Analyses of transcriptomes and global histone modification patterns in Arabidopsis hybrids at early developmental stages

Anyu Zhu

A thesis summited for the Doctor of Philosophy

February 2016

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Acknowledgements

This research project was undertaken at CSIRO Plant Industry, Black Mountain.

First and foremost, I would like to thank my supervisors Pro Elizabeth Dennis (CSIRO, UTS), Pro Jim Peacock (CSIRO, UTS) and Pro Elizabeth Harry (UTS) for the support and guidance that they have given me throughout these three years. They always patiently gave me advice on the ideas, the experiments and the thesis. I have really learned many things from them, in particular, critical thinking.

Furthermore, I would like to especially thank Dr Ian Greaves for guiding me on the bioinformatic analyses, for great suggestions on experimental design and for his comments on my thesis.

I would also like to thank Dr Ming-Bo Wang for his advice and thank Limin Wu, Pei-Chuan (Tina) Liu and Dr Maria M. Alonso-Peral for their assistance with the technical aspects of the project.

To all the other members in the laboratory – Dr Li Wang, Bjorg Sherman, Dr Michael Groszmann, Dr Rebeca Gonzalez-Bayon, Dr Marina Trigueros and Dr Aihua Wang, I thank you for all your help during these years. It is my pleasure to work with you.

I would like to thank my parents for supporting my study in Australia. Finally, to the people who encouraged and supported me during the three years, I thank all of you.

Table of Contents

***CERTIFICATE OF ORIGINAL AUTHORSHIP* .... i**

Acknowledgements........ ii

List of illustrations and tables........ vii

Abbreviations........ xi

Abstract........ xiii

1 Chapter I Introduction 1

1.1 Heterosis in plants........ 1

1.2 Mechanism of heterosis 2

1.3 Gene regulation........ 4

1.4 Epigenetic gene regulation........ 5

1.4.1 DNA methylation 5

1.4.2 Histone modification........ 6

1.4.3 Small RNAs 9

1.4.4 Alterations in hybrid epigenomes 10

1.5 Heterosis at early developmental stages 14

2 Chapter II Research aims........ 16

3 Chapter III Materials and Methods 18

3.1 Plant growth media and conditions........ 18

3.2 Measurements of seed weight and size 19

3.3 Seed germination time assay........ 19

3.4 RNA extraction 20

3.5 Chromatin Immune-Precipitation Sequencing (ChIP-Seq)........ 20

3.5.1 Chromatin immune-precipitation (ChIP) 20

3.5.2 ChIP-Seq library preparation 23

3.6 qRT-PCR 25

3.7 Bioinformatics analyses 27

3.7.1 Read quality control and read mapping 28

3.7.2 RNA-Seq 35

3.7.3 ChIP-Seq 36

3.7.4 SNP analyses 37

4 Chapter IV Results - Phenotypic measurements on parents and hybrids at early developmental stages 39

4.1 Increased seed size and weight in C24 x L*er* but not in L*er* x C24 39

4.2 Hybrids show heterosis in germination time 41

4.3 The L*er* x C24 hybrid grows faster than the C24 x L*er* hybrid following germination 42

4.4 Summary 44

5 Chapter V Results –Transcriptomic analyses on parents and hybrids at 0, 3, 5 and 7 DAS 46

5.1 Differentially expressed genes are identified in both reciprocal hybrids 46

5.2 Early and transient changes in gene activity exist in young hybrid seedlings compared to their parents 49

5.3 The differentially expressed genes are enriched in specific functional groups, especially photosynthesis. 53

5.4 Pathway analyses suggest that the gene activity of the photosynthesis system is enhanced. 57

5.5 Validation of non-additive gene expression 62

5.6 The non-additively expressed genes may be not regulated by common transcription factors. 68

5.7 Summary 70

6 Chapter VI Results – genome-wide histone modification analyses of parents and hybrids at 0 DAS 72

6.1 DNA specific to each histone mark was enriched in the ChIP-Seq libraries of parent and hybrid mature seeds 72

6.2 ChIP-Seq data show the expected distribution patterns of histone modifications over gene regions in parents and hybrids. 74

6.3 Genome-wide histone modification levels are related to gene expression levels in parents and hybrid seeds...78

6.4 Histone modifications have the expected patterns associated with genes. 79

6.5 L*er* and C24 and their hybrids have similar genome-wide histone modifications patterns...82

6.6 Histone modification patterns differ between the parents and the hybrids at specific genes...84

6.7 The changed histone modification levels are not fully correlated with changed gene expression levels in the hybrids. 87

6.8 Summary 91

7 Chapter VII Results – allelic analyses on gene expression and histone modifications in Ler and C24 and their hybrids 93

7.1 Reads across SNPs reflect true levels of gene expression and histone modifications 93

7.2 Changed expression ratios between L*er* and C24 alleles in hybrids at 0 DAS are mainly caused by C24 allele up-regulation 94

7.3 Increased levels of histone modifications at the L*er* alleles are commonly found in the peaks with changed allelic ratios in L*er* x C24 but not in C24 x L*er*................96

7.4 Ler and C24 alleles have equivalent roles in causing non-additive gene expression in hybrids at early development stages except for 0 DAS.........................98

7.5 The changes in allelic expression do not completely correlate with the changes in allelic histone modification in hybrid seeds 100

7.6 The allelic expression patterns in hybrid seeds anticipate the allelic expression patterns in parents in subsequent developmental stages...101

7.7 Summary 105

8 Chapter VII Discussion...106

8.1 Early and transient changes in gene activity are not caused by differences in germination time between hybrids and parents 107

8.2 Biomass heterosis is established in early stages of seedling development 108

8.3 Divergence in reciprocal hybrids 109

8.4 Early and transient changes in gene activity over time in hybrids 109

8.5 Allelic expression patterns in hybrid seeds anticipate future allelic patterns in parents 110

8.6 Regulation of non-additive gene expression at early developmental stages ... 112

8.7 From heterotic transcriptome to heterotic metabolome 113

8.8 Transient transcriptional advantages in hybrids might result in long-lasting heterosis 114

8.9 Future work 115

8.10 Conclusions 116

Appendices 118

Bibliography 128

List of illustrations and tables

Table 3.1. MS medium recipe (in 1 L) (P. 19)

Table 3.2. ChIP lysis buffer stock (P. 22)

Table 3.3. Low salt wash buffer stock (P. 22)

Table 3.4. High salt wash buffer stock (P. 22)

Table 3.5. LiCl wash buffer stock (P. 23)

Table 3.6. TE buffer stock (P. 23)

Table 3.7. Elution buffer stock (P. 23)

Table 3.8. Primers used in ChIP-qPCR and expression level validation (P. 27)

Table 3.9. The numbers of reads in RNA-Seq data mapped to Col genome. (P. 33)

Table 3.10. The numbers of reads in ChIP-Seq data mapped to Col genome. (P. 34)

Table 7.1. Counts of genes and histone modification peaks in SNP analyses. (P. 94)

Figure 1.1. Ler/C24 hybrids show significant heterosis at vegetative stages. (P. 2)

Figure 1.2. The cascade of gene regulation of the cell. (P. 5)

Figure 1.3. Spatial locations of histone modifications in a chromosome. (P. 7)

Figure 1.4. Model shows the relation between histone marks and chromatin states. (P. 8)

Figure 1.5. Model for the association between siRNA and DNA methylation level at different loci in the Arabidopsis F1 hybrids. (P. 12)

Figure 3.1. Workflow of N-ChIP. (P. 24)

Figure 3.2. Workflow of statistical analyses for RNA-Seq and ChIP-Seq data. (P. 28)

Figure 3.3. The RNA-Seq data of Ler x Ler showing high quality scores per base (A) and per read (B). (P. 30)

Figure 3.4. The RNA-Seq data of Ler x Ler showing constant percentage of sequence content over reads (A) and expected distribution of GC content (B). (P. 31)

Figure 3.5. Dendrogram shows the similarity of RNA-Seq libraries in two biological replicates of the parents and the hybrids at 0 DAS. (P. 33)

Figure 4.1. Ler/C24 reciprocal hybrids showing significant maternally influenced seed weight and seed area. (P. 40)

Figure 4.2. Seed germination time for the Ler and C24 and their hybrids. (P. 41)

Figure 4.3. Ler/C24 hybrids showing heterosis in plant size at early vegetative developmental stages. (P. 44)

Figure 5.1. Ler and C24 showing dynamic differences in gene expression over time. (P. 47)

Figure 5.2. Ler/C24 hybrids showing various patterns at different time points in young seedlings. (P. 49)

Figure 5.3. Transient and early changes in gene activity in the reciprocal hybrids at young seedling stages. (P. 51)

Figure 5.4. The distributions of the five types of non-additive genes at 0, 3 and 5 DAS. (P. 53)

Figure 5.5. Eight clusters of non-additive genes based on expression patterns over time and the significant GO terms common to both reciprocal hybrids. (P. 56)

Figure 5.6. Transcriptional activities of the genes in the light reactions is transiently up-regulated in both reciprocal hybrids at 3 DAS. (P. 58)

Figure 5.7. Transcriptional activities of the genes in Calvin cycle are transiently up-regulated in both reciprocal hybrids at 3 DAS. (P. 59)

Figure 5.8. Two distinct expression patterns over time exist in the genes involved in the light reactions. (P. 60)

Figure 5.9. The transcriptional activity is transiently changed in young seedlings of both reciprocal hybrids in the biosynthesis pathways of chlorophyll, TAG, carotenoids and sulfur-containing glucosinolates. (P. 61)

Figure 5.10. Ler/C24 hybrids sown 7 hours later are in similar developmental stages as the parents from 3 to 7 DAS. (P. 65)

Figure 5.11. qPCR validation for the expression levels of 10 selected genes involved in the light reactions and the chlorophyll biosynthesis in parents and hybrids from 3 to 7 DAS. (P. 68)

Figure 5.12. Motif analyses on the upstream 1 kb promoter regions of non-additive genes. (P. 70)

Figure 6.1. qPCR verifications for ChIP DNA. (P. 73)

Figure 6.2. Bioanalyser results showing quality of ChIP-Seq libraries. (P. 74)

Figure 6.3. Global patterns of the four histone marks in C24 x C24 in protein-coding genes. (P. 76)

Figure 6.4. Global patterns of the four histone marks in C24 x C24 in TEs. (P. 77)

Figure 6.5. Correlations between histone modification and gene expression in C24 x C24. (P. 79)

Figure 6.6. The correlation patterns between histone modification peaks and between peaks and genes in C24 x C24. (P. 81)

Figure 6.7. The distributions of four histone marks over the five chromosomes in Ler and C24 and their hybrids. (P. 83)

Figure 6.8. The numbers of differentially histone-modification-targeted genes (DHGs) for each histone mark between Ler and C24 in seeds. (P. 84)

Figure 6.9. The DHGs identified in Ler/C24 hybrids in seeds. (P. 86)

Figure 6.10. Correlations between the changes in histone modifications between parents

and between parents and hybrids. (P. 87)

Figure 6.11. Correlations between histone modifications and gene expression in the parents and the hybrids in seeds. (P. 89)

Figure 6.12. Relative average levels of gene expression and histone modifications in the non-additive genes in both hybrids at 0 DAS. (P. 90)

Figure 6.13. Correlations between histone modifications at 0 DAS and gene expression at 3 DAS in the hybrids. (P. 91)

Figure 7.1. Patterns of allelic expression in hybrids at four time points. (P. 96)

Figure 7.2. Patterns of allelic changes of the four histone marks in hybrid seeds. (P. 98)

Figure 7.3. Contributions of Ler and C24 alleles to non-additive gene expression in hybrids. (P. 99)

Figure 7.4. Changes of gene expression against changes of histone marks at Ler and C24 alleles in Ler x C24. (P. 100)

Figure 7.5. Correlations of parental (A) and allelic (B) proportions in Ler x C24 between 0 DAS and the following three time points. (P. 103)

Figure 7.6. Allelic proportions against parental proportions in Ler x C24 at the four time points. (P. 104)

Figure 7.7. Venn-diagram showing common genes to Group A and the total non-additive genes from the four time points. (P. 104)

Figure 8.1. Models to describe the early ‘‘set-up’’ of allelic expression patterns in hybrid seeds. (P. 111)

Abbreviations

BMG	better-than-MPV gene
BPG	better-than-best-parent gene
ChIP	chromosome immuno-precipitation
DAS	day after sowing
DHG	differentially histone-modification-enriched genes
EV	expected value
GO	gene ontology
HAS	hour after sowing
IAA	indole-3-acetic acid
MPV	mid-parent value
NGS	next-generation sequencing
PTM	post-translational modification
RdDM	RNA-directed DNA Methylation
RHL	relative heterosis level
RISC	RNA-induced silencing complex
RNAi	RNA interference
SA	salicylic acid
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
sRNA	small RNA
ssRNA	single-stranded RNA
TAG	triacylglycerol
TCdM	trans chromosomal demethylation

TCM - trans chromosomal methylation

TE - transposable element

TSS - transcription start site

TTS - transcription termination site

Abstract

Heterosis has been used for decades in the crop industry, especially in the production of rice and maize. Hybrids usually exceed their parents in plant biomass, seed number and seed weight. Previous findings suggested that heterosis could be associated with altered gene expression in hybrids. In some cases, alterations in gene expression are associated with the alterations in epigenetic factors, such as DNA methylation and histone modifications. Although biomass heterosis has been shown in hybrids at relatively late developmental stages, the timing of heterosis establishment is not clear.

In this project, the transcriptomes and global histone modification patterns were analysed in Arabidopsis hybrids at early stages of seedling development. The results suggested that biomass heterosis was present in young seedlings of L*er*/C24 hybrids. This early heterosis was associated with transient changes in the hybrids relative to the parents in the activities of genes involved in critical pathways, including photosynthesis pathways, responsible for plant growth. A limited role for histone modifications in regulating the differentially expressed genes in hybrid seeds was shown. Finally, our results demonstrated that allelic expression patterns in hybrid seeds anticipate those in parents at later developmental stages.