NETWORK ANALYSIS TO REVEAL JASMONIC ACID PATHWAY RELATED MECHANISTIC CHANGES FOR DROUGHT-TOLERENT MAIZE

A Thesis

by

FAN LIU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Xiaoning Qian
Co-Chair of Committee,	Ulisses Braga-Neto
Committee Members,	James Cai
	Tie Liu
Head of Department,	Miroslav Begovic

December 2018

Major Subject: Electrical Engineering

ABSTRACT

Jasmonic Acid (JA) is one of the important substances that associates with stomatal closure under drought stress for maize. In this project, we performed multiple network analyses to identify the expression patterns of JA pathway genes in drought responses of different maize genotypes where certain genes from the related pathway were knocked out. Differential gene expression analysis was performed to identify differentially expressed genes across different water conditions for each maize genotype, respectively. As the number of samples is limited in this project, we used partial correlation for co-expression network construction for different genotypes using the identified DEGs and a weighted LASSO regression method that applies to all samples for detailed network connectivity analyses of the JA pathway. The results reveal the changes of behaviors and regulatory relationships of JA pathway genes in different genotypes. We discussed the shifting of behaviors between different genotypes from the analysis results.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Xiaoning Qian [advisor], Professor Ulisses Braga-Neto and Professor Tie Liu of the Department of Electrical and Computer Engineering and Professor James Cai of the Department of Veterinary Medicine and Biomedical Sciences. The data to be analyzed was provided by Professor Mike Kolomiets from the Department of Plant Pathology and Microbiology. There are no outside funding contributions to acknowledge related to the research and compilation of this document. The work conducted for the thesis is completed by the student independently.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Maize is one of the primary crops in Texas and drought is one of the most harmful environmental factors that affects the growth and production of maize, especially in Texas. In 2017, Texas maize growers planted almost 2.24 million acres of maize, while only 40 percent of total maize acres in Texas are irrigated. Drought tolerance, therefore, is a highly appreciated trait for maize breeding. A clearer understanding of the molecular mechanisms of maize's drought response is very important for the goal of a successful improvement on this crop. One of the first responses to drought stress for plants is stomatal closure which controls the dehydration of plant and Jasmonic Acid (JA) is one of the important substances that associates with stomatal closure under drought stress [1]. Specifically, 12-OPDA, the JA precursor, is the molecule promoting stomatal closure and JA-Ile is the molecule that promotes the opening of stomates. Therefore, the understanding of 12-OPDA-JA synthetic pathway behavior is important. Various research had been conducted about JA pathway in other crops [2-5], however the detailed molecular mechanism for JA pathway in maize is still elusive. In recent years, network-based methods have been widely adopted in genomic data analysis. These methods can jointly analyze genes in the same pathway or subnetwork module, predict the activity levels of a given pathway and relate it with different phenotypes based on the analysis [6-12].

In this project, we performed multiple network analysis to identify the expression patterns of JA pathway genes in drought responses of different maize genotypes where

certain genes from the related pathway were knocked out. This study will help to further elucidate the shifting of JA pathway and the downstream effect of other genes' behaviors in maize drought response, therefore provide valuable candidate genes that could be used to develop maize genotypes with better drought tolerance.

Specifically, we have performed differential gene expression analysis to identify significant differentially expressed genes across different water conditions for wild-type as well as different mutated/knockout maize genotypes, respectively. The main obstacle in this project is the limit of samples: there are no more than 9 samples for each genotype and the well adapted robust methods, like WGCNA [13], are not suitable for such small sample size. To overcome this difficulty, co-expression network construction for different genotypes were constructed with the identified DEGs using partial correlation. Detailed network connectivity analysis of the JA pathway was performed for understanding the pathway shifting. Because the sample size for each genotype is too small for applying regular linear regression methods to each genotype's data, we used a weighted LASSO regression method that applies to all samples to reveal the changes of regulatory relationships of JA pathway genes in different genotypes.

This thesis consists of four chapters and appendix. The first chapter introduced the background and a summary of all the works, the second chapter described the materials and all the methodology for this study in detail. The third chapter discussed all the results and the fourth chapter is the conclusion. We provided additional analysis results in the appendix as supplementary materials.

2. MATERIALS AND METHODOLOGY

The overall working pipeline of this project is illustrated in Figure 1. First, the RNA-seq data of every maize genotype were compared between well-watered status and drought stressed for differential expression analysis and differentially expressed genes for each genotype were identified. Second, the drought response co-expression networks were constructed with the previously identified differentially expressed genes using partial correlation. Next, we compared the local and global centralities of the JA pathway genes for behavior analysis and finally, we used a weighted LASSO regression method to find the regulatory network of JA pathway for each genotype and reveal the changes of regulatory relationships of JA pathway genes. In the following subsections, we will introduce the data we used and analysis methodology in detail.

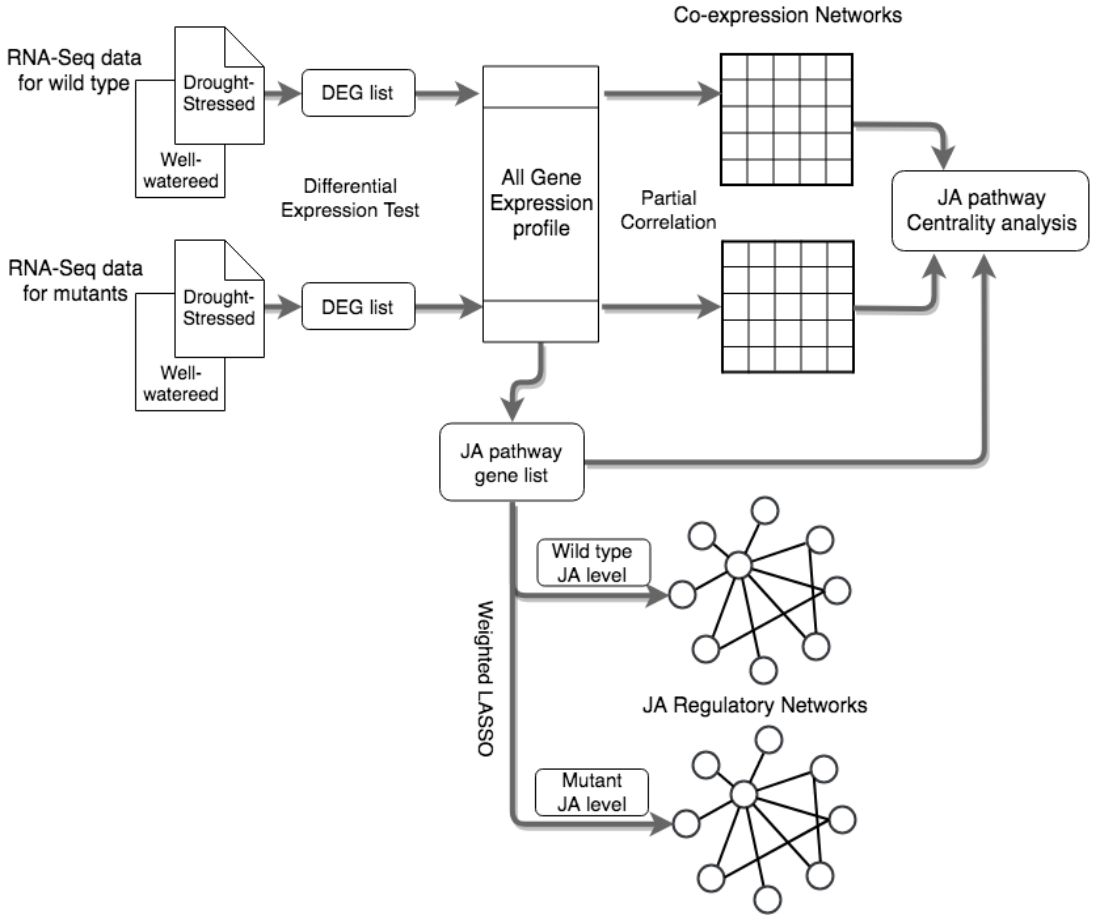

Figure 1. The working pipeline for the whole project

2.1 Materials

We focus on the study of drought tolerance for the maize mutations with single gene knockouts (*lox2*, *lox4* with genes *LOX2* and *LOX4* knocked out respectively) and two-gene knockouts (*lox2lox4, opr7opr8* with *LOX2* and *LOX4*, *OPR7* and *OPR8* knocked out respectively) respectively, comparing to the wild-type maize (B73). The *lox4* and *opr7opr8* mutants are considered as drought tolerant genotypes as they have significantly reduced water loss through transpiration, while the *lox2* mutant had much higher transpirational water loss. The leave samples of these different maize genotypes

were treated in well-watered condition (WW), drought stress 4 days (D4) and drought stress 6 days (D6). The gene read extraction from plant leaves were performed by HiSeq HO-50 with 50 bp single-end. There are 3 biological replicates for each condition and each genotype, leading to 42 samples in total. In addition, the JA levels for each maize genotype at well-watered, drought 4 days and drought 6 days timepoints were collected for determining the characterization of each sample, as shown in Table 1.

JA-Ile/nmol	Well-Watered/D0	Drought-Stressed/D4	Drought-Stressed/D6
B73	0.008678111	0.043974411	0.309116225
lox2	0.009767507	0.017042564	0.499436978
lox4	0.015794705	0.022293994	0.008952668
lox2lox4	0.003230448	0.023636207	0.164577413
opr7opr8	0	0	0

Table 1. JA-Ile content for each genotype at each timepoint

The alignments for reading were processed with Cufflinks [14] which assembles transcript isoforms and the gene expression values were quantified as Fragments Per Kilobase of transcript per Million mapped reads (FPKM) using the Maize B73 Reference Genome from MaizeGDB [15] as the reference annotation.

2.2 Differential expression analysis

Differential expression analysis was performed using the Cuffdiff toolkit [16]. In Cuffdiff software, the change of a transcript's expression level is measures by using a statistical approach that compares the distribution of fragment count in each condition. First, for each condition, the fragment count k_{ij} in each replica j is geometric

normalized with the scale $s_j = \text{median}_i \left(\frac{k_{ij}}{(\prod_{v=1}^{m} k_{iv})^{\frac{1}{m}}} \right)$, which denotes the size of library in replica j. Second, for each replicate, the Cuffdiff algorithm assume that the number for each transcript x_t^j in each replica j approximately follows a negative binomial distribution $x_t^j \sim NB(\mu_t^j, {\sigma_t^j}^2)$. The variances for these distributions are estimated through fitting a generalized linear model with the normalized fragment counts from the previous step. The fragment count uncertainty for each transcript is modeled with a beta distribution and fitted through maximum likelihood estimation. Next, for each condition, the Cuffdiff algorithm combines the overdispersion of fragment numbers and the uncertainty of each transcript's fragment count into one beta negative binomial distribution which reflects both sources of variability in an isoform's measured expression level. After getting the estimation of distribution for every transcript, the posterior distribution of each gene is calculated by combining the distribution of all the transcripts of the gene's isoforms into a posterior distribution for the expression level of the gene [16]. Cuffdiff then compute the $\log_2$ fold change, the p-value and adjusted p-value for multiple test statistics. To better address the significance of the differentially expressed genes and cut down the number of variables in the co-expression network construction, we determine the differentially expressed genes (DEGs) as the genes that both have p-value smaller than a 0.0001 threshold and a $\log_2$ fold change that is larger than 0.58 or smaller than -0.58.

2.3 Co-expression network construction and network topological analysis for JA genes

We assumed all the differentially expressed genes (DEGs) between well-watered condition and drought-stressed condition are the only genes that involved in the drought response. Five co-expression networks for each maize genotype respectively were constructed within the significantly co-expressed DEGs. As we only have no more than nine samples for each genotype from our collaborator, instead of using pairwise Pearson correlation (PCC)[1], here we used partial correlation to infer the co-expression networks for each genotype. Unlike pairwise correlation that only reflects the marginal dependency between variables, partial correlation can measure the strength of association between a pair of genes while eliminating the effect of other genes in the dataset, therefore reduce false positive detection for connections when the number of samples is small, which many previous researches preferred [17-19].

For an observed data matrix $\mathbb{X} = \left[\mathbf{X}_1, \dots, \mathbf{X}_p\right] = \left[\mathbf{X}_{(1)}^T, \dots, \mathbf{X}_{(p)}^T\right]^T$ where $\mathbf{X_i} = [X_{1i}, \dots, X_{ni}]$ and $\mathbf{X}_{(i)} = \left[X_{i1}, \dots, X_{ip}\right]$ denote the i-th column and row respectively, the sample mean of the i-th column is given by $\overline{X_\iota} = \frac{1}{n}\sum_{j=1}^{n} X_{ji}$, the vector of sample means is given by $\overline{\mathbf{X}} = \left[\overline{X_1}, \dots, \overline{X_p}\right]$. Then the covariance matrix is given by $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\mathbf{X}_{(i)} - \overline{\mathbf{X}}\right)^{\boldsymbol{T}}\left(\mathbf{X}_{(i)} - \overline{\mathbf{X}}\right)$ and the Pearson correlation matrix is defined as $\mathbf{R} = \mathbf{D_S}^{-\frac{1}{2}}\,\mathbf{S}\mathbf{D_S}^{-\frac{1}{2}}$, where $\mathbf{D_S}$ is the diagonal of $\mathbf{S}$. For a positive definite correlation matrix $\mathbf{R}$, the partial correlation

[1] The co-expression network construction and analysis results based on PCC can be found in the Appendix.

matrix can be defined as $\mathbf{P} = \mathbf{D}_{\mathbf{R^{-1}}}^{-\frac{1}{2}} \mathbf{R^{-1}} \mathbf{D}_{\mathbf{R^{-1}}}^{-\frac{1}{2}}$ where $\mathbf{D}_{\mathbf{R^{-1}}}$ is the diagonal of the inverse of $\mathbf{R}$. When the correlation matrix is not full rank, then the partial correlation can be estimated by replacing the inverse as pseudo inverse matrices: $\mathbf{P} = \mathbf{D}_{\mathbf{R^{\dagger}}}^{-\frac{1}{2}} \mathbf{R^{\dagger}} \mathbf{D}_{\mathbf{R^{\dagger}}}^{-\frac{1}{2}}$, where $\mathbf{R}^{\dagger}$ denotes the Moore–Penrose pseudoinverse of $\mathbf{R}$[19].

For the co-expression network, we only included the edges that exceeding a specific threshold as significant connections. In our analysis, we considered three different high threshold levels (0.90, 0.95 and 0.98) for comparison and selected the most appropriate one. After finding the co-expression network based on partial correlation, we calculated the clustering coefficient and the degree exponent, which are two important features in understanding the topological structure of biological networks [20], for each genotype-specific network and decide an appropriate threshold to obtain a graph for further topological analysis.

The degree exponent is widely used for deciding whether the network is scale-free. Real biological networks are scale-free networks where a small number of hubs have large numbers of connections to other nodes and the rest of nodes have few interactions [20]. The distribution of degrees in scale-free networks fits the power law $P(k) \sim k^{\gamma}$, where k is the degree of nodes and γ is the degree exponent. For a real biological network, the degree exponent normally lies in the range of $2 < \gamma < 3$.

The clustering coefficient, or transitivity, is a measure of the degree that nodes in a graph tend to connect with each other and form clusters. The local clustering coefficient of a node is defined as the ratio between the number of interactions between

its neighboring nodes divided by the number of all possible interactions they could have. The global clustering coefficient is defined as the ratio between the number of interactions of all closed triplets in the graph and all the possible interactions for all triplets in the graph. Here we compared the global clustering coefficients between the network for all genotypes.

After deciding the co-expression network graph with appropriate threshold, we checked the betweenness and degree centrality of the JA pathway genes in each network to understand their potential roles in signal transduction. The betweenness centrality of a node measures the degree that a node stands between each other nodes. The betweenness for a node v is given by $g(v) = \sum_{s \neq v \neq t} \sigma_{st}(v)/\sigma_{st}$, where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those shortest paths that pass through node v. A high betweenness means that this node is a key point in many pathways, therefore having a key role in the signaling network. The degree of node is another important measurement of local centrality which reflects the influence of a node on its direct neighbors.

2.4 Construction of the regulatory networks around JA pathway

To better reveal the details of the shifting of the JA pathway, we constructed five regulatory networks of the JA pathway for each genotype respectively. The general idea of regulatory network construction is to fit a linear relationship between genes in the pathway. Considering the number of samples (6-9) for each genotype is too small for performing a standard LASSO linear regression, we used the Adaptive NetworkProfiler

method [21], which is a weighted LASSO regression and implement this locally linear regression fitting for all 4 variants of maize. The Adaptive NetworkProfiler method can utilize all sample data into the regularized linear regression and it enables us to fit the regulatory network of every condition even if we have too few samples in some conditions to perform a standard LASSO linear regression.

The objective function for Adaptive NetworkProfiler in this project is:

$$L(\boldsymbol{\beta}_{l\alpha}) = argmin\left\{\frac{1}{2}\sum_{i=1}^{n}\left[\left(T^i - \boldsymbol{\beta}_{l\alpha}^T\mathbf{X^i}\right)^2 K(m_i - m_\alpha|b^{\mathrm{KNN}}, RG(M))\right]\right\} + P_{\delta_\alpha\lambda_\alpha}(\boldsymbol{\beta}_{l\alpha})$$

Where $K(d_{i-\alpha}|b^{\mathrm{KNN}}, RG(M)) = \exp\left(-\frac{-(m_i-m_\alpha)^2}{b^{\mathrm{KNN}}\cdot RG(M)}\right)$,

$$P_{\lambda_\alpha}(\boldsymbol{\beta}_{l\alpha}) = \lambda_\alpha[\delta_{l\alpha}\|\boldsymbol{\beta}_{l\alpha}\|_1]$$

$$RG(M) = \max(m_i|i = 1,2,\dots n) - \min\,(m_i|1 = 1,2,\dots n)$$

$\mathbf{X^i}$ is the expression profile for the direct neighbors of the l-th JA pathway gene of the i-th sample, T_l^i is the expression profile for the l-th JA pathway gene of the i-th sample. $\boldsymbol{\beta}_{l\alpha}$ is the coefficients of the direct neighbors of the l-th JA pathway gene for the samples of the α-th maize genotype. $P_{\lambda_\alpha}(\boldsymbol{\beta}_{l\alpha})$ is the regularization term.

$K(m_i - m_\alpha|b^{\mathrm{KNN}}, RG(M))$ is the Gaussian kernel that determines the weight of each sample point in the linear regression model. This Gaussian kernel imposes weights on each sample in the fitting of the linear model according to their "similarity" to the target's type. The more the sample and the target is "similar in type", the closer the weight imposed is to 1. m_i denotes the "modulator value" that characterizes the i-th sample. In this project, this characteristic is the drought response for every type of maize

and the JA level is used as the modulator value. $b^{\mathrm{KNN}} = \sqrt{(m_i - m_{\alpha^{Kth}})^2}$ measures the distance of modulator values between condition α and its K-th nearest neighbor. In this project, we use K=2. RG(M) is the distance between the largest and smallest modulator values.

After getting the coefficients $\boldsymbol{\beta}_{l\alpha}$ for every direct neighbor gene of every gene l inside the JA pathway, we can construct corresponding co-regulatory networks around the JA pathway for each type of maize. For each gene, the modified linear regression fits a linear combination of every other genes' expression data. The larger the absolute value of the coefficient for a gene is, the greater influence it is on the target gene it "points to". From the previous formulation, $\boldsymbol{\beta}_{l\alpha}$ is sparse and the coefficients that are too small are suppressed. Therefore, we decide the regulation exist for one gene to its target when the corresponding coefficient is non-zero. If the coefficient is above zero, it means that this gene can promote the expression of its target gene; if the coefficient is below zero, it means that this gene can suppress the expression of its target gene. By assembling all these regulatory relationships of genes that belong to or directly connected to JA pathway, we can get the regulatory network around JA pathway for each maize type. By comparing these networks for each maize type, we can find out the shifting of the regulatory network around JA pathway across different maize genotypes.

3. RESULTS

3.1 Differential expression analysis

The differential expression analysis is carried out using the Cufflinks toolkit and a threshold of p-value < 0.0001 and $\log_2$ fold change < -0.58 or > 0.58 is employed to decide the DEGs. To find out the similarities and differences of gene expression for drought response, the differential expressed genes for the five maize genotypes under 4 days and 6 days of drought stress compared with well-watered status were analyzed.

After 4 days of drought stress, there are 4342, 5481, 5418, 4715 and 955 significant DEGs detected in B73(wild type), *lox2*, lox4, *lox2lox4*, and *opr7opr8* mutants respectively. After 6 days of drought stress, 6744, 8967, 8283, and 3119 genes were identified as DEGs in B73, Lox2, Lox4 and opr7opr8 respectively. (Table.1) The results show that the DEG numbers in opr7opr8 mutant is significantly smaller than other genotypes, indicating that this mutant has a relatively smaller change in behavior under drought-stressed condition, which means a better drought tolerant.

	Drought Stressed Day 4			**Drought Stressed Day 6**		
	Up-regulated	**Down-regulated**	**Sub-total**	**Up-regulated**	**Down-regulated**	**Sub-total**
B73	1996	2346	4342	2776	3968	6744
lox2	3024	2457	5481	5030	3937	8967
lox4	2903	2515	5418	4772	3511	8283
lox2lox4	2526	2189	4715	-	-	-
opr7opr8	558	397	955	2046	1073	3119

Table 2. Summary of differential expressed genes in all maize genotypes under 4d and 6d drought stress with respective well-watered control comparison

3.2 Co-expression network and topology analysis

Five co-expression networks were constructed within the previously selected significant DEGs respectively for each maize genotype. Three different thresholds were used for deciding the edges for each genotype. We tested the degree exponent and the clustering coefficient for each graph. The results in Table 3 showed that the threshold 0.98 eliminated too many edges and the threshold 0.90 produced ill-condition graphs with too large degree exponent. We preferred the threshold 0.95 for the co-expression network construction[2].

Genotype		**B73**	***lox2***	***lox4***	***lox2lox4***	***opr7opr8***
Threshold	**# of Nodes**	7611	9832	9183	4681	3333
0.90	**Connections**	306953	334022	248546	331811	62304
	Degree Exponent	2.492	2.112	17.773	44.593	3.200
	Clustering Coefficient	0.4662	0.4827	0.4498	0.4928	0.6366
0.95	**Connections**	59881	63427	43929	98752	19190
	Degree Exponent	2.162	1.894	1.881	11.416	2.286
	Clustering Coefficient	0.4309	0.4728	0.4286	0.4605	0.7475
0.98	**Connections**	6298	7169	4249	18353	4934
	Degree Exponent	7.280	1.797	1.873	11.365	1.854
	Clustering Coefficient	0.4195	0.4787	0.4361	0.4337	0.7404

Table 3. The construction of co-expression network for each genotype using partial correlation with different thresholds

[2] The co-expression network analysis results based on the 0.90 threshold can be found in the Appendix.

We performed betweenness and degree centrality analysis for the JA pathway genes with the co-expression networks constructed above for each genotype respectively. Here in Table 4 we selected the top 5 genes for both centralities.

Genotypes	Top 5 JA genes for betweenness centrality	Normalized betweenness value	Top 5 JA genes for degree centrality	Degree Value
B73	*ZmLOX9*	2.102310e-03	*ZmLOX10*	117
	ZmAOC1	1.930861e-03	*ZmLOX9*	80
	ZmJAR1a	1.526549e-03	*ZmOPR7*	71
	ZmLOX10	1.475413e-03	*ZmACX1b*	50
	ZmKAT2b	1.246712e-03	*ZmKAT1*	43
lox2	*ZmACX1b*	1.386132e-03	*ZmLOX7*	86
	ZmKAT2a	1.381361e-03	*ZmKAT3d*	48
	ZmLOX7	1.025668e-03	*ZmOPR8*	41
	ZmLOX8	8.295204e-04	*ZmKAT1*	17
	ZmAOS1c	6.242498e-04	*ZmKAT2d*	16
lox4	*ZmACX3*	1.466588e-03	*ZmACX3*	82
	ZmKAT3d	1.046364e-03	*ZmKAT2d*	22
	ZmKAT2d	8.192535e-04	*ZmKAT3d*	17
	ZmAOS1b	8.190933e-04	*ZmLOX13*	15
	ZmLOX13	7.638071e-04	*ZmLOX9*	12
lox2lox4	*ZmKAT1*	2.290444e-03	*ZmKAT2c*	101
	ZmOPR8	1.499362e-03	*ZmAOS1a*	92
	ZmKAT3b	1.459457e-03	*ZmKAT2b*	87
	ZmACX1b	1.372179e-03	*ZmAOC2*	83
	ZmKAT2c	1.292097e-03	*ZmAOC1*	73
opr7opr8	*ZmKAT2c*	9.074133e-03	*ZmAOS1b*	21
	ZmLOX10	8.939360e-03	*ZmLOX11*	16
	ZmAOS1a	6.925383e-03	*ZmAOS1a*	15
	ZmAOS1b	4.034165e-03	*ZmLOX10*	14
	ZmLOX8	3.738651e-03	*ZmKAT2c*	12

Table 4. The centrality analysis of 12OPDA-JA pathway genes for each genotype in partial correlation network with threshold of 0.95

We also compared the profile of wild type with each mutant and selected the top 5 genes that had the largest change (mutant – wild type) in betweenness centrality and degree centrality for each mutant, respectively. The results are listed in Table 5. A positive difference means that the gene's centrality in mutant is larger than wild type and negative means smaller.

Genotypes	Top 5 JA genes for change in betweenness centrality	Change in normalized betweenness value	Top 5 JA genes for change in degree centrality	Change in degree value
lox2	*ZmLOX9*	-2.102310e-03	*ZmLOX10*	-117
	ZmAOC1	-1.930861e-03	*ZmLOX7*	83
	ZmLOX10	-1.475413e-03	*ZmLOX9*	-80
	ZmKAT2a	1.381361e-03	*ZmOPR7*	-71
	ZmKAT2b	-1.246712e-03	*ZmKAT3d*	45
lox4	*ZmLOX9*	-1.880077e-03	*ZmLOX10*	-108
	ZmJAR1a	-1.526549e-03	*ZmACX3*	77
	ZmAOC1	-1.468023e-03	*ZmOPR7*	-71
	ZmACX3	1.388285e-03	*ZmLOX9*	-68
	ZmKAT2b	-1.246712e-03	*ZmACX1b*	-48
lox2lox4	*ZmKAT1*	1.864502e-03	*ZmLOX10*	-92
	ZmOPR8	1.499362e-03	*ZmAOS1a*	90
	ZmLOX9	-1.454940e-03	*ZmAOC2*	83
	ZmAOC1	-1.140428e-03	*ZmKAT2c*	79
	ZmKAT3b	1.097716e-03	*ZmKAT2b*	74
opr7opr8	*ZmKAT2c*	8.032531e-03	*ZmLOX10*	-103
	ZmLOX10	7.463947e-03	*ZmLOX9*	-80
	ZmAOS1a	6.732398e-03	*ZmOPR7*	-71
	ZmAOS1b	3.928742e-03	*ZmACX1b*	-42
	ZmLOX8	2.762226e-03	*ZmKAT1*	-35

Table 5. The changes of 12OPDA-JA pathway genes' centrality for each mutant from wild type

From the comparing results, we can see that comparing to wild type, the degree centrality of gene *ZmLOX10* is lower in all 4 mutants and the degree centrality of gene *ZmLOX9* is lower in *lox2*, *lox4* and *opr7opr8*. The degree centrality of *ZmLOX7* and *ZmKAT3d* went up in *lox2* mutant and the degree centrality of *ZmACX3* went up in the *lox4* mutant. In the mutant *lox2lox4* we observed a rise in the degree of genes *ZmAOS1a*, *ZmAOC2*, *ZmKAT2c* and *ZmKAT2b*, while in the mutant *opr7opr8*, the number of direct neighbors went down for all genes.

For the changes in the betweenness centrality, we found that the gene *ZmLOX9* and *ZmAOC1* went down for *lox2*, *lox4* and *lox2lox4* mutants, while in the *opr7opr8*, the betweenness centrality of genes ZmKAT2c, ZmLOX10, ZmAOS1a, ZmAOS1b and ZmLOX8 all went up.

3.3 JA Regulatory network construction

Five regulatory networks were constructed within the 12-OPDA-JA pathway genes respectively for each maize genotype, as Figure 2-6. The penalizing term were adjusted for a sparse network where for any gene there would be no more than 3 other genes interacting with. In the network, the gene with green arrow pointing into the target gene means the gene promotes the target gene and red arrow means suppressing.

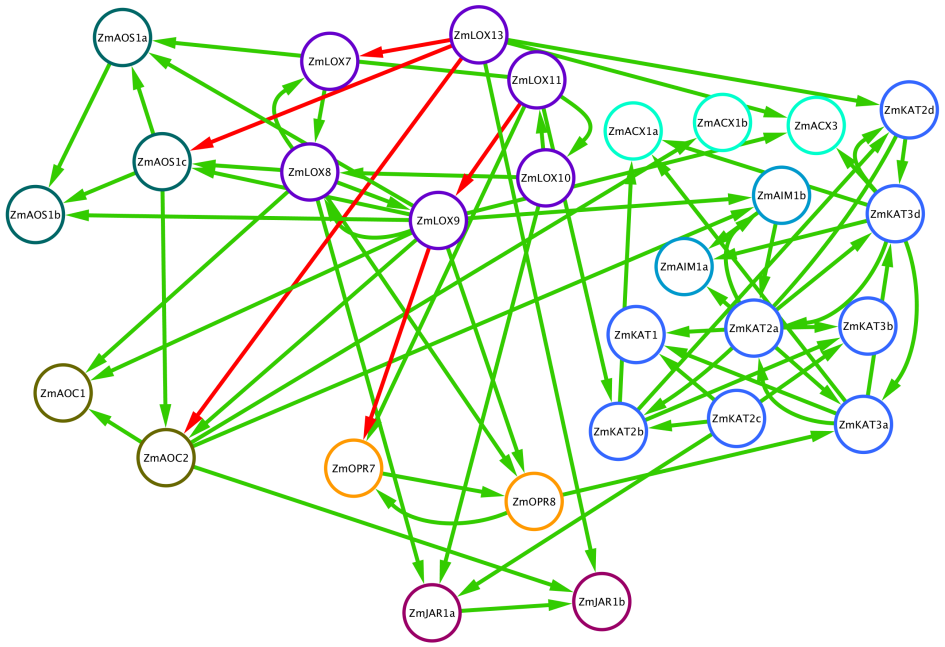

Figure 2. JA regulatory network for B73

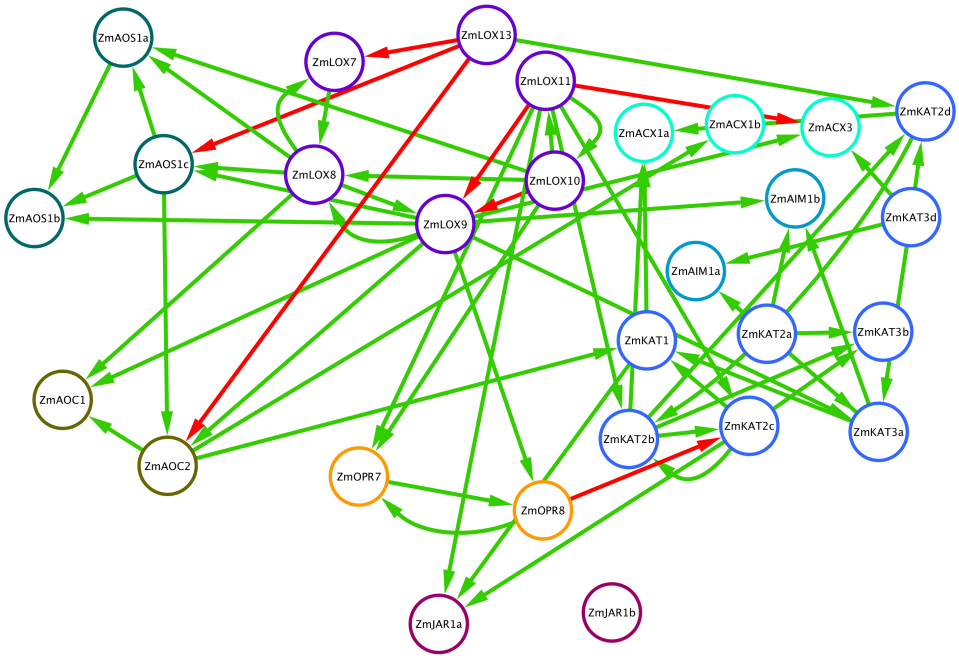

Figure 3. JA regulatory network for *lox2*

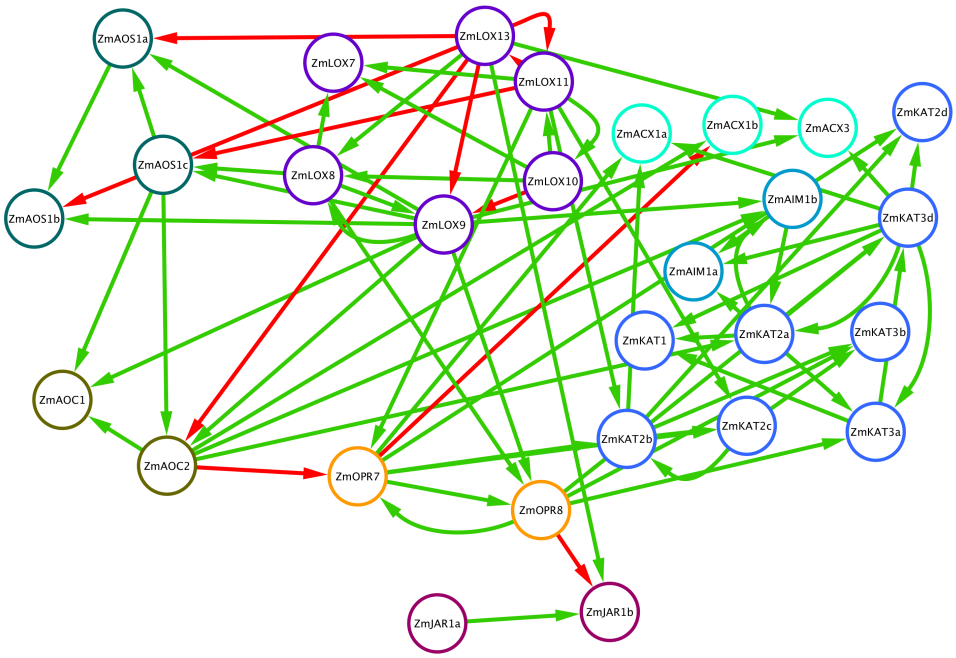

Figure 4. JA regulatory network for *lox4*

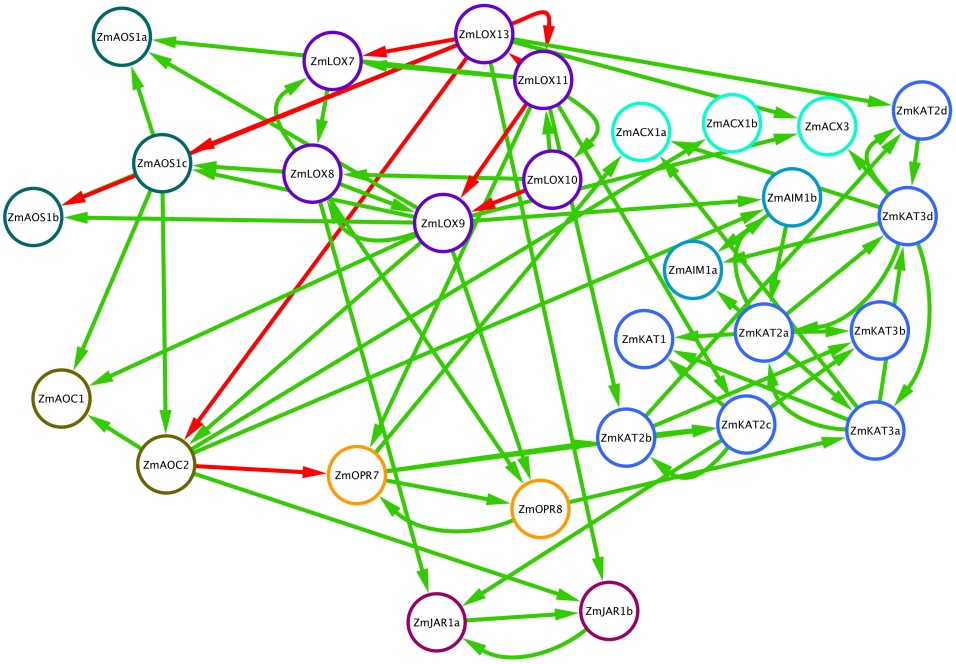

Figure 5. JA regulatory network for *lox2lox4*

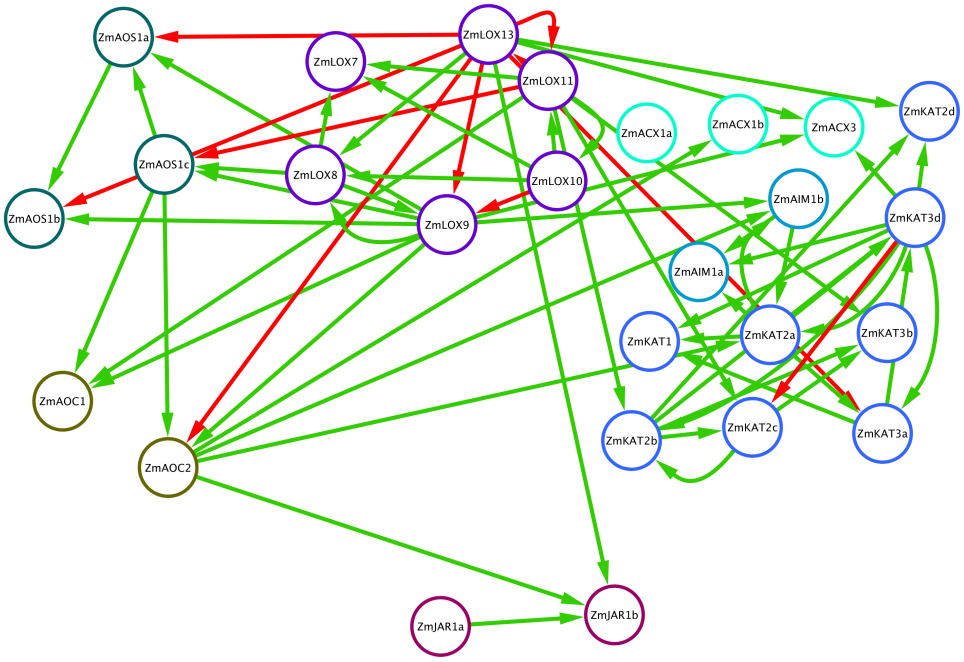

Figure 6. JA regulatory network for *opr7opr8*

The regulatory network construction displayed comparable results from previous part. Comparing with wild type, in the *lox2*, *lox4* and *lox2lox4* mutant the gene *ZmLOX9* does not have any suppressive effect on the ZmOPR7 gene. The gene ZmLOX10 has a suppressive effect on *ZmLOX9* in every mutant comparing with wild type. This is consistent with the behavior changes identified in the previous part.

In *lox2* mutant, apart from the behavior change of *ZmLOX9* and *ZmOPR7*, *ZmLOX11* has additional suppressive effect on *ZmACX3* and *ZmOPR8* suppresses *ZmKAT2c*.

Compared with wild type, there is significant changes in the JA regulatory network of *lox4* mutant. In *lox4*, the gene *ZmLOX13* has additional suppressive effect on *ZmAOS1a*, *ZmAOS1b*, *ZmLOX11* and *ZmLOX9*. The gene *ZmAOC2* suppresses

ZmOPR7 and *ZmOPR7* suppresses *ZmACX1b*. *ZmOPR8* has suppressive effect on *ZmJAR1b*.

In *lox2lox4 mutant*, *ZmLOX13* still suppresses *ZmAOS1b*, *ZmLOX11* and *ZmAOC2* has suppressive effect on *ZmOPR7*.

For the *opr7opr8* mutant, the absence of *ZmOPR7* and *ZmOPR8* made it an exception where no JA were produced during drought stress. In this mutant, *ZmLOX13* has more genes to suppress including *ZmLOX9*, *ZmAOS1a*, *ZmAOS1b*, *ZmAOC2* and *ZmKAT3a*, *ZmKAT3d* suppress *ZmKAT2c*, which is consistent with the discovery form previous part that *ZmKAT2c*, *ZmAOS1a*, *ZmAOS1b* have significant changes in betweenness centrality.

From the JA-Ile content at drought day 6 for each genotype, we can see that *opr7opr8* and *lox4* are the mutants with lowest JA level. Their JA pathways' behaviors have some similarities especially on the behavior of *ZmLOX13* gene and its neighborhoods.

4. CONCLUSION

In this paper, we performed a network-based analysis of the drought stress response behaviors of different drought-tolerant maize mutations. To overcome the challenge of a large number of genes and a small number of samples, we performed differential expression analysis with the obtained RNA-seq data to identify key genes that associate with drought response. We inferred the co-expression networks for each genotype through partial correlation to identify the significant relationships between genes that might be responsible for the behavior difference in drought response. We performed local and global centrality analysis for the JA pathway genes with inferred networks to find out the behavior shifting of JA genes in each genotype. We performed regulatory network analysis on the JA pathway genes with weighted LASSO method to reveal the shifting of the JA pathway. From the results, we identified two genes *ZmLOX9* and *ZmOPR7* that behaves significantly different in all mutants and the behavior change of *ZmLOX13*, *ZmAOS1c*, *ZmAOC2* genes in the *lox4* mutant. This will help further understand the mechanism of the JA pathway in maize drought response and potentially help searching for new candidate genes for crop improvements.

REFERENCES

[1] David, Teresa Soares, et al. "Water-use strategies in two co-occurring Mediterranean evergreen oaks: surviving the summer drought." Tree physiology 27.6 (2007): 793-803.

[2] Creelman, Robert A., and John E. Mullet. "Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress." Proceedings of the National Academy of Sciences 92.10 (1995): 4114-4119.

[3] Akter, Nasima, et al. "Effects of depletion of glutathione on abscisic acid-and methyl jasmonate-induced stomatal closure in Arabidopsis thaliana." Bioscience, biotechnology, and biochemistry 76.11 (2012): 2032-2037.

[4] Daszkowska-Golec, Agata, and Iwona Szarejko. "Open or close the gate–stomata action under the control of phytohormones in drought stress conditions." Frontiers in plant science 4 (2013): 138.

[5] Huang, Liyu, et al. "Comparative transcriptome sequencing of tolerant rice introgression line and its parents in response to drought stress." BMC genomics 15.1 (2014): 1026.

[6] Bild, Andrea H., et al. "Oncogenic pathway signatures in human cancers as a guide to targeted therapies." Nature 439.7074 (2006): 353.

[7] Su, Junjie, Byung-Jun Yoon, and Edward R. Dougherty. "Accurate and reliable cancer classification based on probabilistic inference of pathway activity." PloS one 4.12 (2009): e8161.

[8] Lee, Eunjung, et al. "Inferring pathway activity toward precise disease classification." PLoS computational biology 4.11 (2008): e1000217.

[9] Wang, Yixin, et al. "Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer." The Lancet 365.9460 (2005): 671-679.

[10] Tian, Lu, et al. "Discovering statistically significant pathways in expression profiling studies." Proceedings of the National Academy of Sciences 102.38 (2005): 13544-13549.

[11] Subramanian, Aravind, et al. "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proceedings of the National Academy of Sciences 102.43 (2005): 15545-15550.

[12] Khunlertgit, Navadon, and Byung-Jun Yoon. "Identification of robust pathway markers for cancer through rank-based pathway activity inference." Advances in bioinformatics 2013 (2013).

[13] Zhang, Bin, and Steve Horvath. "A general framework for weighted gene co-expression network analysis." Statistical applications in genetics and molecular biology 4.1 (2005).

[14] Trapnell, Cole, et al. "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nature biotechnology28.5 (2010): 511.

[15] Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalusu VV, Huerta M, Cho KT, Wimalanathan K, Richter JD, Mauch ED, Rao BS, Birkett SM, Sen TZ, Lawrence-Dill

CJ. (2015) MaizeGDB update: new tools, data and interface for the maize model organism database. Nucleic acids research; D1: 44, D1195-D1201.

[16] Trapnell, Cole, et al. "Differential analysis of gene regulation at transcript resolution with RNA-seq." Nature biotechnology31.1 (2013): 46.

[17] Peng, Jie, et al. "Partial correlation estimation by joint sparse regression models." Journal of the American Statistical Association 104.486 (2009): 735-746.

[18] De La Fuente, Alberto, et al. "Discovery of meaningful associations in genomic data using partial correlation coefficients." Bioinformatics 20.18 (2004): 3565-3574.

[19] Hero, Alfred, and Bala Rajaratnam. "Hub discovery in partial correlation graphs." IEEE Transactions on Information Theory58.9 (2012): 6064-6078.

[20] Barabasi, Albert-Laszlo, and Zoltan N. Oltvai. "Network biology: understanding the cell's functional organization." Nature reviews genetics 5.2 (2004): 101.

[21] Park, Heewon, et al. "Adaptive NetworkProfiler for Identifying Cancer Characteristic-Specific Gene Regulatory Networks." Journal of Computational Biology 25.2 (2018): 130-145.

APPENDIX

For the co-expression network construction, we also used pairwise Pearson correlation coefficients (PCCs) with the DEGs obtained from the previous step. The threshold for PCC cut-off is calculated through an estimation based on the sample size and the gene numbers. This estimation is based on the hub screening framework from Hero et.al [19] where the correlation matrix is computed with a small number of samples and a great number of genes. A hub is defined as a vertex of the correlation graph with at least degree δ at threshold ρ. The framework gives out an asymptotic expression for the false positive rate $P(N_{\delta,\rho} > 0)$ and a phase transition in the mean number of hub discovery $E[N_{\delta,\rho}]$. The mean number of hub discoveries of degree δ depends on the applied threshold ρ and there is a critical phase transition threshold $\rho_{c,\delta}$ that if the screening threshold for edge ρ decreases below $\rho_{c,\delta}$, the number of hub discoveries of degree δ abruptly increases to the maximum numbers of vertexes p that the graph has. The mathematical form of this critical phase transition threshold is:

$$\rho_{c,\delta} = \sqrt{1 - \left(c_{n,\delta}(p-1)\right)^{-\frac{2\delta}{\delta(n-2)-2}}}$$

Where $c_{n,\delta} = a_n \delta J_{p,\delta}$, p denotes the number of variables (genes), n denotes the number of samples and $a_n = 2B(\frac{n-2}{2}, \frac{1}{2})$ with $B(i,j)$ denoting the beta function. In the framework, $J_{p,\delta} = 1 + O\left(\left(\frac{k}{p}\right)^{\gamma\delta}\right)$ where $\gamma\delta = \delta + 1$ for screening on correlation matrices, and this $J_{p,\delta}$ can approximately equal to 1 under the assumption of the

framework that the network is sparse. This threshold works same for both correlation graphs and partial correlation graphs and here we can directly apply it for edge screening when we set $\delta = 1$.

Genotype	**B73**	***lox2***	***lox4***	***lox2lox4***	***opr7opr8***
# of DEGs	7611	9832	9183	4681	3333
Threshold	0.9889342	0.990017	0.9897391	0.9990	0.9845685
Connections	62075	113954	80819	5596	55125
Degree exponent	6.908139	2.997043	2.553736	16.6253	27.149
Clustering coefficients	0.449563	0.492679	0.4691549	0.418476	0.8259683

Table 6. The construction of co-expression network for each genotype using pairwise Pearson Correlation (PCC)

From Table 6, we can see the degree exponential measured for PCC is quite unstable across each genotype, which varies from 2.5 to over 27. The clustering coefficients for PCC networks are larger than the partial correlation networks and we found the opr7opr8 genotype has a really large clustering coefficient of 0.826.

Genotype	**B73**	***lox2***	***lox4***	***lox2lox4***	***opr7opr8***
Top5 JA genes for betweenness centrality	*ZmKAT3a*	*ZmLOX9*	*ZmAIM1b*	*ZmKAT3a*	*ZmLOX13*
	ZmAIM1a	*ZmACX3*	*ZmKAT3a*	*ZmACX3*	*ZmKAT3a*
	ZmKAT3d	*ZmAIM1a*	*ZmAOC1*	*ZmKAT2a*	*ZmJAR1a*
	ZmACX3	*ZmKAT3a*	*ZmAIM1a*	*ZmAIM1a*	-
	ZmOPR8	*ZmACX1a*	*ZmKAT3d*	*ZmAIM1b*	-
Top 5 JA genes for degree centrality	*ZmAOC2*	*ZmAIM1b*	*ZmAIM1b*	*ZmKAT3a*	*ZmLOX13*
	ZmOPR8	*ZmKAT3a*	*ZmKAT3a*	*ZmKAT2a*	*ZmKAT3a*
	ZmLOX9	*ZmLOX9*	*ZmAIM1a*	*ZmAIM1a*	*ZmJAR1a*
	ZmKAT3a	*ZmKAT3d*	*ZmACX3*	*ZmAOC2*	*ZmJAR1b*
	ZmAIM1a	*ZmACX3*	*ZmLOX9*	*ZmAIM1b*	-

Table 7. Network topological analysis result for the PCC co-expression network

Genotypes	B73	*lox2*	*lox4*	*opr7opr8*	*lox2lox4*
ZmLOX8	1	1	-	-	-
ZmLOX9	39	156	31	-	1
ZmLOX10	-	3	1	-	
ZmLOX11	4	1	-	-	-
ZmLOX13	3	-	-	47	
ZmAOC1	7	-	13	-	
ZmAOC2	76	1	-	-	9
ZmACX1a	-	17	-	-	
ZmACX3	15	128	34	-	7
ZmAIM1a	22	108	35	-	10
ZmAIM1b	1	273	181	-	8
ZmKAT1	11	10	8	-	3
ZmKAT2a	19	47	2	-	14
ZmKAT2d	1	114	6	-	1
ZmKAT3a	29	190	49	8	19
ZmKAT3d	15	142	14	-	4
ZmOPR8	64	1	22	-	3
ZmJAR1a	-	-	-	3	-
ZmJAR1b	2	-	-	1	-

Table 8. The degree of 12OPDA-JA pathway genes for each genotype

Table 7 and 8 showed the result of the centrality analysis and degree analysis for these networks. showed that the behavior of the *ZmLOX9*, *ZmAOC2* and *ZmAIM1b* genes has significant change when the *ZmLOX2* or *ZmLOX4* gene is absent. Specially, the interaction between genes *ZmACX3*, *ZmAIM1a*, *ZmAIM1b*, *ZmKAT2d*, *ZmKAT3a, ZmKAT3d* with other genes have considerably increased in the *lox2* mutant where *ZmLOX2* is knocked out. For the *opr7opr8* mutant, most genes in the 12-OPDA-JA pathway were not differentially expressed and the only behavior alter is the *ZmLOX13*

gene. For the *lox2lox4* mutant, the interactions of gene *ZmLOX9*, *ZmAOC2* and *ZmOPR8* is lower compared to wild type.

Here in Table 9 and 10 we also added the centrality analysis results for the co-expression networks constructed with the other two thresholds 0.90 and 0.98:

Genotypes	Top 5 JA genes for betweenness centrality	Normalized Betweeness value	Top 5 JA genes for degree centrality	Degree Value
B73	*ZmAOS1a*	1.193810e-03	*ZmLOX10*	416
	ZmAOS1b	8.846237e-04	*ZmLOX9*	305
	ZmAIM1b	8.042178e-04	*ZmOPR7*	280
	ZmAOS1c	6.378607e-04	*ZmACX1b*	261
	ZmACX1b	6.308290e-04	*ZmAIM1b*	227
lox2	*ZmLOX8*	7.841538e-04	*ZmLOX7*	376
	ZmOPR8	6.985335e-04	*ZmKAT3d*	283
	ZmLOX7	5.964007e-04	*ZmOPR8*	183
	ZmJAR1a	5.622205e-04	*ZmKAT1*	168
	ZmJAR1b	4.744887e-04	*ZmKAT2d*	158
lox4	*ZmACX1b*	7.516941e-04	*ZmACX3*	292
	ZmLOX7	7.177028e-04	*ZmKAT3d*	121
	ZmOPR7	6.650055e-04	*ZmLOX13*	93
	ZmAOS1b	6.389806e-04	*ZmKAT2d*	92
	ZmACX1a	6.386675e-04	*ZmLOX9*	83
lox2lox4	*ZmACX1b*	0.0009843214	*ZmAOS1a*	245
	ZmKAT3b	0.0007858551	*ZmKAT2b*	236
	ZmLOX7	0.0007811912	*ZmKAT2c*	227
	ZmAOS1b	0.0007083171	*ZmAOS1b*	226
	ZmLOX10	0.0007059783	*ZmAOC1*	223
opr7opr8	*ZmAOS1b*	0.0043094060	*ZmAOS1b*	90
	ZmAIM1a	0.0035531362	*ZmAOS1a*	71
	ZmLOX8	0.0024652655	*ZmLOX11*	68
	ZmAOS1a	0.0023286530	*ZmACX1b*	62
	ZmKAT3a	0.0022949657	*ZmKAT2c*	56

Table 9. The centrality analysis of 12OPDA-JA pathway genes for each genotype in partial correlation network with threshold of 0.90

Genotypes	Top 5 JA genes for weighted degree centrality	Degree Value
B73	*ZmLOX10*	3481.562
	ZmLOX9	3464.289
	ZmACX1b	3414.676
	ZmAIM1b	3407.932
	ZmOPR7	3384.121
lox2	*ZmLOX7*	4209.996
	ZmKAT3d	4046.592
	ZmKAT2d	4024.182
	ZmACX1b	3946.869
	ZmAOS1c	3876.646
lox4	*ZmACX3*	3643.215
	ZmAIM1b	3553.993
	ZmKAT3d	3533.379
	ZmAOS1a	3454.440
	ZmLOX10	3453.177
lox2lox4	*ZmACX1b*	2063.072
	ZmKAT3b	2046.524
	ZmKAT2b	2018.412
	ZmOPR8	2018.210
	ZmACX3	2005.956
opr7opr8	*ZmKAT2c*	1373.7421
	ZmAOS1b	1365.7982
	ZmLOX8	1346.7508
	ZmACX1b	1342.9854
	ZmAOS1a	1339.2611

Table 10. The weighted degree analysis of 12OPDA-JA pathway genes for each genotype in partial correlation network with threshold of 0.90