ah_A/B_09. In addition, NCBI-BLASTn search using probe sequences against a library of all annotated transcripts of *A. duranensis* and *A. ipaensis* (Additional files: Table 2.S9) captured only the two-allele hits of every probe except for P_01, which captured extra three sequences, i.e., Ad_25, Ad_20 or Ai_20. The expression profile of genes included in this group, EP-I, was confirmed using northern blot analysis (Additional files: Figure 2.S9) since both Ah_A/B_8 and Ah_A/B_9 generated signals only in the two seed tissues, out of five tested tissues, Immature and mature, pericarp, leaves and root.

Southern blot analysis (Additional files: Figure 2.S8-middle panel) of Ah_A/B_06 and Ah_A/B_17 (EP-II) showed two bands with all restriction enzymes for the two cultivars, except some low intensity bands in Ah_A/B_17. However, the analysis of Ah_A/B_14 and Ah_A/B_15 (Additional files: Figure 2.S8-bottom panel) produced three bands with two restriction enzymes, out of three. Among the 83,749 entries in combined BLASTn library of *A. duranensis* and *A. ipaensis*, BLASTn search using the probes captured only the two pairs of both genes (Additional files: Table 2.S9). The probes were very specific based on BLASTn results, even different probes for the same gene, P_03a_Alox6 and P_03b_Alox6 (Additional Files: Figure 2.S7), yet the extra bands may have resulted from low but sufficient level of similarity to hybridize.

Northern blot analysis (Additional files: Figure 2.S9) confirmed the RNA-seq profiling of genes in these two groups (EP-II and EP-III) since no bands were generated for tested genes of the EP-II group and the tested genes of the EP-III group produced bands only in root and leaf tissues. However, the pericarp had high expression in RNA-seq profiling and no bands in northern blot analysis. This may have resulted from cultivar differences since RNA-seq data were produced from Tifrunner tissues and RNAs for northern blot analysis were extracted from GTC-20.

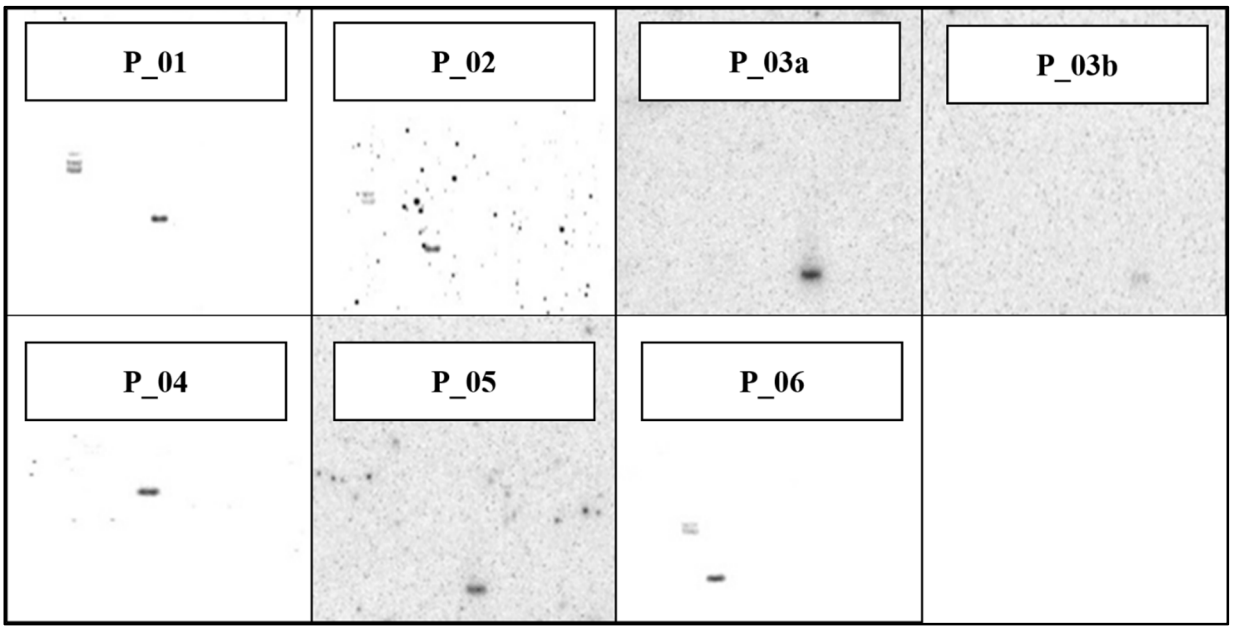


Figure 2.S9: LOX probe specificity every membrane has marker (first lane) and 7 probes.

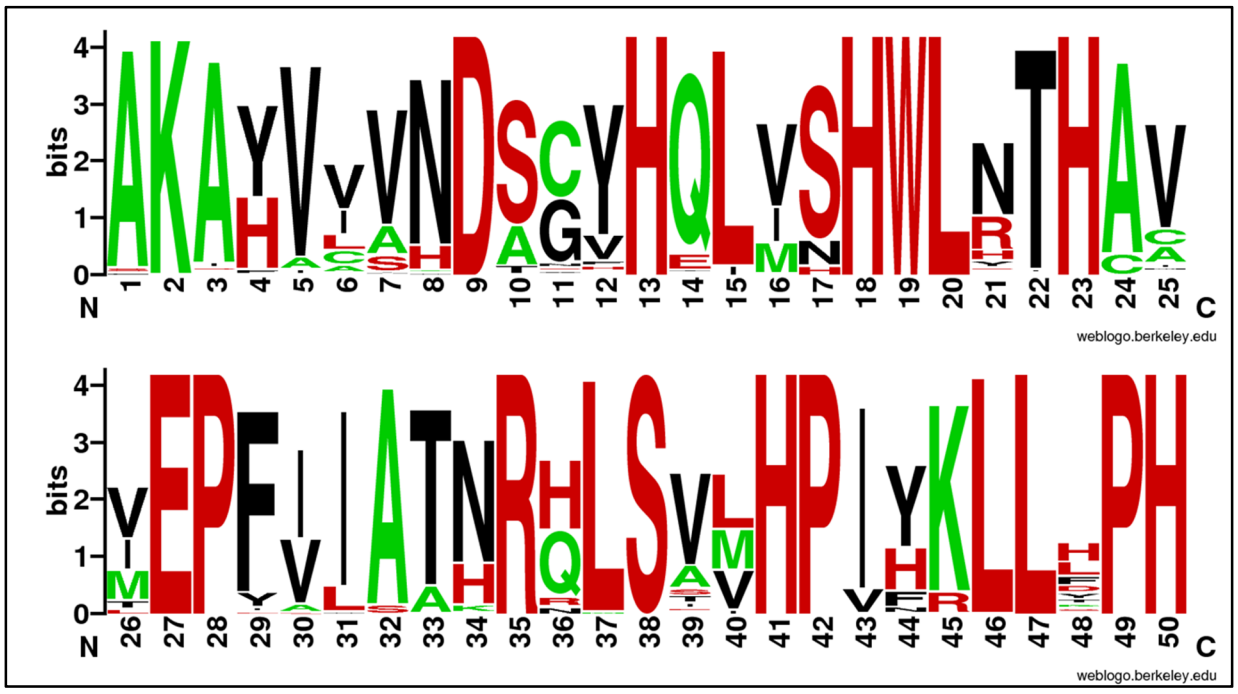


Figure 2.S10: The probability signature of amino acid sequences for LOX-Core.

CHAPTER 3

GENOTYPIC REGULATION OF AFLATOXIN ACCUMULATION BUT NOT *ASPERGILLUS FLAVUS* GROWTH UPON POST-HARVEST INFECTION OF PEANUT (*ARACHIS HYPOGAEA* L.) SEEDS

[1]W. A. Korani, Y. Chu, C. C. Holbrook, J. P. Clevenger, and P. Ozias-Akins. To be submitted to *Toxins*

Abstract

Aflatoxin contamination is a major economic and food safety concern for the peanut industry that largely could be mitigated by genetic resistance. To screen peanut for aflatoxin resistance, ten genotypes were infected with a green fluorescent protein (GFP)-expressing *Aspergillus flavus* strain. Percentages of fungal infected area and fungal GFP signal intensity were documented by visual ratings every 8 hours for 72 hours after inoculation. Significant genotypic differences in fungal growth rates were documented by repeated measures and area under the disease progress curve (AUDPC) analyses. SICIA (Seed Infection Coverage and Intensity Analyzer), an image processing software, was developed to digitize fungal GFP signals. Data from SICIA image analysis confirmed visual rating results validating its utility for quantifying fungal growth. Among the tested peanut genotypes, NC 3033 and GT-C20 supported the lowest and highest fungal growth on the surface of peanut seeds, respectively. Although differential fungal growth was observed on the surface of peanut seeds, total fungal growth in the seeds was not significantly different across genotypes based on a fluorometric GFP assay. Significant differences in aflatoxin B levels were detected across peanut genotypes. ICG 1471 had the lowest aflatoxin level whereas Florida-07 had the highest. Two-year aflatoxin tests under simulated late-season drought confirmed the reduced aflatoxin production of ICG 1471 under pre-harvest field conditions. These results suggest that all studied peanut genotypes support *A. flavus* fungal growth yet differentially influence aflatoxin production.

Introduction

Aspergillus flavus is an opportunistic pathogen with a wide host range including peanut, corn, wheat, barley, rice, tree nuts, and cotton seeds (Chang and Markakis, 1982; Rajasekaran *et al.,* 2008; Taheri *et al.,* 2012; Ostadrahimi *et al.,* 2014; Elzupir *et al.,* 2015; Abbas *et al.,* 2015).

Peanut is one of the most susceptible crops to *A. flavus* and *A. parasiticus* infection either in the field (pre-harvest) or during storage (post-harvest) (Diener *et al.,* 1987; Xue *et al.,* 2004). *A. flavus* and *A. parasiticus* produce aflatoxins as secondary metabolites under conducive environmental conditions. Aflatoxins cause toxicosis, cancer, and immunosuppressive diseases in animals and humans (Scheidegger and Payne, 2003; Verma 2004) and alfatoxin B1 level is highly regulated worldwide (Smith *et al.,* 2016). Aflatoxin contamination incurs an average $20 million annual cost to the U.S. peanut industry (Lamb and Sternitzke, 2001).

Aspergillus spp. conidia are abundant in the soil and can survive through harsh weather conditions. Since peanut pods develop underground, direct contact of developing peanut pods with fungal mycelium provides the main entry for fungal invasion (Cole *et al.,* 1986; Torres *et al.,* 2014). Peanut pods damaged by insects and nematodes were shown to have elevated levels of aflatoxin contamination (Lynch and Wilson, 1991; Timper *et al.,* 2004). Another possible route of fungal infection is through flowers (Styer *et al.,* 1983). Heat and drought stress in the field exacerbates aflatoxin contamination (Holbrook *et al.,* 2009). Peanut host genes altered by *A. flavus* contamination were discovered by a genome wide RNA-seq analysis (Clevenger *et al.,* 2016). Disruption of peanut ABA signaling pathway by *A. flavus* invasion was suggested to facilitate aflatoxin accumulation. As for minimizing post-harvest aflatoxin contamination, clean, dry and temperature controlled storage conditions with protection from insect and rodent infestation are recommended (Torres *et al.,* 2014). In developing countries, appropriate storage conditions are often inaccessible or unaffordable.

Development of aflatoxin resistant peanut cultivars has been one of the most challenging goals due to the large variation in pre-harvest aflatoxin contamination. Even aflatoxin resistant lines accumulated widely different levels of aflatoxin when grown in the same or different

environments (Blankenship *et al.,* 1984). A recent gene profiling study comparing an aflatoxin resistant line and a susceptible line revealed multiple biological pathways enriched in the resistant line upon *A. flavus* challenge (Houmiao *et al.,* 2016). However, the resistant line accumulated over 20,000 ppb of aflatoxin over the 10-d time frame of the experiment, which even though 1% of the aflatoxin in the susceptible line, far exceeds the U.S. action level of 20 ppb for human food consumption and is not actually resistant.

To circumvent the high variation in field aflatoxin evaluation of genetic resistance, *in vitro* inoculation has been used to ensure more uniform fungal infection of seeds. Several wild diploid peanut relatives and interspecific tetraploids were reported to be resistant to aflatoxin based on *in vitro* inoculation of seeds and analysis 8-d post inoculation (Xue *et al.,* 2004). Since complete inhibition of fungal growth or aflatoxin contamination is unlikely, a time-course to monitor fungal growth during the disease progression could reveal differential fungal-host interactions among peanut genotypes. Surrogate parameters estimating fungal growth such as β-l-3-glucanase activity previously were used in *in vitro* studies (Liang *et al.,* 2005; Sharaf *et al.,* 2011), but are destructive assays. Green fluorescent protein (from *Aequorea victoria*) transformed *A. flavus* allows for the non-destructive measurement of fungal growth (Lorang *et al.,* 2001) and AF-70-GFP, a GFP-expressing *A. flavus* strain, was used in cotton to identify resistant cotton lines (Rajasekaran *et al.,* 2008). Since dissociation of *in vitro* aflatoxin resistance and pre-harvest aflatoxin contamination was reported previously (Nigam *et al.,* 2009), it is possible that *in vitro* and field tests were interrogating different aspects of aflatoxin resistance mechanisms in the genotypes examined.

In this study, ten peanut genotypes were selected based on their resistance to aflatoxin and or drought tolerance. ICGV88145 (Rao *et al.,* 1995) and ICG 1471 (Dwivedi and Varma,

2002; Waliyar *et al.,* 1994; Asis *et al.,* 2005; Hamidou *et al.,* 2014) are aflatoxin resistant lines released by ICRISAT and Senegal, respectively. ICG 1471 also is drought tolerant (Mehan 1989). GT-C20 is a Chinese cultivar (Liang *et al.,* 2006) reported to retard *A. flavus* fungal growth and prevent aflatoxin production (Zhang *et al.,* 2015). C76-16 is a drought tolerant USDA breeding line with field aflatoxin resistance (Holbrook *et al.,* 2007). Tifguard (Holbrook *et al.,* 2008), NC 3033 (Beute *et al.,* 1976), Tifrunner (Holbrook and Culbreath, 2007) and Florida-07 (Gorbet and Tillman, 2009) showed less aflatoxin contamination compared to susceptible control breeding line A72 (Luis *et al.,* 2016). Another susceptible breeding line, A69, was selected from the cross NCV-11 x GFA-2 from the USDA peanut breeding program in Tifton, GA. In this study, these ten peanut genotypes were inoculated with AF-70-GFP and evaluated for fungal growth by both visual rating and image analysis as well as for aflatoxin contamination over a 3-d time course. Genotypic differences in aflatoxin production per unit of fungus were documented.

Results and Discussion

It is known that aflatoxin contamination increases in seeds with low viability and germination rate as Aspergilli produce metallo and serine proteases that reduce seed vaiablity to facilitate their accession (Klich and Lee, 1982; Asis *et al.,* 2009). In this study, instead of using seed sources from storage, all genotypes were grown in a common field and harvested at their respective optimum maturity dates. Sound and mature seeds were selected to reduce the variable of maturity. Germination tests of these freshly harvested seeds showed greater than 98% germination rate (data not shown). Since the presence or absence of testa can affect the level of colonization by *Aspergilli* (Xue *et al.,* 2005), and tannins (Lansden 1982) and antioxidants (Shiow Chyn *et al.,* 2003) have been identified in peanut testa, we adopted a method of surface

sterilization to preserve testa integrity. Therefore, peanut kernels were UV sterilized prior to inoculation which protected the intactness of peanut testa and minimized the confounding effect of *A. flavus* or other microbiota persisting from the field. The more extensive testa surface area of Florida-07 seeds, which were approximately twice the size of ICG 1471, remained intact with UV sterilization.

The AF-70-GFP strain does not differ from wild-type *A. flavus* in terms of pathogen aggressiveness and aflatoxin production (Rajasekaran *et al.,* 2008). The GFP signal produced by AF-70-GFP is a good indicator of fungal growth allowing non-destructive, real-time monitoring of fungal development. In this study, *A. flavus* fungal growth rates were significantly different among the tested peanut genotypes determined by visual ratings of fungal GFP signal on the seed surface (Figure 3.1). Fungal growth curves based on visual ratings and reported as percentage of infected area (Figure 3.1A) and intensity of fungal GFP signal (Figure 3.1B) gave similar patterns suggesting either parameter can be applied to monitor fungal development on the seed surface. GFP signal was not detected on seed surfaces freshly inoculated with conidial suspension whereas newly grown hypha and conidia emitted strong GFP signals. Fungal GFP expression became visible across peanut genotypes between 8 to 24 hours after inoculation (HAI) and rapidly increased throughout the 72-h time course. Therefore, under current testing conditions, AF-GFP-70 had a short incubation period of 8-16 h prior to initiating a vigorous growth phase. Tukey's range test showed that GT-C20 is the genotype most conducive to fungal growth followed by C76-16 and A69. NC 3033 had the least fungal growth on seed surfaces, however it was not significantly different from Tifrunner, Tifguard, A72, ICGV88145 and Florida-07. ANOVA analysis of area under the disease progress curve (AUDPC) of these two parameters gave similar results (Figure 3.S1).

To objectively determine fungal GFP signal on seed surfaces, seed infection coverage and intensity analyzer (SICIA), a Matlab software, was developed. The software analysis flow chart (Figure 3.2A) and examples of the fungal infected area and fungal intensity quantification by SICIA (Figure 3.2B) are provided. Both percentage of infected area, and intensity of fungal GFP signal, estimated by SICIA (Figure 3.3), confirmed visual rating scores (Figure 3.1) in that GT-C20, C76-16, and A69 supported the highest level of fungal growth and NC 3033 had the least surface fungal growth. Visual rating across a time course is tedious, which limits the number of genotypes that can be included in a study. On the contrary, image analysis is automated, removes subjectivity, and is fast. In addition, visual rating scores are categorical with limited interpolation whereas image analysis quantifies GFP signal by continuous measurements which explains the better separation of means by SICIA than that of visual rating.

GT-C20, NC 3033, ICG 1471, Tifrunner, and Florida-07 were selected for quantification of GFP and aflatoxin B by fluorometric and VICAM assays, respectively, since these five genotypes represented the range of fungal growth in this study. GFP and aflatoxin analyses were performed on a single-seed basis for best comparison between aflatoxin production and fungal growth. No statistically significant differences were found for GFP expression among genotypes (Figure 3.4A). Although NC 3033 appeared to support significantly less fungal growth on the seed surface in the previous analysis, vigorous fungal growth beneath the testa negates the difference on seed surfaces. Previously, tannins extracted from the testa were shown to inhibit *A. flavus* growth [43]. It is possible that testa tannin or phenolic content of NC 3033 may retard fungal growth on the seed surface which could be of interest for further investigation. In addition, this shows that the progressing fungal growth pattern can be different across peanut

genotypes as it tends to be on the surface in some genotypes and more penetrated inside the seeds in others.

ICG 1471 was found to accumulate significantly less aflatoxin than other genotypes and Florida-07 had the highest level of aflatoxin (Figure 3.4A). Interaction plots between GFP and aflatoxin level indicated that ICG 1471 produced minimum amounts of aflatoxin regardless of the amount of fungal GFP accumulation (Figure 3.4B). Florida-07 is the only high oleic genotype [38] among the tested genotypes and high oleic, low linoleic, acid in peanut has been suggested to increase post-harvest aflatoxin accumulation (Xue *et al.,* 2003; Xue *et al.,* 2005). Among the other four tested genotypes with normal oleic acid content, ICG 1471 accumulated significantly less aflatoxin.

To further investigate genotypic differences in pre-harvest aflatoxin accumulation, all genotypes were tested in rain exclusion shelters inoculated with *A. flavus* and *A. parasiticus* (Table 3.1). From the two-year data, ICG 1471 had low aflatoxin production whereas most tested lines accumulated greater than 20 ppb of aflatoxin. High variation of field aflatoxin accumulation documented here is common for field studies (Holbrook *et al.,* 2009), and aflatoxin resistant lines need to be tested for multiple years in multiple environments. Although our *in vitro* data is consistent with field studies in regard to aflatoxin resistance of ICG 1471, additional field and laboratory studies are needed to define the resistance mechanisms for this genotype. GT-C20 was previously reported to limit *A. flavus* fungal growth and inhibit aflatoxin production (Zhang *et al.,* 2015), which is contrary to our findings. Both our field and *in vitro* studies demonstrated the high susceptibility of GT-C20 to *A. flavus* fungal growth and aflatoxin production. As mentioned earlier, a truly aflatoxin resistant line should not be claimed unless it withstands multiple tests

due to the highly variable nature of the *A. flavus*/peanut host/secondary product biosynthesis interaction.

Highly reproducible results from this study support the robustness of these methods applied to quantify *A. flavus* growth. Differential genotypic responses to *A. flavus* fungal growth and aflatoxin production were revealed. Peanut germplasm ICG 1471 was found to inhibit aflatoxin production without restraining fungal growth. Rainout shelter testing supported the reduced aflatoxin production in ICG 1471, suggesting that this germplasm and its underlying genetic mechanisms for resistance may be useful in breeding for pre- and post-harvest aflatoxin reduction. Future RNA-seq study is required to reveal the genetic mechanism controlling aflatoxin contamination in this genotype.

Materials and Methods

Plant materials

Ten peanut genotypes were planted on the Tifton Campus of the University of Georgia in June and harvested in October 2015. Harvest was according to respective maturity dates as follows: ICG88145, ICG 1471, GT-C20: 115 days; C76-16, A72, A69, Tifguard, NC 3033: 135 days; and Tifrunner, Florida-07: 150 days. Mature seeds were selected from each genotype. Thirty seeds per genotype were used to determine seed viability by a germination test, and another 150 seeds were used for fungal infection and aflatoxin analysis studies described below. The seeds were stored at 4°C during the experiment progress. In addition, the seeds were dried at 30°C for 7-10 days.

In vitro A. flavus inoculation using AF-70-GFP

This experiment was conducted with a randomized complete block (RCB) design; 3-4 blocks were used. Each block had five individual seeds as replicates, and the experiment was

replicated three times (11 total blocks of infected seeds). Seeds were surface sterilized for 15 min under UV light (LABCONCO purified class II biosafety cabinet, Kansas City, MO). The AF-70-GFP strain (Rajasekaran *et al.,* 2008) was used for infection. This strain had Enhanced Green Fluorescent Protein (EGFP) under the control of glyceraldehyde phosphate dehydrogenase (gpdA) promoter and *A. parasiticus* nmt-1 terminator (Rajasekaran *et al.,* 2008). The fungus was grown on potato dextrose agar (PDA) medium in petri dishes for two weeks at 30°C, and conidia were suspended in 0.01% Tween-20 solution. Conidia concentration was estimated using a Fuchs-Rosenthal Counting Chamber (Hausser Scientific, Horsham, PA) and was adjusted to 1000 conidia/ml. Twenty ml of conidial suspension was used to inoculate every five seeds by immersion for 30 min with quick vortexing every 5 min (Burow *et al.,* 2000). A wide range of conidia concentrations (10^4 to 10^6 conidia/ml) had been used previously (Burow *et al.,* 2000; Tsitsigiannis *et al.,* 2005; Muller *et al.,* 2014). A low concentration of conidia was chosen for this study to avoid an overwhelming level of colonization. After inoculation, the seeds were placed in 12-well tissue culture plates (Fisher Scientific, Suwanee, GA) at a density of one seed per well with the two middle wells filled with sterilized water to increase humidity. The ten seeds on each plate came from the ten peanut genotypes included in this study and were randomly distributed on the plate, therefore, each plate is a randomized replicate. Five plates were included in this study which formed one experimental block. The plates were sealed and incubated in dark at 30°C for 72 hours.

Visual tracking of GFP expression

Fungal growth on the seeds was visually monitored by fungal GFP expression under a microscope (STEMI SV 11 ZEISS equipped with a HBO 100 microscope illumination system, Carl Zeiss Microscopy, Thornwood, NY). GFP signal was observed at 480 ± 30 nm excitation;

545 nm emission wavelengths. Fungal GFP expression on the seed surface was scored visually every 8 hours for up to 72 hours after inoculation. Percentage of fungal infected area over the surface area of peanut kernels visualized under the microscope was estimated based on a scale of 0 to 5 as follows: 0: no infection to 5%; 1: 5-25%; 2: 25-50%; 3: 50-75%; 4: 75-90%; and 5: > 90%. The intensity of fungal GFP signal was estimated based on a scale of 0 to 3 as follows: 0: no infection; 1: low intensity; 2: medium intensity; and 3: high intensity. Initiation of sporulation was estimated by documenting the time point that the first spore became visible on the surface of infected seeds.

Image processing

At 72 hours, all seeds that had a percentage of infected area higher than 5% were photographed using an Axiocam CCD camera (Carl Zeiss Microscopy, Thornwood, NY). Two images were taken for each seed; one image was taken for GFP visualization and the other image was taken under white light.

Seed Infection Coverage and Intensity Analyzer (SICIA) software was designed with MATLAB R2016a (The University of Georgia campus-wide site licensing agreement). The software layout is presented (Figure 3.2A). SICIA extracts the outline of a seed using the image taken under white light. GFP signal was captured from the image taken under fluorescent light and superimposed to the seed outline to calculate seed size, infected area, GFP signal density, and infection coverage ratio. The program generated a .csv output file containing the calculated values and produced processed images in .jpg format. SICIA is freely available to the public under MIT license and can be downloaded from https://github.com/w-korani/SICIA. SICIA is also calibrated to analyze images from small seeds, *e.g.,* rice, and leaves.

GFP and aflatoxin analysis

All seeds were harvested at 72 hours after inoculation and ground in liquid nitrogen. A portion of the pulverized tissue was used for total protein extraction in a sucrose-Tris solution (0.5 M and 0.1 M respectively, pH 7.5 containing 1 mM PMSF) (Chen and Dickman, 2004). The tube was then stored on ice for 15 min and centrifuged at 17530 xg for 15 min at 4 °C. The supernatant containing extracted protein was quantified by a BCA assay kit (Pierce, Rockford, IL). The GFP relative fluorescence units (RFU) were detected in 100 µl of the protein extract at 485 nm excitation and 535 nm emission wavelengths (Synergy HT Bio-Tek instrument, Bio-Tek Instruments Inc, Winooski, VT). RFU was normalized to 1 mg of the total extracted protein. Aflatoxin was extracted from a second aliquot of pulverized tissue by adding 200 µl of 25% NaCl and 800 µl of 100% methanol (Anderson *et al.,* 1995). The tube was vortexed and incubated at room temperature for 30 min. One hundred µl of supernatant was collected after centrifugation at 9,000 xg for 10 min at room temperature, and 400 µl of HPLC-grade water was added. VICAM Afla-B column (VICAM, Nixa, MO) was used to extract aflatoxin B1, and the VICAM Fluorometer (VICAM, Nixa, MO) was used to detect the quantity according to the manufacturer's instructions.

Rainout shelter study

Pre-harvest susceptibility to aflatoxin accumulation was evaluated in rainout shelters established in Tifton, GA using a randomized block design (Holbrook *et al.,* 2000). The fields were inoculated with *A. flavus* Link ex Fries (NRRL 3357) and *A. parasiticus* (NRRL 2999) at mid-bloom as both strains produce aflatoxin B1 which is our main target in this work and it is the most regulated type (Smith *et al.,* 2016). Drought stress was imposed by moving the rainout

shelters over test plots 40 d before harvest. Peanuts were shelled and measured for aflatoxin levels using the VICAM columns as described above.

Statistical analysis

Statistical analysis was carried out by R3.2.2 using ‘stats’ package. Agricolae package in R3.2.2 (Core Team 2011) was used to perform Duncan’s multiple range test and to calculate the area under disease progress curve (AUDPC).

Tables

Table 3.1: Aflatoxin levels tested in the rainout shelter

Genotype	Year 2014		Year 2015	
	Aflatoxin range (ppb)	Average aflatoxin B (ppb)	Aflatoxin range (ppb)	Average aflatoxin B (ppb)
ICG 1471	5 to 35	15	2 to 7	3
Florida-07	5 to 18	11	23 to 641	214
Tifrunner	9 to 48	26	4 to 1200	381
A69	3 to 230	61	3 to 54	20
A72	3 to 679	216	3 to 2100	425
ICGV88145	1 to 1034	226	1 to 11	4
C76-16	25 to 535	246	22 to 599	200
Tifguard	12 to 734	272	n/a	n/a
NC 3033	220 to 360	303	n/a	n/a
GT-C20	220 to 39000	4581	n/a	n/a

Figures

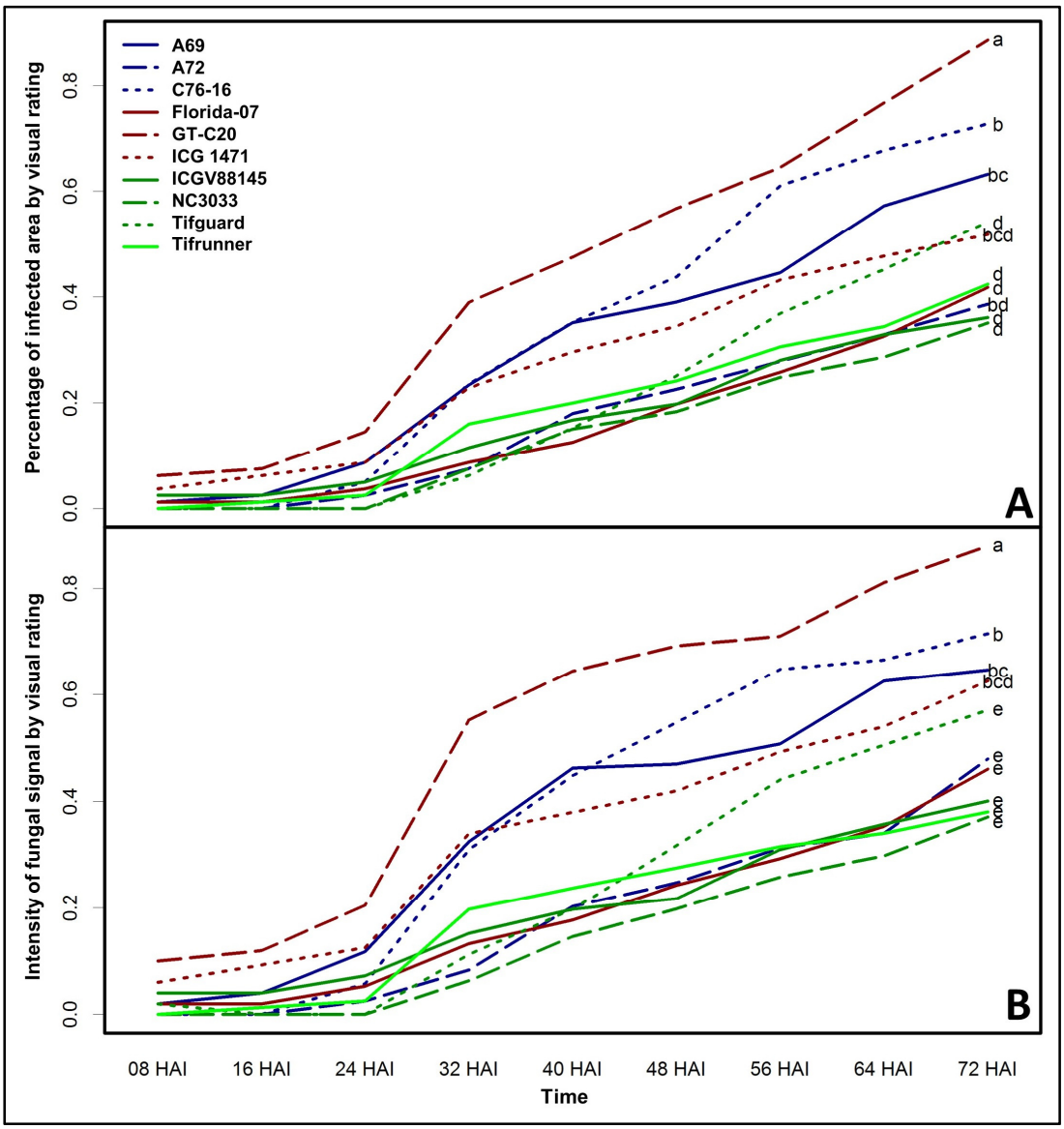


Figure 3.1: Repeated measure analysis of ten peanut genotypes inoculated with AF-70-GFP strain across nine time points from 8 to 72 hours after inoculation (HAI) determined by visual rating. Log-transformed percentage of colonized area (upper panel) and fungal GFP intensity (lower panel) were presented. Different letters indicate significant differences at $p < 0.05$ level determined by Tukey's range test.

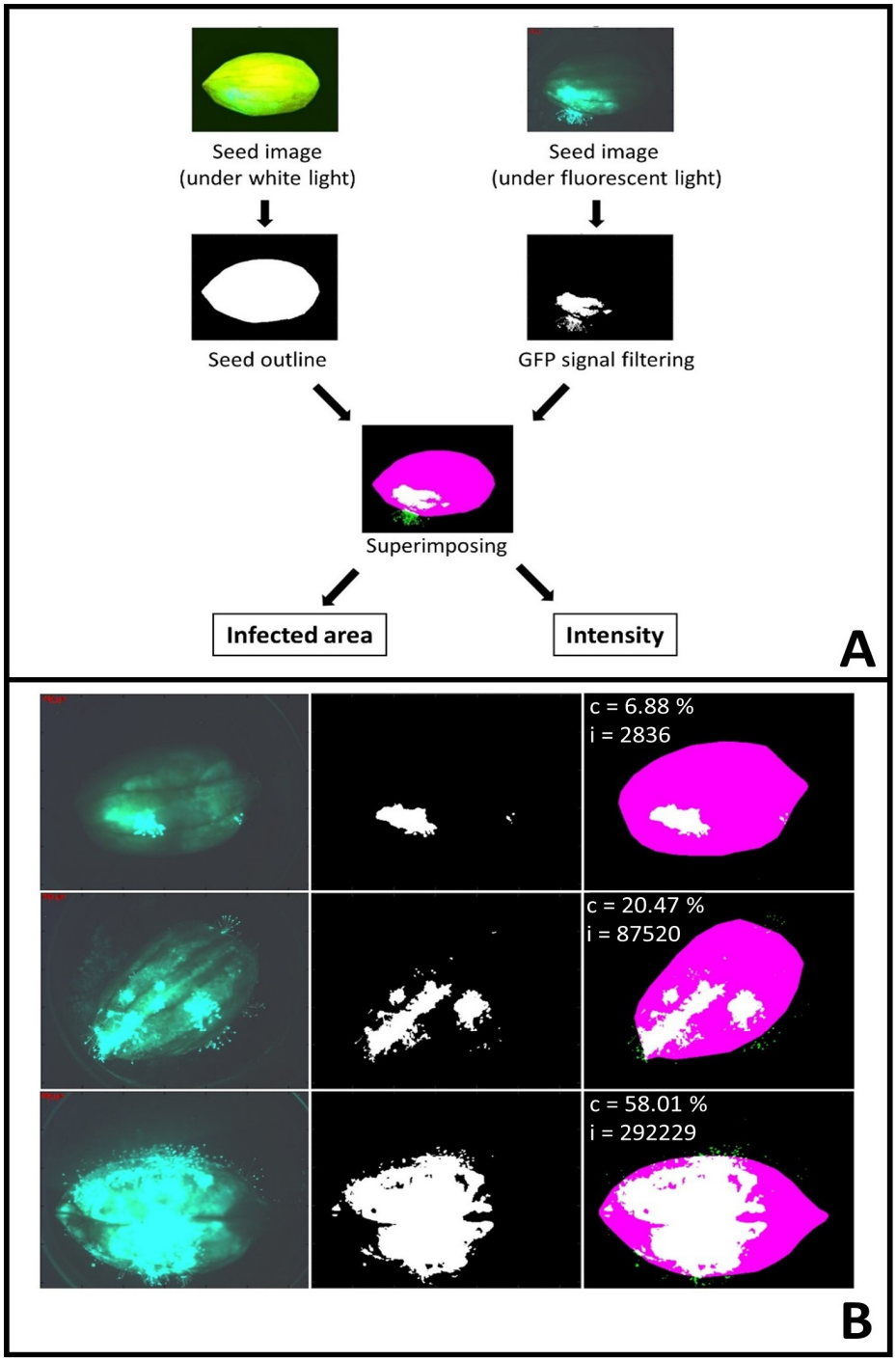


Figure 3.2: SICIA. A. SICIA data analysis flow chart; B. Three examples of SICIA output of percentage of infected area (c) and intensity of GFP signal (i) for seeds with varied levels of fungal infection.

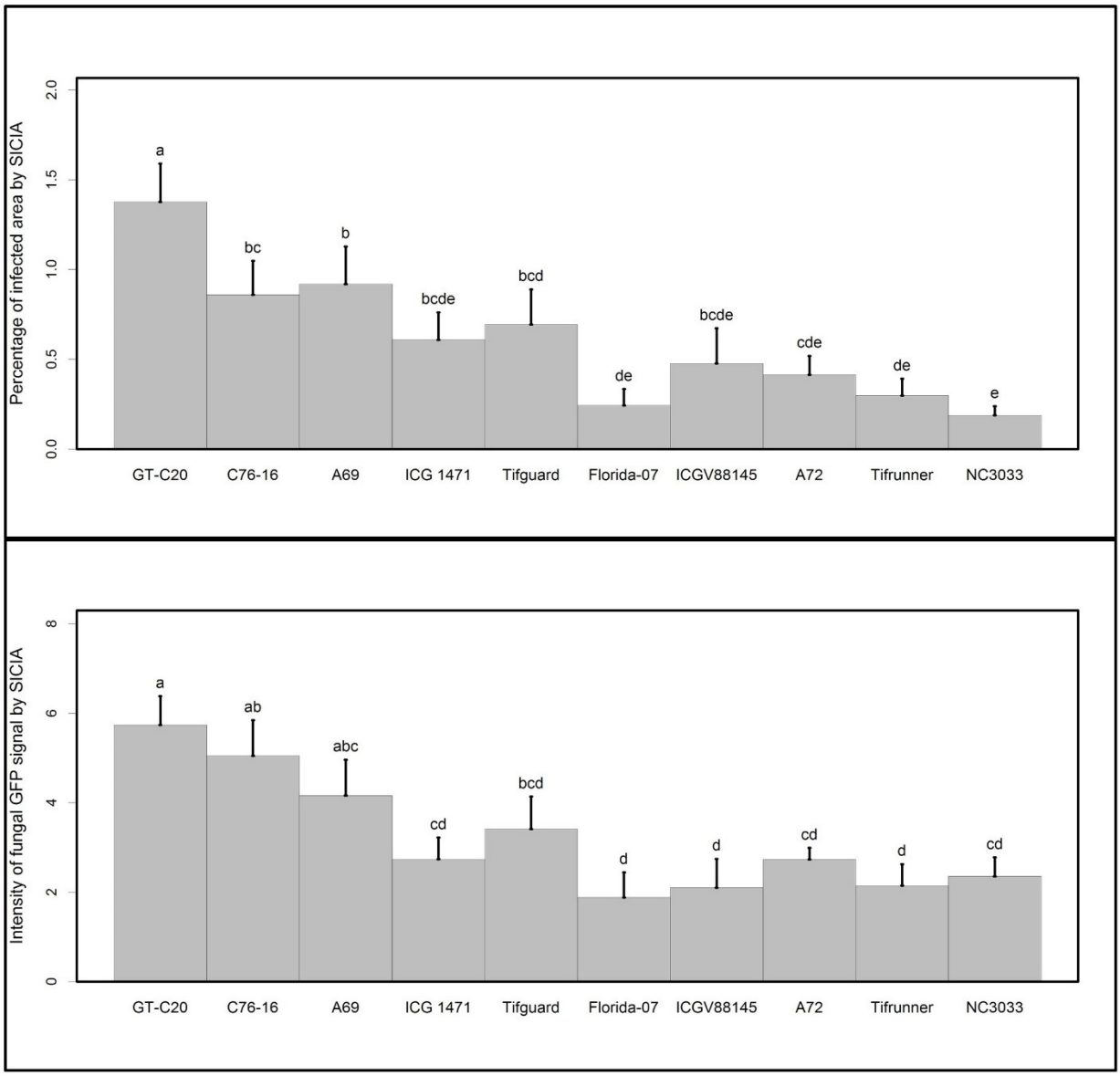


Figure 3.3: ANOVA analysis of log-transformed values of percentage of the infected area (upper panel) and intensity of fungal GFP signal (lower panel) determined by SICIA. Different letters indicate significant differences at $p < 0.05$ level determined by Duncan's multiple range test.

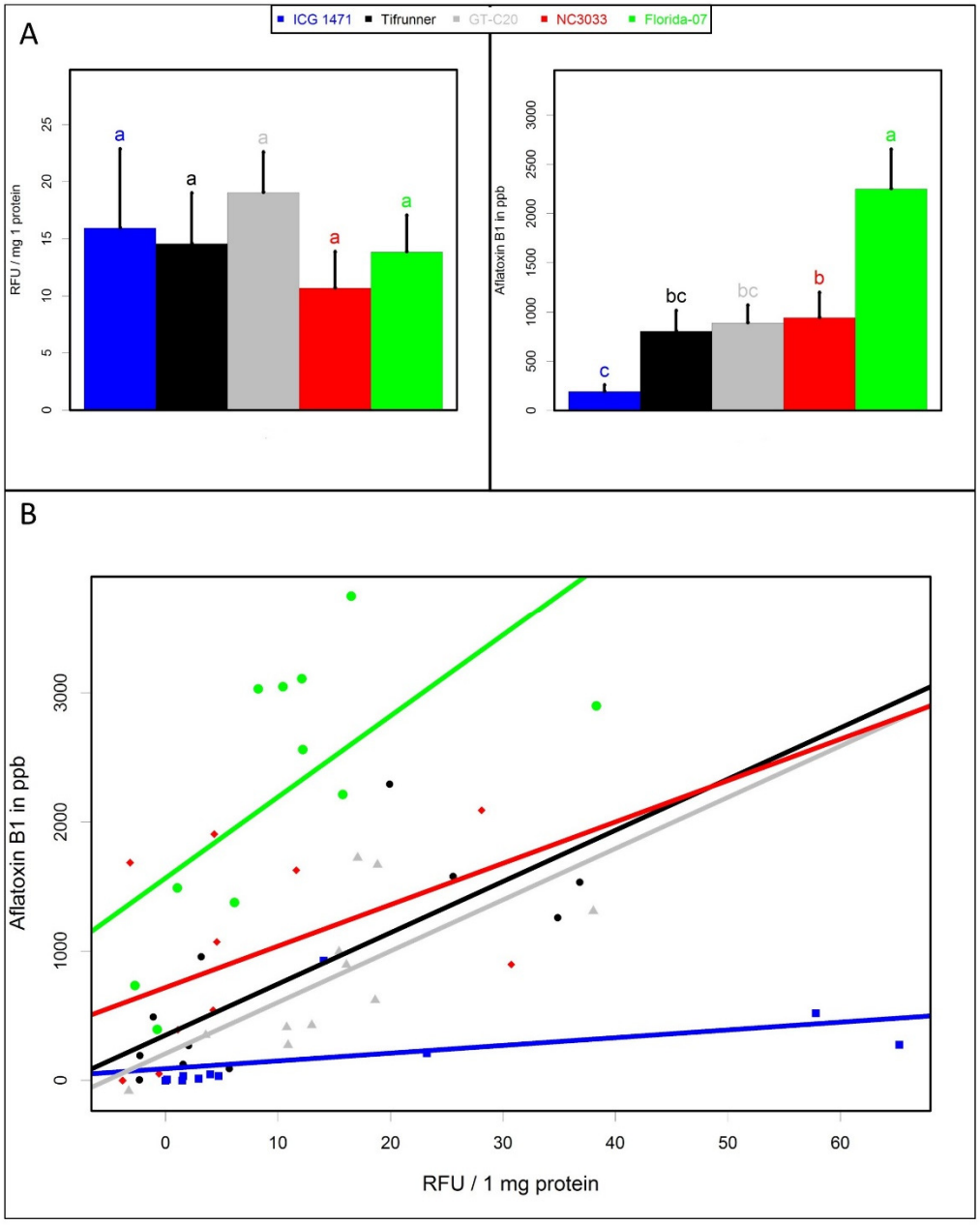


Figure 3.4: Fungal GFP expression and aflatoxin accumulation in five selected genotypes measured by fluorometric and VICAM assays. A. ANOVA analysis of fungal GFP level, left bar graph, and aflatoxin B level, right bar graph. Different letters indicate significant differences at $p < 0.05$ level determined by Duncan's multiple range test; B. Interaction between GFP and aflatoxin levels, each data point is the average of an experimental block. RFU is GFP relative fluorescence units.

References

Abbas HK, Zablotowicz, RM, Shier, WT, Johnson, BJ, Phillips, NA, *et al.* (2015) Aflatoxin and fumonisin in corn (*Zea mays*) infected by common smut *Ustilago maydis*. Plant Disease **99:** 1236-1240

Anderson WF, Holbrook, CC, Wilson, DM, Matheron, ME (1995) Evaluation of preharvest aflatoxin contamination in several potentially resistant peanut genotypes. Peanut Science **22:** 29-32

Asis R, Barrionuevo, DL, Giorda, LM, Nores, ML, Aldao, MA (2005) Aflatoxin production in six peanut (*Arachis hypogaea* L.) genotypes infected with *Aspergillus flavus* and *Aspergillus parasiticus*, isolated from peanut production areas of Cordoba, Argentina. Journal of Agricultural and Food Chemistry **53:** 9274-9280

Asis R, Muller, V, Barrionuevo, DL, Araujo, SA, Aldao, MA (2009) Analysis of protease activity in *Aspergillus flavus* and *A. parasiticus* on peanut seed infection and aflatoxin contamination. European Journal of Plant Pathology **124:** 391-403

Beute MK, Wynne, JC, Emery, DA (1976) Registration of NC3033 peanut germplasm. Crop Science **16:** 887-887

Blankenship P, Cole, R, Sanders, T, Hill, R (1984) Effect of geocarposphere temperature on pre-harvest colonization of drought-stressed peanuts by *Aspergillus flavus* and subsequent aflatoxin contamination. Mycopathologia **85:** 69

Burow GB, Gardner, HW, Keller, NP (2000) A peanut seed lipoxygenase responsive to *Aspergillus* colonization. Plant Molecular Biology **42:** 689-701

Chang HG, Markakis, P (1982) Effect of temperature on aflatoxin production in barley by *Aspergillus parasiticus*. Korean Journal of Food Science and Technology **14:** 162-167

Chen S, Dickman, MB (2004) Bcl-2 family members localize to tobacco chloroplasts and inhibit programmed cell death induced by chloroplast-targeted herbicides. Journal of Experimental Botany **55:** 2617

Clevenger J, Marasigan, K, Liakos, V, Sobolev, V, Vellidis, G, *et al.* (2016) RNA sequencing of contaminated seeds reveals the state of the seed permissive for pre-harvest aflatoxin contamination and points to a potential susceptibility factor. Toxins **8:** 317

Cole RJ, Hill, RA, Blankenship, PD, Sanders, TH (1986) Color mutants of *Aspergillus flavus* and *Aspergillus parasiticus* in study of preharvest invasion of peanuts. Applied and Environmental Microbiology **52:** 1128-1131

Core Team R (2011) R: A language and environment for statistical computing. Vienna, Austria: the R Foundation for Statistical Computing

Diener UL, Cole, RJ, Sanders, TH, Payne, GA, Lee, LS, *et al.* (1987) Epidemiology of aflatoxin formation by *Aspergillus flavus*. Annual Review of Phytopathology **25:** 249-270

Dwivedi SL, Varma, TS (2002) Molecular diversity among groundnut varieties differing in drought tolerance traits. International Arachis Newsletter **22:** 34-36

Elzupir AO, Alamer, AS, Dutton, MF (2015) The occurrence of aflatoxin in rice worldwide: a review. Toxin Reviews **34:** 37-42

Gorbet DW, Tillman, BL (2009) Registration of 'Florida-07' peanut. Journal of Plant Registrations **3:** 14-18

Guo Y, Khanal, S, Tang, S, Bowers, JE, Heesacker, AF, *et al.* (2012) Comparative mapping in intraspecific populations uncovers a high degree of macrosynteny between a-and b-genome diploid species of peanut. BMC genomics **13:** 608.

Hamidou F, Rathore, A, Waliyar, F, Vadez, V (2014) Although drought intensity increases aflatoxin contamination, drought tolerance does not lead to less aflatoxin contamination. Field Crops Research **156:** 103-110

Holbrook CC, Culbreath, AK (2007) Registration of 'Tifrunner' peanut. Journal of Plant Registrations **1:** 124-124

Holbrook CC, Guo, BZ, Wilson, DM, Timper, P (2009) The U.S. breeding program to develop peanut with drought tolerance and reduced aflatoxin contamination. Peanut Science **36:** 50-53

Holbrook CC, Kvien, CK, Rucker, KS, Wilson, DM, Hook, JE, *et al.* (2000) Preharvest aflatoxin contamination in drought-tolerant and drought-intolerant peanut genotypes. Peanut Science **27:** 45-48

Holbrook CC, Timper, P, Culbreath, AK, Kvien, CK (2008) Registration of 'Tifguard' peanut. Journal of Plant Registrations **2:** 92-94

Holbrook C, Cantonwine, E, Sullivan, D, Guo, B, Wilson, D, *et al.* (2007) Development of peanut germplasm with improved drought tolerance. Proceeding of American Peanut Research and Education Society **39:** 21

Houmiao W, Yong, L, Liyun, W, Liying, Y, Jianwei, L, *et al.* (2016) Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by *Aspergillus flavus*. BMC Plant Biology **16:** 16

Klich MA, Lee, LS (1982) Seed viability and aflatoxin production in individual cottonseed naturally contaminated with *Aspergillus flavus*. Journal of the American Oil Chemists' Society **59:** 545-545

Lamb MC, Sternitzke, DA (2001) Cost of aflatoxin to the farmer, buying point, and sheller segments of the southeast United States peanut industry. Peanut Science **28:** 59-63

Lansden JA (1982) Aflatoxin inhibition and fungistasis by peanut tannins. Peanut Science **9:** 17-20

Liang XQ, Holbrook, CC, Lynch, RE, Guo, BZ (2005) beta-1,3-glucanase activity in peanut seed (*Arachis hypogaea*) is induced by inoculation with *Aspergillus flavus* and copurifies with a conglutin-like protein. Phytopathology **95:** 506-511

Liang XQ, Luo, M, Holbrook, CC, Guo, BZ (2006) Storage protein profiles in Spanish and runner market type peanuts and potential markers. BMC Plant Biology **6:** 10

Lorang JM, Tuori, RP, Martinez, JP, Sawyer, TL, Redman, RS, *et al.* (2001) Green fluorescent protein is lighting up fungal biology. Applied & Environmental Microbiology **67:** 1987-1994

Luis JM, Ozias-Akins, P, Holbrook, CC, Kemerait, RC, Jr Snider, JL, *et al.* (2016) Phenotyping peanut genotypes for drought tolerance. Peanut Science **43:** 36-48

Lynch RE, Wilson, DM (1991) Enhanced infection of peanut, *Arachis hypogaea* L., seeds with *Aspergillus flavus* group fungi due to external scarification of peanut pods by the lesser cornstalk borer, *Elasmopalpus lignosellus* (Zeller). Peanut Science **18:** 110-116

Mehan VK (1989) Screening groundnuts for resistance to seed invasion by *Aspergillus flavus* and to aflatoxin production. In “aflatoxin Contamination of Groundnut: Proceeding of International workshop”; McDonald, D, Mehan, V K, Eds; ICRISAT: Patancheru, India, 323-334

Muller V, Ame, MV, Carrari, V, Gieco, J, Asis, R (2014) Lipoxygenase activation in peanut seed cultivars resistant and susceptible to *Aspergillus parasiticus* colonization. Phytopathology **104:** 1340-1348

Nigam SN, Waliyar, F, Aruna, R, Reddy, SV, Lava Kumar, P, *et al.* (2009) Breeding peanut for resistance to aflatoxin contamination at ICRISAT. Peanut Science **36:** 42-49

Ostadrahimi A, Ashrafnejad, F, Kazemi, A, Sargheini, N, Mahdavi, R, *et al.* (2014) Aflatoxin in raw and salt-roasted nuts (pistachios, peanuts and walnuts) sold in markets of Tabriz, Iran. Jundishapur. Journal of Microbiology **7:** e8674

Rajasekaran K, Cary, JW, Cotty, PJ, Cleveland, TE (2008) Development of a GFP-expressing *Aspergillus flavus* strain to study fungal invasion, colonization, and resistance in cottonseed. Mycopathologia **165:** 89-97

Rao MJV, Upadhyaya, HD, Mehan, VK, Nigam, SN, Mcdonald, D, *et al.* (1995) Registration of peanut germplasm ICGV 88145 and ICGV 89104 resistant to seed infection by *Aspergillus flavus*. Crop Science **35:** 1717-1717

Scheidegger K, Payne, G (2003) Unlocking the secrets behind secondary metabolism: A review of *Aspergillus flavus* from pathogenicity to functional genomics. Journal of Toxicology -- Toxin Reviews **22:** 423-459

Sharaf AN, Gaber, A, Abdelhadi, AA, Ragab, AI, Korani, WA (2011) Induction, characterization and genetic analysis of *Aspergillus flavus* resistant mutants in *Arachis hypogaea.* African Journal of Biotechnology **10:** 17095-17105

Shiow Chyn H, Gow-Chin, Y, Lee-Wen, C, Wen-Jye, Y, Pin-Der, D (2003) Identification of an antioxidant, ethyl protocatechuate, in peanut seed testa. Journal of Agricultural and Food Chemistry **51:** 2380-2383

Smith MC, Madec, S, Coton, E, Hymery, N (2016) Natural co-occurrence of mycotoxins in foods and feeds and their *in vitro* combined toxicological effects. Toxins **8:** 94

Styer CH, Cloe, RJ, Hill, RA (1983) Inoculation and infection of peanut flowers by *Aspergillus flavus*. Proceeding of American Peanut Research and Education Society **15:** 91

Taheri N, Semnani, S, Roshandel, G, Namjoo, M, Keshavarzian, H, *et al.* (2012) Aflatoxin contamination in wheat flour samples from golestan province, northeast of Iran. Iranian Journal of Public Health **41:** 42-47

Timper P, Wilson, DM, Holbrook, CC, Maw, BW (2004) Relationship between *Meloidogyne arenaria* and aflatoxin contamination in peanut. Journal of Nematology **36:** 167-170

Torres AM, Barros, GG, Palacios, SA, Chulze, SN, Battilani, P (2014) Review on pre- and post-harvest management of peanuts to minimize aflatoxin contamination. Food Research International **62:** 11-19

Tsitsigiannis DI, Kunze, S, Willis, DK, Feussner, I, Keller, NP (2005) *Aspergillus* infection inhibits the expression of peanut 13S-HPODE-forming seed lipoxygenases. Molecular Plant-Microbe Interactions **18:** 1081-1089

Verma RJ (2004) Aflatoxins cause DNA damage. International Journal of Human Genetics **4:** 231-236

Waliyar F, Hassan, H, Bosc, J (1994) Sources of resistance to *Aspergillus flavus* and aflatoxin contamination in groundnut genotypes in West Africa. Plant Disease **78:** 704-708

Xue HQ, Isleib, TG, Payne, GA, Novitzky, WFG, OBrain, G (2005) Aflatoxin production in peanut lines selected to represent a range of linoleic acid concentrations. J Food Prot **68:** 126-132

Xue HQ, Isleib, TG, Payne, GA, Wilson, RF, Novitzky, WP, *et al.* (2003) Comparison of aflatoxin production in normal- and high-oleic backcross-derived peanut lines. Plant Disease **87:** 1360-1365

Xue HQ, Obrian, G, Payne, GA, Isleib, TG (2004) Evaluation of post-harvest aflatoxin production in peanut germplasm with resistance to seed colonization and pre-harvest aflatoxin contamination. Peanut Science **31:** 124-134

Zhang H, Scharfenstein, LL, Carter-Wientjes, C, Chang, PK, Zhang, D, *et al.* (2015) Lack of aflatoxin production by *Aspergillus flavus* is associated with reduced fungal growth and delayed expression of aflatoxin pathway genes. World Mycotoxin Journal **8:** 335-340

Appendix 3.A

Supplemental Materials

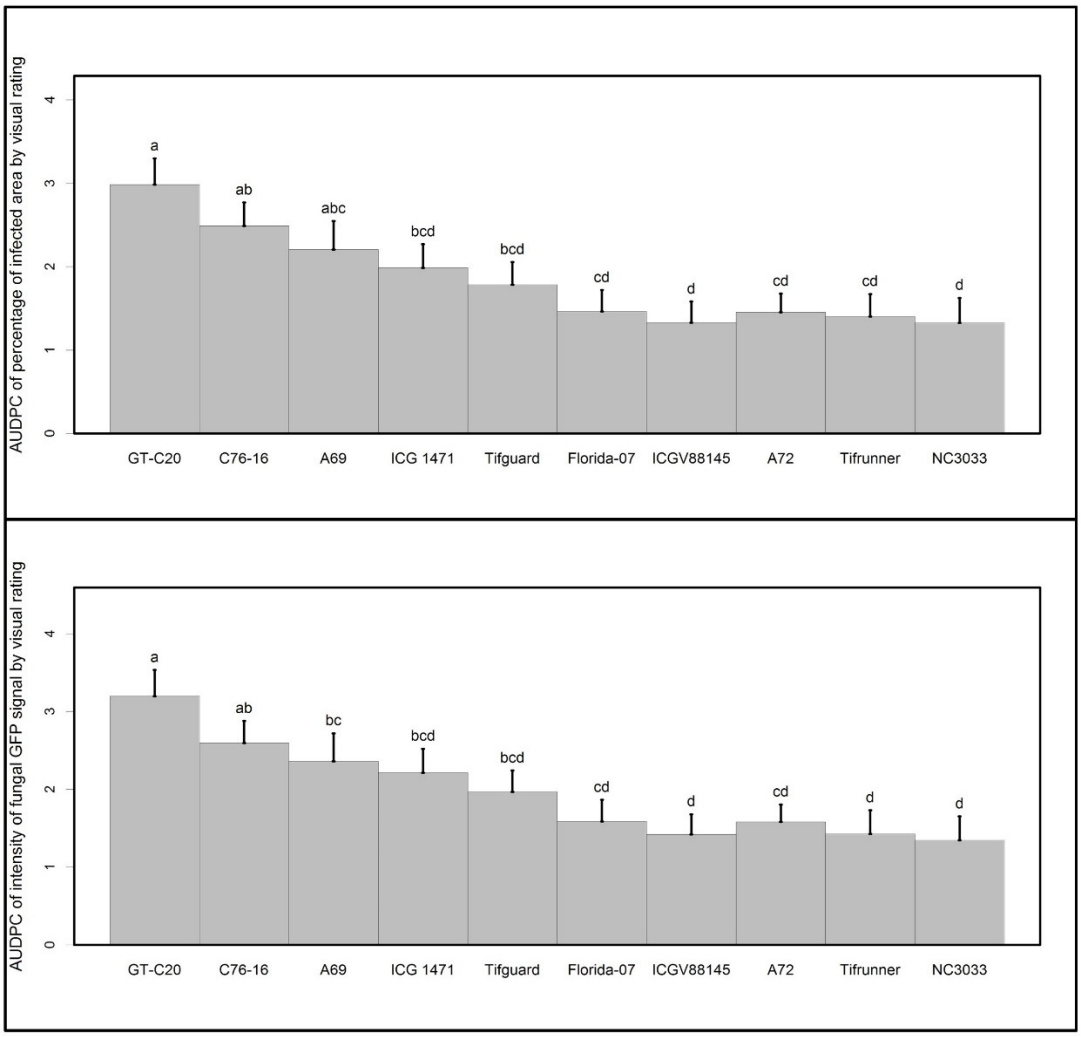


Figure S1: ANOVA analysis of log-transformed AUDPC values of the ten inoculated peanut genotypes. Percentage of infection area (upper panel) and the intensity of the fungal GFP signal (lower panel); Different letters indicate significant differences at $p < 0.05$ level determined by Tukey's range test.

CHAPTER 4

MACHINE LEARNING AS AN EFFECTIVE METHOD FOR IDENTIFYING TRUE SNPS IN POLYPLOID PLANTS

[1]W. A. Korani, J. P. Clevenger, and P. Ozias-Akins. To be submitted to *BMC Bioinformatics.*

Abstract

Single Nucleotide Polymorphisms (SNPs) have many advantages as molecular markers since they are ubiquitous and co-dominant. However, the discovery of true SNPs especially in polyploid species is difficult. Peanut is an allopolyploid, which has a very low rate of true SNP calling. A large set of true and false SNPs identified from the *Arachis* 58k Affymetrix array was leveraged to train machine learning models to select true SNPs straight from sequence data. These models achieved accuracy rates of above 80% using real peanut RNA-seq and whole genome shotgun (WGS) re-sequencing data, which is higher than previously reported for polyploids. Using methods to simulate SNP variation in peanut, cotton, wheat, and strawberry, we show that models built with our parameter sets achieve above 98% accuracy in selecting true SNPs. Additionally, models built with simulated genotypes were able to select true SNPs at above 80% accuracy using real peanut data, demonstrating that our model can be used even if real data are not available to train the models. A novel tool was developed for predicting true SNPs from sequence data, designated as SNP-ML (SNP-Machine Learning, pronounced "snip mill"), using the aforesaid models. SNP-ML additionally provides functionality to train new models not included in this study for customized use, designated SNP-MLer (SNP-Machine Learner, pronounced "snip miller"). SNP-ML is freely available for public use.

Introduction

Single Nucleotide Polymorphisms (SNPs) are a major source of variation across plant genotypes. Therefore, the demand for discovery of a large number of SNPs increased after the advent of Next-Generation Sequencing (NGS). However, the extraction of true SNPs in polyploid organisms is challenging. Cultivated peanut is an allotetraploid, which poses an exceptional challenge for the discovery of true SNPs since the two parental diploid genomes (A

and B) are very similar and the natural polymorphisms among peanut genotypes are very low (Bertioli *et al.,* 2016; Kochert *et al.,* 1991).

Using Restriction-site-Associated DNA (RAD) sequencing, a large number of SNPs were identified in peanut diploid species; however, very few SNPs were discovered in cultivated peanuts (Gupta *et al.,* 2015). Generally, the true SNP discovery in tetraploid peanut using NGS data is very low (Zhou *et al.,* 2014; Khera *et al.,* 2013; Peng *et al.,* 2016). Sliding Window Extraction of Explicit Polymorphisms (SWEEP) was developed to improve the SNP calling by filtering out the polymorphisms between the two parental subgenomes (Clevenger and Ozias-Akins, 2015). However, SNP calling in tetraploid peanut requires additional improvement. An Affymetrix SNP array was designed using the SWEEP pipeline and showed promising genotyping results among cultivated peanuts (Clevenger *et al.,* 2017). The chip showed that SWEEP identified ~40% true SNPs in tetraploid peanut genotypes. The array provided an unprecedented number of validated true and false SNP calls that can be leveraged with machine learning to increase the accuracy of selection of true SNPs straight from sequence data. The ability to have confidence in *in-silico* SNP calls gives researchers access to all avenues of sequence-based genotyping methods.

Machine learning applies sets of different algorithms that facilitate pattern recognition and classification leading to prediction by creating models using existing data (Tarca *et al.,* 2007). Machine learning algorithms are divided into two major classes; i.e. supervised and unsupervised. Supervised algorithms train previously well classified existing objects to predict the classes of new objects based on available features. Unsupervised algorithms cluster objects depending on their features without providing pre-defined classes. Both algorithms are used widely in different biological fields; e.g. coding region recognition, signal peptide prediction,

biomarker identification, disease gene recognition, metabolic network detection, and protein-protein interaction (Bostan *et al.,* 2009; Lingner, *et al.,* 2011; Swan *et al.,* 2013; Jowkar and Mansoori, 2016; Roche-Lima, 2016; Melo *et al.,* 2016). For SNP calling, neural networks were used to differentiate between true SNPs and sequence errors and this method showed promising results for human SNPs (Unneberg *et al.,* 2005). In plants, neural networks also were used to classify called SNPs as true or false positives and the approach showed a positive prediction rate of 84.8% on the testing sets of soybean (Matukumalli *et al.,* 2006). However, there has been little application of machine learning in polyploid organisms where the occurrence of more than one subgenome with high similarity to each other increases the complexity of read mapping and confounds the calling of true SNPs.

In this study, different supervised machine learners were used to improve the discovery of tetraploid peanut SNPs, utilizing the information of sequencing features and mapping data of the validated true- and false-positive SNP data sets extracted from analysis of the *Arachis* Affymetrix array. Simulated SNP variant data from peanut, cotton, wheat, and strawberry also were used to extend the functionality of machine learning to other allopolyploids. Models trained with simulated data then were used to select SNPs from real peanut data with an accuracy exceeding 80%. This result has implications for using machine learning to select true SNPs in polyploid crops where no large validation sets are available. A tool was created, SNP-MLer (SNP-Machine Learner), which allows users to train models for use in selecting true SNPs from sequence data. The user can completely customize parameter sets used in training the models or default to the complete set used to train the peanut models. The models then can be implemented in SNP-ML (SNP-Machine Learning) to select true SNPs in new data sets.

Results and discussion

Identification and evaluation of attributes for the model

A set of 18,057 validated true-positive SNPs and 26,050 false-positive SNPs were collected from the Axiom *Arachis* 58K SNP array (Clevenger *et al.,* 2017). These SNPs had been identified using SWEEP from 21 tetraploid peanut genotypes. The true-positive rate achieved was 40%, which was higher than previous efforts in peanut, but still inadequate. All of the mapping data in vcf form was available from the initial SNP calling, which provided the ability to test the hypothesis that machine learning would increase the accuracy of true SNP selection.

Seventy percent of the array-validated true- and false-positive SNPs (12,640 and 18,235, respectively) were randomly selected to train the machine learning model. Seventeen different attributes to be used in the model were calculated from sequences surrounding these SNPs (Table 1). These attributes were categorized into two groups, i.e. sequence and map features. The first machine learning approach used in biological applications was neural networks where it was used for recognizing the transcriptional start sites in Escherichia coli (Tarca *et al.,* 2007). Since that time, it has become one of the most common machine learning approaches. In addition, neural networks have many advantages such as detection of all possible interactions between predictor variables, the ability to detect complex nonlinear relationships between independent and dependent variables, and applicability for different types of data sets (Tu 1996). Therefore, we used neural networks to build our first model and to select the most effective attributes. Sequence features previously were used for genome wide *de novo* prediction purposes such as the prediction of coding regions, and to build a reliable neural network model for SNP calling in humans (Unneberg *et al.,* 2005). Thermodynamics of nucleic acids are important for diagnostic genetic markers for diseases, SNP sequencing on a genome-wide scale, designing PCR primers

and creating probes for cloning and hybridization experiments (Wu *et al.,* 2002). Since thermodynamic parameters give indications for DNA molecule stability, they were used widely to predict the DNA secondary structure (SantaLucia and Hicks, 2004). Therefore, we calculated the thermodynamic parameters deltaH, deltaS and deltaG for the SNP locations and flanking seven nucleotides (15 nucleotide segments) and incorporated the highest values from each pair of alternate SNP segments into the model. The higher value is associated with less stable states. Melting temperature (Tm) also was used in the same manner as it shares the primary components of deltaH and deltaS. Molecular weight was included since the change of a nucleotide affects the molecular weight of the DNA molecule. Lower GC contents decrease the stability of the DNA molecule. Therefore, we used the lower GC percentage of the two 15 nucleotide segments (the one with reference nucleotide versus the one with alternative nucleotide). In addition, frequency of the reference and alternate nucleotides in the sequences adjacent to the SNP location were calculated (for the seven nucleotides before and after the SNP location) and included in the model. We hypothesized that higher abundance of a particular nucleotide (reference or alternate nucleotide) would lower the probability of a true SNP.

The map features represent the quality of the mapping process and sequence data. Nine mapping parameters were selected to be used in the training model, namely quality features, i.e. mq (mapping quality) and qual (SNP quality); read abundance features, i.e. dp (depth of reads covering the SNP), af (minor allele frequency), n1 (reads with a reference base), n2 (reads with an alternate base) and n1/n2 (ratio of reference reads to alternate reads). In addition, a probability feature of homozygous reference genotypes (lg) was included. Some of these attributes, i.e. dp, n1, n2 and qual, were successfully used to create a neural network classifier for SNP calling for

soybean (Matukumalli *et al.,* 2006). Therefore, we assumed that these attributes and related features are good candidates for building a classifier in polyploids.

Twenty percent of the array-validated SNPs (3,611 true- and 5,210 false-positive SNPs) were used to test the model. Neural network models were applied to every one of the seventeen attributes independently and the relationship between false positives to false negatives was plotted for every model (Figure 4.1A). Interestingly, eight out of 17 attributes, all eight being map attributes, strongly affected the trainer (Figure 4.1A). These eight attributes were used for building one model, which showed a high reliability in classification of true- and false-positive SNPs (Figure 4.1B). The neural network score output of the testing data was applied to different neural network score cutoffs, from 0.1 to 0.9 by 0.1 intervals. The confusion matrices (predicted vs. actual) showed a gradual increase in the percentage of true negative (TN; false-positive SNP on the array and not called by SNP-ML) and decrease in the percentage of true positives (TP; true-positive SNP on the array and called by SNP-ML) as the cutoff increased (Figure 4.2). Increasing the cutoff over 0.5 dramatically decreased the percentage of TP SNPs, and also led to loss of a large number of valid SNPs (FN; true-positive SNP on the array but missed by SNP-ML). On the other hand, decreasing the cutoff below 0.5 increased the occurrence of a large number of false-positive SNPs (FP; false-positive SNP on the array and called by SNP-ML), an undesirable result. The cutoff of 0.5 showed a reasonable trade-off for recovery of the largest possible number of TP while minimizing FP and FN SNPs. These confusion matrices confirmed the efficiency of the eight selected attributes to build a reliable classifier.

Comparison among different supervised machine learning models using the selected attributes

The training data set was used to build training models by applying different supervised algorithms, i.e. Logistic Regression (LR), Discriminant Analysis (DA), K-nearest Neighbors

(KN), Naïve Bayes (NB), Decision Trees (DT) and TreeBagger (TB). The testing data set was applied for these trainers along with the neural network output of 0.5 cutoff (Figure 4.3). All models showed 60 to 80% true-positive rates relative to the number of SNPs extracted by a respective model, or between 25.0 to 33.4% of the total number of SNPs in the testing set. KN showed the highest false-positive rate and the neural network gave the lowest rate. Conversely, NB showed the lowest true-positive rate while TB produced the best rate. However, both TB and neural network showed the best trade-off between the two rates (Figure 4.4 and Supplemental Material, Table 4.S1). Therefore, we combined these two models to increase accuracy. TB was first described around 50 years after the first neural network approach was proposed (Breiman 1996). It reduces the variance among observations and avoids overfitting, which are two limitations for neural network, thus it works as a complementary model to neural network to overcome its drawbacks.

To further test the model, the remaining 10% of the original data set, 1,806 validated true-positive and 2,605 false-positive SNPs, was used as a validation set. This data set was applied to the combined NN + TB model. A total of 1,510 SNPs was extracted by the model and 1,214 of those were true-positive SNPs. Therefore, the combined model efficiency increased to 80% versus 73% (1,271 out of 1,792) and 76% (1,369 out of 1,797) of using only neural network or TB, respectively. However, 33% of validated SNPs were lost through the prediction process using the combined model.

Building models for RNA-seq

Unlike the re-sequencing data, RNA-seq provides data that measure gene expression and can produce a very high depth at specific loci (Lopez-Maestre *et al.,* 2016). The values of the attributes are different from the genomic re-sequencing data. For this reason, a specific model

was built for RNA-seq data using sequence from nine tetraploid peanut genotypes. The analysis of this data set with SWEEP produced 3,525 SNP-chip overlapped SNPs, 2,143 true and 1,382 false SNPs.

Eighty percent of the array-validated SNPs were used for training the models, 1,714 true- and 1,104 false-positive SNPs, and the remaining 20% of SNPs were used as a testing set, 429 true- and 278 false-positive SNPs. Two models were built, i.e. neural network and TB, and the scored results were combined. The combined model extracted 371 SNPs (using the cutoff of 0.5 for neural network model). Of the SNPs extracted, 328 of them were true SNPs. The accuracy of true SNP discovery was raised to 88%. However, 101 SNPs were lost (~24%).

Application in other polyploids

In the absence of validation SNP sets for allotetraploid cotton (*Gossypium hirsutum*), allohexaploid wheat (*Triticum aestivum*), or allo-octoploid strawberry (*Fragaria x ananassa*), a simulation experiment was carried out to generate allelic variation. Genome sequence for each species was downloaded and five genotypes were simulated in one of the subgenomes while keeping the other subgenomes constant. The locations of true-positive SNPs thus were known due to the *in-silico* mutation of the sequence and any other SNPs called by the program were considered false-positive. Because only one subgenome was mutated to derive the genotypes, all true SNPs were subgenome-specific. The true and false SNP calls were randomly categorized as training set (70%) and testing set (30%). The training set was used to train neural network models which were then used to select SNPs from the testing set. Simulations for all three species achieved accuracy of greater than 99% at five different sequence coverage depths (10X, 20X, 30X, 40X and 50X) (Table 4.2 and Supplemental Material, Table 4.S2). A peanut simulation was also included for comparison.

Application of simulation trained models on real data

Next, it was tested if models trained with simulated genotypes could achieve high accuracy in predicting true SNPs from real data, using the validation SNP sets available for peanut. Models that were trained in the simulations discussed above were used to select SNPs from the 21 genotypes of peanut (Supplemental Material, Table 4.S3). Each run of SNP-ML was performed three times to show variation between runs. For peanut, the models trained with simulated data were able to select true SNPs with accuracy on average of 78%. This result strongly suggests that this method can be used effectively in species where there are no large validation sets to train the models, but some reference sequence is available. This result, combined with the simulation results and results on real peanut data led us to construct a novel tool, SNP-ML, to carry out these analyses. The tool is designed to be highly flexible so that it can be used effectively in the broadest sense.

SNP-MLer

All of the models discussed in this work are provided in the SNP-ML subdirectory "/db". They include the peanut WGS and RNA-seq-trained models from real data and the models trained from simulated cotton, wheat, and strawberry data. The binary executable tool, SNP-MLer, will take two files as input, a vcf file containing true-positive SNP calls and a vcf file containing false-positive SNP calls. By default, SNP-MLer will train a neural network model using these sets of SNP calls and the eight parameters used in this work. The user has the ability with '-skip' to not use one or more (up to seven) of the parameters if they wish or use '-custom' to specify selected parameters in a comma-delimited sequence. Additionally, the user can use '-m' to train a treebagging model as well. Most importantly, the user can add customized parameters to include in the model training by invoking '-addnew1' and '-addnew2'. These

options take csv files that include one or more new parameter lists for the true-positive SNP calls (-addnew1) and the false-positive calls (-addnew2). The user also needs to add the prefix name for the new model using '-o'.

SNP-ML

If the user has trained new models using SNP-MLer or will use the models trained in this study, all models are located in the '\db' folder for use with SNP-ML. SNP-ML is the tool that will take as input (-i) a vcf file of the SNP calls of interest. It is recommended to first use SWEEP to filter most of the false-positive SNP calls, but it is not required. The name of model to be used for SNP selection (-iM) should also be given as input to SNP-ML. The program contains currently two models, "peanut_DNA" for use with WGS data, and "peanut_RNA" for use with RNA-seq data. Any new models trained with SNP-MLer by the user will be included in this folder as well. Users can submit any newly trained models to be included in new versions of SNP-ML by emailing the author. SNP-ML has similar options as SNP-MLer to skip (-skip) or customize (-custom) parameter sets for SNP prediction, and to invoke the treebagging model (-m) or add new parameters (-addnew; for custom trained models). An additional option (-c) allows the user to increase or decrease the stringency of true-positive selection from the default of 0.5. As discussed above and in Figure 4.2, increasing this cutoff will decrease false positives (decreasing selection of false SNPs) while increasing false negatives (limiting recovery of validated true SNPs) while decreasing the cutoff has the opposite effect.

The program is freely available for public use under MIT license and can be downloaded from https://github.com/w-korani/SNP-ML. A help file containing detailed information about using the program can be accessed by typing SNP-ML -h.

Conclusions

We introduce a highly reliable method for calling SNPs for polypoid species using machine learning. To have a good classifier, the most effective attributes should be determined. Many attributes were tested and the best were selected for creating the model. In addition, different supervised machine learning algorithms were tested and the best ones for the data sets, neural network and bagging, were combined. We built and tested our method on peanut, an allotetraploid for which identifying true SNPs has been difficult. The method was then used on simulated data from three other allopolyploids with different ploidy levels and achieved high accuracy. Most importantly, we showed that simulated data can be used to train models that achieve similar accuracy in selecting true SNPs using real data as do models trained with real data. The implication is that for species where there are no large validation sets available, our method can still be used to efficiently select true SNPs. With this important result in mind, SNP-MLer was developed; a tool that will train new neural network or treebagging models with user inputted data. Subsequently, SNP-ML can be used with newly trained models or included peanut models to select true SNPs for two different data set types, re-sequencing and RNA-seq. The flexibility and functionality of these tools allow the user a completely customizable experience, giving the ability to use the power of machine learning to researchers of all expertise levels.

Materials and Methods

Data sets

The re-sequencing data set was created using 21 tetraploid *A. hypogaea* genotypes described in Clevenger *et al.* (2017) and deposited publically at ncbi.nlm.nih.gov (Bio Project PRJNA340877 and Bio Samples SAMN05721179 to SAMN05721198). The RNA-seq data set has information from nine tetraploid peanut genotypes described in Clevenger *et al.* (2014,

2016a, 2016b). Validated true and false-positive SNP sets were based on testing the *Arachis* Affymetrix array with 384 peanut genotypes (Clevenger *et al.,* 2017). Mapping parameters were extracted from the vcf files used for the original design of the array. All positions of SNPs and surrounding sequence are based on the *A. duranensis* and *A. ipaensis* v1 pseudomolecules (peanutbase.org; Bertioli *et al.,* 2016).

Creating and testing the machine learning models

The data sets were prepared by R statistical software, e.g. extracting the attributes, randomly created training and testing sets and preparing fasta files for SNP flanking segments. Various toolboxes in MATLAB R2015b (the University of Georgia campus-wide site licensing agreement) were used for different purposes. Bioinformatics Toolbox was used for calculating the thermodynamic parameters, molecular weights and GC contents, Statistics and Machine Learning Toolbox was used for creating and testing the different models of supervised machine learning and Graphics functions were used for producing the bar plots and ROC (Receiver Operating Characteristic) graphs.

SNP-ML construction

We built paired (neural network and TB) specific trainer models for the two data types, WGS re-sequencing and RNA-seq. The models were built and stored in four files by a python script. In addition, three C++ classes were built, vcf.h, csv_write.h and csv.h, to process vcf and csv files. The SNP-ML main steps are illustrated in Figure 4.5. It uses C++ class vcf.h to extract the eight selected attributes from the input file, which is a vcf file of the output of SNP calling by mpileup of SAMtools, either directly or after primary filtration by SWEEP. The output is saved using the C++ class csv_write.h into a csv file, which is read by a python script to be applied to one pair of stored models (two files, one for neural network and the other for TB) depending on

the data type. The two score sets are saved to a csv file, which is read by C++ class csv.h. The scores are filtered by passing only SNPs that have a value higher than the cutoff of neural network, which can be selected by the user (the default is 0.5), and occurred in the two score sets (shared SNPs in the output of the neural network and TB score file), in case the user selects that option. The scores are stored in csv files and the corresponding SNPs are stored in a vcf file.

To extend the program applications, a second tool was designed, designated SNP-MLer (pronounced 'snip miller') to allow users to create predictors that are suitable for interested species/experimental conditions. SNP-MLer uses reading/writing approach as described above, it takes validated true-positive and false-positive vcf files as input and generates predictor models as outputs.

Both tools, SNP-ML and SNP-MLer allow the user to skip or select some of the eight attributes, and to apply new user defined attributes as csv files.

SNP-ML requirements

The script was written by C++ and python 2.7.1. C++ was used for processing the data, input, output and filtering. The binary file was created by GCC 4.1.2 that was run on Red Hat 4.1.2-55 linux system. Python was used for creating the neural network and bagging machine learning models and applying the prediction using them. Different python packages were used for these purposes, i.e. numpy-1.11.0 (SciPy.org), scipy-0.17.1 (SciPy.org), pandas-0.18.1 (pandas.pydata.org), python-dateutil-2.0 (pypi.python.org), pytz-2016.4 (pypi.python.org), scikit-learn-0.17.1 (scikit-learn.org) and pyrenn 0.1 (pyrenn.readthedocs.io).

Creating and testing models using simulated data

The pseudo molecule assembly AD1_BGI of cotton (cottongen.org; Li *et al.,* 2015), the pseudomolecule assembly of the 3B chromosome of wheat (Choulet *et al.,* 2014), the contigs of

TGACv1 wheat genome (plants.ensembl.org), the pseudomolecule assembly of Fragaria vesca Genome v1.1 (rosaceae.org; Shulaev *et al.,* 2011), and the contigs of F. nipponica Genome v1.0 (FNI_r1.1), F. nubicola Genome v1.0 (FNU_r1.1) and F. orientalis Genome v1.0 (FOR_r1.1) (rosaceae.org; Hirakawa *et al.,* 2014) were downloaded.

10,000 random loci were assigned in Chromosomes Aradu.A01, At_chr1, 3B and LG1, of peanut, cotton, wheat and strawberry, respectively. The loci were randomly mutated five times to form five synthetic genotypes using ART tool (Huang *et al.,* 2012). HiSeq 125 bp paired end sequences with different depths, 10x to 50x, were generated. The fastq produced files were mapped using BWA 0.7.10 (Li and Durbin, 2009) with default parameters to synthetic references as follows: a synthetic tetraploid reference containing Aradu.A01 and Araip.B01 chromosomes for peanut, a synthetic tetraploid reference containing At_chr1 and Dt_chr1 for cotton, a synthetic hexaploid reference containing 3B chromosome and the contigs of A and D genomes for wheat, and a synthetic octoploid reference containing LG1 chromosome and the contigs of FNI, FNU and FOR genomes for strawberry. SNPs were called using samtools mpileup 1.2 and bcftools 1.2.1 with default parameters without filtration. The SNP calling was carried out twice for every species. SNPs between two genotypes were called in the first instance and SNPs among the five genotypes were called in the second.

For each species, the SNPs located among the 10,000 loci were extracted in a separate vcf file, and considered to be True-positive (TP) SNPs. Any others identified by the program were extracted in another vcf file, and considered to be False-positive (FP) SNPs. Seventy percent of each one were randomly selected, and combined to be used as training sets, and the remaining 30% were used as testing sets for Neural Network models using Matlab R2015b (the University of Georgia campus-wide site licensing agreement).

Testing simulated data against the real data

For peanut, 21 synthetic genotypes with 10X depth were generated and SNPs were called in four batches (three with five and one with six genotypes). The simulated data were used to train the model to mimic the conditions of the real data.

All sets of the TP and FP simulated data were used to train the models, to increase the strength, and the testing sets of the real data were re-applied to these simulated models. The generation of synthetic genotypes and carrying out the machine learning (training and testing) were applied as described above.

Tables

Table 4.1: The attributes that were used for building the machine learning models. *group1: sequence features, group2: map features, bold records: the selected attributes.

Attribute abbreviation	Attribute description	group
Gc	Lowest GC contents of the segment of SNP and seven flanking nucleotides	1
Mw	Highest molecular weight of the segment of SNP and seven flanking nucleotides	1
Tm	Highest melting temperature of the segment of SNP and seven flanking nucleotides [32]	1
Dh	Highest enthalpy (in kilocalories per mole) of the segment of SNP and seven flanking nucleotides [32]	1
Ds	Highest entropy (in calories per mole-degrees Kelvin) of the segment of SNP and seven flanking nucleotides [32]	1
Dg	Highest free energy (in kilocalories per mole) of the segment of SNP and seven flanking nucleotides [32]	1
Dp	**The number of reads that cover the SNP**	**2**
n1	**The number of reads with the reference nucleotide**	**2**
n2	**The number of reads with the alternate nucleotide**	**2**
Mq	**The Root Mean Square (RMS) mapping quality**	**2**
Af	**EM estimate of the site allele frequency of the strongest non-reference allele**	**2**
qual	**Phred-scaled probability of all samples being homozygous reference**	**2**
No	SNP counts in the segment of SNP and 150 flanking nucleotides	2
Lg	**The mean of middle phred-scaled data likelihoods of all homozygous reference genotypes**	**2**
n1/n2	**The ratio of the number of reads with the reference nucleotide to the alternate one**	**2**
freq1	Frequency of the reference nucleotide in the segment of SNP and 150 flanking nucleotides	1
freq2	Frequency of the alternate nucleotide in the segment of SNP and 150 flanking nucleotides	1

Table 4.2: The SNP-ML calling accuracy on different polyploid simulated data.

Depth	Cotton	Wheat	Strawberry	Peanut
	True positive %	True positive %	True positive %	True positive %
10X	100.00	99.64	99.72	99.85
20X	99.85	99.65	99.49	99.96
30X	100.00	99.88	99.73	99.96
40X	99.93	100.00	99.57	99.92
50X	99.96	99.88	98.15	99.96

Figures

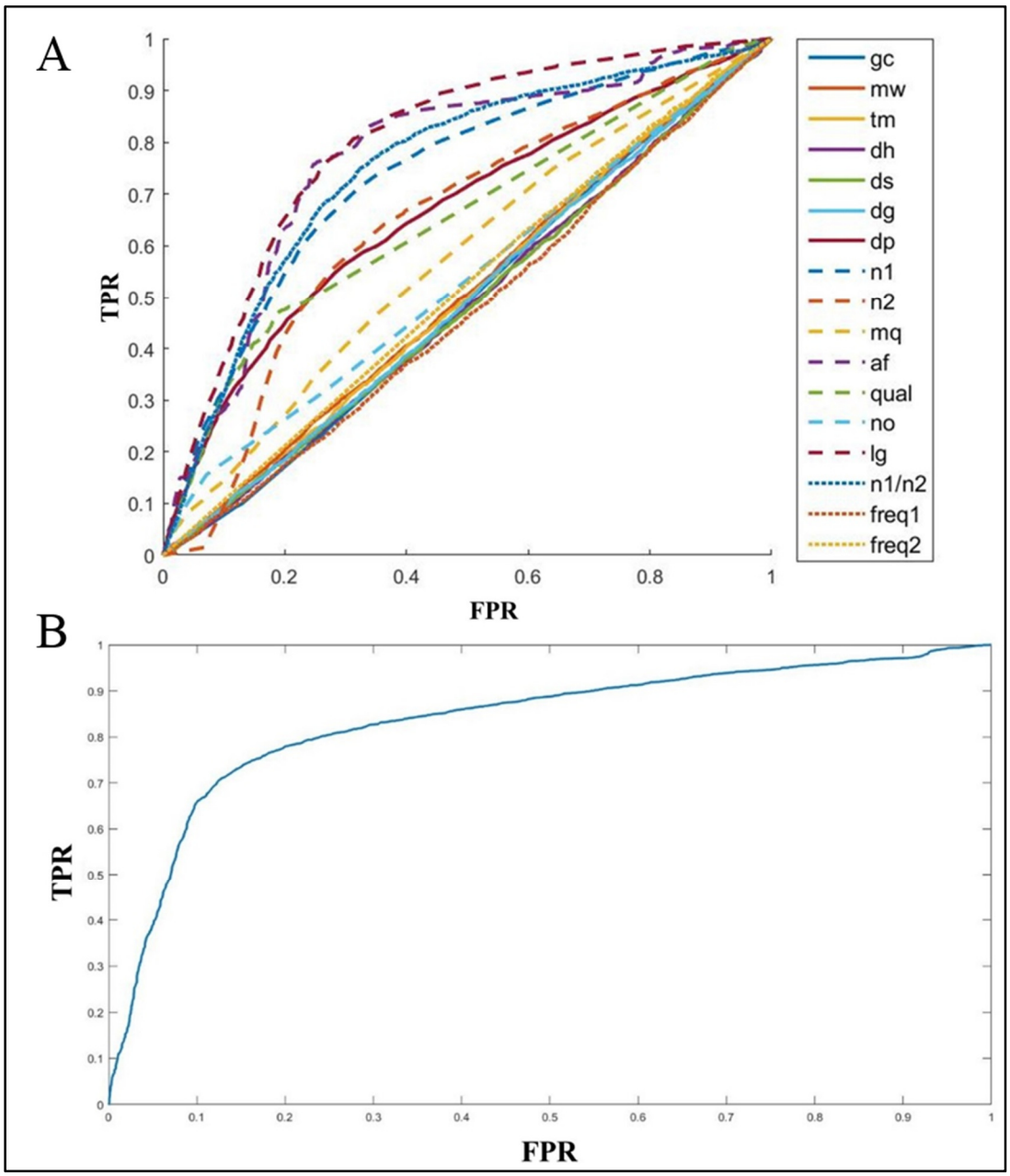


Figure 4.1: Receiver Operating Characteristic (ROC) curve of the attributes used in the neural network model trainer A. Independent applications of the 17 attributes. B. The combined application of the selected eight attributes.

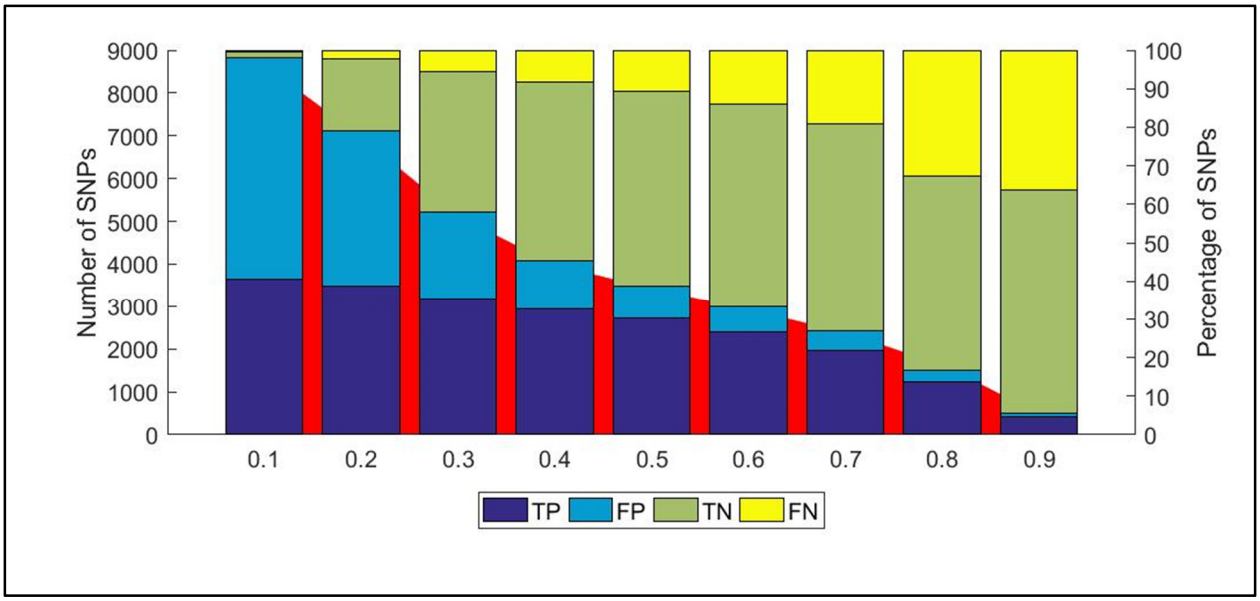


Figure 4.2: Bar plots representing the confusion matrices of the testing data using different cutoffs in neural network model, TP: True Positive (validated as a true SNP on the array and called by SNP-ML), FP: False Positive (not a true SNP according to array data but called by SNP-ML), TN: True Negative (validated as a true SNP on the array and not called by SNP-ML), FN: False Negative (not a true SNP according to array data and not called by SNP-ML). The red area shows the number of SNPs which are recognized by the model. The left Y scale presents the number of SNPs within every class and the right Y scale presents the percentage SNPs of every class to the total SNPs.

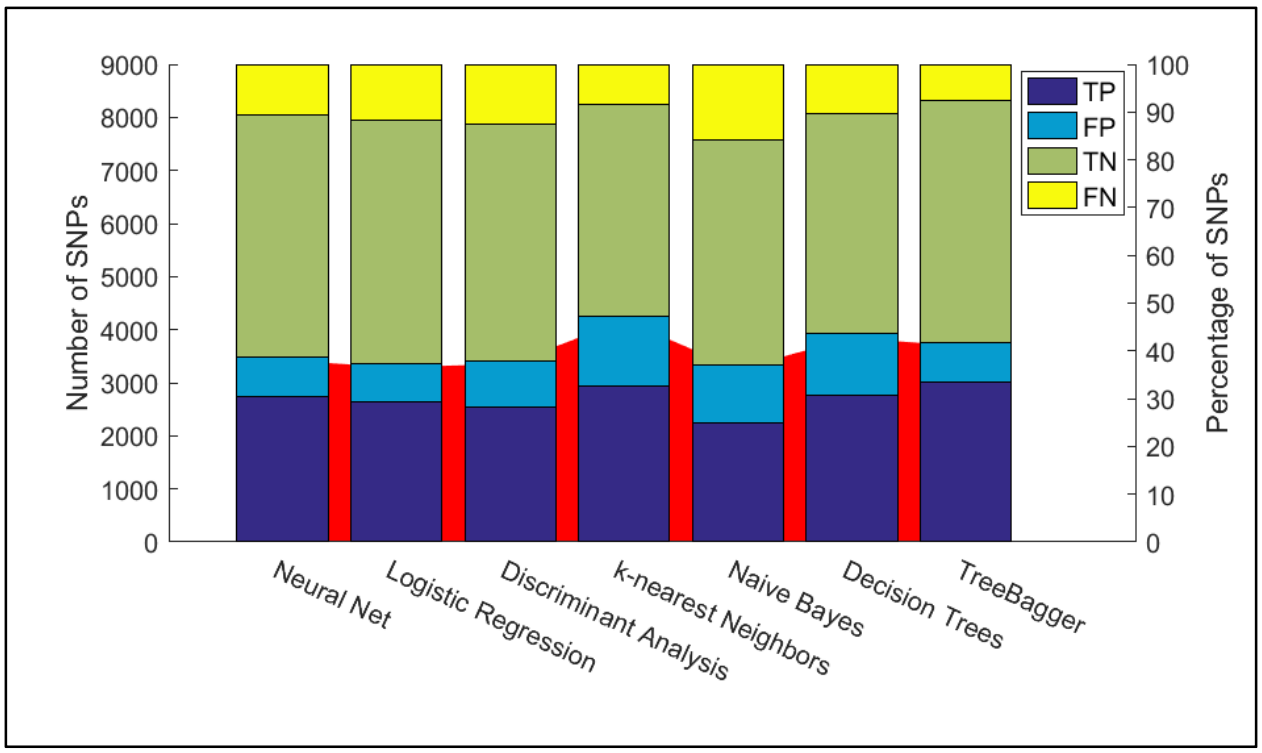


Figure 4.3: Bar plots represented the confusion matrices of the testing data using supervised machine learning algorithms, TP: True Positive, FP: False Positive, TN: True Negative, FN: False Negative. The red area shows the number of SNPs which are recognized by the model. The left Y scale presents the number of SNPs within every class and the right Y scale presents the percentage SNPs of every class to the total SNPs.

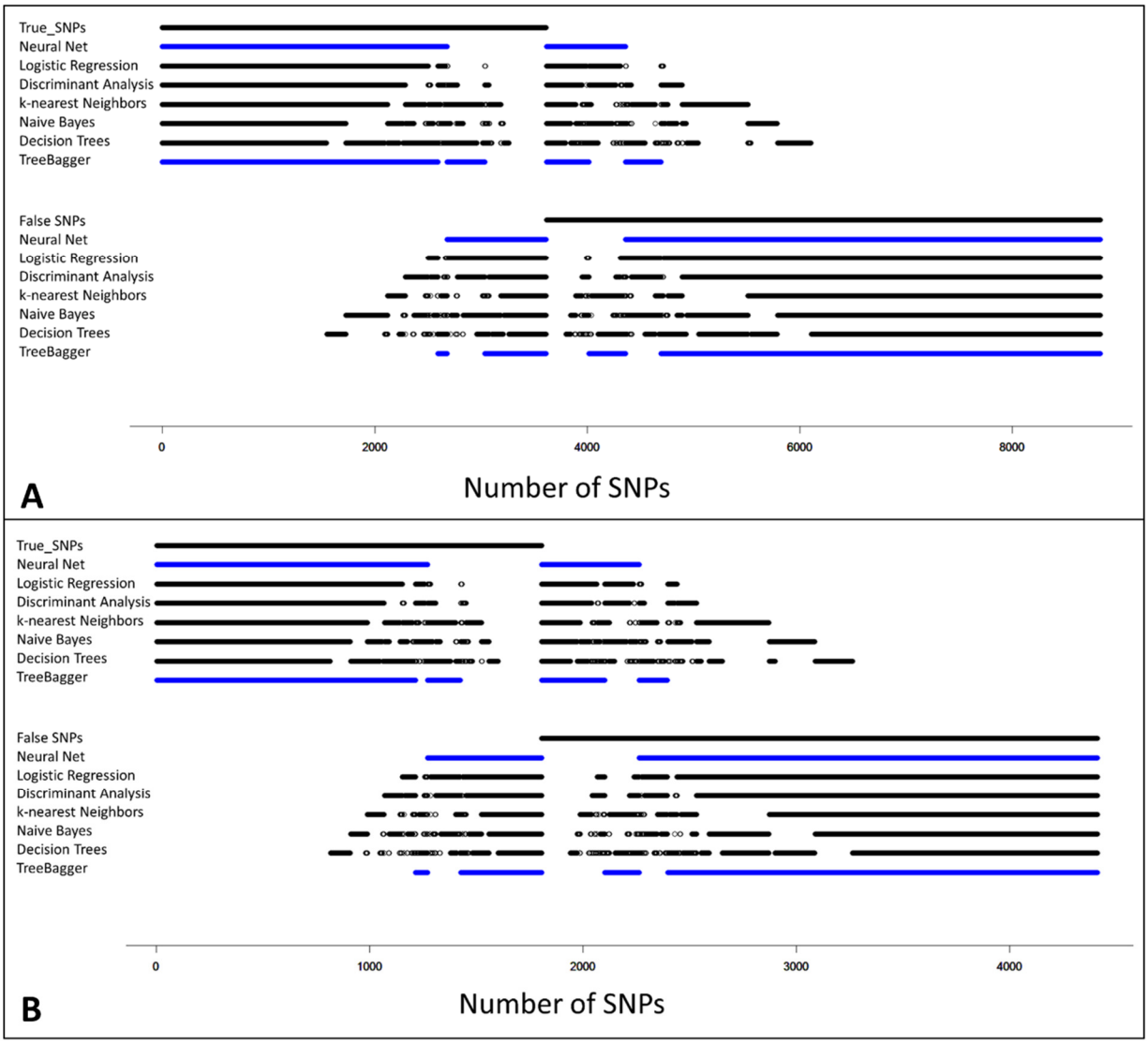


Figure 4.4: A dotplot of the trade-off combination between the different machine learning algorithms on the testing set (A) and validation set (B). Every dot shows a single true or false SNP (upper black line) and corresponding dots shows if this SNP was called using different machine learning algorithms. SNPs that were called by neural network and bagging are represented in blue dots and those that were called by other machine learning approaches are represented in black dots.

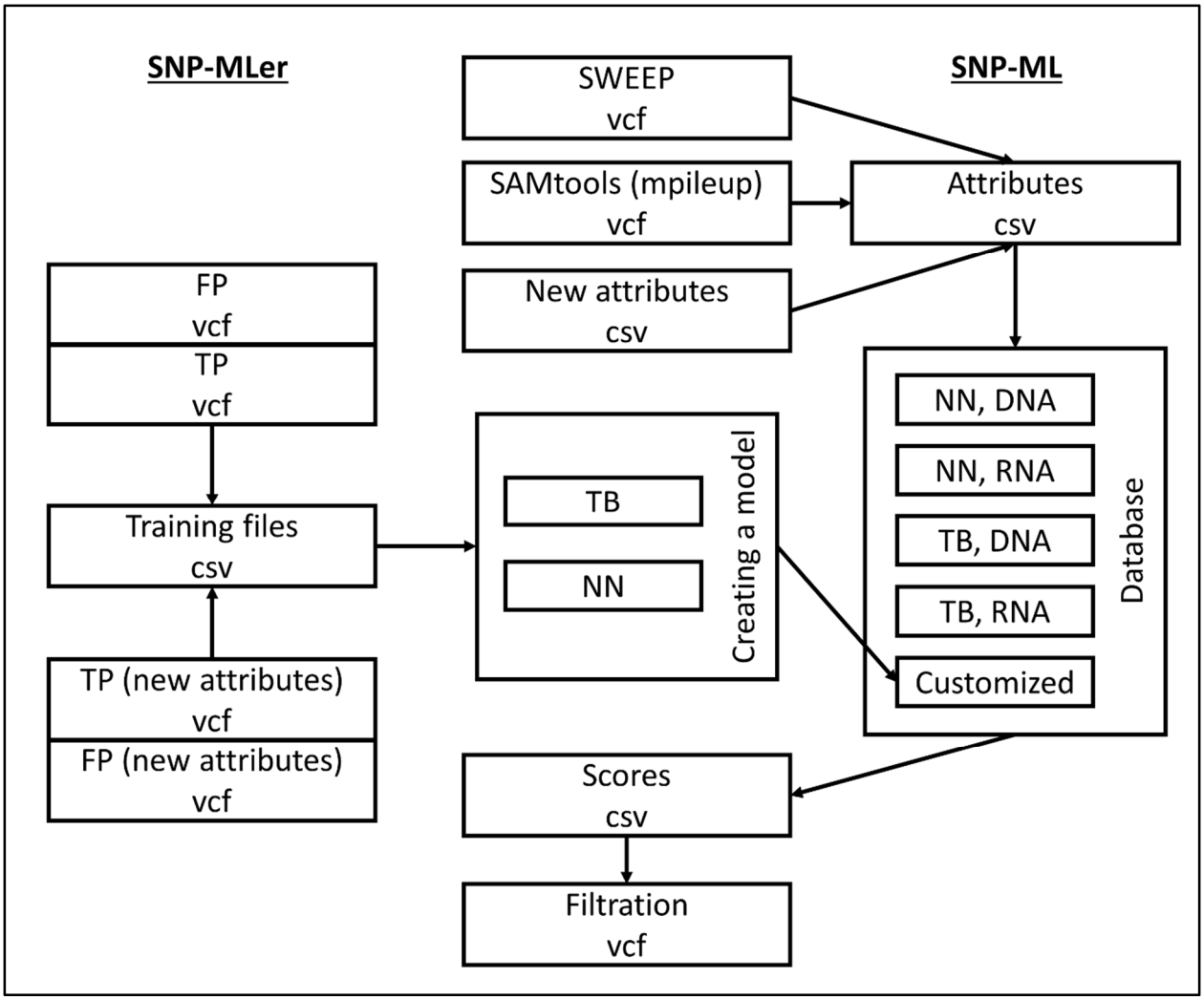


Figure 4.5: SNP-ML/SNP-MLer infrastructure.

References

Bertioli DJ, Cannon, SB, Froenicke, L, Huang, G, Farmer, AD, *et al.* (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet **48:** 438-446

Bostan B, Greiner, R, Szafron, D, Lu, P (2009) Predicting homologous signaling pathways using machine learning. Bioinformatics **25:** 2913-2920

Breiman L (1996) Bagging Predictors. Machine Learning **24:** 123-140

Choulet F, Alberti, A, Theil, S, Glover, N, Barbe, V, *et al.* (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science **345:** 1249721

Clevenger J, Chavarro, C, Pearl, SA, Ozias-Akins, P, Jackson, SA (2015) Single nucleotide polymorphism identification in polyploids: A Review, Example, and Recommendations. Mol Plant **8:** 831-846

Clevenger J, Chu, Y, Chavarro, C, Agarwal, G, Bertioli, DJ, *et al.* (2017) Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. Mol Plant **10:** 309-322

Clevenger J, Chu, Y, Scheffler, B, Ozias-Akins, P (2016) A Developmental transcriptome map for allotetraploid *Arachis hypogaea*. Front Plant Sci **7:** 1446

Clevenger J, Marasigan, K, Liakos, V, Sobolev, V, Vellidis, G, *et al.* (2016) RNA sequencing of contaminated seeds reveals the state of the seed permissive for pre-harvest aflatoxin contamination and points to a potential susceptibility factor. Toxins (Basel) **8:** 317

Clevenger JP, Ozias-Akins, P (2015) SWEEP: A tool for filtering high-quality SNPs in polyploid crops. G3 (Bethesda) **5:** 1797-1803

Gupta SK, Baek, J, Carrasquilla-Garcia, N, Penmetsa, RV (2015) Genome-wide polymorphism detection in peanut using next-generation restriction-site-associated DNA (RAD) sequencing. Molecular Breeding **35:** 145

Hirakawa H, Shirasawa, K, Kosugi, S, Tashiro, K, Nakayama, S, *et al.* (2014) Dissection of the octoploid strawberry genome by deep sequencing of the genomes of Fragaria species. DNA Res **21:** 169-181

Huang W, Li, L, Myers, JR, Marth, GT (2012) ART: a next-generation sequencing read simulator. Bioinformatics **28:** 593-594

Jowkar GH, Mansoori, EG (2016) Perceptron ensemble of graph-based positive-unlabeled learning for disease gene identification. Comput Biol Chem **64:** 263-270

Khera P, Upadhyaya, HD, Pandey, MK, Roorkiwal, M, Sriswathi, M, *et al.* (2013) Single nucleotide polymorphism–based genetic diversity in the reference set of peanut (*Arachis spp*.) by developing and applying cost-effective kompetitive allele specific polymerase chain reaction genotyping assays. The Plant Genome **6:** 1-11

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

Li F, Fan, G, Lu, C, Xiao, G, Zou, C, *et al.* (2015) Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol **33:** 524-530

Li H, Durbin, R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:** 589-595

Lingner T, Kataya, AR, Antonicelli, GE, Benichou, A, Nilssen, K, *et al.* (2011) Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. Plant Cell **23:** 1556-1572

Lopez-Maestre H, Brinza, L, Marchet, C, Kielbassa, J, Bastien, S, *et al.* (2016) SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. Nucleic Acids Res **44:** e148

Matukumalli LK, Grefenstette, JJ, Hyten, DL, Choi, IY, Cregan, PB, *et al.* (2006) Application of machine learning in SNP discovery. BMC Bioinformatics **7:** 4

Melo R, Fieldhouse, R, Melo, A, Correia, JD, Cordeiro, MN, *et al.* (2016) A machine learning approach for hot-spot detection at protein-protein interfaces. Int J Mol Sci **17:**

Peng Z, Gallo, M, Tillman, BL, Rowland, D, Wang, J (2016) Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L.). Mol Genet Genomics **291:** 363-381

Roche-Lima A (2016) Implementation and comparison of kernel-based learning methods to predict metabolic networks. Netw Model Anal Health Inform Bioinform **5:** 26

Santalucia J, Hicks, D (2004) The thermodynamics of DNA structural motifs. Annu Rev Biophys Biomol Struct **33:** 415-440

Shulaev V, Sargent, DJ, Crowhurst, RN, Mockler, TC, Folkerts, O, *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet **43:** 109-116

Sugimoto N, Nakano, S, Yoneyama, M, Honda, K (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. Nucleic Acids Res **24:** 4501-4505

Swan AL, Mobasheri, A, Allaway, D, Liddell, S, Bacardit, J (2013) Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. Omics **17:** 595-610

Tarca AL, Carey, VJ, Chen, XW, Romero, R, Draghici, S (2007) Machine learning and its applications to biology. PLoS Comput Biol **3:** e116

Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol **49:** 1225-1231

Unneberg P, Stromberg, M, Sterky, F (2005) SNP discovery using advanced algorithms and neural networks. Bioinformatics **21:** 2528-2530

Wu P, Nakano, S, Sugimoto, N (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. Eur J Biochem **269:** 2821-2830

Zhou X, Xia, Y, Ren, X, Chen, Y, Huang, L, *et al.* (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics **15:** 351

Appendix 4.A

Supplemental Materials

Table 4.S1:

Testing set						
SNPs	**3613**			**5218**		
	Correctly called true	**Miscalled true**	**Efficiency %**	**Miscalled false**	**Correctly called false**	**Efficiency %**
Neural Net	2679	934	74.1	744	4474	14.3
Logistic Regression	2587	1026	71.6	705	4513	13.5
Discriminant Analysis	2503	1110	69.3	854	4364	16.4
k-nearest Neighbors	2878	735	79.7	1298	3920	24.9
Naive Bayes	2207	1406	61.1	1056	4162	20.2
Decision Trees	2704	909	74.8	1141	4077	21.9
TreeBagger	2948	665	81.6	739	4479	14.2
Validating set						
SNPs	**1806**			**2605**		
	Correctly called true	**Miscalled true**	**Efficiency %**	**Miscalled false**	**Correctly called false**	**Efficiency %**
Neural Net	1271	535	70.4	458	2147	17.6
Logistic Regression	1218	588	67.4	451	2154	17.3
Discriminant Analysis	1177	629	65.2	512	2093	19.7
k-nearest Neighbors	1338	468	74.1	699	1906	26.8
Naive Bayes	1209	597	66.9	799	1806	30.7
Decision Trees	1276	530	70.7	690	1915	26.5
TreeBagger	1369	437	75.8	428	2177	16.4

Table 4.S2:

Cotton

				True SNPS				False positive					
Number of genotypes	**Depth**	**Total calls**	**samtools calls**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**samtools**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**Accuracy %**
2	10X	9662	7669	2332	2311	21	99.1	1993	566	562	4	99.3	**99.83**
2	20X	8035	7663	2301	2288	13	99.4	372	109	102	7	93.6	**99.69**
2	30X	8025	7679	2300	2294	6	99.7	346	107	101	6	94.4	**99.74**
2	40X	8419	7692	2326	2306	20	99.1	727	199	193	6	97.0	**99.74**
2	50X	12532	7685	2327	2315	12	99.5	4847	1432	1427	5	99.7	**99.78**
5	10X	11019	8894	2665	2656	9	99.7	2125	640	637	3	99.5	**99.89**
5	20X	10099	8892	2655	2640	15	99.4	1207	374	369	5	98.7	**99.81**
5	30X	11545	8897	2735	2725	10	99.6	2648	728	728	0	100.0	**100.00**
5	40X	11092	8918	2688	2681	7	99.7	2174	639	638	1	99.8	**99.96**
5	50X	13118	8896	2661	2653	8	99.7	4222	1274	1273	1	99.9	**99.96**

Wheat

				True SNPS				False positive					
Number of genotypes	**Depth**	**Total calls**	**samtools calls**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**samtools**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**Accuracy %**
2	10X	4210	3502	1055	530	525	50.2	708	208	177	31	85.1	**94.47**
2	20X	7950	6308	1886	1874	12	99.4	1642	499	493	6	98.8	**99.68**
2	30X	6184	5181	1536	1526	10	99.3	1003	319	56	263	17.6	**85.30**
2	40X	6093	5184	1571	1559	12	99.2	909	257	255	2	99.2	**99.87**
2	50X	4185	3538	1060	1052	8	99.2	647	195	195	0	100.0	**100.00**
5	10X	4997	2791	843	841	2	99.8	2206	656	653	3	99.5	**99.64**
5	20X	2489	1925	575	570	5	99.1	564	172	170	2	98.8	**99.65**
5	30X	6158	2795	813	804	9	98.9	3363	1034	1033	1	99.9	**99.88**
5	40X	2648	2103	629	622	7	98.9	545	165	165	0	100.0	**100.00**
5	50X	13455	2817	848	841	7	99.2	10638	3189	3188	1	100.0	**99.88**

Strawberry

				True SNPS				False positive					
Number of genotypes	**Depth**	**Total calls**	**samtools calls**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**Samtools**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**Accuracy %**
2	10X	7298	7220	2168	2158	10	99.5	78	21	15	6	71.4	**99.72**
2	20X	7300	7226	2170	2162	8	99.6	74	20	10	10	50.0	**99.54**

2	30X	7467	7280	2189	2179	10	99.5	187	51	42	9	82.4	**99.59**
2	40X	7415	7314	2196	2194	2	99.9	101	29	8	21	27.6	**99.05**
2	50X	7373	7300	2194	72	2122	3.3	73	18	18	0	100.0	**100.00**
5	10X	8771	8469	2532	2523	9	99.6	302	99	92	7	92.9	**99.72**
5	20X	8694	8459	2542	2537	5	99.8	235	66	53	13	80.3	**99.49**
5	30X	8875	8502	2556	2553	3	99.9	373	107	100	7	93.5	**99.73**
5	40X	8940	8555	2567	2564	3	99.9	385	115	104	11	90.4	**99.57**
5	50X	8735	8550	2573	2277	296	88.5	185	48	5	43	10.4	**98.15**

Peanut

				True SNPS					**False positive**				
Number of genotypes	**Depth**	**Total calls**	**samtools calls**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**samtools**	**Test set**	**Correctly called SNPs**	**uncorrectly called SNPs**	**%**	**Accuracy %**
2	10X	8694	7664	2324	1696	628	73.0	1030	284	0	284	0.0	**85.66**
2	20X	8839	7676	2312	2309	3	99.9	1163	340	0	340	0.0	**87.16**
2	30X	8804	7669	2312	1102	1210	47.7	1135	329	313	16	95.1	**98.57**
2	40X	8777	7680	2302	2201	101	95.6	1097	331	0	331	0.0	**86.93**
2	50X	8748	7674	2291	2275	16	99.3	1074	333	331	2	99.4	**99.91**
5	10X	11746	8894	2679	2661	18	99.3	2852	845	841	4	99.5	**99.85**
5	20X	10009	8898	2674	2665	9	99.7	1111	329	328	1	99.7	**99.96**
5	30X	11124	8902	2651	2635	16	99.4	2222	686	685	1	99.9	**99.96**
5	40X	9766	8905	2669	2654	15	99.4	861	261	259	2	99.2	**99.92**
5	50X	13905	8907	2694	2677	17	99.4	4998	1477	1476	1	99.9	**99.96**

Table 4.S3:

Depth	Test set	Correctly called SNPs	True SNPs not recovered	%	Test set	False SNPs not called	False positive SNPs called	%	Accuracy %
10X	16,250	2,338	13,912	14.4	23,445	22,884	561	97.6	80.65
10X	16,250	2,003	14,247	12.3	23,445	22,679	766	96.7	72.34
10X	16,250	1,151	15,099	7.1	23,445	23,164	281	98.8	80.38

CHAPTER 5

TRANSCRIPTIONAL PROFILING PROVIDES INSIGHT INTO GENETIC FACTORS REGULATING POST-HARVEST AFLATOXIN CONTAMINATION OF TETRAPLOID PEANUT

[1]W. A. Korani, Y. Chu, C. C. Holbrook, and P. Ozias-Akins. To be submitted to *Genetics*

Abstract

Aflatoxin contamination is the most challenging issue that affects peanut quality. Aflatoxin is produced by fungi belonging to the Aspergilli group, and it is known as an acutely toxic, carcinogenic and immune-suppressing class of mycotoxins. Evidence for several host genetic factors that may impact aflatoxin contamination has been reported, *e.g.*, genes for lipoxygenase, ROS and WRKY; however, their role is still tentative. Therefore, we conducted an RNA-seq experiment to differentiate gene response to the infection by *Aspergillus flavus* between resistant (ICG 1471) and susceptible (Florida-07) peanut genotypes. In addition, gene expression profiling analysis between the two genotypes was based on biological replication and was designed to reveal differentially expressed genes in response to the infection (infected vs mock-treated seeds) and the variation within treatments (high vs low contaminated seeds). Moreover, the differential expression of the fungal genes was profiled. The study revealed the complexity of the interaction between the fungus and peanut seeds as expression of a large number of genes was altered including some in the process of plant defense to aflatoxin accumulation. Since peanut is tetraploid and currently lacks a reference genome sequence, an R package, designated 'keggseq', was designed to carry out KEGG enrichment analysis for polyploids using combined subgenome sequence. Analysis of the experimental data with 'keggseq' showed the importance of alpha-linolenic acid metabolism, protein processing in endoplasmic reticulum, spliceosome and carbon fixation and metabolism pathways in the resistance. In addition, co-expression network analysis was carried out to reveal the correlation of gene expression among peanut and fungal genes. The results showed the importance of WRKY, TIR-NBS-LRR, ethylene and heat shock proteins in the resistance mechanism.

Introduction

Peanut seeds provide a suitable substrate for fungal growth and mycotoxin production, and it is the most suscptable species for aflatoxin production as compared with other oilseed crops such as soybean (Bean *et al.,* 1972). Different mycotoxins are formed on peanuts, *e.g.*, cyclopiazonic acid, zearalenone, trichothecene-toxins and aflatoxin (Chang *et al.,* 2013). The latter is the most common and destructive mycotoxin for peanut and other crops such as corn, cottonseed, rice, wheat, oat and barley (Stubblefield *et al.,* 1967; Jaime-Garcia and Cotty 2003; Mateo *et al.,* 2011; Suárez-Bonnet *et al.,* 2013; Dunham *et al.,* 2017). Aflatoxin has received widespread attention since the discovery that it was the causative agent of "Turkey X disease", a disease that killed 100,000 young turkeys on English poultry farms in 1960 (Spensley 1963). Aflatoxin products are an acutely toxic, carcinogenic and immunosuppressive class of mycotoxins affecting animals including humans (Scheidegger and Payne, 2005). In addition, aflatoxins are considered mutagenic agents as they cause oxidative damage to DNA (Verma 2004). Aflatoxins are classified in four major classes, B1, B2, G1, and G2 (Ehrlich *et al.,* 2004). However, aflatoxin B1 is the most potent and carcinogenic naturally occurring substance known (Squire 1981).

Aflatoxin is produced in agricultural products mainly through contamination by *A. flavus* or *A. parasiticus*. Not only are the fungal products harmful, the fungus *A. flavus* is an ascomycetous fungus that can infect humans, plants, animals and insects (Klich 2007). In humans, it is the second leading cause of invasive aspergillosis disease after *A. fumigatus* (Hedayati *et al.,* 2007). *A. flavus* may infect peanut and lead to aflatoxin accumulation in the field (pre-harvest aflatoxin contamination) or during storage (post-harvest aflatoxin contamination). Pre-harvest aflatoxin contamination also may occur by flower infection as the

fungus invades the flowers and then travels down the pegs to become established in the developing seeds (Styer *et al.,* 1983).

Abiotic stress is an important factor contributing to pre-harvest aflatoxin accumulation. Drought conditions and heat stress exacerbate aflatoxin contamination (Kisyombe *et al.,* 1985; Holbrook *et al.,* 2000a; Craufurd *et al.,* 2006; Nigam *et al.,* 2009). Therefore, drought tolerant genotypes, sufficient irrigation and best management practices may reduce pre-harvest aflatoxin contamination. However, understanding the resistance mechanism and developing resistant genotypes for post-harvest aflatoxin contamination are needed.

Different genetic factors that may affect *Aspergillus spp.* infection and/or aflatoxin accumulation have been proposed, however, the exact role of such factors remains unclear. Lipoxygenase (LOX) is a gene super family that encodes dioxygenases. It was found to have a critical role in many disease response mechanisms of plants such as nematode (Gao *et al.,* 2008; Ozalvo *et al.,* 2014), rust (Choi *et al.,* 2008), downy mildew (Shivakumar *et al.,* 2003; Babitha *et al.,* 2004; Babitha *et al.,* 2006) and insects (Wang *et al.,* 2008; Tang *et al.,* 2009; Yan *et al.,* 2013). However, in peanut it has received most attention for its potential role in resistance to *A. flavus*. Burow *et al.,* (2000) isolated the first peanut LOX, PnLOX1 from a seed cDNA library and found expression to be enhanced after infection by *A. parasiticus*. An opposite result was obtained by Tsitsigiannis *et al.,* (2005) while studying two more LOXs, PnLOX2 and PnLOX3, where they observed reduced expression upon infection by *A. flavus* (Tsitsigiannis *et al.,* 2005). Another two LOXs subsequently were discovered which showed various responses to *A. flavus* inoculation (Muller *et al.,* 2014). In addition, LOX expression differences have been observed upon interaction of *Aspergillus spp.* with plants other than peanut, *e.g.*, soybean (Bean *et al.,*

1972; Doehlert *et al.,* 1993; Boue *et al.,* 2005), maize (Huang *et al.,* 2013; Gao *et al.,* 2009), cottonseeds (Zeringue 1996) and almond (Mita *et al.,* 2007).

Additionally, β-1,3-glucanases, chitinases, pathogenesis-related proteins 10 and 10.1, ribosome inactivating proteins (RIPs), and zeamatin may be related to *A. flavus* resistance (Fountain *et al.,* 2014) along with WRKY transcription factors (Fountain *et al.,* 2015a). Furthermore, the drought stress-responding compounds such as reactive oxygen species (ROS) are highly associated with aflatoxin production (Jayashree and Subramanyam, 2000; Reverberi *et al.,* 2012; Fountain *et al.,* 2015b) and antioxidant enzymes are highly co-expressed with fungal growth under infection conditions (Fountain *et al.,* 2016a).

Cultivated peanut, *Arachis hypogaea*, is an allo-tetraploid (2n = 4x = 40) that was formed from spontaneous doubling of a cross between two diploid progenitors, *A. duranensis* and *A. ipaensis* (Seijo *et al.,* 2004). The whole genome sequence of tetraploid peanut is not yet available. However, high quality, well-annotated genomes of *A. duranensis* and *A. ipaensis* have been released (Bertioli *et al.,* 2016; https://peanutbase.org). The two subgenomes together have a size of ~2.7 Gb and potentially express 88,876 predicted proteins. The whole genome sequence of *A. flavus* also has been released (https://www.aspergillusflavus.org/genomics/). The genome is 40 Mbp containing 13,478 predicted genes on 8 chromosomes. Aflatoxin biosynthesis is encoded by a 70 Kbp gene cluster and has been extensively studied for *A. flavus* and *A. parasiticus* (Yu *et al.,* 2004; Ehrlich *et al.,* 2005; Georgianna and Payne, 2009). Although only these two fungi are responsible for aflatoxin production in food products, the cluster region is conserved with other species such as *A. bombycis* and *Emericella astellata* (Amaike and Keller 2011). The aflatoxin biosynthetic pathway is sensitive to environmental conditions, *e.g.*, temperature, stress, lipids

and salts (Bhatnagar *et al.,* 2003), which makes breeding for resistance to aflatoxin production challenging.

In this study, we utilized the published peanut and *A. flavus* genomes to study the genes and gene networks that respond to *A. flavus* infection and are differentially expressed during fungal interaction with resistant vs susceptible peanut genotypes. Extended analysis comprising self-organizing maps, GO term enrichment, KEGG enrichment, and co-expression network analysis was conducted.

Materials and Methods

Plant material and infection

ICG 1471 and Florida-07 seeds were collected from the field in the season immediately preceding the experiment. Thirty seeds from each genotype were inoculated with the fungal spores, alongside 10 mock-treated seeds. The seeds were harvested after 16 hours (16 HAI, Hours After Inoculation), 32 and 64 HAI time points. The experiments were conducted using randomized complete block (RCB) designs (10 seeds per block). Every individual seed was ground in liquid nitrogen and divided equally into three aliquots. The first portion was used for GFP quantification, a second for aflatoxin analysis and the third for RNA-seq analysis. Sterilization, all infection procedures, GFP and aflatoxin analysis were carried out according to the methods described in our previous work (Chapter 3). A t-test was used to test the differences in GFP expression and aflatoxin contamination between the two genotypes under infection conditions for every time point (R v3.2.2) (R Core Team, 2014).

RNA extraction

For every time point and genotype, the third pulverized portion of six mock-treated seeds and six infected seeds, three with high and three with low aflatoxin contamination, were used for

RNA extraction using Qiagen RNeasy Plant Mini kit (QIAGEN Inc. Valencia, CA) according to the manufacturer's instructions. The quality of RNA was checked with an Agilent 2100 Bioanalyzer (Georgia Genomics Facility, University of Georgia, Athens, GA).

RNA sequencing

DNA was eliminated from the extracted RNA using DNase I, Amplification Grade (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Seventy-two RNA libraries were constructed using KAPA RNA-Seq library Preparation kit (KR0934-v1.13; Kapa Biosystems, Wilmington, MA, USA) and the Illumina set B indexes (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The integrity analysis and quantification of the libraries were carried out using an Agilent 2100 Bioanalyzer and Qubit 2.0 Fluorometer (Georgia Genomics Facility, University of Georgia, Athens, GA). Sequencing was done on an Illumina HiSeq2500 (Illumina, San Diego, CA, USA) in six lanes with 12 samples pooled per lane (HudsonAlpha Institute for Biotechnology. Huntsville, AL, USA).

Differential expression analysis

The sequence quality for all libraries was determined using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, v. 0.11.4 2015). Trimmomatic v0.36 (Bolger *et al.,* 2014) was used to trim the low quality bases and filter out low quality sequences. The cleaned paired-end reads were aligned to a Bowtie (v1.1.0, Langmead *et al.,* 2009) indexed peanut synthetic tetraploid reference genome, containing the genomes of *A. duranensis* and *A. ipaensis* (Bertioli *et al.,* 2016, https://peautbase.org/), using Tophat v2.0.14 (Trapnell *et al.,* 2009). Only the cleaned paired-end reads of infected libraries were aligned to a Bowtie (v1.1.0, Langmead *et al.,* 2009) indexed *A. flavus* NRRL3357 reference genome (NCBI, txid5059). The raw counts were calculated using HTSeq v0.6.1p1 (Anders *et al.,* 2015).

Differential expression analysis of counts was carried out using edgeR (Robinson *et al.,* 2010). DESeq2 (Love *et al.,* 2014) and Next MaSigPro (Nueda *et al.,* 2014) were used to repeat the analysis for *in-silico* validation. Table 5.S1 shows different models used in the analysis. Two models were applied to test the differences between the genotypes (resistant and susceptible) due to the infection versus control treatments, and high vs low aflatoxin-contaminated treatments. A third model was applied to test the differences between the responses of the fungal genes during fungal growth on the two genotypes for only high aflatoxin contaminated treatments, since control treatments had no fungal growth and the low aflatoxin contaminated treatment had limited fungal growth.

Cufflinks v2.2.1 (Trapnell *et al.,* 2010) was used to calculate FPKM then the Z-score was calculated using R v3.2.2 (R Core Team, 2014). The expression profile of differentially expressed genes was clustered using Self-Organization Maps (SOM) of the kohonen package (R Core Team, 2014).

GO enrichment analysis

Libraries GenomicFeatures (Lawrence *et al.,* 2013) and biomaRt (Durinck *et al.,* 2005) were used to extract gene lengths and GO terms from annotation files, respectively. GO enrichment analysis of differentially expressed genes was implemented using GOseq v2.12 (Young *et al.,* 2010) with a correction for gene length bias.

KEGG enrichment analysis

KEGG enrichment analysis was carried using 'keggseq' package for the three models described above. The KEGG enrichment analysis for a synthetic tetraploid genome requires merging the two sub-genomes in one analysis. However, the available tools for KEGG analysis do not support combining two species. Therefore, we designed R packages to carry out this type

of analysis designated 'keggseq'. The p-value was calculated according to Yang *et al.,* (2015) within the 'keggseq' package. 'keggseq' is freely available to the public under MIT license and can be downloaded from https://github.com/w-korani/keggseq.

The package 'keggseq' provides some other advantages over the available tools. 1. It allows application of KEGG enrichment analysis for diploids or polyploids with any level of genome duplication; 2. It generates ready-to-publish plots and produces graphs of interested pathways that have differentially expressed enzymes marked; 3. It generates csv files containing detailed information of enzymes included in pathways of interest; 4. It allows editing of gene IDs if the user wants to use an annotation different from Kyoto Encyclopedia of Genes and Genomes annotation; 5. It is a run-time package since the data is downloaded directly from Kyoto Encyclopedia of Genes and Genomes database so it does not require an internal database for specific species; 6. It is step-by-step and easily implemented.

De novo assembly of transcripts

The unmapped reads of ICG 1471 controls, remaining after alignment with *A. duranensis*, *A. ipaensis* and *A. flavus* genomes, were converted back to paired-end fastq files using bamtools v2.25.0 (Barnett *et al.,* 2011) and concatenated. Trinity v. 2.0.6 (Haas *et al.,* 2013) was used to assemble the concatenated reads with normalization to maximum coverage of 50x. The transcripts were given IDs starting with RC.

The process was repeated for ICG 1471 treatments, Florida-07 controls and treatments with given IDs starting with RT, SC and ST, respectively. The four assemblies were combined and the redundant transcripts were filtered out using EvidentialGene pipeline (http://arthropods.eugenes.org/). Since the assembly contained sequence from peanut and sequences from *A. flavus*, BLAST+ (Camacho *et al.,* 2009) was used to cluster the assembly into

peanut and fungal transcripts by applying BLASTn for the transcript against NCBI nucleotide database. Transcripts that matched plant sequences were identified as peanut transcripts and those that had fungal matches were defined as *A. flavus* transcripts. The peanut filtered assembly was merged with the peanut tetraploid assembly produced by Clevenger *et al.,* (2016a). Blast2GO was used to annotate GO terms of the new transcripts (https://www.blast2go.com/). Differential expression analysis, GO and KEGG enrichment analyses were carried out as described above in the first model (differences between resistant and susceptible genotypes due to the infection versus control) (Table 5.S1).

Co-expression network analysis

Differentially expressed gene analysis of the fungal response to the infection on both genotypes was carried out separately for each genotype using edgeR (Robinson *et al.,* 2010). Since a fungal control treatment was lacking, the 16-hour treatment was used as control. In addition, differential expression analysis was carried out for both genotypes to test the treatment effect (controls vs treatments) for time points 32 and 64 HAI for each genotype separately. The Z-score fungal and peanut genes were combined in one matrix/genotype. Pearson correlation analysis was done using R v3.2.2 package (R Core Team, 2014) as described by Musungu *et al.,* (2016). Only pairs that showed expression correlation > 0.99 were loaded into cytoscape network v3.4.0 (Shannon *et al.,* 2003). As the dataset that contained the correlated paired genes of the susceptible genotype was huge, the network was clustered only for the resistant genotype using MCODE app (Bader and Hogue 2003) and then the genes that matched those of the susceptible genotype matrix were excluded from the clusters of the resistant genotype.

Results and Discussion

Fungal growth and aflatoxin accumulation

It was shown previously that the peanut genotype ICG 1471 is a strong candidate for resistance to aflatoxin accumulation upon *in vitro* inoculation of mature peanut seeds with *A. flavus* (Chapter 3). In addition, ICG 1471 has been reported as a resource for resistance to pre- (Waliyar *et al.,* 2003; Nigam *et al.,* 2009) and post- (Waliyar *et al.,* 2008) harvest aflatoxin contamination. Therefore, it was used in this study along with Florida-07, which was recognized as a susceptible genotype for both types of aflatoxin accumulation (Clevenger *et al.,* 2016b; Chapter 3).

To estimate the dynamic change in gene expression, infected seeds and their controls were harvested at three different time points, 16, 32, and 64 hours after infection (HAI). Figure 5.1 shows the interaction between aflatoxin B produced by *A. flavus* and the progression of fungal growth estimated indirectly by relative fluorescence units (RFU) of GFP protein signal. The data was in consensus with our previous observations (Chapter 3) as ICG 1471 supported a very low amount of aflatoxin per unit of fungus as compared with Florida-07 across the three time points. A t-test revealed no significant differences between the two genotypes for GFP relative fluorescence for the three time points, yet aflatoxin showed significant differences for the respective time points for all three of them.

The interaction plots revealed that not only peanut genotypes interact differently with the fungus, but also every individual seed may produce different levels of aflatoxin within the same genotype/treatment/time point. This supports previous reports that aflatoxin accumulation is very sensitive to environment (Blankenship *et al.,* 1984; Kisyombe *et al.,* 1985; Bhatnagar *et al.,* 2003; Craufurd *et al.,* 2006). The samples that were chosen for RNA-seq analysis (Table 5.S2)

were clustered in high/low aflatoxin contamination groups (Figure 5.1). Picking such highly variant samples gives a realistic representation of the biological replication; however, it increases standard deviation. Therefore, six biological replicates were used to study the differentially expressed genes due to genotypic effect. In addition, this design allows testing the effect of the variations within treatments on accumulated aflatoxin amount.

Peanut genotypic differential expression analysis

The cleaned paired-end reads that were mapped to the tetraploid peanut synthetic and the *A. flavus* genomes are presented in Figure 5.2. Except for the highly fungal contaminated libraries (treatments of 64 HAI of Florida-07), 3.3 to 9.9 million paired-end reads were correctly mapped to the peanut genome for every sample. This gave an average of 6 million paired-end reads/library and a total average of 5.6 million paired-end reads/library, including the highly contaminated libraries (which had 2.4 to 4.4 million mapped fragments/library). These results showed a reasonable depth to the peanut genome that has a total size of ~2.7 Gb (Bertioli *et al.,* 2016).

4272 genes were differentially expressed between the two genotypes (resistant vs susceptible) due to the infection by *A. flavus* (treatment vs control). Since the expression profile included the dynamic change across the three time points, some SOM clusters of these genes may have similar general trends and differ slightly in the dynamic change from one time point to another. The general pattern groups represented by SOM clusters are shown in Figure 5.S1. The clustering showed that some genes were down-regulated due to the infection in the susceptible genotype and up-regulated (Figure 5.S1A) or unaffected (Figure 5.S1B) in the resistant one and on the other hand, some genes were up-regulated due to the infection in the susceptible genotype

and down-regulated (Figure 5.S1C) or unaffected (Figure 5.S1D) in the resistant one. A fifth group was up-regulated in both genotypes but more so in the resistant genotype (Figure 5.S1E).

GO enrichment analysis of the differentially expressed genes generated 146 significant GO terms out of the GO terms found in the annotation of the two subgenomes of peanut. The 20 most significant GO terms (Figure 5.3A) included several for protein processing, *i.e.,* protein polymerization, protein complex, unfolded protein binding, protein folding, protein heterodimerization activity, protein binding. The latter had 529 differentially expressed genes. On the other hand, KEGG enrichment analysis (Figure 5.3B) only generated five significant pathways, *i.e,*, alpha-linolenic acid metabolism, protein processing in endoplasmic reticulum, spliceosome and carbon fixation and carbon metabolism (Figures 5.S2 to S6, respectively).

The environmental effects, especially water deficit and heat stress, were reported to affect pre-harvest aflatoxin production by *Aspergillus spp.* on peanuts, which can be extended to also affect post-harvest aflatoxin production (Blankenship *et al.,* 1984; Kisyombe *et al.,* 1985; Sanders *et al.,* 1985; Waliyar *et al.,* 2003; Craufurd *et al.,* 2006; Guo *et al.,* 2006; Cotty and Jaime-Garcia, 2007; Shan *et al.,* 2011). This increases the variation of aflatoxin production within treatments. Therefore, the samples were clustered in two groups of high and low aflatoxin accumulation levels in both genotypes. To test the differentially expressed genes between the two genotypes due to the variation within the biological samples, the differential expression analysis was repeated for only high- vs low-contamination groups (Figure 5.1), excluding controls. The analysis generated 3485 differentially expressed genes. The SOM clustering (Figure 5.S7) and GO term enrichment analysis revealed 194 significant GO terms. Out of the 20 most significant GO terms (Figure 5.4A) the most abundant and highly significant were metabolic process, catalytic activity, oxidation-reduction process, protein phosphorylation,

protein kinase activity and ATP binding with the latter having 372 differentially expressed genes. Therefore, it was the GO term with the most differentially expressed genes after protein binding (414 differentially expressed); however, protein binding had a much lower significance level (0.02) compared with ATP binding (1.6^{-12}). Unlike the KEGG enrichment analysis of the genotypic differences irrespective of aflatoxin level which generated only five significant pathways, genotypic differences taking into account high- vs low-aflatoxin showed 14 significant pathways (Figure 5.4B), *i.e.,* the metabolism pathways of galactose, phenylalanine, taurine and hypotaurine, glutathione, linoleic acid, butanoate, and thiamine, the biosynthesis pathways of ubiquinone and other terpenoid-quinone, phenlypropanoid, flavonoid, and isoflavonoid, in addition to metabolic pathways, biosynthesis of secondary metabolites and circadian rhythm.

To mitigate post-harvest aflatoxin contamination of peanut, the resistance mechanisms have to be understood. Therefore, genes that are responsive to the infection (infection vs mock-treatment) have to be studied in addition to those genes that may be affected by the variations within treatments (high vs low contaminated clusters). The large number of significant differentially expressed genes and GO terms generated by both analyses revealed the complexity of the interaction between *A. flavus* and peanut in terms of aflatoxin production and the extreme effect of individual seed physiology on the process. In addition, it can be misleading, even after clustering/grouping, to assign a resistance function to a particular gene or small group of genes of interest; however, KEGG enrichment analysis gives a clue to pathways that control either the resistance or the physiological response.

As only five KEGG pathways were identified as significant for resistance (excluding individual seed effects), it can be assumed that they are the main keys controlling the defense

mechanism in ICG 1471. The carbon and alpha-linolenic acid pathways may be of particular importance for resistance.

The most interesting significant pathway is alpha-linolenic metabolism, which contains different components that have been reported as important or responsive for biotic and abiotic stresses of plants. The pathway catabolizes alpha-linolenic acid into different important products including jasmonates, *i.e.,* jasmonate and methyl-jasmonate. Jasmonates are synthesized though this pathway in two main cellular compartments, the chloroplast where alpha-linolenic acid is converted to 12-oxo-phytodienoic acid (12-OPDA) in a process initiated by chloroplast 13S-lipoxygenase (Bell *et al.,* 1995), and the peroxisome where 12-OPDA is localized and converted to jasmonates (Stintzi 2000).

Lipoxygenases were documented to play a role in *Aspergillus spp.* infection and the subsequent aflatoxin contamination of different crops such as peanut (Burow *et al.,* 2000; Tsitsigiannis *et al.,* 2005; Kumari *et al.,* 2012; Muller *et al.,* 2014), soybean (Bean *et al.,* 1972; Doehlert *et al.,* 1993; Boue *et al.,* 2005), maize (Gao *et al.,* 2009; Huang *et al.,* 2013), cottonseeds (Zeringue 1996) and almond (Mita *et al.,* 2007). Figure 5.5 shows nine lipoxygenases that were found among the differentially expressed genes; eight are predicted to generate 13-hydroxyperoxides and six have features of plastidial enzymes, whereas one was not classified (Chapter 2).

Additionally, both 12-OPDA and jasmonates were documented to play independent roles in the wound response of *Arabidopsis* and they can change the expression of the same and different responsive genes (Taki *et al.,* 2005; Sham *et al.,* 2015). In addition, numerous reports showed the importance of jasmonates in plant response to biotic and abiotic stresses, *e.g.*, insects (Thaler *et al.,* 1996; McConn *et al.,* 1997; Kessler *et al.,* 2004), fungi (Vijayan *et al.,* 1998;

Thomma *et al.,* 2000; Zeneli *et al.,* 2006; Mei *et al.,* 2006), and wounding (Baldwin *et al.,* 1997) and during development (Creelman and Mullet, 1997). In particular, methyl-jasmonate was found to delay spore germination, and inhibit mycelial pigment formation and aflatoxin poduction of *A. flavus* (Goodrich-Tanrikulu *et al.,* 1995). Interestingly, it was found to enhance aflatoxin production by *A. parasiticus* (Vergopoulou *et al.,* 2001). However, Meimaroglou *et al.,* (2009) showed that methyl-jasmonate might enhance or reduce aflatoxin production by *A. parasiticus* depending on its concentration. Moreover, fungal-pathogens can manipulate, enhance or suppress jasmonate signaling in plant-hosts (Zhang *et al.,* 2017).

The alpha-linolenic acid metabolism pathway also can catalyze alpha-linolenic acid by 9S-lipoxygenase to different important products such as 10-oxo-11-phytodienoic acid (10-OPDA) which has a high phytotoxicity (Sherif *et al.,* 2016).

Another route through the alpha-linolenic acid metabolism pathway produces 8,11,14-heptadecatrienoic acid using alpha-dioxygenase 1 (DOX1) without lipoxygenase activity. Both enzyme and its product were documented to increase in tobacco during interaction with *Pseudomonas syringae* pv *syringae* (Hamberg *et al.,* 2003). In addition, DOX1 was up-regulated in *Arabidopsis* after 12-OPDA treatment (Sham *et al.,* 2015) which functioned to protect the plant from oxidative stress (De Leon *et al.,* 2002).

Therefore, it can be concluded that, regardless of the direction of the effect, jasmonates and 12-OPDA produced by the alpha-linolenic pathway, in addition to other pathway components, have an important role in aflatoxin biosynthesis of *Aspergillus spp*. ICG 1471, as a resistant genotype, may regulate the synthesis of jasmonates to reduce aflatoxin production. On the other hand, Florida-07, as a susceptible genotype, does not respond to infection with jasmonate production, thereby resulting in elevated aflatoxin accumulation.

Protein processing in endoplasmic reticulum also was a significant pathway that may contribute to the resistance mechanism. However, its role may be integrated with the alpha-linolenic pathway since endoplasmic reticulum contributes to the formation of peroxisomes (Hoepfner *et al.,* 2005). Fountain *et al.,* (2016b) showed that alternative carbon sources have different effects on aflatoxin and kojic acid production; kojic acid has an important role in remediating damage resulting from Reactive Oxygen Species (ROS). These results reveal the importance of carbon fixation and metabolism pathways in aflatoxin production by *Aspergillus spp.* and protection of the fungus against oxidative damage.

As a synthetic reference genome of tetraploid peanut was used in our differential expression analysis, some genes/transcripts having roles in resistance to aflatoxin accumulation may not be represented within the two sub-genomes. Therefore, *de novo* assemblies were constructed to capture such novel transcripts. Four assemblies were created for ICG 1471 control and treatments, and Florida-07 controls and treatments, which generated 61176, 67813, 90543 and 109068 total transcripts, respectively, and among them 413, 457, 551 and 505 were new transcripts, respectively.

To validate the genes and pathways involved in resistance, analysis of differential expression between peanut genotypes, was repeated using a combined reference transcriptome (88,626 transcripts) that included the 2026 novel transcripts (Supplemental materials: afla_new_transcripts.fasta) and the previously published tetraploid peanut transcriptome (86,600 transcripts) (Clevenger *et al.,* 2016a). The differential expression analysis generated 3879 significant genes. The expression profile of the novel transcripts is given (Figure 5.S8), out of the 2026 novel genes, 66 were differentially expressed. GO enrichment analysis identified 406, out of 8530, significant GO terms (Figure 5.S9A). Most significant GO terms resulting from

genomic analysis (using predicted transcripts) also were significant in transcriptomic analysis. However, interestingly, KEGG enrichment analysis generated four of the same significant pathways as with genomic analysis (Figure 5.S9B), except for alpha-linolenic acid metabolism, which was near the significance threshhold with a q-value of 0.06. These outputs confirmed the key role of these five pathways and their respective genes in peanut resistance to aflatoxin produced by Aspergilli.

The relatively large number of biosynthetic and metabolic KEGG pathways (13 out of 14) that were significant due the physiological condition (Figure 5.4B), may help interpret published results of variation underlying aflatoxin production in peanut. For instance, Xue *et al.,* (2005) reported that peanut lines with elevated linoleic and low oleic acid are more resistant to post-harvest aflatoxin production. However, Holbrook *et al.,* (2000b) found that oleic/linoleic levels do not have a significant effect on pre-harvest aflatoxin production. In this analysis, the linoleic acid pathway was found among the significant pathways (Figure 5.4B). Therefore, the ratio of oleic/linoleic acid may increase variation among individuals yet not be the primary determinant of resistance or susceptibility.

Differential expression of fungal genes and co-expression network analysis

The interaction between peanut seeds and Aspergilli encompasses responsive pathways inside the plant and those inside the fungi, and genes regulating the signaling between organisms. Furthermore, some fungal genes may be affected differentially by growth of the fungus on different peanut genotypes. To investigate host-pathogen interaction differences, differential expression analysis was carried out for fungal genes, which generated 1197 significant genes. SOM clusters of the expression patterns of these genes (Figure 5.S10) and GO term enrichment analysis showed 97 significant GO terms, out of 4918 total (Figure 5.6A). KEGG enrichment

analysis identified eight significant pathways (Figure 5.6B), one interconversion pathway (pentose and glucuronate) one degradation pathway (valine, leucine and isoleucine), and six metabolic pathways (fructose/mannose, galactose, starch/sucrose, glycerolipid, carbon and metabolic pathways). Interestingly, seven of these pathways include carbohydrate processing. These results are in agreement with previous studies that showed changes in aflatoxin production by *A. flavus* or *A. parasiticus* using different sugar sources (Abdollahi and Buchanan 1981; Davis and Diener 1986). Growth of *A. flavus* on ICG 1471 may result in the production of different sugars than growth on Florida-07, leading to lower aflatoxin production by the fungus. A further consequence may be reduced kojic acid production and subsequent increase in the sensitivity of the fungus to ROS. These two hypotheses need to be validated in future work.

To further investigate the differential response of fungal genes due to host genotype, co-expression network analysis based on Pearson correlation was conducted (Figure 5.7). 1265 and 1111 differentially expressed peanut and fungal genes, respectively, were found in *A. flavus*/ ICG 1471 interaction (for the time points of 64 and 32 HAI for the comparison of treatments vs controls), which formed a matrix of 0.5 million correlated pairs (edges). More (6795 peanut and 1265 fungal genes) were differentially expressed in *A. flavus*/Florida-07 interaction, which created a huge matrix of 14 million correlated pairs (edges). Figure 5.7 shows the interspecies peanut/*A. flavus* co-expression network for ICG 1471 (A) and Florida-07 (D). The MCODE cluster analysis of the ICG 1471 co-expression network generated 45 clusters (Supplemental materials, ICG 1471_sub_networks.txt). The most interesting clusters (sub-networks) were 1 and 15; sub-network 1 had 1037 peanut genes and 8 *A. flavus* genes (Figure 5.7B), including gene10037 (AflNA) and the sub-network 15 had 28 peanut genes and only one *A. flavus* gene, gene10043 (AflH) (Figure 5.7C). AflNA and AflH encode two enzymes regulating the aflatoxin

biosynthetic pathway (very upstream enzymes), *i.e.,* averantin hydroxylase (EC 1.14.13.174) and versiconal hemiacetal acetate reductase (EC 1.1.1.353), respectively.

Out of the 1037 and 28 ICG 1471 peanut genes whose expression was highly correlated with gene10037 and gene10043 of *A. flavus*, 640 and 24 genes were not found in the Florida-07/*A. flavus* matrix. Among these genes, eight WRKY family transcription factors, nine TIR-NBS-LRR, six ethylene signaling and one heat shock protein were up-regulated, and expression was correlated with gene10037. One heat shock and an ethylene signaling gene were up-regulated and expression was correlated with gene10043. Figure 5.8 represents the expression profile of these genes. Although gene expression was up-regulated in both genotypes for all genes, ICG 1471 genes were co-expressed with gene10037 or gene10043 of *A. flavus.*
Many plant disease resistance genes encode NBS-LRR proteins (McHale *et al.,* 2006; Sekhwal *et al.,* 2015). Ethylene signaling genes were significantly up-regulated in response to *A. flavus* infection of maize (Musungu *et al.,* 2016). Heat shock proteins may play a role in plant defense by affecting R protein stability and their regulation (Lee *et al.,* 2012). WRKY transcription factors were differentially expressed in the response of resistant and susceptible genotypes of maize to infection by *A. flavus* (Fountain *et al.,* 2015a). In addition, they were found to affect on the ethylene-jasmonate-mediated defense (Birkenbihl *et al.,* 2012), plant response to heat stress (Li *et al.,* 2010) and defense triggered by jasmonates, either negatively (Gao *et al.,* 2011) or positively (Journot-Catalino *et al.,* 2006). Therefore, these eight WRKY genes may be important in controlling jasmonate defense mechanisms. In addition, the high correlations between expression of these genes in ICG 1471 and gene10037 of *A. flavus* reveals their importance in the defense mechanism and suggests that they may be involved in regulation of the alpha-linolenic acid metabolism pathway.

In-silico validation of differential expression analysis

In this study, different complex factors were involved in the RNA-seq experiment, *e.g.,* genotypic effect, physiological effect, *A. flavus* infection and time-course dynamic change. Therefore, three analytical models were compared (Figure 5.9). Across all analyses, DESeq2 showed similar results to edgeR for identifying differentially expressed genes. On the other hand, Next maSigPro identified many genes that were not discovered by the other two methods and failed to extract many other genes that were determined to be differently expressed by the other two methods.

EdgeR is one of the most common methods used for differential expression analysis of RNA-seq data. However, it is not a standard method to handle the time course experiments as it uses a negative binomial model, which deals with time points as independent factors (Robinson *et al.,* 2010). Methods have been designed to account for time course experiments that used different models such as Next maSigPro (polynomial regression model) (Neuda *et al.,* 2014), DyNB (non-parametric gaussian processes regression negative binomial likelihood model) (Aijo *et al.,* 2014), TRAP (beta-negative binomial model) (Jo *et al.,* 2014), SMARTS (input-output hidden Markov model) (Wise and Bar-Joseph 2014), EBSeq-HMM (empirical Bayes mixture model) (Leng *et al.,* 2015), FunPat (different distribution models) (Savania *et al.,* 2015) and timeSeq (negative binomial mixed-effect model) (Sun *et al.,* 2016). All these methods had limitations and none was standardized to this type of analysis. Next maSigPro was initially designed to analyze microarray data using polynomial regression and later was updated to handle RNA-seq data (Neuda *et al.,* 2014). This method relies on R^2 factor to extract the significant differentially expressed genes, which is considered a drawback since the threshold is user-defined (Spies and Ciaudo 2015). Although both edgeR and DESeq2 use a negative binomial

model, DESeq2 has different implementation, tests and normalization (Love *et al.,* 2014). Both gave a reasonable level of analysis validation.

Conclusions

The objective of this study was to identify genetic factors and biochemical pathways that function to limit aflatoxin production in resistant peanut genotypes. Differential expression analysis revealed five important biochemical pathways regulating resistance. In addition, results captured pathways involved in physiological interaction with aflatoxin formation and the fungal pathways that are differentially affected by fungal infection and aflatoxin production on resistant vs susceptible peanut genotypes. The study highlighted the critical role of the alpha-linolenic acid metabolism pathway and certain WRKY genes likely regulating the jasmonate-based defense pathways to mitigate aflatoxin production. These results provide key information and identify materials that can be used in breeding of peanut lines resistant to aflatoxin production.

Figures

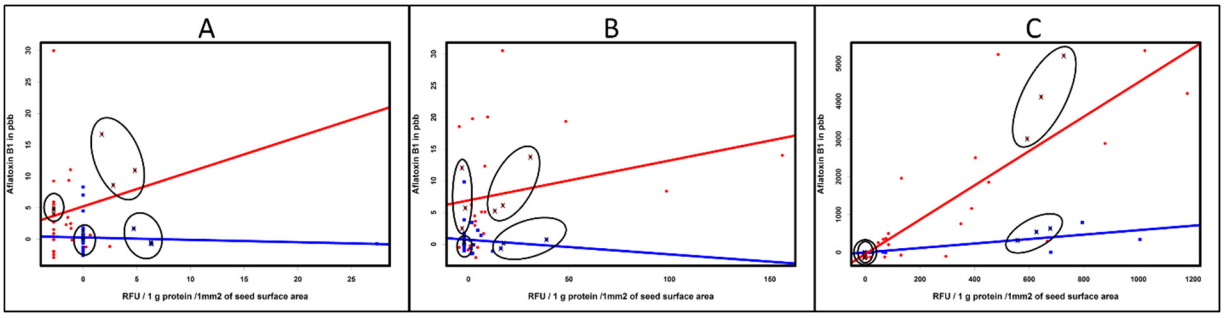


Figure 5.1: Interaction between GFP signals and aflatoxin levels for 16 (A), 32 (B), 64 (C) HAI; the red line and points represent Florida-07 data; the blue line and points represent ICG 1471 data; the cross marks show the samples that were chosen for RNA-seq analysis, the circles/ovules reveal the high/low contaminated clusters.

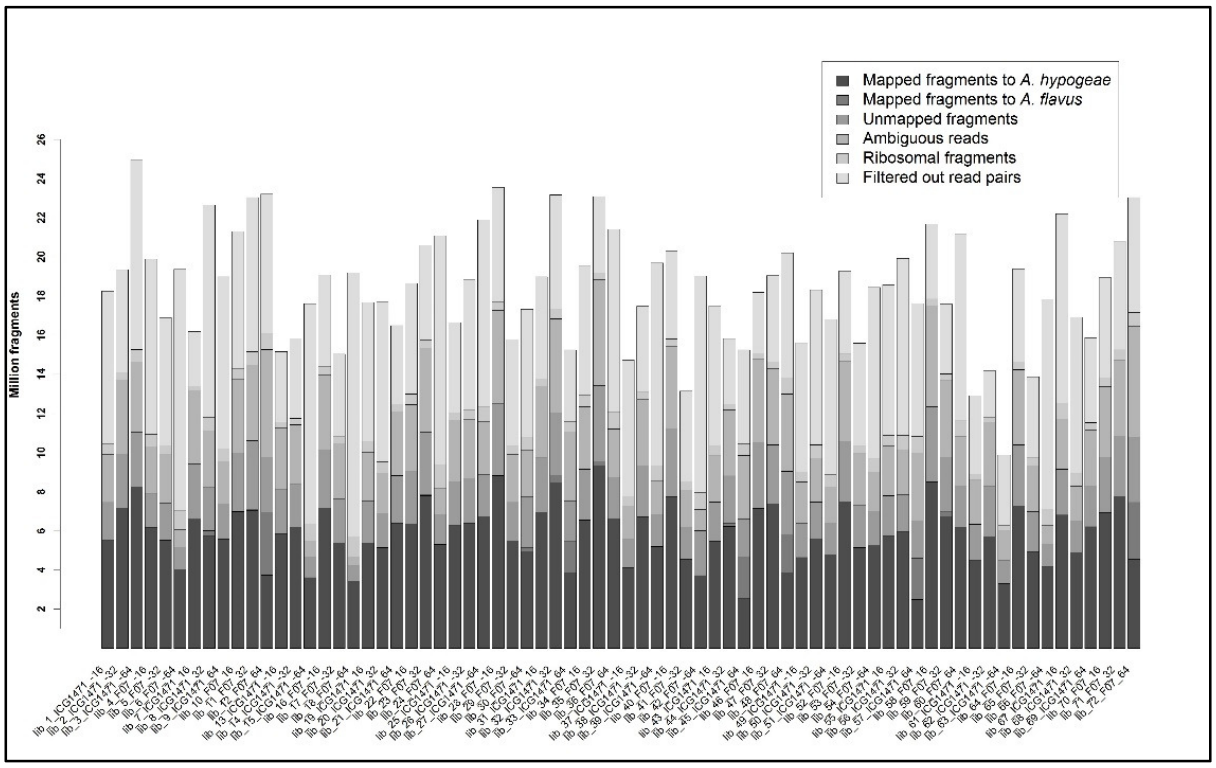


Figure 5.2: Sequence read integrity and mapping results to the synthetic tetraploid peanut and *A. flavus* genomes.

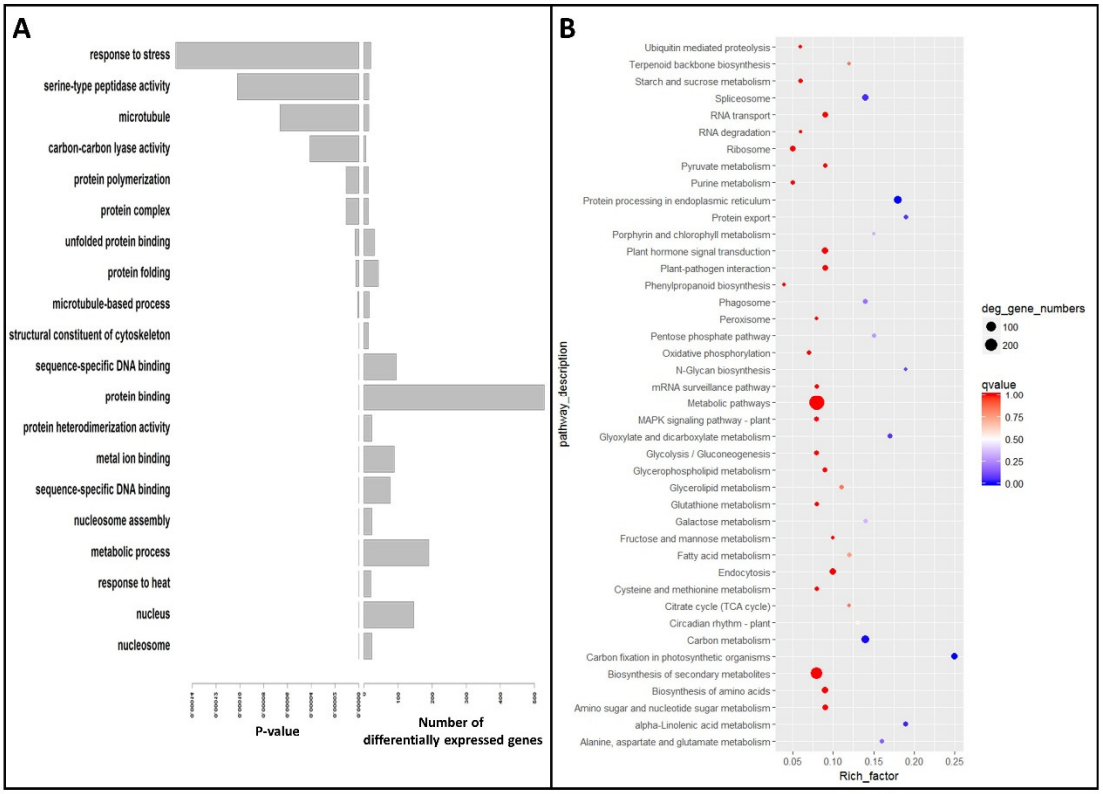


Figure 5.3: GO/KEGG enrichment analysis of peanut differently expressed genes between genotypes due to infection vs control. **A**. The 20 most significant GO terms extracted by GO enrichment analysis; **B.** KEGG enrichment analysis carried out by keggseq package; Rich_factor: the ratio of differentially expressed genes to the all genes that were annotated in the pathway.

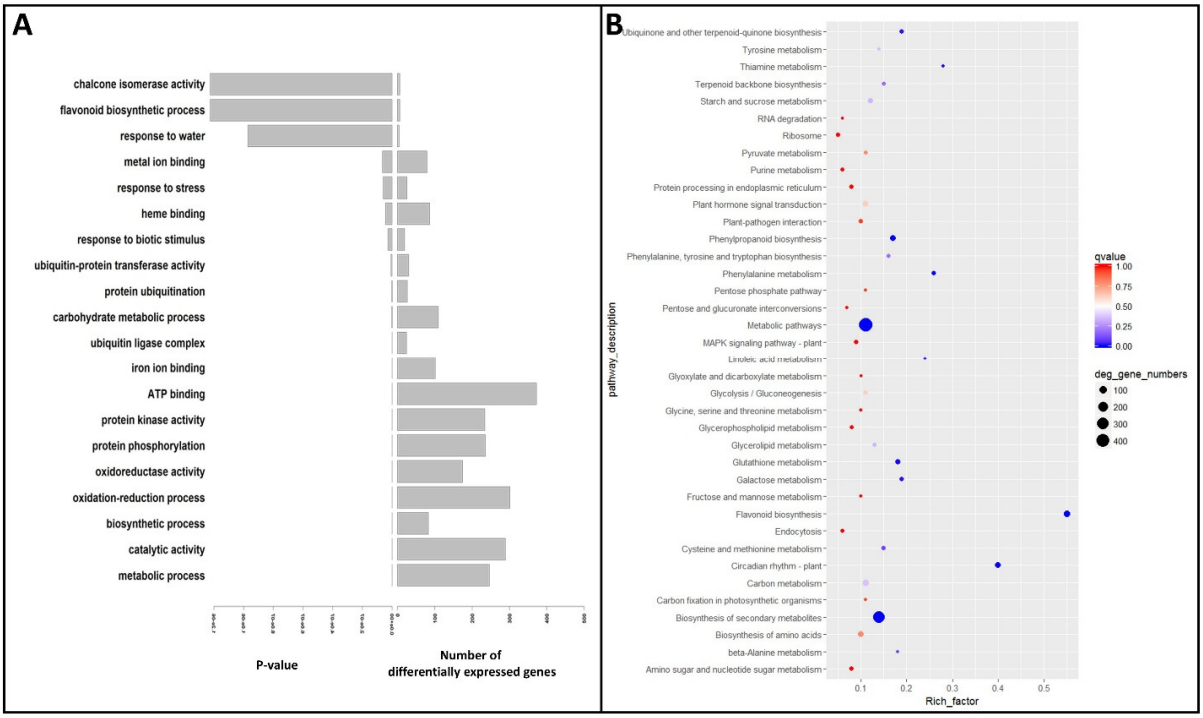


Figure 5.4: GO/KEGG enrichment analysis of differently expressed genes between peanut genotypes due to high vs low aflatoxin levels. **A.** The 20 most significant GO terms extracted by GO enrichment analysis; **B.** KEGG enrichment analysis carried out by keggseq package; Rich_factor: the ratio of the differentially expressed genes to the all genes that were annotated in the pathway.

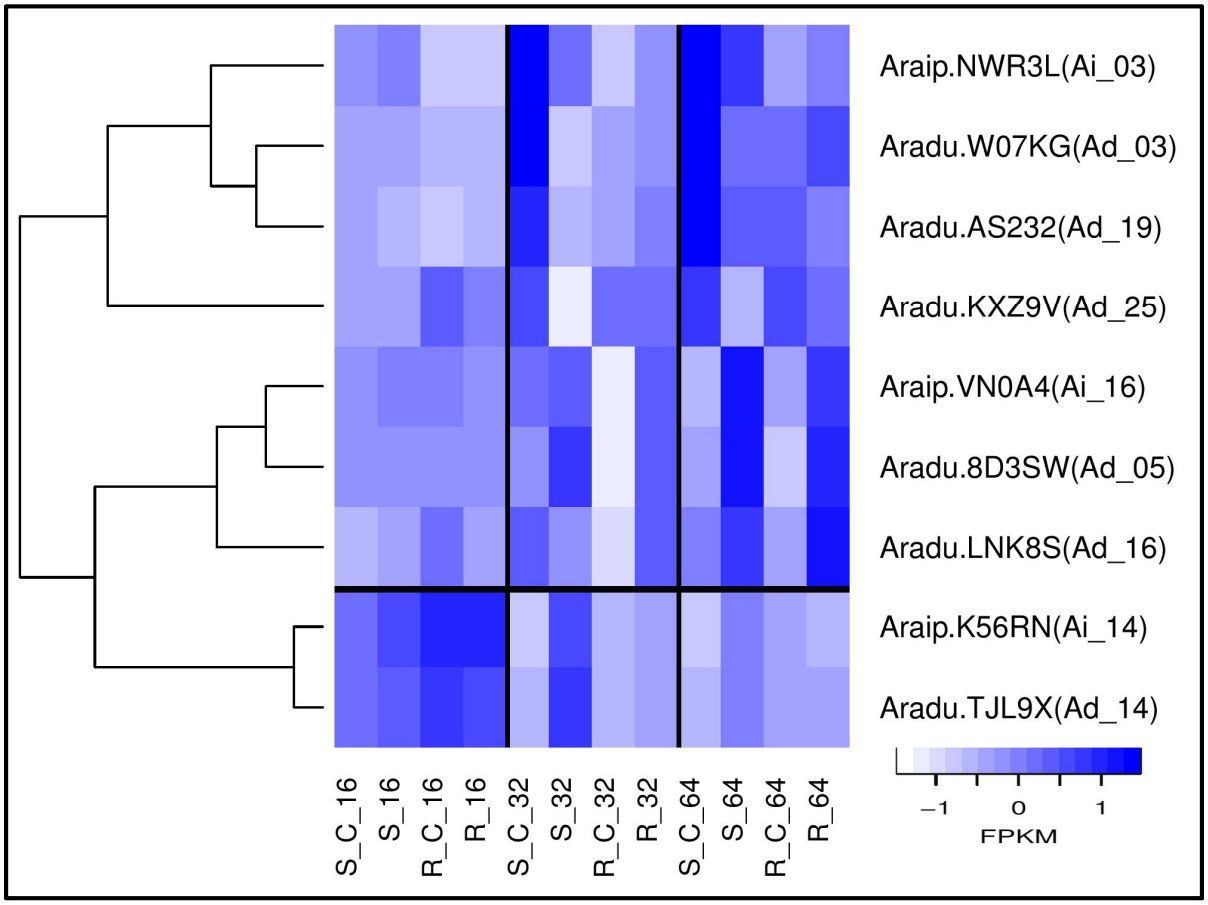


Figure 5.5: Differentially expressed lipoxygenases. Left, middle and right panels are 16, 32 and 64 HAI, respectively. The upper panel is extra-plastidial genes, and the lower panel is plastidal genes. All genes genes except Ad_25 are predicted to generate 13-S-hydroxyperoxides. Ad_25 was not classified. FPKM: Z-scores of Fragments Per Kilobase of transcript per Million mapped reads.

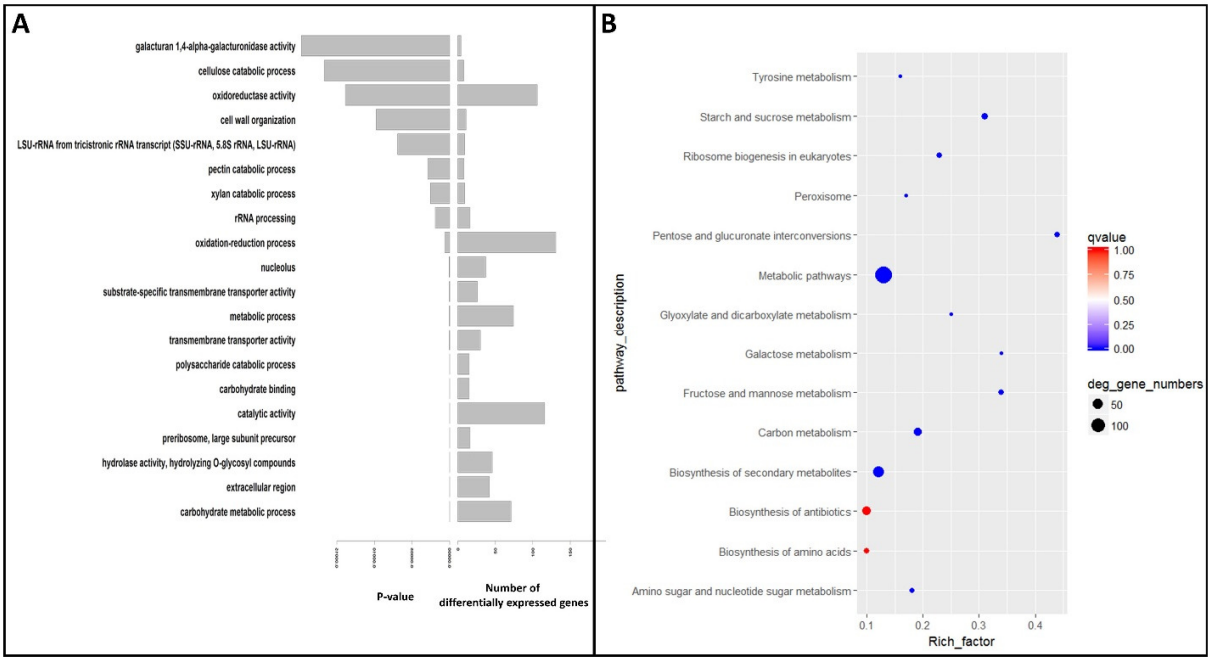


Figure 5.6: GO/KEGG enrichment analysis of *A. flavus* differently expressed genes due to growth of the fungus on resistant vs susceptible genotypes. **A.** The 20 most significant GO terms extracted by GO enrichment analysis; **B.** KEGG enrichment analysis carried out by keggseq package; Rich_factor: the ratio of the differentially expressed genes

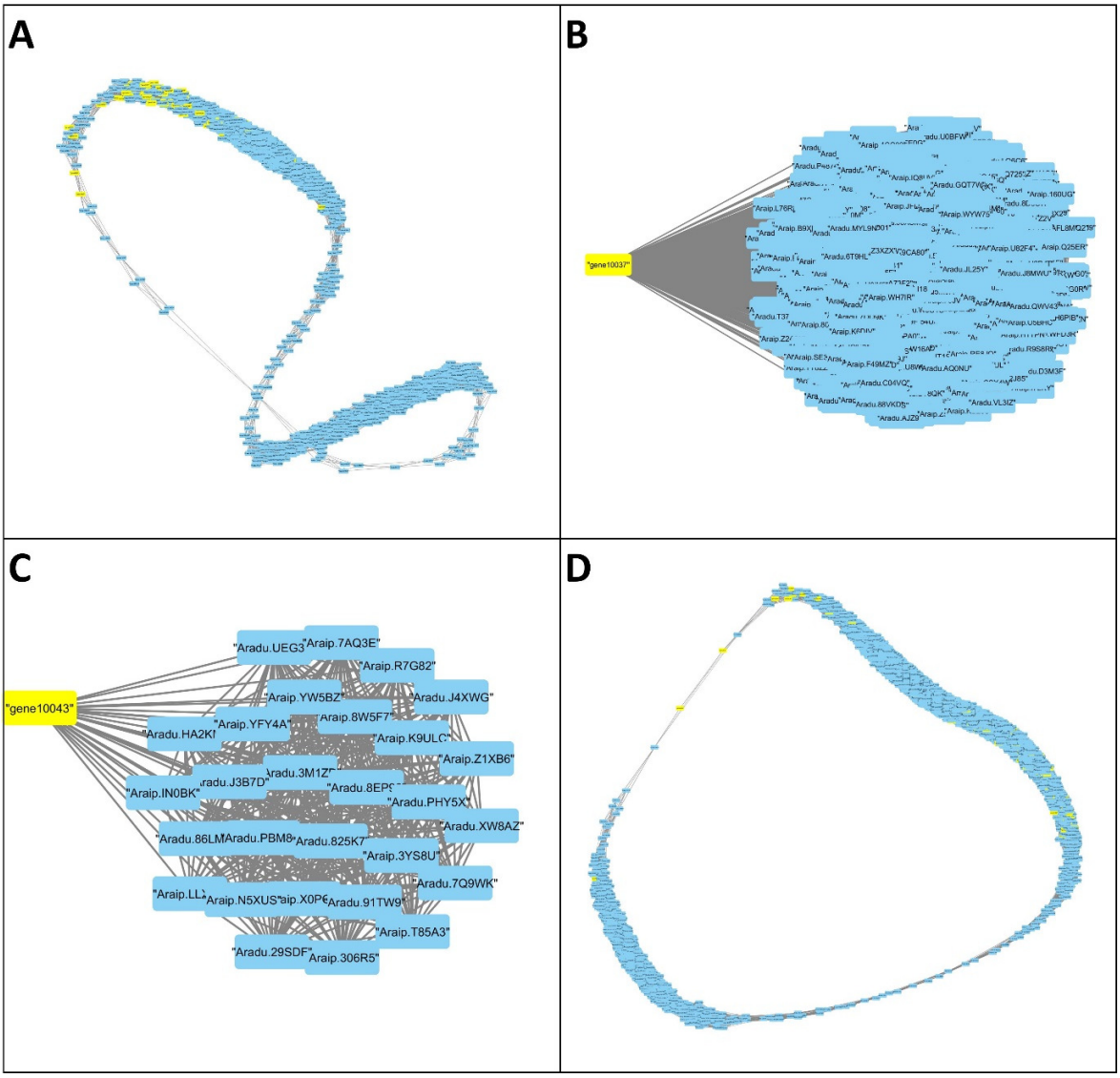


Figure 5.7: Co-expression network analysis of peanut/*A. flavus* genes. A. ICG 1471/*A. flavus* network. B. Sub-network 1 of ICG 1471/*A. flavus* network. C. Sub-network 15 of ICG 1471/*A. flavus* network. D. Florida-07/*A. flavus* network; lines represent edges, blue rectangles are peanut nodes, yellow rectangles are *A. flavus* nodes.

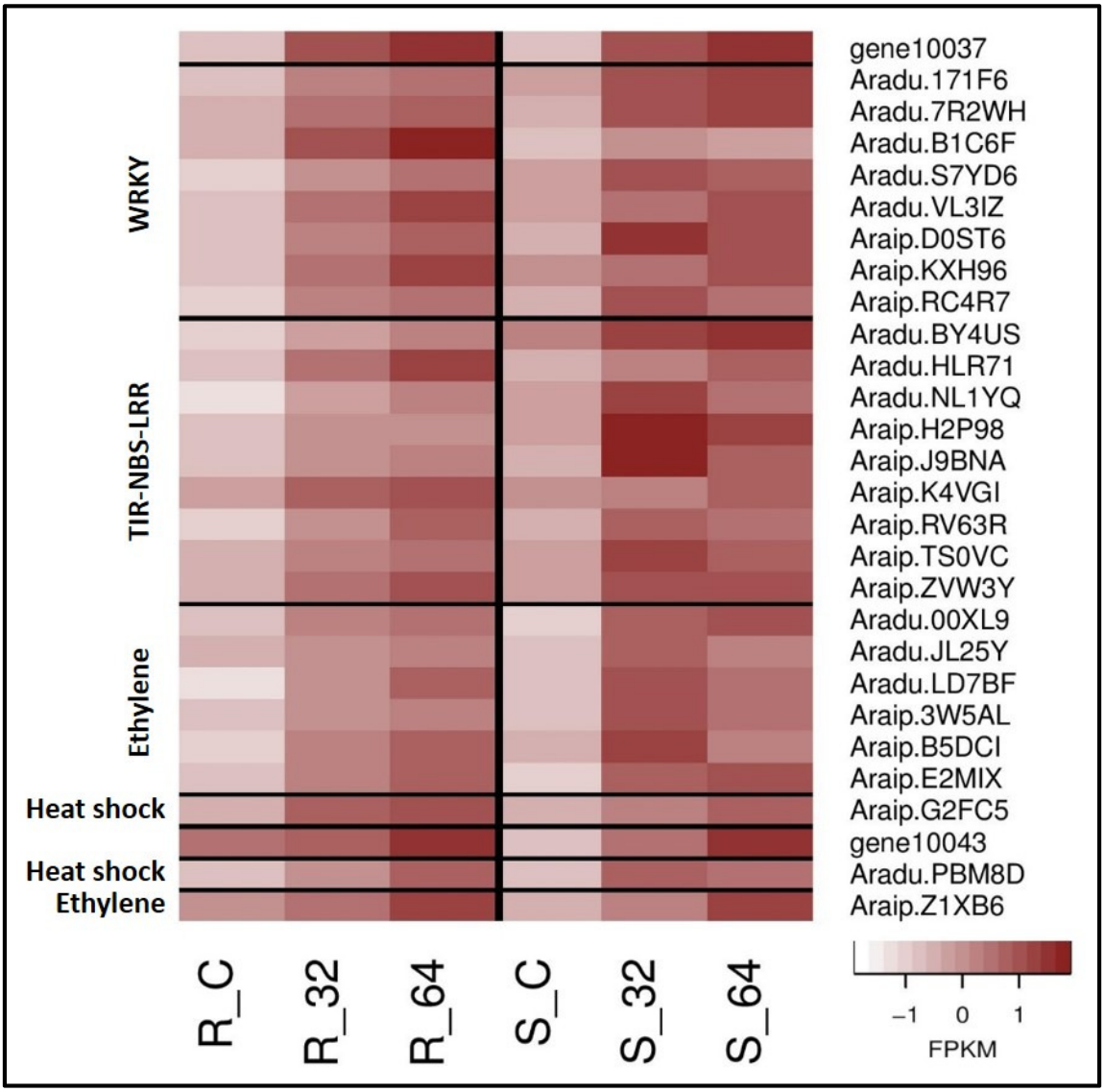


Figure 5.8: Peanut co-expressed genes with gene10037 and gene10043 of *A. flavus*; FPKM: Z-scores of Fragments Per Kilobase of transcript per Million mapped reads.

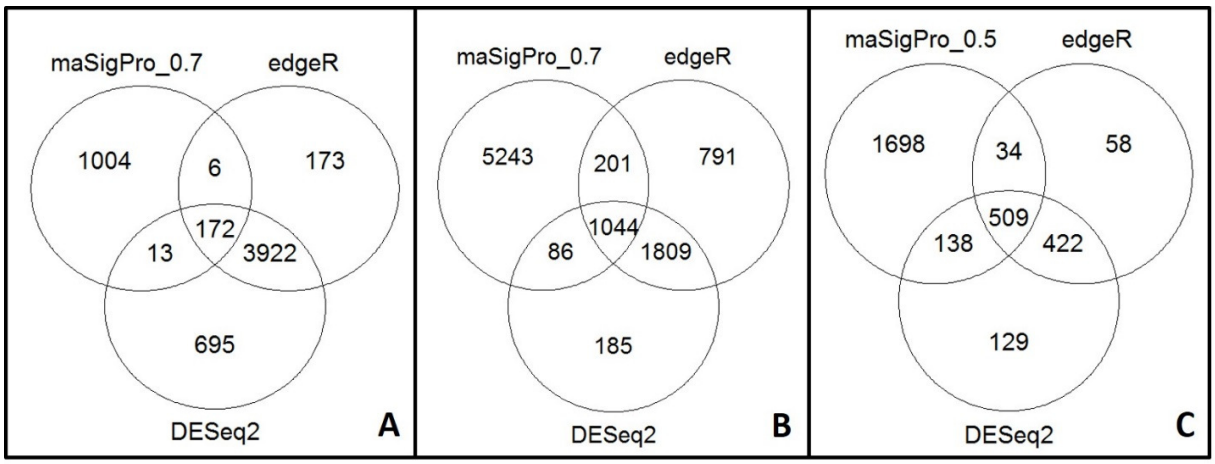


Figure 5.9: Differential expression analysis with multiple programs for *in-silico* validation of peanut genotypic differences due to *A. flavus* infection **(A)**, physiological effects **(B)**, and fungal/peanut-genotype effects **(C)**.

References

Abdollahi A, Buchanan, R (1981) Regulation of aflatoxin biosynthesis: induction of aflatoxin production by various carbohydrates. Journal of Food Science **46:** 633-635

Äijö T, Butty, V, Chen, Z, Salo, V, Tripathi, S, *el al.* (2014) Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. Bioinformatics **30:** i113-i120

Amaike S, Keller, NP (2011) *Aspergillus flavus*. Annual Review of Phytopathology **49:** 107-133

Anders S, Pyl, PT, Huber, W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics **31:** 166-169

Babitha M, Prakash, H, Shetty, HS (2006) Induction of lipoxygenase in downy mildew resistant seedlings of pearl millet in response to inoculation with *Sclerospora graminicola*. International Journal of Agriculture and Biology **8:** 560-564

Babitha MP, Prakash, HS, Shetty, HS (2004) Purification and properties of lipoxygenase induced in downy mildew resistant pearl millet seedlings due to infection with *Sclerospora graminicola*. Plant Science **166:** 31-39

Bader GD, Hogue, CW (2003) An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics **4:** 2

Baldwin IT, Zhang, ZP, Diab, N, Ohnmeiss, TE, Mccloud, ES, *et al.* (1997) Quantification, correlations and manipulations of wound-induced changes in jasmonic acid and nicotine in *Nicotiana sylvestris*. Planta **201:** 397-404

Barnett DW, Garrison, EK, Quinlan, AR, Stromberg, MP, Marth, GT (2011) BamTools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics **27:** 1691-1692

Bean GA, Schillinger, JA, Klarman, WL (1972) Occurrence of aflatoxins and aflatoxin-producing strains of *Aspergillus spp.* in soybeans. Applied Microbiology **24:** 437-439

Bell E, Creelman, RA, Mullet, JE (1995) A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in *Arabidopsis*. Proceedings of the National Academy of Sciences **92:** 8675-8679

Bertioli DJ, Cannon, SB, Froenicke, L, Huang, G, Farmer, AD, *et al.* (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet **48:** 438-446

Bhatnagar D, Ehrlich, K, Cleveland, T (2003) Molecular genetic analysis and regulation of aflatoxin biosynthesis. Applied Microbiology and Biotechnology **61:** 83-93

Birkenbihl RP, Diezel, C, Somssich, IE (2012) *Arabidopsis* WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward *Botrytis cinerea* infection. Plant Physiology **159:** 266-285

Blankenship PD, Cole, RJ, Sanders, TH, Hill, RA (1984) Effect of geocarposphere temperature on pre-harvest colonization of drought-stressed peanuts by *Aspergillus flavus* and subsequent aflatoxin contamination. Mycopathologia **85:** 69-74

Bolger AM, Lohse, M, Usadel, B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114-2120

Boué SM, Shih, BY, Carter-Wientjes, CH, Cleveland, TE (2005) Effect of soybean lipoxygenase on volatile generation and inhibition of *Aspergillus flavus* mycelial growth. Journal of agricultural and food chemistry **53:** 4778-4783

Burow GB, Gardner, HW, Keller, NP (2000) A peanut seed lipoxygenase responsive to *Aspergillus* colonization. Plant Mol Biol **42:** 689-701

Camacho C, Coulouris, G, Avagyan, V, Ma, N, Papadopoulos, J, *et al.* (2009) BLAST+: Architecture and applications. BMC bioinformatics **10:** 421

Chang AS, Sreedharan, A, Schneider, KR (2013) Peanut and peanut products: A food safety perspective. Food Control **32:** 296-303

Choi J, Alkharouf, N, Schneider, K, Matthews, B, Frederick, R (2008) Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. Functional & Integrative Genomics **8:** 341

Clevenger J, Chu, Y, Scheffler, B, Ozias-Akins, P (2016a) A developmental transcriptome map for allotetraploid *Arachis hypogaea*. Front Plant Sci **7:** 1446

Clevenger J, Marasigan, K, Liakos, V, Sobolev, V, Vellidis, G, *et al.* (2016b) RNA sequencing of contaminated seeds reveals the state of the seed permissive for pre-harvest aflatoxin contamination and points to a potential susceptibility factor. Toxins **8:** 317

Cotty PJ, Jaime-Garcia, R (2007) Influences of climate on aflatoxin producing fungi and aflatoxin contamination. International Journal of Food Microbiology **119:** 109-115

Craufurd P, Prasad, P, Waliyar, F, Taheri, A (2006) Drought, pod yield, pre-harvest *Aspergillus* infection and aflatoxin contamination on peanut in Niger. Field Crops Research **98:** 20-29

Creelman RA, Mullet, JE (1997) Biosynthesis and action of jasmonates in plants. Annual Review of Plant Biology **48:** 355-381

Davis ND, Diener, UL (1968) Growth and aflatoxin production by *Aspergillus parasiticus* from various carbon sources. Applied Microbiology **16:** 158

De León IP, Sanz, A, Hamberg, M, Castresana, C (2002) Involvement of the *Arabidopsis*α-DOX1 fatty acid dioxygenase in protection against oxidative stress and cell death. The Plant Journal **29:** 61-72

Doehlert DC, Wicklow, DT, Gardner, HW (1993) Evidence implicating the lipoxygenase pathway in providing resistance to soybeans against *Aspergillus flavus*. Phytopathology **83:** 1473-1477

Dunham NR, Peper, ST, Downing, CD, Kendall, RJ (2017) Aflatoxin contamination in corn sold for wildlife feed in Texas. Ecotoxicology **26:** 516-520

Durinck S, Moreau, Y, Kasprzyk, A, Davis, S, De Moor, B, *et al.* (2005) BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. Bioinformatics **21:** 3439-3440

Ehrlich KC, Chang, P-K, Yu, J, Cotty, PJ (2004) Aflatoxin biosynthesis cluster gene cypA is required for G aflatoxin formation. Applied and Environmental Microbiology **70:** 6518-6524

Ehrlich KC, Montalbano, BG, Cotty, PJ (2005) Divergent regulation of aflatoxin production at acidic pH by two *Aspergillus strains*. Mycopathologia **159:** 579-581

Fountain JC, Bajaj, P, Nayak, SN, Yang, L, Pandey, MK, *et al.* (2016a) Responses of *Aspergillus flavus* to oxidative stress are related to fungal development regulator, antioxidant enzyme, and secondary metabolite biosynthetic gene expression. Frontiers in Microbiology **7:** 2048

Fountain JC, Bajaj, P, Pandey, M, Nayak, SN, Yang, L, *et al.* (2016b) Oxidative stress and carbon metabolism influence *Aspergillus flavus* transcriptome composition and secondary metabolite production. Scientific Reports **6:** 38747

Fountain JC, Khera, P, Yang, L, Nayak, SN, Scully, BT, *et al.* (2015b) Resistance to *Aspergillus flavus* in maize and peanut: Molecular biology, breeding, environmental stress, and future perspectives. The Crop Journal **3:** 229-237

Fountain JC, Raruang, Y, Luo, M, Brown, RL, Guo, B, *et al.* (2015a) Potential roles of WRKY transcription factors in regulating host defense responses during *Aspergillus flavus* infection of immature maize kernels. Physiological and Molecular Plant Pathology **89:** 31-40

Fountain JC, Scully, BT, Ni, X, Kemerait, RC, Lee, RD, *et al.* (2014) Environmental influences on maize-*Aspergillus flavus* interactions and aflatoxin production. Front Microbiol **5:** 44

Gao Q-M, Venugopal, S, Navarre, D, Kachroo, A (2011) Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the WRKY50 and WRKY51 proteins. Plant Physiology **155:** 464-476

Gao X, Brodhagen, M, Isakeit, T, Brown, SH, Göbel, C, *et al.* (2009) Inactivation of the lipoxygenase *ZmLOX3* increases susceptibility of maize to *Aspergillus spp*. Molecular Plant-Microbe Interactions **22:** 222-231

Gao X, Starr, J, Göbel, C, Engelberth, J, Feussner, I, *et al.* (2008) Maize 9-lipoxygenase *ZmLOX3* controls development, root-specific expression of defense genes, and resistance to root-knot nematodes. Molecular Plant-Microbe Interactions **21:** 98-109

Georgianna DR, Payne, GA (2009) Genetic regulation of aflatoxin biosynthesis: From gene to genome. Fungal Genetics and Biology **46:** 113-125

Goodrich-Tanrikulu M, Mahoney, NE, Rodriguez, SB (1995) The plant growth regulator methyl jasmonate inhibits aflatoxin production by *Aspergillus flavus*. Microbiology **141:** 2831-2837

Guo B, Xu, G, Cao, Y, Holbrook, C, Lynch, R (2006) Identification and characterization of phospholipase D and its association with drought susceptibilities in peanut (*Arachis hypogaea*). Planta **223:** 512-520

Haas BJ, Papanicolaou, A, Yassour, M, Grabherr, M, Blood, PD, *et al.* (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc **8:** 1494-1512

Hamberg M, Sanz, A, Rodriguez, MJ, Calvo, AP, Castresana, C (2003) Activation of the fatty Acid α-dioxygenase pathway during bacterial infection of tobacco leaves, formation of oxylipins protecting against cell death. Journal of Biological Chemistry **278:** 51796-51805

Hedayati M, Pasqualotto, A, Warn, P, Bowyer, P, Denning, D (2007) *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. Microbiology **153:** 1677-1692

Hoepfner D, Schildknegt, D, Braakman, I, Philippsen, P, Tabak, HF (2005) Contribution of the endoplasmic reticulum to peroxisome formation. Cell **122:** 85-95

Holbrook C, Kvien, C, Rucker, K, Wilson, D, Hook, J, *et al.* (2000a) Preharvest aflatoxin contamination in drought-tolerant and drought-intolerant peanut genotypes. Peanut Science **27:** 45-48

Holbrook CC, Wilson, DM, Matheron, ME, Hunter, JE, Knauft, DA, *et al.* (2000b) *Aspergillus* colonization and aflatoxin contamination in peanut genotypes with reduced linoleic acid composition. Plant Disease **84:** 148-150

Huang P, Franco, O, Yan, Y, Keller, N, Kolomiets, M (2013) Maize lipoxygenase *LOX2* regulates pathogenesis of mycotoxin-producing *Aspergillus flavus.* APS-MSA Joint meeting, August 10-14, Austin, Texas

Jaime-Garcia R, Cotty, PJ (2003) Aflatoxin contamination of commercial cottonseed in South Texas. Phytopathology **93:** 1190-1200

Jayashree T, Subramanyam, C (2000) Oxidative stress as a prerequisite for aflatoxin production by *Aspergillus parasiticus*. Free Radical Biology and Medicine **29:** 981-985

Jo K, Kwon, H-B, Kim, S (2014) Time-series RNA-seq analysis package (TRAP) and its application to the analysis of rice, *Oryza sativa* L. *ssp. Japonica*, upon drought stress. Methods **67:** 364-372

Journot-Catalino N, Somssich, IE, Roby, D, Kroj, T (2006) The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in *Arabidopsis thaliana*. The Plant Cell **18:** 3289-3302

Kessler A, Halitschke, R, Baldwin, IT (2004) Silencing the jasmonate cascade: induced plant defenses and insect populations. Science **305:** 665-668

Kisyombe CT, Beute, M, Payne, G (1985) Field evaluation of peanut genotypes for resistance to infection by *Aspergillus parasiticus*. Peanut Science **12:** 12-17

Klich MA (2007) *Aspergillus flavus*: the major producer of aflatoxin. Molecular Plant Pathology **8:** 713-722

Kumari A, Sucharitha, A, Devi, P (2012) Impact of 13-lipoxygenase on induction of resistance in peanut against aflatoxigenic *Aspergillus*. Continental Journal of Biological Sciences **5:** 4-13

Langmead B, Trapnell, C, Pop, M, Salzberg, SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology **10:** R25-R25

Lawrence M, Huber, W, Pages, H, Aboyoun, P, Carlson, M, *et al.* (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol **9:** e1003118

Lee J-H, Yun, HS, Kwon, C (2012) Molecular communications between plant heat shock responses and disease resistance. Molecules and Cells **34:** 109-116

Leng N, Li, Y, Mcintosh, BE, Nguyen, BK, Duffin, B, *et al.* (2015) EBSeq-HMM: a Bayesian approach for identifying gene-expression changes in ordered RNA-seq experiments. Bioinformatics btv193

Li S, Zhou, X, Chen, L, Huang, W, Yu, D (2010) Functional characterization of *Arabidopsis thaliana* WRKY39 in heat stress. Molecules and Cells **29:** 475-483

Love MI, Huber, W, Anders, S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15:** 550

Mateo EM, Gil-Serna, J, Patino, B, Jimenez, M (2011) Aflatoxins and ochratoxin A in stored barley grain in Spain and impact of PCR-based strategies to assess the occurrence of aflatoxigenic and ochratoxigenic *Aspergillus spp*. Int J Food Microbiol **149:** 118-126

McConn M, Creelman, RA, Bell, E, Mullet, JE (1997) Jasmonate is essential for insect defense in *Arabidopsis*. Proceedings of the National Academy of Sciences **94:** 5473-5477

McHale L, Tan, X, Koehl, P, Michelmore, RW (2006) Plant NBS-LRR proteins: Adaptable guards. Genome Biology **7:** 212

Mei C, Qi, M, Sheng, G, Yang, Y (2006) Inducible overexpression of a rice allene oxide synthase gene increases the endogenous jasmonic acid level, PR gene expression, and host resistance to fungal infection. Molecular Plant-Microbe Interactions **19:** 1127-1137

Meimaroglou DM, Galanopoulou, D, Markaki, P (2009) Study of the effect of methyl jasmonate concentration on aflatoxin B 1 biosynthesis by *Aspergillus parasiticus* in yeast extract sucrose medium. International Journal of Microbiology **2009:** 842626

Mita G, Fasano, P, De Domenico, S, Perrone, G, Epifani, F, *et al.* (2007) 9-Lipoxygenase metabolism is involved in the almond/*Aspergillus* carbonarius interaction. Journal of Experimental Botany **58:** 1803-1811

Müller V, Amé, M, Carrari, V, Gieco, J, Asis, R (2014) Lipoxygenase activation in peanut seed cultivars resistant and susceptible to *Aspergillus parasiticus* colonization. Phytopathology **104:** 1340-1348

Musungu BM, Bhatnagar, D, Brown, RL, Payne, GA, OBrain, G, *et al.* (2016) A network approach of gene co-expression in the *Zea mays*/*Aspergillus flavus* pathosystem to map host/pathogen interaction pathways. Front Genet **7:** 206

Nigam S, Waliyar, F, Aruna, R, Reddy, S, Kumar, PL, *et al.* (2009) Breeding peanut for resistance to aflatoxin contamination at ICRISAT. Peanut Science **36:** 42-49

Nueda MJ, Tarazona, S, Conesa, A (2014) Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics **30:** 2598-2602

Ozalvo R, Cabrera, J, Escobar, C, Christensen, SA, Borrego, EJ, *et al.* (2014) Two closely related members of *Arabidopsis* 13-lipoxygenases (13-LOXs), *LOX3* and *LOX4*, reveal distinct functions in response to plant-parasitic nematode infection. Molecular Plant Pathology **15:** 319-332

Reverberi M, Punelli, M, Smith, CA, Zjalic, S, Scarpari, M, *et al.* (2012) How peroxisomes affect aflatoxin biosynthesis in *Aspergillus flavus*. PLoS One **7:** e48097

Robinson MD, Mccarthy, DJ, Smyth, GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26:** 139-140

Sanavia T, Finotello, F, Di Camillo, B (2015) FunPat: function-based pattern analysis on RNA-seq time series data. BMC Genomics **16:** S2

Sanders TH, Cole, RJ, Blankenship, PD, Hill, RA (1985) Relation of environmental stress duration to *Aspergillus flavus* invasion and aflatoxin production in preharvest peanuts. Peanut Science **12:** 90-93

Scheidegger KA, Payne, GA (2003). Unlocking the secrets behind secondary metabolism: A review of *Aspergillus flavus* from pathogenicity to functional genomics. Journal of Toxicology **22:** 423-459

Seijo JG, Lavia, GI, Fernández, A, Krapovickas, A, Ducasse, D, *et al.* (2004) Physical mapping of the 5S and 18S–25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (*Leguminosae*). American Journal of Botany **91:** 1294-1303

Sekhwal MK, Li, P, Lam, I, Wang, X, Cloutier, S, *et al.* (2015) Disease resistance gene analogs (RGAs) in plants. International Journal of Molecular Science **16:** 19248-19290

Sham A, Moustafa, K, Al-Ameri, S, Al-Azzawi, A, Iratni, R, *et al.* (2015) Identification of *Arabidopsis* candidate genes in response to biotic and abiotic stresses using comparative microarrays. PLoS One **10:** e0125666

Shan SH, Zhang, T, Li, C, Yang, C, Yan, C, *et al.* (2011) Cloning and analysis of a NBS-LRR disease resistance gene candidate PnAG1 from peanut (*Arachis hypogaea* L.). Electronic Journal of Biotechnology **14:** 6-6

Shannon P, Markiel, A, Ozier, O, Baliga, NS, Wang, JT, *et al.* (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res **13:** 2498-2504

Sherif M, Becker, E, Herrfurth, C, Feussner, I, Karlovsky, P, *et al.* (2016) Volatiles emitted from maize ears simultaneously infected with two *Fusarium* species mirror the most competitive fungal pathogen. Frontiers in Plant Science **7:** 1406

Shivakumar P, Geetha, H, Shetty, HS (2003) Peroxidase activity and isozyme analysis of pearl millet seedlings and their implications in downy mildew disease resistance. Plant Science **164:** 85-93

Spensley PC (1963) Aflatoxin, the active principle in turkey 'X' disease. Endeavour **22:** 75-79

Spies D, Ciaudo, C (2015) Dynamics in transcriptomics: advancements in RNA-seq course and downstream analysis. Computational and Structural Biotechnology Journal **12:** 469-477

Squire RA (1981) Ranking animal carcinogens: a proposed regulatory approach. Science **214:** 877-880

Stintzi A (2000) The *Arabidopsis* male-sterile mutant, opr3, lacks the 12-oxophytodienoic acid reductase required for jasmonate synthesis. Proceedings of the National Academy of Sciences **97:** 10625-10630

Stubblefield RD, Shotwell, OL, Hesseltine, CW, Smith, ML, Hall, HH (1967) Production of aflatoxin on wheat and oats: measurement with a recording densitometer. Appl Microbiol **15:** 186-190

Styer CH, Cloe, RJ, Hill, RA (1983) Inoculation and infection of peanut flowers by *Aspergillus flavus*. Proceeding of American Peanut Research and Education Society **15:** 91

Suarez-Bonnet E, Carvajal, M, Mendez-Ramirez, I, Castillo-Urueta, P, Cortes-Eslava, J, *et al.* (2013) Aflatoxin (B1 , B2 , G1 , and G2) contamination in rice of Mexico and Spain, from local sources or imported. J Food Sci **78:** T1822-1829

Sun X, Dalpiaz, D, Wu, D, Liu, JS, Zhong, W, *et al.* (2016) Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model. BMC Bioinformatics **17:** 324

Taki N, Sasaki-Sekimoto, Y, Obayashi, T, Kikuta, A, Kobayashi, K, *et al.* (2005) 12-oxo-phytodienoic acid triggers expression of a distinct set of genes and plays a role in wound-induced gene expression in *Arabidopsis*. Plant Physiology **139:** 1268-1283

Tang Q, Zhang, Y, Liu, B, Yu, Z, Wu, Y (2009) Study on the relationship between lipoxygenase-3 and the characteristic of resisting storage insects of rice grain. Journal of Food, Agriculture & Environment **7:** 334-338

Thaler JS, Stout, MJ, Karban, R, Duffey, SS (1996) Exogenous jasmonates simulate insect wounding in tomato plants (*Lycopersicon esculentum*) in the laboratory and field. Journal of Chemical Ecology **22:** 1767-1781

Thomma BP, Eggermont, K, Broekaert, WF, Cammue, BP (2000) Disease development of several fungi on *Arabidopsis* can be reduced by treatment with methyl jasmonate. Plant Physiology and Biochemistry **38:** 421-427

Trapnell C, Pachter, L, Salzberg, SL (2009) TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics **25:** 1105-1111

Trapnell C, Williams, BA, Pertea, G, Mortazavi, A, Kwan, G, *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28:** 511-515

Tsitsigiannis DI, Kunze, S, Willis, DK, Feussner, I, Keller, NP (2005) *Aspergillus* infection inhibits the expression of peanut 13 S-HPODE-forming seed lipoxygenases. Molecular Plant-Microbe Interactions **18:** 1081-1089

Vergopoulou S, Galanopoulou, D, Markaki, P (2001) Methyl jasmonate stimulates aflatoxin B1 biosynthesis by *Aspergillus parasiticus*. Journal of Agricultural and Food Chemistry **49:** 3494-3498

Verma R (2004) Aflatoxin cause DNA damage. International Journal of Human Genetics **4:** 231

Vijayan P, Shockey, J, Lévesque, CA, Cook, RJ (1998) A role for jasmonate in pathogen defense of *Arabidopsis*. Proceedings of the National Academy of Sciences **95:** 7209-7214

Waliyar F, Kumar, PL, Traoré, A, Ntare, BR, Diarra, B, *et al.* (2008) Pre-and post-harvest management of aflatoxin contamination in peanuts. Mycotoxins: Detection methods, management, public health and agricultural trade, Leslie, JF, Bandyopadhyay, R and Visconti, A. 209–218

Waliyar F, Traore, A, Fatondji, D, Ntare, B (2003) Effect of irrigation interval, planting date, and cultivar on *Aspergillus flavus* and aflatoxin contamination of peanut in a sandy soil of Niger. Peanut Science **30:** 79-84

Wang R, Shen, W, Liu, L, Jiang, L, Liu, Y, *et al.* (2008) A novel lipoxygenase gene from developing rice seeds confers dual position specificity and responds to wounding and insect attack. Plant Molecular Biology **66:** 401-414

Wise A, Bar-Joseph, Z (2014) SMARTS: Reconstructing disease response networks from multiple individuals using time series gene expression data. Bioinformatics btu800

Xue H, Isleib, T, Payne, G, Novitzky, W, OBrian, G (2005) Aflatoxin production in peanut lines selected to represent a range of linoleic acid concentrations. Journal of Food Protection **68:** 126-132

Yan L, Zhai, Q, Wei, J, Li, S, Wang, B, *et al.* (2013) Role of tomato lipoxygenase D in wound-induced jasmonate biosynthesis and plant immunity to insect herbivores. PLoS Genet **9:** e1003964

Yang J, Chen, X, Zhu, C, Peng, X, He, X, *et al.* (2015) RNA-seq reveals differentially expressed genes of rice (*Oryza sativa*) spikelet in response to temperature interacting with nitrogen at meiosis stage. BMC Genomics **16:** 959

Young MD, Wakefield, MJ, Smyth, GK, Oshlack, A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology **11:** R14

Yu J, Bhatnagar, D, Cleveland, TE (2004) Completed sequence of aflatoxin pathway gene cluster in *Aspergillus parasiticus*. FEBS Letters **564:** 126-130

Zeneli G, Krokene, P, Christiansen, E, Krekling, T, Gershenzon, J (2006) Methyl jasmonate treatment of mature Norway spruce (*Picea abies*) trees increases the accumulation of terpenoid resin components and protects against infection by *Ceratocystis polonica*, a bark beetle-associated fungus. Tree Physiology **26:** 977-988

Zeringue Jr H (1996) Possible involvement of lipoxygenase in a defense response in aflatoxigenic *Aspergillus*–cotton plant interactions. Canadian Journal of Botany **74:** 98-102

Zhang L, Zhang, F, Melotto, M, Yao, J, He, SY (2017) Jasmonate signaling and manipulation by pathogens and insects. Journal of Experimental Botany erw478

Appendix 5.A

Supplemental Materials

Table 5.S1: Statistical models that were used for differential expression analysis of the RNA-seq data.

Analysis	Package	Test	Model
R vs S due to T vs C Peanut	edgeR	glmLRT	time16_effect = (R_T_16-R_C_16)-(S_T_16-S_C_16) time32_effect = (R_T_32-R_C_32)-(S_T_32-S_C_32) time64_effect = (R_T_64-R_C_64)-(S_T_64-S_C_64)
	Deseq2	LRT	full = ~ genotype (R/S)+ afla_trt16 (T/C) + **genotype:afla_trt16**) reduced = ~ genotype (R/S)+ afla_trt16 (T/C) full = ~ genotype (R/S)+ afla_trt32 (T/C) + **genotype:afla_trt32**) reduced = ~ genotype (R/S)+ afla_trt32 (T/C) full = ~ genotype (R/S)+ afla_trt64 (T/C) + **genotype:afla_trt64**) reduced = ~ genotype (R/S)+ afla_trt64 (T/C)
	Next MaSigPro	LLRT	R_C, R_T, S_C, S_T
R vs S due to H vs L Peanut	edgeR	glmLRT	time16_effect = (R_H_16-R_L_16)-(S_H_16-S_L_16) time32_effect = (R_H_32-R_L_32)-(S_H_32-S_L_32) time64_effect = (R_H_64-R_L_64)-(S_H_64-S_L_64)
	Deseq2	LRT	full = ~ genotype (R/S)+ afla_trt16 (H/L) + **genotype:afla_trt16**) reduced = ~ genotype (R/S)+ afla_trt16 (H/L) full = ~ genotype (R/S)+ afla_trt32 (H/L) + **genotype:afla_trt32**) reduced = ~ genotype (R/S)+ afla_trt32 (H/L) full = ~ genotype (R/S)+ afla_trt64 (H/L) + **genotype:afla_trt64**) reduced = ~ genotype (R/S)+ afla_trt64 (H/L)
	Next MaSigPro	LLRT	time16, time32, time64; R_H, R_L, S_H, S_L

R vs S *A. flavus*	edgeR	etLRT	time16_effect = R_H_16-S_H_16 time32_effect = R_H_32-S_H_32 time64_effect = R_H_64-S_H_64
	Deseq2	WT	time16_effect = ~ genotype (R_H_16/S_H_16) time32_effect = ~ genotype (R_H_32/S_H_32) time64_effect = ~ genotype (R_H_64/S_H_64)
	Next MaSigPro	LLRT	time16, time32, time64; R_H, S_H
Peanut (R)*	edgeR	etLRT	time32_effect = R_T_32-R_C_32 time64_effect = R_T_64-R_C_64
Peanut (S)*		etLRT	time32_effect = S_T_32-S_C_32 time64_effect = S_T_64-S_C_64
A. flavus (R)*		etLRT	time32_effect = R_T_32-R_T_16 time64_effect = R_T_64-R_T_16
A. flavus (S)*		etLRT	time32_effect = S_T_32-S_T_16 time64_effect = S_T_64-S_T_16

* Design for network analysis; R, Resistant cultivar; S, susceptible cultivar; T, Treatment; C, Control; H, High-level of aflatoxin; L, Low-level aflatoxin; 16, 16 hours after infection; 32, 32 hours of infection; 64, 64 hours after infection; LRT, Likelihood Ratio Test; WT, Wald Test; glmLRT, general linear model - Likelihood Ratio Test; etLRT, exact test - Likelihood Ratio Test; LLRT, Log Likelihood Ratio Test.

Table 5.S2: RNA integrity, GFP and aflatoxin analysis of samples that were sequenced.

Genotype	Trt	GFP RFU/1mg	Aflatoxin B (ppb)	Aflatoxin level	Rep.	Index	Lane	RIN
Florida-07	-16	0.00	5.61	Low	1	TruSeq Adapter, Index 5	1	7.4
Florida-07	-32	2.69	3.96	Low	1	TruSeq Adapter, Index 13	1	6.7
Florida-07	-64	2.54	5.21	Low	1	TruSeq Adapter, Index 18	1	6.5
Florida-07	+16	-2.74	4.66	Low	1	TruSeq Adapter, Index 6	1	6.9
Florida-07	+32	-1.54	5.71	Low	1	TruSeq Adapter, Index 14	1	8.4
Florida-07	+64	-3.72	-128.10	Low	1	TruSeq Adapter, Index 19	1	7.1
ICG 1471	-16	0.00	2.59	Low	1	TruSeq Adapter, Index 2	1	7.4
ICG 1471	-32	0.00	0.51	Low	1	TruSeq Adapter, Index 7	1	6.8
ICG 1471	-64	0.00	1.80	Low	1	TruSeq Adapter, Index 15	1	7.2
ICG 1471	+16	0.00	0.39	Low	1	TruSeq Adapter, Index 4	1	na
ICG 1471	+32	-2.10	0.52	Low	1	TruSeq Adapter, Index 12	1	7
ICG 1471	+64	-3.53	0.49	Low	1	TruSeq Adapter, Index 16	1	6.5
Florida-07	-16	3.99	2.83	Low	2	TruSeq Adapter, Index 5	2	6.7
Florida-07	-32	1.29	3.31	Low	2	TruSeq Adapter, Index 13	2	7.3
Florida-07	-64	0.00	3.22	Low	2	TruSeq Adapter, Index 18	2	6.6
Florida-07	+16	-2.74	4.05	Low	2	TruSeq Adapter, Index 6	2	7
Florida-07	+32	-3.16	12.03	Low	2	TruSeq Adapter, Index 14	2	7.8
Florida-07	+64	-2.06	-136.22	Low	2	TruSeq Adapter, Index 19	2	6.7
ICG 1471	-16	0.00	2.45	Low	2	TruSeq Adapter, Index 2	2	7
ICG 1471	-32	0.00	2.52	Low	2	TruSeq Adapter, Index 7	2	7.6
ICG 1471	-64	0.00	2.01	Low	2	TruSeq Adapter, Index 15	2	7.1
ICG 1471	+16	0.00	0.05	Low	2	TruSeq Adapter, Index 4	2	7.3
ICG 1471	+32	-2.10	0.27	Low	2	TruSeq Adapter, Index 12	2	7.1
ICG 1471	+64	-3.53	0.52	Low	2	TruSeq Adapter, Index 16	2	8.3
Florida-07	-16	3.56	4.27	Low	3	TruSeq Adapter, Index 5	3	6.6
Florida-07	-32	3.59	4.42	Low	3	TruSeq Adapter, Index 13	3	7.1
Florida-07	-64	5.65	1.96	Low	3	TruSeq Adapter, Index 18	3	7.5
Florida-07	+16	-2.74	4.98	Low	3	TruSeq Adapter, Index 6	3	7.4
Florida-07	+32	-3.12	2.51	Low	3	TruSeq Adapter, Index 14	3	8.4
Florida-07	+64	-1.40	-128.82	Low	3	TruSeq Adapter, Index 19	3	7.6
ICG 1471	-16	0.00	1.83	Low	3	TruSeq Adapter, Index 2	3	7
ICG 1471	-32	0.00	1.27	Low	3	TruSeq Adapter, Index 7	3	7.6
ICG 1471	-64	0.00	1.92	Low	3	TruSeq Adapter, Index 15	3	7
ICG 1471	+16	0.00	1.21	Low	3	TruSeq Adapter, Index 4	3	7
ICG 1471	+32	-2.10	0.31	Low	3	TruSeq Adapter, Index 12	3	6.6
ICG 1471	+64	-3.53	0.47	Low	3	TruSeq Adapter, Index 16	3	6.5
Florida-07	-16	9.46	1.79	High	4	TruSeq Adapter, Index 5	4	6.8

Florida-07	-32	0.00	4.47	High	4	TruSeq Adapter, Index 13	4	7.6
Florida-07	-64	1.89	3.76	High	4	TruSeq Adapter, Index 18	4	7.1
Florida-07	+16	2.81	8.59	High	4	TruSeq Adapter, Index 6	4	7
Florida-07	+32	17.20	6.13	High	4	TruSeq Adapter, Index 14	4	8
Florida-07	+64	726.71	5208.14	High	4	TruSeq Adapter, Index 19	4	8.1
ICG 1471	-16	0.00	0.44	High	4	TruSeq Adapter, Index 2	4	7
ICG 1471	-32	0.00	1.78	High	4	TruSeq Adapter, Index 7	4	7
ICG 1471	-64	0.00	1.88	High	4	TruSeq Adapter, Index 15	4	6.3
ICG 1471	+16	4.71	1.67	High	4	TruSeq Adapter, Index 4	4	na
ICG 1471	+32	16.19	-0.67	High	4	TruSeq Adapter, Index 12	4	6.3
ICG 1471	+64	627.20	541.97	High	4	TruSeq Adapter, Index 16	4	7.2
Florida-07	-16	0.00	8.07	High	5	TruSeq Adapter, Index 5	5	7
Florida-07	-32	3.52	5.83	High	5	TruSeq Adapter, Index 13	5	6.6
Florida-07	-64	4.84	4.64	High	5	TruSeq Adapter, Index 18	5	6.1
Florida-07	+16	1.74	16.71	High	5	TruSeq Adapter, Index 6	5	86.9
Florida-07	+32	30.86	13.75	High	5	TruSeq Adapter, Index 14	5	8
Florida-07	+64	644.20	4116.44	High	5	TruSeq Adapter, Index 19	5	7.9
ICG 1471	-16	0.00	2.71	High	5	TruSeq Adapter, Index 2	5	6.6
ICG 1471	-32	0.00	2.77	High	5	TruSeq Adapter, Index 7	5	6.7
ICG 1471	-64	0.00	1.53	High	5	TruSeq Adapter, Index 15	5	6.4
ICG 1471	+16	6.40	-0.72	High	5	TruSeq Adapter, Index 4	5	6.8
ICG 1471	+32	38.81	0.75	High	5	TruSeq Adapter, Index 12	5	6.4
ICG 1471	+64	557.88	320.23	High	5	TruSeq Adapter, Index 16	5	7
Florida-07	-16	1.61	5.76	High	6	TruSeq Adapter, Index 5	6	6.9
Florida-07	-32	3.62	4.72	High	6	TruSeq Adapter, Index 13	6	6.3
Florida-07	-64	2.74	5.05	High	6	TruSeq Adapter, Index 18	6	6.5
Florida-07	+16	4.82	10.96	High	6	TruSeq Adapter, Index 6	6	7
Florida-07	+32	13.19	5.25	High	6	TruSeq Adapter, Index 14	6	7.9
Florida-07	+64	592.85	3006.58	High	6	TruSeq Adapter, Index 19	6	8.3
ICG 1471	-16	0.00	3.07	High	6	TruSeq Adapter, Index 2	6	6.8
ICG 1471	-32	0.00	3.45	High	6	TruSeq Adapter, Index 7	6	8.7
ICG 1471	-64	0.00	2.31	High	6	TruSeq Adapter, Index 15	6	6.3
ICG 1471	+16	6.31	-0.55	High	6	TruSeq Adapter, Index 4	6	6.7
ICG 1471	+32	17.48	0.18	High	6	TruSeq Adapter, Index 12	6	6.5
ICG 1471	+64	677.24	636.33	High	6	TruSeq Adapter, Index 16	6	7.1

Trt: treatments; Rep: Replicate; +16, +32 and +64: the infection treatments of 16, 32 and 64 HAI; -16, -32 and -64: the control treatments of 16, 32 and 64 HAI.

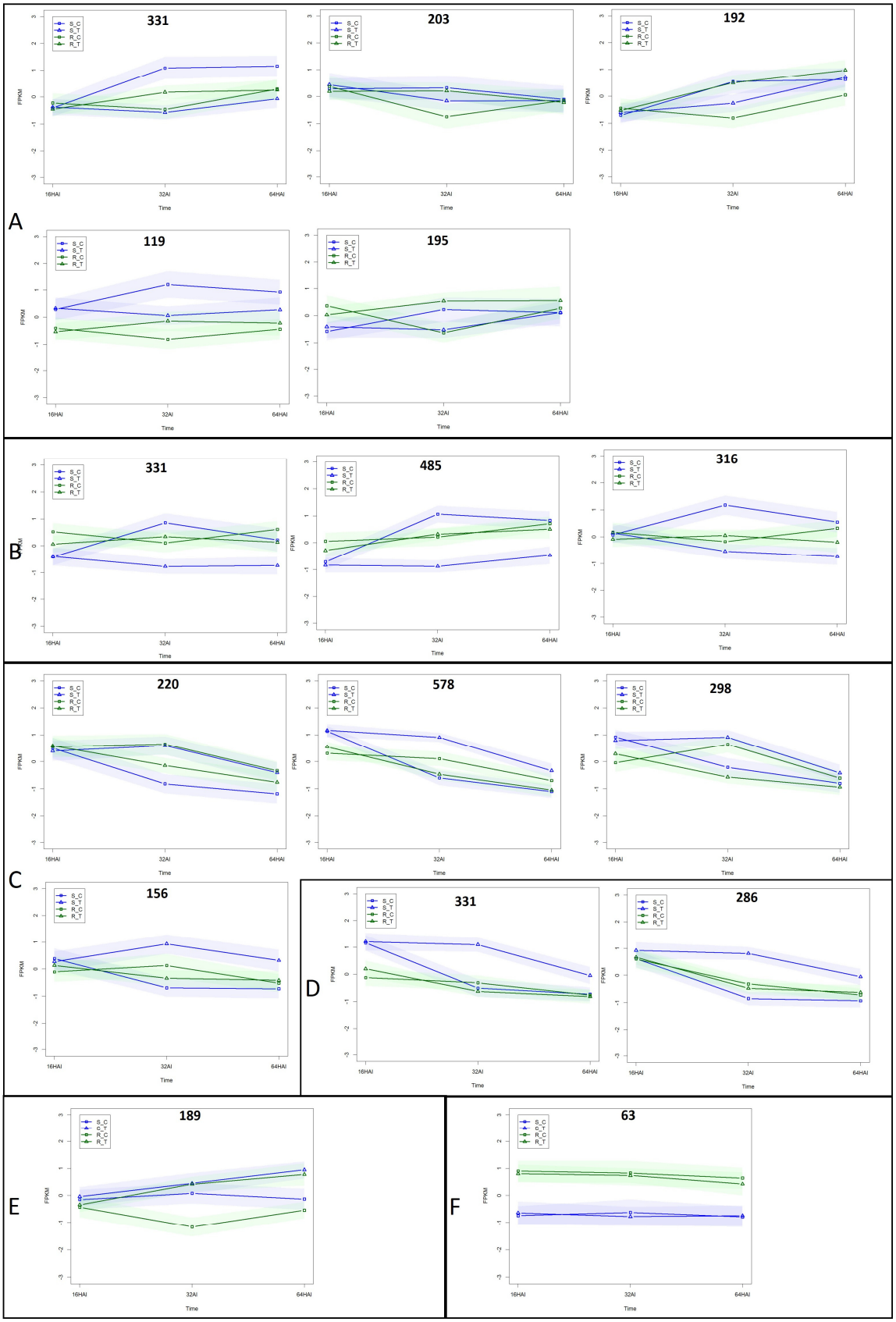

A
331
203
192
119
195
B
331
485
316
C
220
578
298
156
D
331
286
E
189
F
63
S_C
S_T
R_C
R_T
FPKM
Time
16HAI
32AI
64HAI

Figure 5.S1: SOM clusters of differentially expressed genes between genotypes due to the infection effect. S: Susceptible genotype (Florida-07); R: Resistant genotype (ICG 1471); C: Control treatments; T: Infection treatments; Red lines: Resistant genotype; Blue lines: susceptible genotype; Rectangles: Controls; Triangles: Treatments; A: Genes that were down-regulated due to the infection in the susceptible genotype and up-regulated in the resistant one; B: Genes that were down-regulated due to the infection in the susceptible genotype and unaffected in the resistant one; C: Genes that were up-regulated due to the infection in the susceptible genotype and down-regulated in the resistant one; D: Genes that were up-regulated due to the infection in the susceptible genotype and unaffected in the resistant one; E: Genes that were up-regulated in both genotypes but more so in the resistant one; F: Genes that were highly expressed in the resistant genotype and lowly expressed in the susceptible one either for the infection or control treatments.

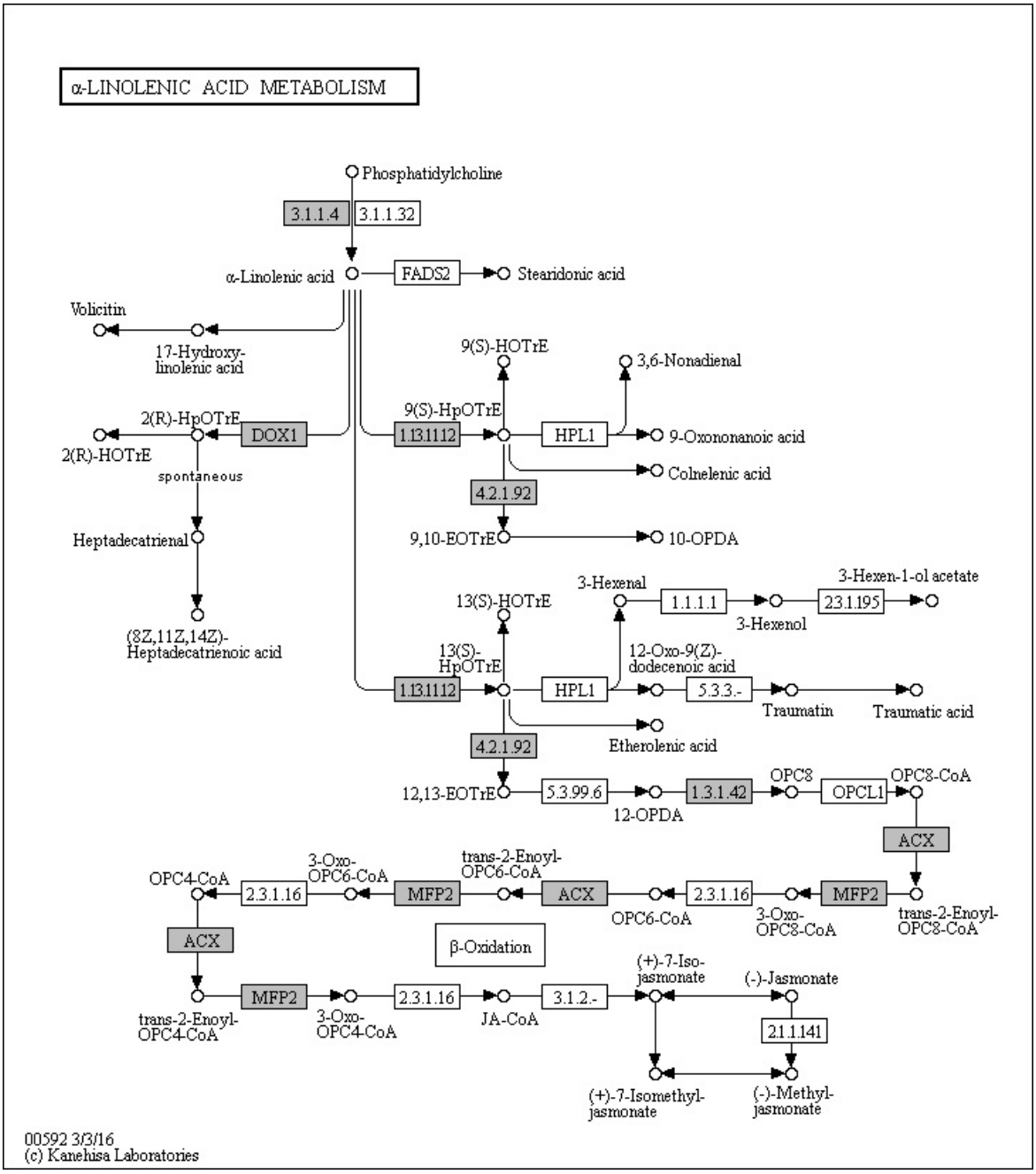


Figure 5.S2: KEGG enrichment significant pathways; shaded boxes are the enzymes coded by differentially expressed genes.

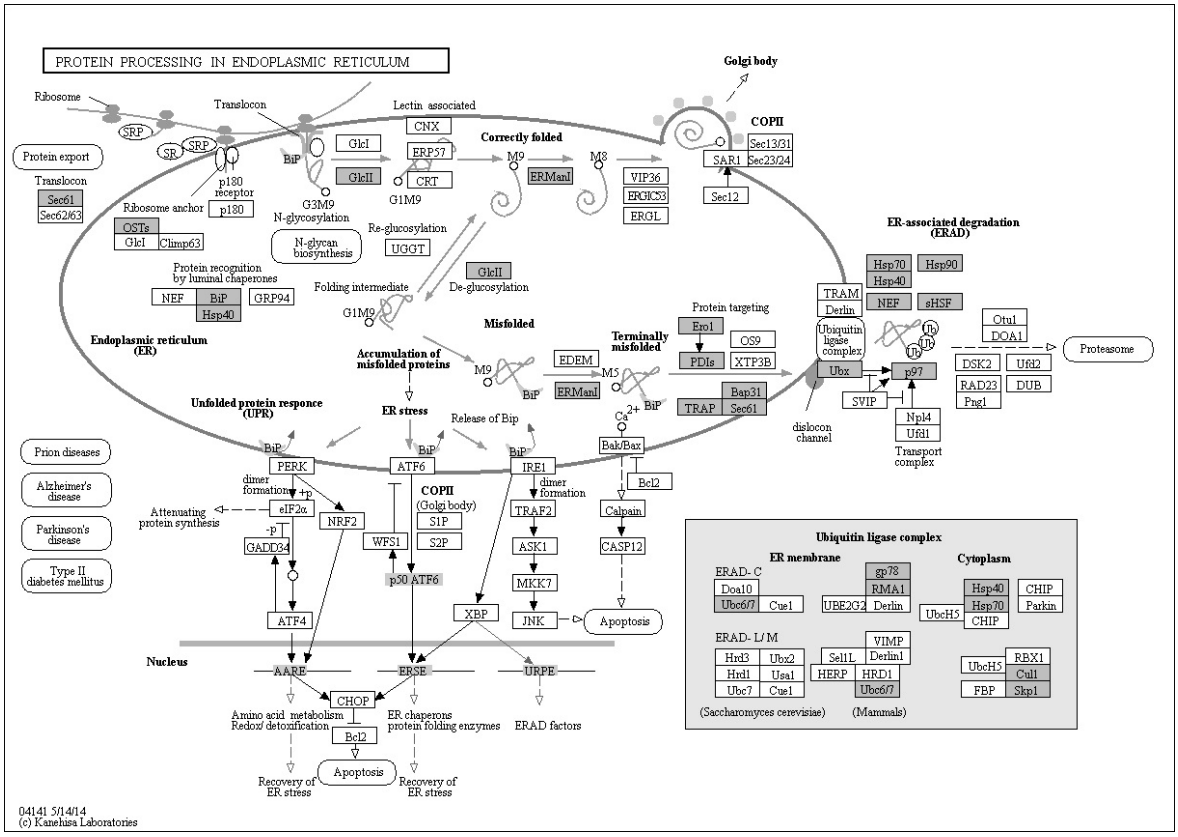


Figure 5.S3: KEGG enrichment significant pathways; shaded boxes are the enzymes coded by differentially expressed genes.

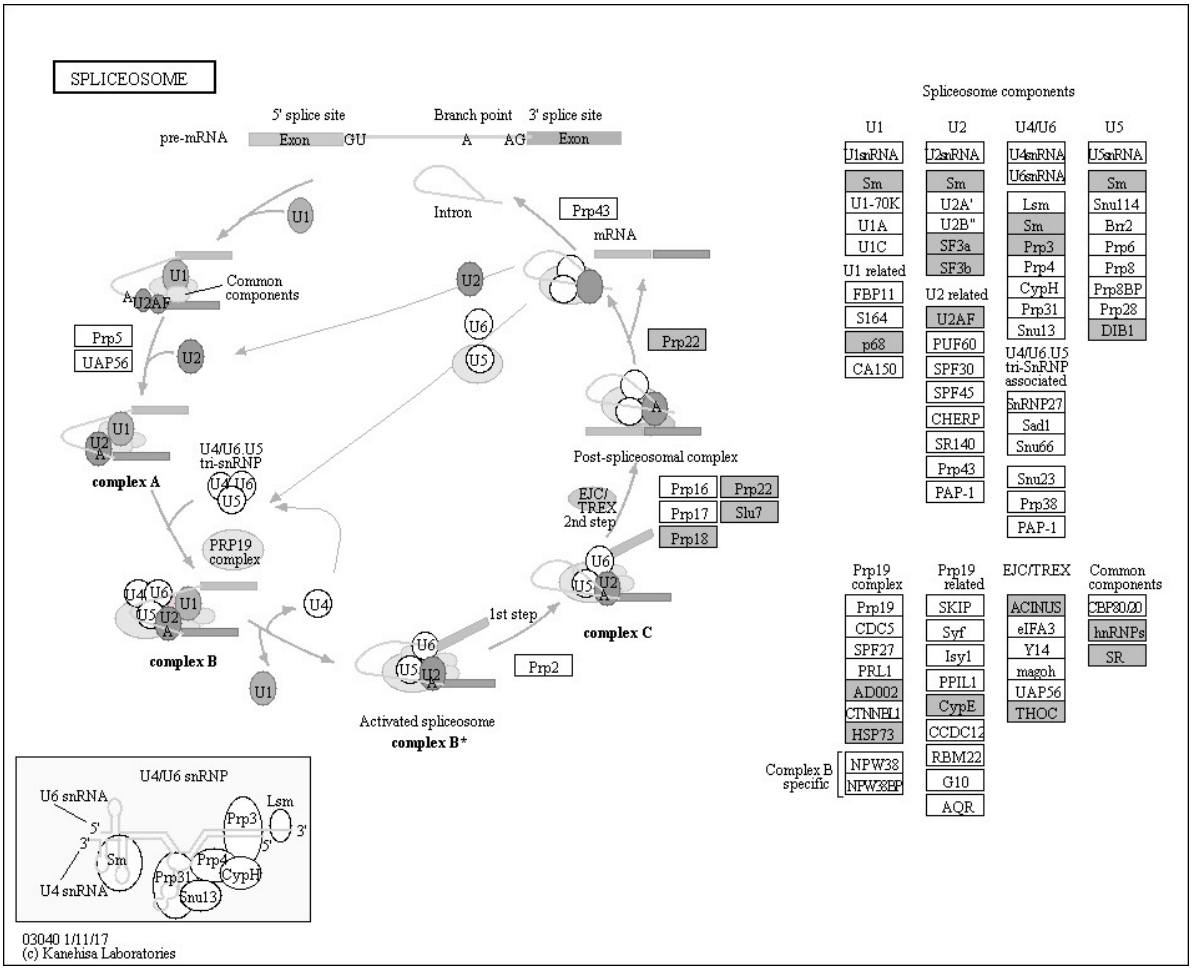


Figure 5.S4: KEGG enrichment significant pathways; shaded boxes are the enzymes coded by differentially expressed genes.

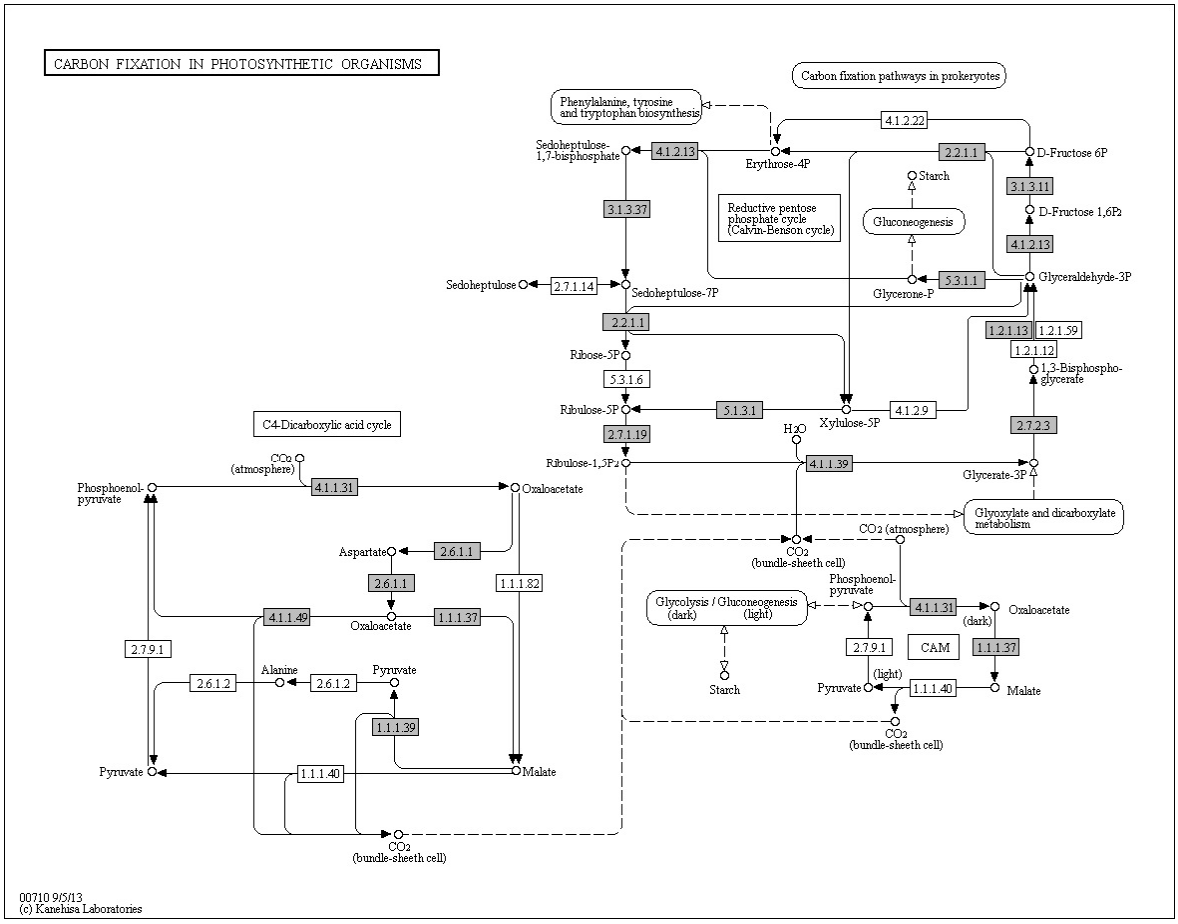


Figure 5.S5: KEGG enrichment significant pathways; shaded boxes are the enzymes coded by differentially expressed genes.

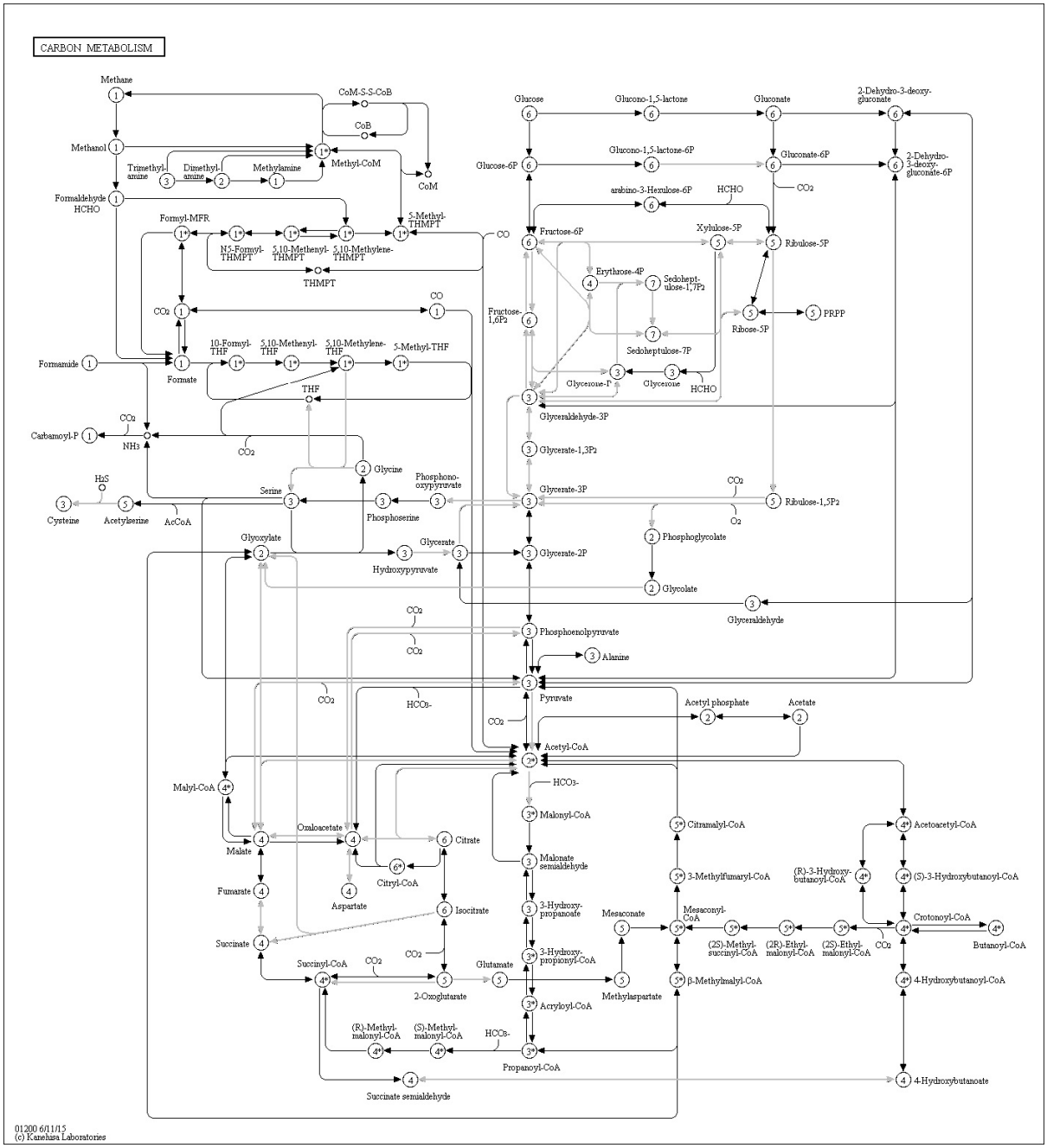


Figure 5.S6: KEGG enrichment significant pathways.

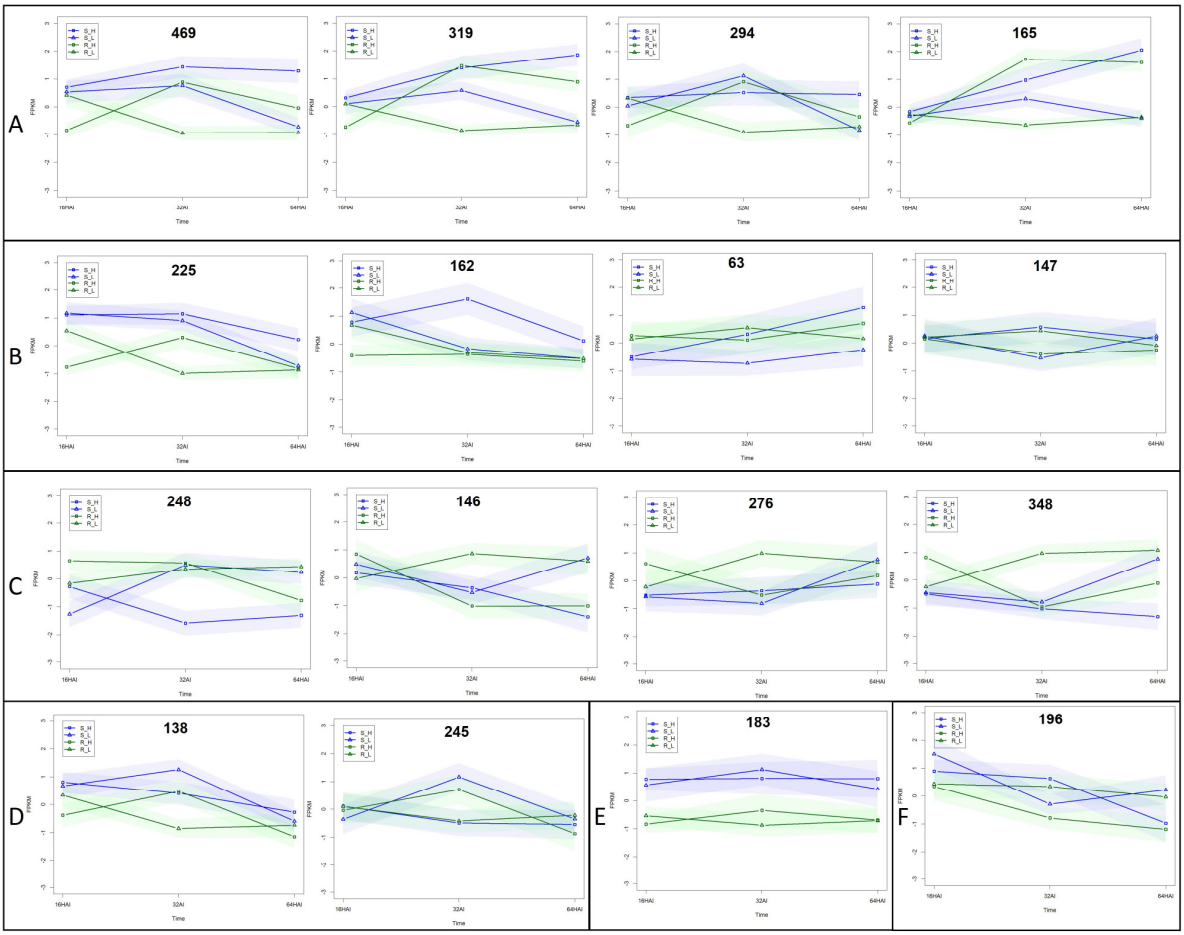


Figure 5.S7: SOM clusters groups of differentially expressed genes between genotypes due to the infection effect. S: susceptible genotype (Florida-07); R: resistant genotype (ICG 1471); H: high aflatoxin-producing cluster; L: low aflatoxin-producing cluster; Red lines: Resistant genotype; Blue lines: Susceptible genotype; Rectangles: Highly contaminated seeds; Triangles: lowly contaminated seeds; A: Genes that were highly up-regulated in the susceptable genotype due to highly contaminated seeds; B: Genes that were up-regulated in susceptible genotype and down-regulated in the resistant one due to highly contaminated seeds; C: Genes that were down-regulated in both genotypes due to the highly contaminated seeds; D: Genes that were up-regulated in the resistant genotype and down-regulated in the susceptible one due to the highly contaminated seeds; E: Genes that were highly expressed in the susceptible genotype and lowly

expressed in the resistant one either for high or low contaminated seeds; F: Genes that were down-regulated by infection progress in both genotypes either for high or low contaminated seeds.

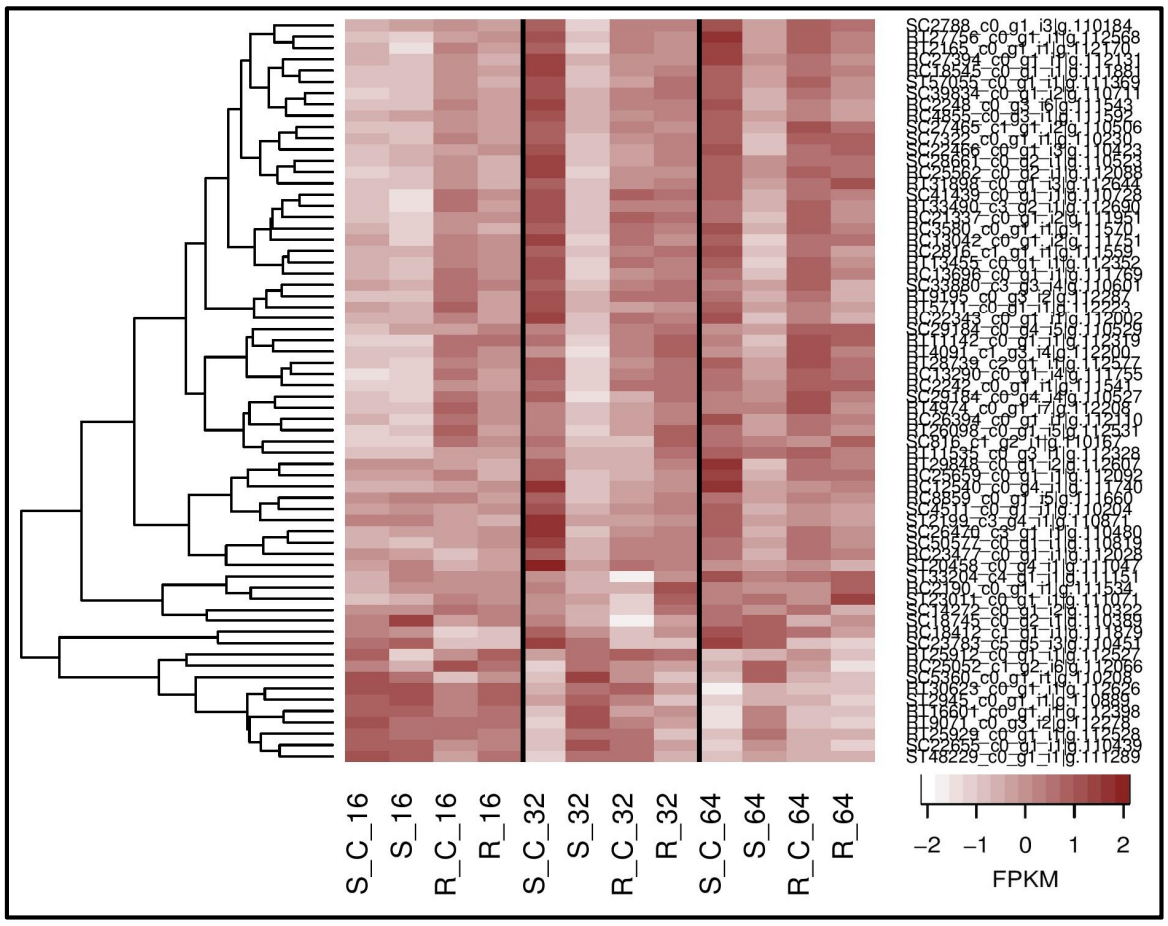


Figure 5.S8: Expression profile of the new transcripts discovered from the de-novo assembly, FPKM: Z-scores of Fragments Per Kilobase of transcript per Million mapped reads.

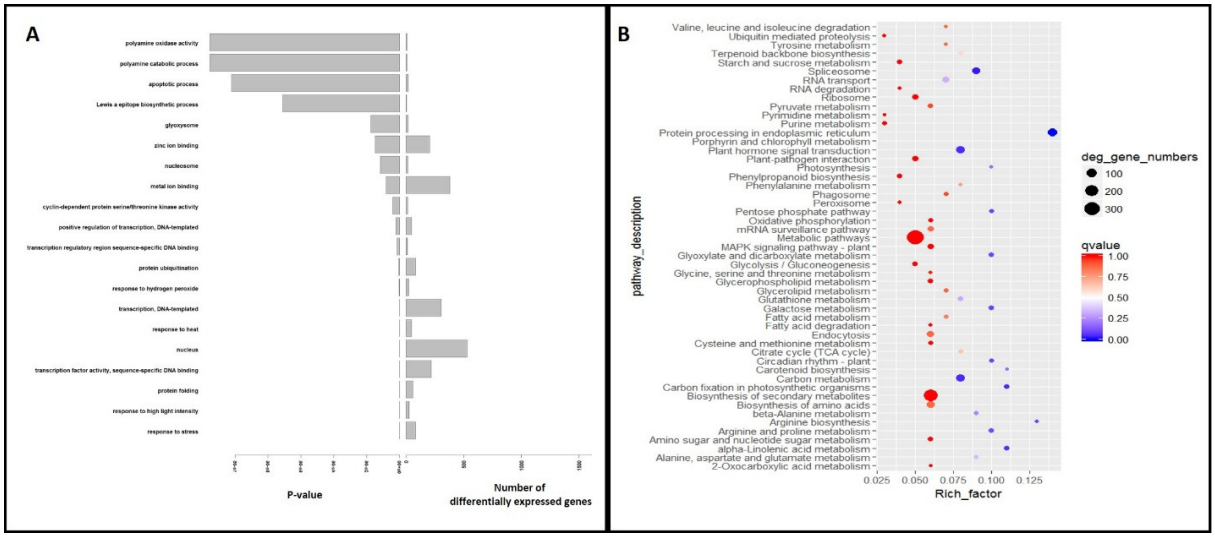


Figure 5.S9: GO/KEGG enrichment analysis of differently expressed genes between peanut genotypes due to the infection vs control of the newly assembled transcripts. A. The 20 most significant GO terms extracted by GO enrichment analysis; B. KEGG enrichment analysis carried out by keggseq package; Rich_factor: the ratio of the differentially expressed genes to all genes that were annotated in the pathway.

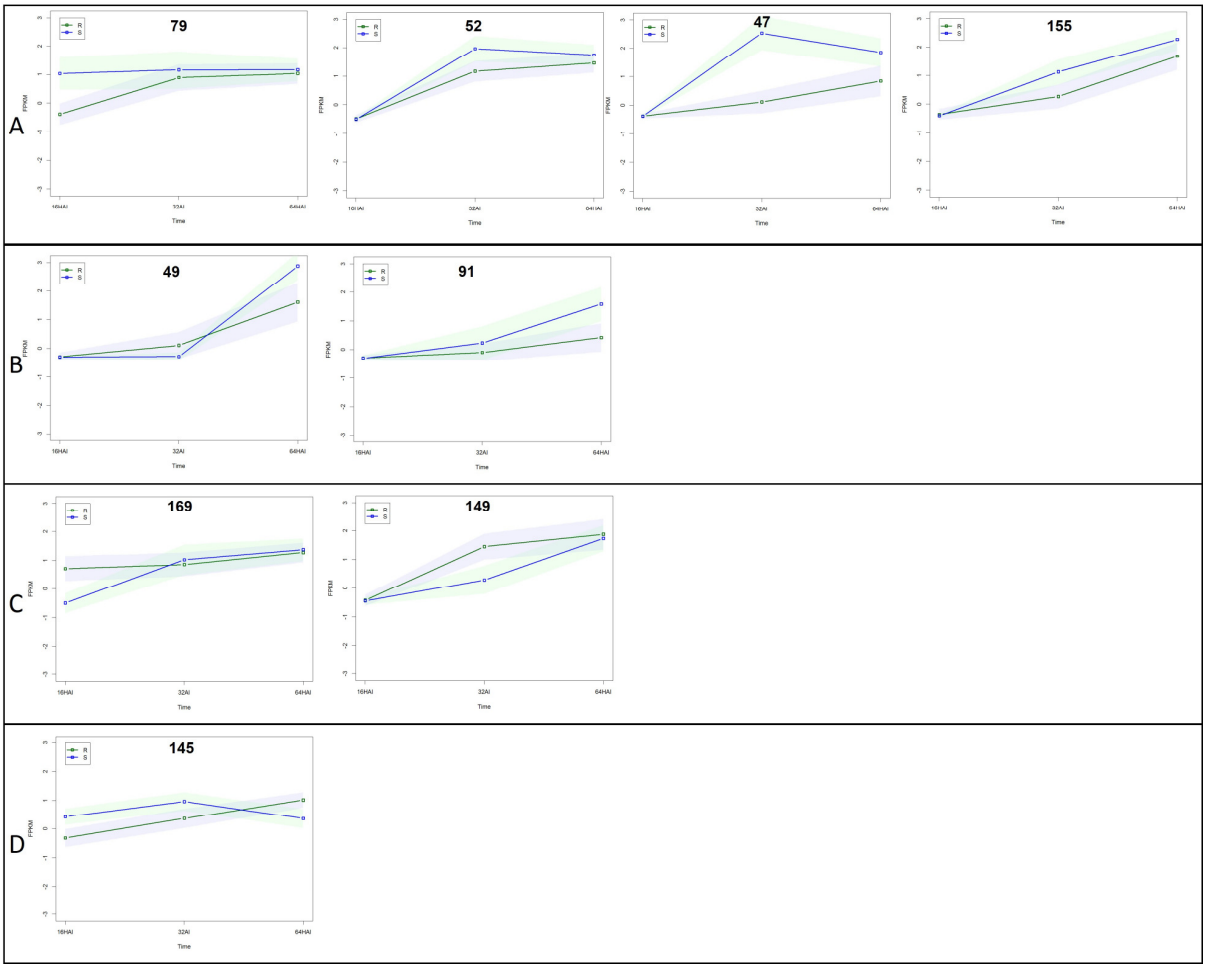


Figure 5.S10: SOM clusters of fungal differentially expressed genes; S: susceptible genotype (Florida-07); R: resistant genotype (ICG 1471); Red lines: Fungus was grown on the resistant genotype; Blue lines: Fungus was grown on the susceptible genotype; A,B: Different patterns of up-regulated genes in case of susceptible genotype; C, D: Different patterns of up-regulated genes in case of resistant genotype

CHAPTER 6

SUMMARY

Post-harvest aflatoxin accumulation is a serious problem for peanut production. The interaction between peanut and aflatoxin-producing fungi, *Aspergilli*, is complex and strongly affected by environmental conditions. Pre-harvest aflatoxin contamination can be mitigated by irrigation, drought tolerant genotypes or informed cultural practices. Post-harvest aflatoxin reduction could benefit from resistant peanut lines in addition to proper storage. To breed for genetic resistance, heritable factors underlying resistance mechanisms should be identified. Up to now, a small number of genes, *e.g.*, lipoxygenases and WRKY, were shown to be responsive to infection by *Aspergilli*. However, the results are tentative and the exact roles of these genes in the resistance mechanism are not yet known. Therefore, the present work was carried out to characterize interactions between the fungus and resistant and susceptible peanut genotypes to identify genes potentially involved in the resistance. Although, the whole genome sequence of cultivated peanut, which is allo-tetraploid, is not available, the genomes of the two diploid progenitors, *A. duranensis* and *A. ipaensis,* recently were released. In addition, the *A. flavus* genome also has been sequenced. These genomes provided essential resources for the project to progress.

Lipoxygenases (LOXs) are a gene superfamily of dioxygenases that work on cis-1,4-pentadiene structure of fatty acids, and their role in *Aspergillus spp.* –peanut interaction has long been debated. In this project, lipoxygenases of cultivated peanut were assigned utilizing gene models from the two diploid progenitor genomes in which 24 and 25 LOX genes were found in

A. duranensis and *A. ipaensis*, respectively. A study of orthology showed that 20 genes of one subgenome had orthologous counterparts with the second genome. In addition, the genes were distributed across peanut chromosomes and tended to be located near to telomeres within rearranged chromosomal regions. However, not all these genes are likely to be functional LOXs since some were truncated or disrupted. Therefore, the study included a detailed structural analysis, which revealed 38 full-length LOX genes in tetraploid peanut. LOX genes were functionally classified into four clusters utilizing previously published LOXs from different resources; 9S_typeII, 13S_typeII classes that generate 9S- to 13S- hydroperoxy products, respectively, and have chloroplast transit peptides; 9S_typeI, 13S_typeI that produce the same products, however remain as extra-plastdial enzymes. The latter class was clustered into two sub-groups depending on the ratio of 13S- to 9S- hydroperoxy products *i.e.,* 13S-typeI_SG1, which has a high ratio (> 0.4), and 13S-typeI_SG2, which has a moderate ratio (< 0.4). The expression profiles of these LOXs were estimated in a wide range of peanut tissues including 22 different tissues/stages. The profile revealed three prominent patterns: the first group contained genes highly expressed in seed tissues, the second group of genes were expressed in tissues other than seeds and the third group was constitutively expressed across all tissues. Furthermore, the expression profiling was carried out to study the response of LOXs to different biotic stresses, *i.e.,* pre-, post- harvest aflatoxin contamination, nematode invasion, late leaf spot infection, and the interaction with *Bradyrhizobium spp*. Genes that were highly expressed in seed tissues were observed to respond to pre- and post- harvest aflatoxin contamination. No new LOX genes were identified in a *de novo* assembly of the tetraploid transcriptome data, which suggests that all LOXs were duplicated within the progenitor genomes before their hybridization and that no LOXs were further duplicated after tetraploidization.

The first and more important step to breed peanut for post-harvest aflatoxin contamination resistance is to reliably identify sources of resistance. To achieve that goal, an inoculation assay using a GFP-expressing strain of *A. flavus* was developed and used to screen 10 peanut genotypes, *i.e.*, ICGV88145, ICG 1471, GT-C20, C76-16, A72, A69, Tifguard, NC 3033, Tifrunner and Florida-07. The screening included studying the fungal growth on the different genotypes by tracking the GFP expression visually every 8 hours after infection and up to 72 hours. In addition, an image processing software was designed using MATLAB script, designated SICIA (Seed Infection Coverage and Intensity Analyzer), and used to phenotype the fungal growth of the last time point with greater accuracy than visual rating. Upon statistical analysis, NC 3033 demonstrated the lowest surface fungal growth while GT-C20 had the greatest. These two genotypes plus three additional genotypes were selected for single seed GFP quantification and aflatoxin analysis. No significant differences were found in GFP amount across the five genotypes; however, aflatoxin was significantly different among them. Florida-07 and ICG 1471 genotypes showed the highest and lowest aflatoxin contamination, respectively. This result suggests that these genotypes are not able to stop or retard *Aspergillus spp.* growth. However, the genotypes differ in their ability to produce aflatoxin. The susceptible genotypes enhance aflatoxin production and the resistant genotypes only retard the formation of aflatoxin production. In addition, the results support ICG 1471 as a candidate resistant genotype for post-harvest aflatoxin contamination breeding.

As cultivated peanut is tetraploid, extracting true SNPs directly from Next Generation Sequence (NGS) data is difficult. Different methods were proposed to increase the efficiency of calling SNPs from NGS of polyploids. However, the published efficiency in peanut was very low, close to 10%. Recently, a novel tool was created, Sliding Window Extraction of Explicit

Polymorphisms (SWEEP). The tool was implemented to design the Axiom_Arachis 58K SNP Chip, which validated an increase in the efficiency of true SNPs to 40%. In the present study, a data set of true and false SNPs was created based on SNP-array analysis of 21 different tetraploid peanut genotypes. Whole genome re-sequencing data of these genotypes was used to calculate different sequence and mapping traits. Neural network models were built to test the most effective traits. Among the tested traits, eight were selected to build the final training model. In addition, different machine learning approaches were tested alongside the neural network model to determine the most efficient models, *i.e.,* logistic regression, Discriminant analysis, K-nearest neighbors, Naïve Bayes, decision tree and tree bagger. Tree bagger showed the best results after neural network; therefore, the two models were combined to create a model for SNP calling of re-sequencing data of tetraploid peanut. Testing the model showed more than 80% efficiency of calling SNPs. However, when RNA-seq expression data from nine genotypes was tested, the efficiency dropped to 78%, but the higher efficiency was recovered by developing another combined model for this type of data. Since there are no similar datasets readily available for other polyploid crops, simulated genotypes with different sequence depths were created for tetraploid cotton, hexaploid wheat, and octoploid strawberry and used to create models. Comparison of the efficiencies of these models with a model built by peanut simulated data showed similar results; all simulated models showed a very high accuracy (>90). A SNP calling tool was designed using C++/Python script that uses neural network and/or tree bagger models to call SNPs in polyploids, designated SNP-ML; the tool is supported by an application for creating user customized models for new data sets, designated SNP-MLer. Furthermore, an extendable database was created containing all previously described models and future user customized models.

To study the genetic factors and pathways that control post-harvest aflatoxin resistance, an RNA-seq experiment was designed to differentiate the response for the infection and the subsequent aflatoxin accumulation in resistant ICG 1471 and susceptible Florida-07 genotypes. A GFP-expressing *A. flavus* strain was used to infect the seeds *in vitro* and the seeds were harvested at three time points, *i.e.*, 16, 32 and 64 hours after inoculation. A randomized complete block design was used, and the GFP expression and aflatoxin were quantified based on single seed analysis. The statistical analysis confirmed the previous results since no significant differences were found between the two genotypes in GFP expression, yet aflatoxin accumulation was significantly different; ICG 1471 supports low aflatoxin contamination and Florida-07 produces a high amount of aflatoxins.

Based on interaction plots between the two genotypes including both aflatoxin and GFP analysis, six seeds were selected for every genotype/time point as biological replicates. Three seeds were selected from each of two clusters, with clusters representing the environmental variations for infection. In addition, six mock-infected seeds were selected as negative controls/genotype/time point.

The sequence files were cleaned and mapped to a synthetic tetraploid reference genome that contained the combined genomes of *A. duranensis* and *A. ipaensis* to test the peanut response to the infection. In addition, sequences were mapped to the *A. flavus* genome to test the response of the fungus to peanut genotype.

4272 and 3845 genes were differentially expressed between the two peanut genotypes as a result of infection and environmental variations, respectively. In addition, 1197 fungal genes were differentially expressed for the fungus growing on the two genotypes. Different statistical models for RNA-seq time course experimental analysis were used to validate the results.

Moreover, extended analysis was carried out such as SOM (self organizing map) clustering and GO (gene ontology) term enrichment analysis.

To perform KEGG enrichment analysis, an R package was designed, designated 'keggseq', since currently available tools do not support combining subgenomes. In addition to combining genomes, 'keggseq' had many other features such as simplicity, run-time analysis and creating standard graphs for publications. The implementation of the package revealed the importance of alpha-linolenic acid metabolism, protein processing in endoplasmic reticulum, splicesome and carbon fixation and metabolism pathways in peanut resistance to *A. flavus*.

Moreover, co-expression network analysis revealed the importance of some defense related proteins in the resistance. Eight WRKY family transcription factors, nine TIR-NBS-LRR, six ethylene signaling and one heat shock proteins were found significantly up-regulated and highly co-expressed with the AflNA gene of *Aspergillus flavus* only in the interaction with the resistant genotype, ICG 1471. The same scenario was discovered for a heat shock and an ethylene signaling protein with the AflH gene. Both AflNA and AflH are proteins upstream in the aflatoxin biosynthesis pathway.

In summary, this work created a comprehensive study of tetraploid peanut lipoxygenases including their functional classification, developmental expression and disease response. Furthermore, this study identified ICG 1471 as a strong candidate for post-harvest aflatoxin resistance, and revealed important genes and pathways that influence the resistance mechanism. Finally, three computational biology tools were created for phenotyping infection, and SNP calling and KEGG enrichment analysis of polyploids. These results and tools introduce valuable materials for breeding of peanut and other polyploids.