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

List of illustrations and tables

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.

1 Chapter I Introduction

1.1 Heterosis in plants

Hybrids are the progeny of two crossed strains or species of organisms. Compared to their parents, hybrids commonly show advantages for many characters. This phenomenon is called heterosis, or hybrid vigour. Heterosis exists widely in animals and plants.

In plants, heterosis is related to increased biomass and seed yield, which are critical traits in agriculture for high yield in crops. As early as 1908, George Shull showed that vegetative growth and grain yield in maize hybrids exceeded that of the parents (Shull, 1908). The maize yield has increased six-fold in the United States since the introduction of hybrid maize in the 1920's (Crow, 1998). Hybrid breeding has also been well developed in rice, the main food source in Asia, particularly in China, a region with the largest population in the world. Thanks to the achievements in hybrid rice breeding, the yield of rice in China increased nearly two-fold between 1976 and 1995 (Yuan, 1998) and is still steadily increasing.

Apart from crops, heterosis has been researched in the model plant, *Arabidopsis thaliana*. Because Arabidopsis has a relatively short life period, small genome size and well sequenced genome, it is commonly used in plant research. Hybrids from different Arabidopsis accessions show heterosis for many characters, especially in vegetative growth (Barth et al., 2003, Meyer et al., 2004). Among Arabidopsis hybrid combinations, the hybrids derived from the two accessions, Landsberg *erecta* (L*er*) and C24, show

significant biomass heterosis (Figure 1.1), approximating 250% of the average value of the two parents at four weeks after sowing (Groszmann, et al, 2011).

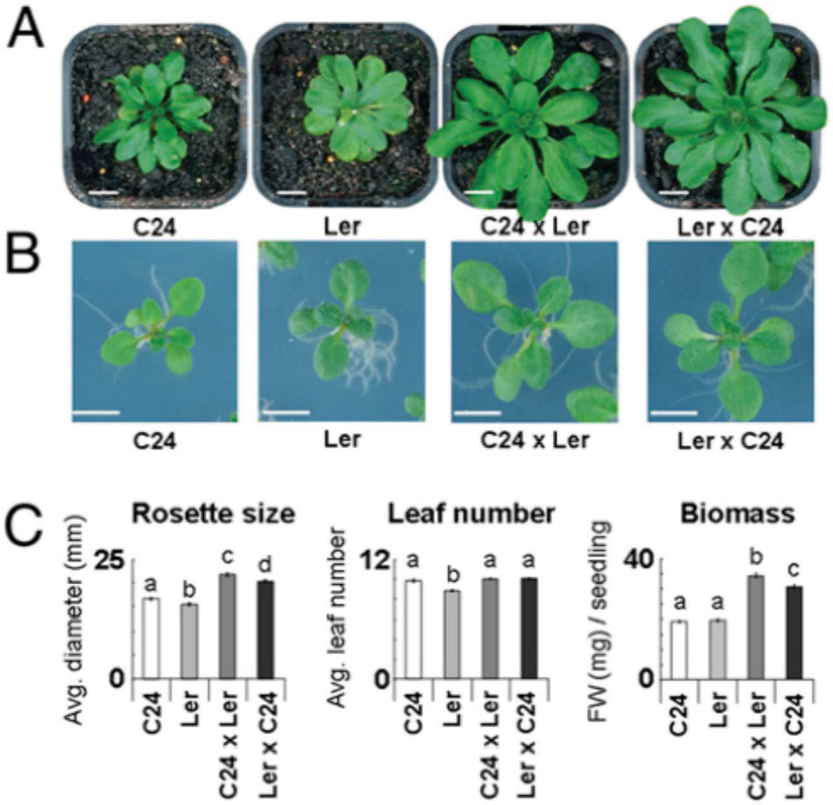


Figure 1.1. Ler/C24 hybrids show significant heterosis at vegetative stages. (A) Heterosis in plant size in 4-week-old hybrids. **(B) and (C)** Heterosis in plant size in hybrids at 14 DAS (day after sowing). Hybrids are similar to parents in leaf number but have increased rosette size and biomass. Images retrieved from (Groszmann et al., 2011b)

1.2 Mechanism of heterosis

Although heterosis has been studied for over a century and has been used worldwide to improve the cropping industry, the molecular basis of heterosis is still unclear. Three genetic models of heterosis have been suggested, dominance, overdominance, and epistasis of gene action. In the dominance hypothesis, heterosis is the consequence of dominant alleles from two parents at multiple loci complementing any unfavourable

alleles in one of the parents (Davenport, 1908, Bruce, 1910). The overdominance hypothesis suggests that the interactions between two parental alleles result in superior traits in heterozygous hybrids compared to either of the homozygous parents (Shull, 1908). Finally, the epistasis hypothesis proposes that a superior phenotypic trait in hybrids is generated from the interactions between two or more non-allelic genes from different parental backgrounds (Powers, 1944). Although each of the three hypotheses has supporting evidence, none of them accounts for all aspects of heterosis.

Quantitative Trait Locus (QTL) analyses indicated a large number of loci as contributing to the final heterotic phenotype (Radoev et al., 2008, Meyer et al., 2010), although it has been demonstrated that single gene heterozygosity can significantly increase yield in tomato (Krieger et al., 2010). Recently, genome-wide gene expression analyses suggested that a large number of genes have changed expression levels in plant hybrids compared to their parents (He et al., 2010, Groszmann et al., 2011b, Fujimoto et al., 2012, Meyer et al., 2012, Miller et al., 2012, Shen et al., 2012, Groszmann et al., 2014). Because hybrids have both parental genomes, the transcript level of a gene in hybrids is predicted to be the average level of this gene in the two parents, which is called the mid-parent value (MPV). In hybrids, genes have additive expression levels when they are equivalent to the MPV, while genes have non-additive expression levels when they are higher or lower than the MPV.

The molecular mechanisms of hybrid vigour are likely to involve alterations of gene expression in hybrids. Understanding gene regulation will be important to help us understand the molecular basis of heterosis.

1.3 Gene regulation

The control of gene regulation can be separated into two distinct phases: transcriptional and post-transcriptional gene regulation (Figure 1.2). As the name implies, transcriptional regulation of a gene occurs at the transcriptional level and determines the amount of messenger RNA (mRNA) transcribed from the DNA. Regulation at this level can be dependent on the DNA sequences located around a gene (e.g promoters and enhancers). In addition, histone modification and DNA methylation may up- or down-regulate the transcription of a gene. Gene regulation mediated by DNA methylation and histone modification is known as epigenetic regulation.

Unlike transcriptional regulation, post-transcriptional regulation influences gene expression by affecting the formation of mature mRNA and its translation into protein products. There are many control points occurring at this layer of regulation: RNA splicing, transport, localization and protein translation. In addition to these classic regulating mechanisms, recent studies have revealed a new layer of post-transcriptional gene regulation which reduces gene expression by cleaving the mRNA of a gene or by blocking protein translation. Because this pathway is directed by small RNAs, this form of gene regulation is known as RNA interference (RNAi).

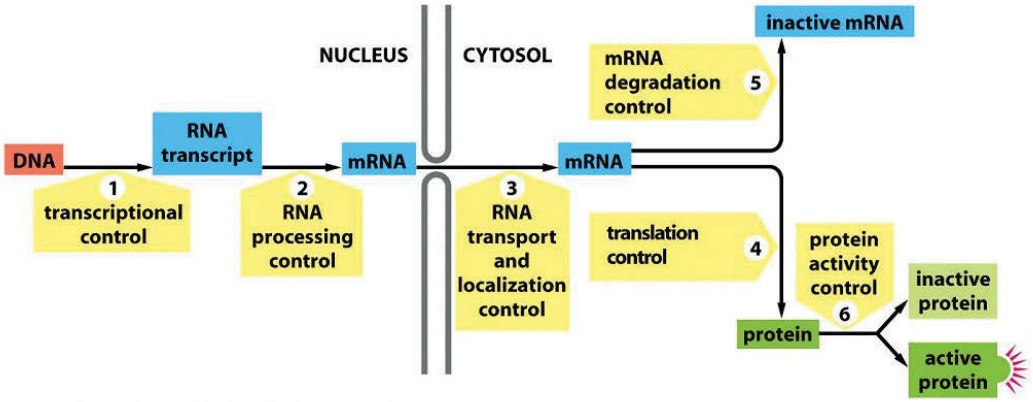


Figure 1.2. The cascade of gene regulation of the cell. Gene regulation can be divided into different layers. The main control point is (1) transcriptional regulation that directly determines if and how much mRNA can be synthesized from DNA. After the generation of mRNA, all the control steps from (2) are post-transcriptional regulation, which includes mRNA maturation, transport and localization, degradation and translational control of mRNA and protein modification. Image retrieved from (Alberts, 2008).

1.4 Epigenetic gene regulation

Epigenetics is a fast developing area that explores how epigenetic systems regulate gene expression. Epigenetic systems refer to DNA methylation, histone modification, histone variants and non-coding RNAs (mainly small RNAs). Epigenetic regulation has been suggested to influence expression of genes involved in many pathways, such as cell differentiation (Reik, 2007), genomic imprinting (Wood and Oakey, 2006) and plant responses to the environment (Chinnusamy and Zhu, 2009).

1.4.1 DNA methylation

As one of the epigenetic systems, DNA methylation has been well studied in animals and plants. In mammals, DNA methylation occurs at cytosine residues in a CG context. In contrast, three methylation contexts have been found in plants: CG, CHG and CHH (H refers to A, T or C). CG and CHG sites are symmetrical, meaning that they occur in both

DNA strands. Their methylation is maintained by METHYLTRANSFERASE I (MET1) (Finnegan and Dennis, 1993) and CHROMOMETHYLASE 3 (CMT3) (Lindroth et al., 2001), respectively. In contrast, CHH sites are asymmetrical and CHH methylation needs to be generated de novo by the RNA-directed DNA Methylation pathway (RdDM, see small RNA section) after each cycle of mitosis and meiosis, and is therefore known as de novo methylation (Bartee et al., 2001, Cao and Jacobsen, 2002). In the Arabidopsis genome, about 24% of CG, 6.7% of CHG and 1.7% of CHH sites are methylated (Cokus et al., 2008). Additionally, DNA methylation accumulates in some particular regions in the genome (mainly in heterochromatin and in transposable elements) rather than being spread evenly.

DNA methylation is important because of its relation to gene expression. In *Arabidopsis thaliana*, DNA methylation has been shown to play critical roles in regulating the expression of many endogenous genes (Finnegan et al., 1993, Kinoshita et al., 2004, Lippman et al., 2004). The expression of these genes is typically repressed when methylation occurs in the promoter regions, and the expression can be recovered by removal of the methylation (Kinoshita et al., 2004, Lawrence et al., 2004, Soppe et al., 2000). Several models suggest that DNA methylation affects gene expression by changing chromatin structure (Segal and Widom, 2009) or preventing transcription factors from binding to their targets (Bell and Felsenfeld, 2000).

1.4.2 Histone modification

Histone modification has been suggested as another epigenetic system participating in epigenetic regulation of gene expression. Chromatin is constructed from nucleosomes, which consist of histone-packed DNA. Five histone proteins associate to form

nucleosomes. H2A, H2B, H3 and H4 are the core histones, while H1 is the linker histone. A nucleosome is assembled by 147 base pairs (bp) of DNA wrapping around the core complex consisting of two molecules of the histone tetramers (H2A-H2B-H3-H4).

All four core histones can undergo post-translational modifications (PTMs). PTMs have been identified for more than 60 amino acid residues in histone core peptides, mostly concentrated on the unstructured N-tails of H3 and H4 (Roudier et al., 2009). These PTMs occurring on histones include methylation, acetylation, ubiquitinylation and phosphorylation (Figure 1.3). In addition, some residues can be methylated by up to three methyl groups, such as dimethylated arginine and tri-methylated lysine residues. The location of a modified residue and the type of modification are referred to as histone marks.

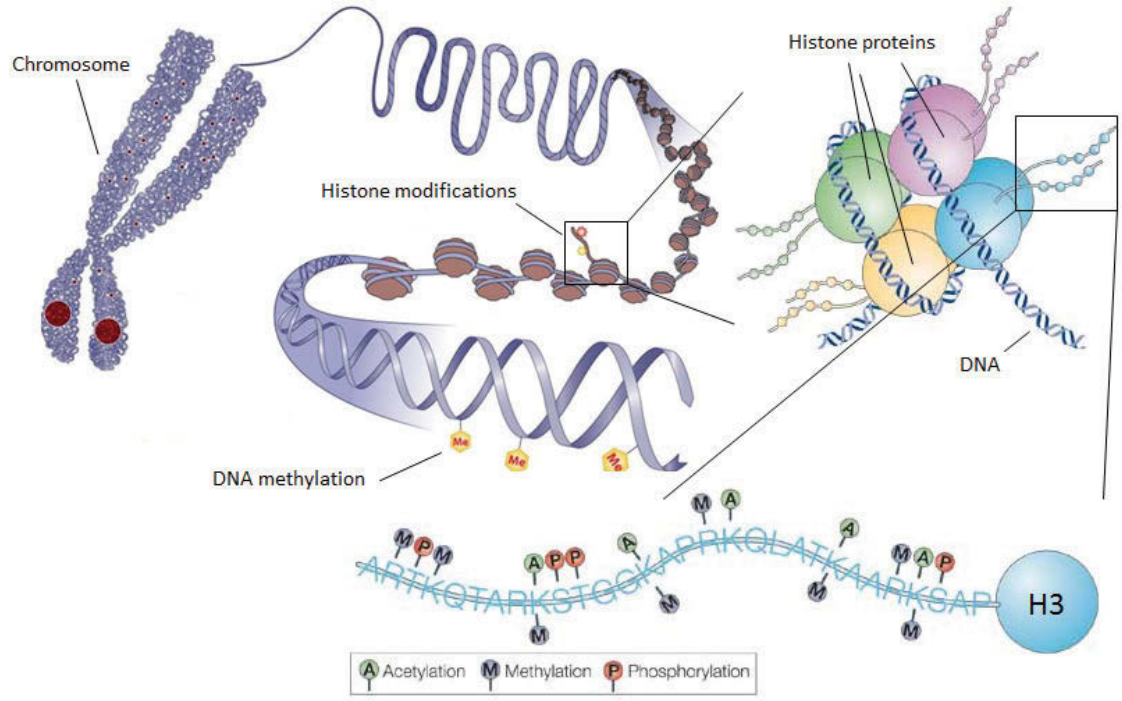


Figure 1.3. The spatial locations of histone modifications in a chromosome. [modified from (Zaidi et al., 2010) and (Levenson and Sweatt, 2005)]

Histone modifications play a role in epigenetic gene regulation. Besides the effect of DNA sequence on gene regulation, the expression level of a gene also depends on the local chromatin state. The local chromatin structure can be open, which permits transcription-related proteins to associate with the exposed DNA resulting in transcription. Chromatin structure can be tight enough to prevent contact between DNA and transcription-related proteins resulting in gene repression or silencing (Figure 1.4). The open and closed structures of chromatin can change in response to histone marks.

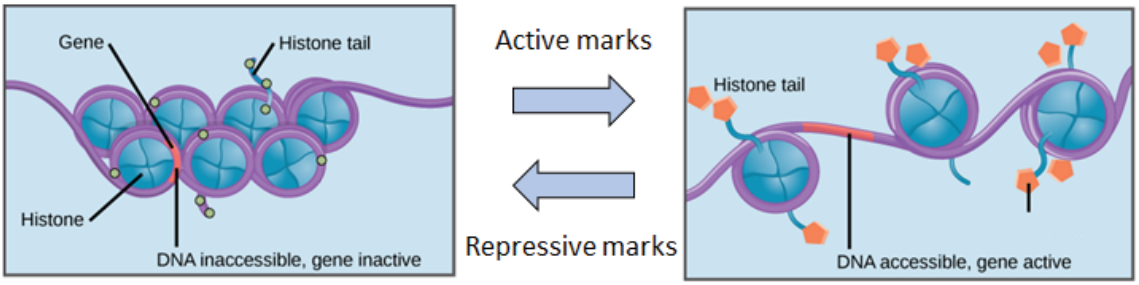


Figure 1.4. Model showing the relation between histone marks and chromatin states. (modified from http://cnx.org/contents/185cbf87-c72e-48f5-b51e-f14f21b5eabd@9.9:82/Biolog-y)

Most of the histone marks have been matched to particular chromatin states, correlating with gene expression patterns. This correlation has been well studied in Arabidopsis. A group of histone marks including histone H3K4 trimethylation (H3K4me3) and histone H3K9 acetylation (H3K9ac) occur in active locations of chromatin in which genes are expressed. In contrast, histone H3K27 dimethylation (H3K27me2) and trimethylation (H3K27me3) are usually located in repressed gene regions. Histone H3K9 dimethylation (H3K9me2) is particularly associated with heterochromatin and with transposable elements in which DNA is hyper-methylated. A recent publication identified four

chromatin states in Arabidopsis, each of which has unique and combinatorial patterns of DNA methylation, histone marks and gene expression (Roudier et al., 2011).

DNA methylation and histone modifications do not always work independently on gene regulation. Interactions between them have been documented. Microarray analysis revealed a genome-wide correlation between H3K9me2 and CHG methylation in Arabidopsis (Bernatavichute et al., 2008). In addition, loss of MET1 leads to redistribution of H3K27me3 and an increased H3K9me2 level in the genome (Deleris et al., 2012). The mechanism of the interactions between these two epigenetic components is still not clear.

1.4.3 Small RNAs

Small RNAs (sRNAs) play a critical role in post-transcriptional gene regulation (RNAi). Small RNA, particularly small interfering RNA (siRNA), is also involved in epigenetic gene regulation through the RdDM pathway, which links small RNA, DNA methylation and gene expression. In RdDM, single-stranded RNA (ssRNA) is transcribed from DNA loci by polymerase IV (POL IV), a plant-specific RNA polymerase. After that, RNA-depended RNA polymerase 2 (RDR2) converts the ssRNA to dsRNA. The dsRNA is then cleaved by Dicer-like protein 3 (DCL3) to yield siRNAs. Then, siRNA (mainly 24-nt siRNA) is loaded into a RNA-induced silencing complex (RISC) consisting of Argonaute proteins, and is used as a guide to trigger DNA methylation at loci with sequences corresponding to that of the siRNA (Zhang and Zhu, 2011). RdDM is usually related to the down-regulation of gene expression if it occurs in the promoter region.

RNAi has been suggested to guide the deposition of histone modifications to establish a

repressive chromatin state (Verdel et al., 2004, Sugiyama et al., 2007, Kloc et al., 2008). In Arabidopsis, it has been reported that small RNAs play a role in accurately maintaining the histone mark H3K9me2 in the genome (Enke et al., 2011). The association of small RNA with DNA methylation and histone modification generates a complex mechanism to epigenetically regulate gene expression in plants.

1.4.4 Alterations in hybrid epigenomes

It has been suggested that the magnitude of heterosis depends on the level of genetic variation between the genomes of the parents (Chen, 2010).However, intra-specific hybrids, such as those in *Arabidopsis thaliana*, show significant heterosis in biomass, indicating that large genetic variation is not required for hybrid vigour. It has been suggested that epigenetic changes may play roles in heterosis, particularly in intra-species hybrids, which have large differences in their epigenomes but limited differences in genomic sequences (Groszmann et al., 2011b, Ha et al., 2009, Herrera and Bazaga, 2010). A recent study has demonstrated that two plant lines having the same genetic background but different epigenetic backgrounds can produce biomass heterosis (Dapp et al., 2015).

Although research on the role of epigenetic changes in intra-species heterosis is still at an early stage, whole genome analyses have revealed alterations in siRNA and DNA methylation in hybrids compared to their parents (Groszmann et al., 2011b, Greaves et al., 2012, Shen et al., 2012). Based on these findings in intra-specific hybrids, epigenetic mechanisms involving DNA methylation and small RNAs have been suggested to play a role in heterosis.

Parents contribute one of two sets of chromosomes and different epigenetic backgrounds

to the hybrids. The two epigenomes may differ in DNA methylation levels at some epi-alleles. One epi-allele may have a higher DNA methylation level than the other. In hybrids, any of the following phenomena could happen in an epi-allele (Greaves et al., 2012): (I) the epi-allele with lower methylation is further methylated through Trans Chromosomal Methylation (TCM); (II) the epi-allele with higher methylation loses its methylation by Trans Chromosomal deMethylation (TCdM); and (III) DNA methylation remains at a similar level as that in the parental genome. (I) and (II) refer to non-additive inheritance with the average methylation level above or below MPV, while (III) refers to additive inheritance with the average methylation level equal to MPV. Non-additive inheritance may result in altered methylation patterns in hybrids, which can be correlated with the changes of transcript levels.

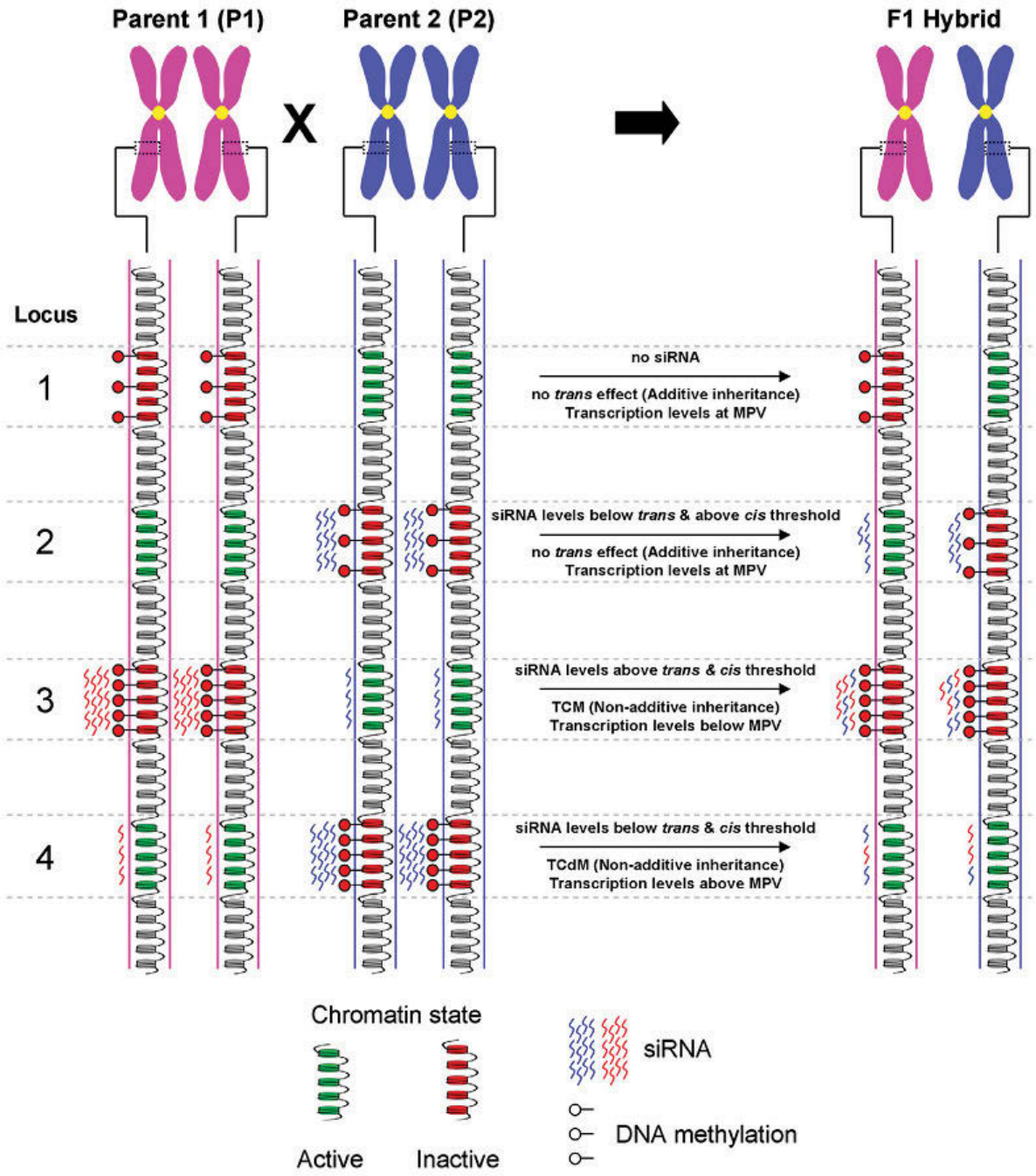


Figure 1.5. Model for the association between siRNA and DNA methylation level at different loci in the Arabidopsis F1 hybrids. (I) If the locus has no associated siRNA, there is no change in DNA methylation level at each epi-allele. The transcription level for the two alleles is at MPV, and additive inheritance happens. (II) If the siRNA level for the epi-allele with high methylation level is above the threshold of maintaining DNA methylation but does not reach the threshold of initiating TCM, this epi-allele maintain its methylation and no change happened at the other epi-allele. The average transcription level is also at MPV, and additive inheritance happens. (III) If the siRNA level for the epi-allele with high methylation level is above the threshold of maintaining DNA methylation and above the threshold of initiating TCM, the other allele undergoes TCM to elevate the methylation level. The average transcription level is below MPV, and non-additive inheritance happens. (IV) If the siRNA level for the epi-allele with high methylation level is below the threshold of maintaining DNA methylation and below the threshold of initiating TCM, this epi-allele undergoes TCdM and loses its methylation. The average transcription level is above MPV, and non-additive inheritance happens. Image retrieved from (Groszmann et al., 2011a).

The mechanism generating TCM and TCdM is not known, but it has been suggested that the trans-chromosomal changes in the genome are related to small RNAs. 24-nt siRNA has been reported to be required for the loci undergoing TCM (Groszmann et al., 2011b). A hypothesis has been suggested that the levels of siRNA are the key to determine whether or not a locus undergoes TCM or TCdM (Figure 1.5).

Histone modifications are another epigenetic system that plays a potential role in heterosis. Patterns of histone modification in hybrid rice and maize have shown correlations between altered gene expression and histone mark changes, compared with the parents (He et al., 2010, He et al., 2013). In Arabidopsis hybrids, the genome-wide histone modification patterns of the parents are additively passed to the F1 generation (Moghaddam et al., 2011, Dong et al., 2012), but non-additive inheritance was also observed at some loci. In allotetraploids derived from crosses between *A*. *thaliana* and *A*. *arenosa*, the key circadian clock genes *CIRCADIAN CLOCK ASSOCIATED1 (CCA1)*, *LATE ELONGATED HYPOCOTYL (LHY)*, *GIGANTEA (GI)*, and *TIMING OF CAB EXPRESSION1 (TOC1)* all have non-additive levels of H3K9ac and H3K4me2 associated with changed levels of expression compared with the parental lines (Ni et al., 2009). Alterations in the expression of these genes may play a role in increasing energy storage, which enables enhanced growth of plants.

Histone modifications have been shown to be linked to DNA methylation and sRNAs (Stroud et al., 2014). Correlations between histone modifications and DNA methylation suggest regions that undergo TCM/TCdM may have changes of histone marks in hybrids. At the Arabidopsis locus At3g43340/50, increases of DNA methylation in the hybrid result in decreases of the active mark H3K9ac, consistent with the observed decrease in expression of the adjacent genes (Greaves et al., 2014).

1.5 Heterosis at early developmental stages

Although biomass heterosis has been shown to be potentially correlated with partially altered transcriptomes and epigenomes in hybrids, the timing of heterosis establishment is still largely unknown. Some findings suggest that biomass heterosis may be established during embryogenesis due to the presence of larger mature seeds in the hybrids derived from certain Arabidopsis ecotypes (Groszmann et al., 2014). However, other hybrid combinations and even some reciprocal hybrids show increased vegetative biomass and do not show increased seed size compared to the parents (Barth et al., 2003, Meyer et al., 2004, Groszmann et al., 2014) indicating that seed size heterosis is not a universal phenomenon and may not be critical for vegetative biomass heterosis.

Other findings suggested that early stages of seedling development are critical for the establishment of biomass heterosis in later stages. Compared to the parents, the hybrid cross C24/Col (Colombia) has a higher growth rate from 0 DAS (day after sowing) to 15 DAS but a similar growth rate from 15 DAS to 25 DAS (Meyer et al., 2004). Measurements of cotyledon area at early seedling stages showed that the highest growth rate occurs at around 3-4 DAS in C24/Col hybrids (Meyer et al., 2012). Transcriptome analyses revealed increased expression levels of photosynthesis- and chlorophyll biosynthesis-related genes in C24/Col hybrids at 3-4 DAS, whereas the expression levels of these genes at 10 DAS are similar to that in the parents (Fujimoto et al., 2012). In addition, fatty acid composition analyses showed that C24/Col hybrids have increased consumption of storage fatty acids and increased production of newly synthesized fatty acids at 6 DAS (Meyer et al., 2012). These transcriptomic and metabolic changes could cause the growth vigour of hybrids, and early developmental stages may be critical for

late biomass heterosis establishment. However, this theory needs to be verified in other hybrid combinations of Arabidopsis.

2 Chapter II Research aims

Biomass heterosis is a common phenomenon in the F1 generation of many intra-specifically crossed plants. Differences at both epigenetic and genetic levels between parents and hybrids might contribute to the biomass heterosis. Although many studies have been done on hybrid transcriptomes and epigenomes, most were concentrated on relatively late development stages. As previous evidence has suggested, biomass heterosis might be determined at early seedling developmental stages. Investigations on hybrid transcriptomes and epigenomes at these early stages are necessary to identify the key genes and/or pathways for biomass heterosis.

In this project, two Arabidopsis ecotypes, L*er* and C24, were used as their hybrids show large levels of heterosis in size and biomass (Figure 1.1). Transcriptomes of the parents and the reciprocal hybrids (L*er* x C24 and C24 x L*er*) at 0, 3, 5, 7 DAS were analysed. Seedling development begins in germinating seeds, and differences in epigenomes between parents and hybrids in seeds possibly cause the differences in transcriptomes at later stages. Genome-wide histone modification patterns of H3K4me3, H3K9ac, H3K27me3 and H3K9me2 were investigated in parent and hybrid seeds (0 DAS).

The aims of this project were:

1. Because heterosis could be established at early seedling stages, the transcriptional alterations between parents and hybrids in the development time course could provide a hint of the timing of heterosis establishment.
2. Through transcriptome analyses on hybrids at early developmental stages, some

genes involved in key pathways to heterosis could be identified.

3. Although global patterns of histone modifications have been shown unchanged in hybrids at late developmental stages in Arabidopsis, potential global and localized changes may still exist in hybrids at early stages.
4. The levels of histone modifications will be compared to gene expression levels to test if histone modifications are involved in regulating non-additive gene expression in hybrids at early stages.

3 Chapter III Materials and Methods

3.1 Plant growth media and conditions

Arabidopsis lines used in this project are L*er*, C24 and their two hybrids, L*er* x C24 and C24 x L*er*. Hybrid seeds were obtained from hand-pollinated crosses between L*er* and C24, and parental seeds were obtained from self-pollinated parents with restricted number of pollinated stigmas. Seeds were sterilized by treatment with chlorine (generated by mixing 100 mL of household bleach with 3 mL concentrated HCl (Hydrocholoric acid) for three hours, washed with 100% ethanol, dried. Sterilized seeds were sown onto Murashige and Skoog (MS) medium (Table 2.1); plates were placed for imbibition at 4 ºC in the dark for three days. Following imbibition, plants were grown at 21 ºC with a 16 hour light and an 8 hour dark. For RNA-Seq, two replicates of 10 seeds or seedlings were harvested at same time of each day at 0, 3, 5 and 7 days after sowing (DAS). For ChIP-Seq, two replicates of 2000 seeds of each plant line were harvested immediately after imbibition. For qPCR validation experiments, hybrid seeds were sown 7-hour later than parent seeds. Two replicates of 10 seeds or seedlings of parents and hybrids were harvested at same time of each day at 3, 4, 5, 6 and 7 DAS. Embryos were isolated from imbibed seeds following the protocol described in previous literature (Perry and Wang, 2003).

Table 3.1. MS medium recipe (in 1 L)

Compound	Mass	Compound	Mass
NH_4NO_3	1.65 g	$CoCl_2.6H_2O$	0.0125 mg
$CaCl_2.2H_2O$	0.44 g	KI	0.115 mg
$MgSO_4.7H_2O$	0.37 g	Na_2EDTA	16.75 mg
KNO_3	1.9 g	$FeCl_3.6H_2O$	13.5 mg
KH_2PO_4	0.17 g	Nicotinic acid	0.05 mg
$MnSO_4.4H_2O$	11.15 mg	Pyridoxine HCl	0.05 mg
$Na_2MoO_4.2H_2O$	0.125 mg	Thiamine HCl	0.01 mg
H_3BO_3	3.11 mg	Glycine	0.2 mg
$ZnSO_4.7H_2O$	4.3 mg	Sucrose	30 g
$CuSO_4.5H_2O$	0.0125 mg	Myoinositol	0.1 g

Combine ingredients and dissolve. Adjust pH to 5.7 with 1N KOH. Add 4.0 g of Difco™ Agar (Bacto Laboratories) to each 500 mL aliquot and autoclave prior to use.

3.2 Measurements of seed weight and size

Seeds were weighed in at least three replicates of 1000 seeds harvested at different time for parents and hybrids. Projected area of seeds were utilized as reflection of seed size. Seeds were placed on white background, and images of seeds were captured by a dissecting scope. Average projected area of seeds was calculated by using Image-J software (Schneider et al., 2012).

3.3 Seed germination time assay

Seeds of parent and hybrids were sowed onto MS medium plates with three-day

imbibition treatment. Emergence of radicles was recorded every two hours after sowing until all seeds germinated. Three biological replicates of 40 seeds for each plant lines were tested.

3.4 RNA extraction

Total RNA was extracted from plant materials by using an RNeasy Plant Mini Kit (Qiagen) following the manufacturer's manual. RNA samples were applied in qRT-PCR analyses and in next-generation sequencing (NGS). RNA library preparation and deep-sequencing were performed by the Australian Genome Research Facility (AGRF).

3.5 Chromatin Immune-Precipitation Sequencing (ChIP-Seq)

3.5.1 Chromatin immune-precipitation (ChIP)

Chromatin immune-precipitation (ChIP) is a technique to immune-precipitate DNA sequences associated with particular proteins, such as transcription factors and histone proteins, in chromatin. There are two methods to perform ChIP based on including a cross-linking step in the experiment or not, named cross-link ChIP (X-ChIP) and native ChIP (N-ChIP), respectively. For histone modification research, cross-linking is normally not needed because of the tight association between histone and DNA. In this project, N-ChIP (Figure 2.1) was applied for this project.

Two replicates of 2000 seeds were ground to fine powder in a 1.5 ml tube using a plastic pestle in 150 µl ChIP lysis buffer. 600 µl ChIP lysis buffer was added into the tube, vortex.

Micrococcal nuclease (MNase; ThermoFisher, 88216) was diluted 1/200 with ChIP lysis buffer with 10 mM $CaCl_2$. After 10 minutes incubation at 37 ^oC, the reaction was stopped by adding EGTA to 10 mM. Samples were sonicated at 40 amplitude for two cycles of 15 seconds by using a UP400S sonicator (Hielscher). Samples were cooled on ice for 30 seconds between the two cycles, vortexed well. Sonicated samples were centrifuged at 20, 000 g for 15 minutes at 4 ^oC. The supernatant was divided into six aliquots. One aliquot was used as Total Input control. Genomic DNA was extracted from Total Input control by using a DNeasy Plant Mini Kit (Qiagen) following the manufacturer's manual. Five aliquots were incubated with 1 µl proteinase inhibitor (Sigma, P9599), 8 µl ChIP lysis buffer, 2.5 µl Magna ChIP™ Protein A Magnetic Beads (Millipore, 16-661) and 1 µl of antibodies for each of four histone marks, K4me3 (Milipore,07-473), K9ac (Milipore, 07-352), K27me3 (Milipore, 07-449) and K9me2 (diagenode, pAb-060-050), or 1 µl ChIP lysis buffer (No Antibody control). Magnetic beads were washed twice in ChIP lysis buffer before use. The mixtures were incubated with rotation by using an Intelli-Mixer RM-2L (ELMI) for four hours at 4 ^oC. After incubation, the supernatant was removed by using MagneSphere® Technology Magnetic Separation Stands (Promega). The magnetic beads were washed for four times using 125 µl of the following solutions in order, low salt buffer, high salt buffer, LiCl wash buffer and TE buffer. Antibody-protein complexes were eluted from beads twice with 75 µl elution buffer (freshly made) for 15 minutes at room temperature. The combined elutes were incubated with 3 µl 0.5M EDTA, 6 µl 1M Tris-Cl (pH 8.0) and 0.5 µl proteinase K (10 mg/ml) for one hour at 4 ^oC, followed by a purification step by using a MinElute Reaction Cleanup Kit (Qiagen). Purified DNA was applied in qRT-PCR analyses and ChIP-Seq library preparation.

Table 3.2. ChIP lysis buffer stock

Compound	Concentration
HEPES	50 mM
NaCl	150 mM
Triton X-100	1 %
Deoxycholate	0.1 %
SDS	0.1 %

Note: Filter sterilise and store at 4 ℃.
Add before use for 1 ml of ChIP lysis buffer:
20 µl proteinase inhibitor
0.5 µl PMSF

Table 3.3. Low salt wash buffer stock

Compound	Concentration
Tris-Cl (8.0)	20 mM
EDTA	2 mM
NaCl	150 mM
SDS	0.1 %
Triton X-100	1 %

Table 3.4. High salt wash buffer stock

Compound	Concentration
Tris-Cl (8.0)	20 mM
EDTA	2 mM
NaCl	500 mM
SDS	0.1 %
Triton X-100	1 %

Table 3.5. LiCl wash buffer stock

Compound	Concentration
Tris-Cl (8.0)	10 mM
EDTA	1 mM
LiCl	250 mM
NP-40	1 %
Sodium deoxycholate	1 %

Table 3.6. TE buffer stock

Compound	Concentration
Tris-Cl (8.0)	10 mM
EDTA	1 mM

Table 3.7. Elution buffer stock

Compound	Concentration
SDS	1 %
$NaHCO_3$	100 mM

3.5.2 ChIP-Seq library preparation

ChIP-Seq library preparation was performed by using NEBNext® ChIP-Seq Library Prep Reagent Set for Illumina® (NEB, E6200) and NEBNext® Multiplex Oligos for Illumina® (NEB, E7335), following manufacturer's manuals. Quality and quantity of ChIP-Seq library samples were measured by using a 2100 Bioanalyzer Instrument (Agilent Technologies). Deep-sequencing was performed by AGRF.

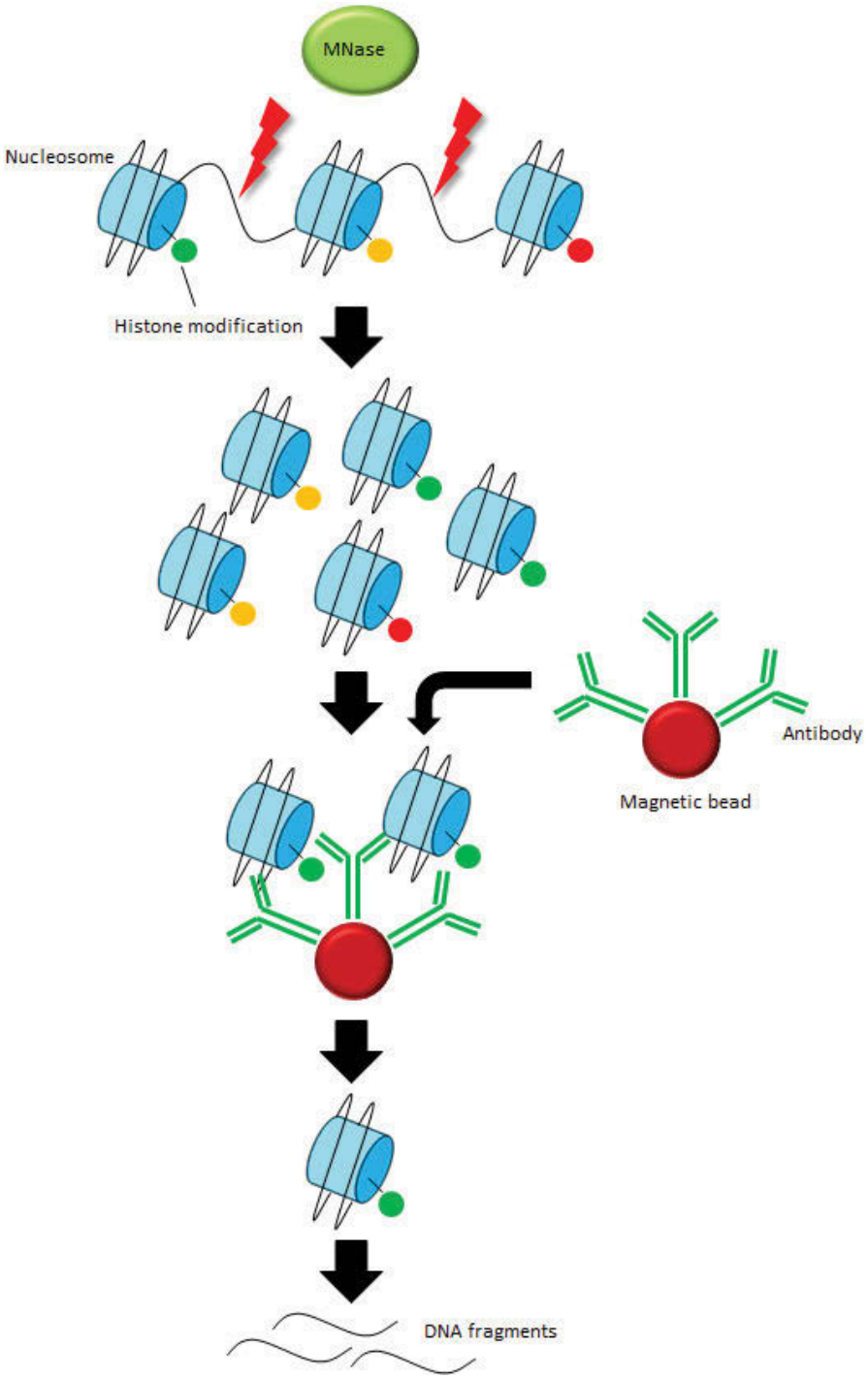


Figure 3.1. Workflow of N-ChIP. Chromatin chains are fragmented into single nucleosomes by micrococcal nuclease (MNase). Single nucleosomes with a specific histone modification can be bound to antibodies combined with magnetic beads, and this is followed by the precipitation of nucleosome-antibody-bead complexes. DNA fragments specific to a histone modification are obtained from a digestion step for the histone paroteins.

3.6 qRT-PCR

Total RNA samples were DNase-treated by incubation at 37ºC for 15 minutes in a reaction containing: 0.05 unit/µL RQ1 RNase-free DNase (Promega), 1x RQ Buffer (Promega), 1 unit/µL RNase-out (Invitrogen) and 0.05 µg/µL total RNA sample. 1 µL of Stop Buffer was added to the reaction and the mixture was then incubated at 65 ºC for 15 minutes. 1 µg of purified DNase-treated total RNA was primed by 5 µM oligo dT at 65ºC for five minutes and immediately transferred to ice and incubated for one minute. The first strand cDNA was synthesized in a reaction containing: 1x First-Strand Buffer (Invitrogen), 5 mM DTT (Invitrogen), 10 unit/µL SuperScript™ III (Invitrogen), 0.5 mM dNTPs and 2 unit/µL RNase-out (Invitrogen). The reaction was terminated by inactivating the enzyme at 85ºC for 5 minutes. The reaction was then diluted to 50 µL with DEPC H2O and stored at -20 ºC.

For ChIP-Seq library verification, qRT-PCR analysis was carried out on a Corbett 2000 Rotor-Gene real-time PCR machine (Corbett Research). qRT-PCR was performed using two biological replicates and four technical replicates for each cDNA sample. The cDNA was amplified in a 20 µL reaction containing: 0.5X Power SYBR® Green (Invitrogen), 0.4 mM dNTP (ThermoFisher), 7 mM $MgCl_2$, Platinum Taq polymerase (Invitrogen), 0.8 µM forward primer, 0.8 µM reverse primer (Table 3.8), 20 ng/µL cDNA template. All qRT-PCR reactions were carried out under the following cycling conditions: 1 cycle of 95 ºC for 2 minutes, 40 cycles of 95 ºC for 15 seconds and 60 ºC for one minute. Fluorescence was acquired at the 60 ºC step and a 55 ºC to 95 ºC melting cycle was then carried out. The results were averaged and normalized to values of Total Input. Standard error of the mean (SEM) was calculated.

To validate expression levels of the selected genes, two biological pools of ten individuals for each cDNA sample from different time points and four technical replicates were used. qRT-PCR was carried out on a 7900HT Fast Real-Time PCR System (AppliedBiosystems), and sample loading was operated by a Genesis Workstation 200 (Tecan). cDNA was amplified in a 5 µL reaction containing: 0.5X Power SYBR® Green (Invitrogen), 0.4 mM dNTP (ThermoFisher), 7 mM $MgCl_2$, Platinum Taq polymerase (Invitrogen), 0.8 µM forward primer, 0.8 µM reverse primer (Table 3.8), 20 ng/µL cDNA template. All qRT-PCR reactions were carried out under the following cycling conditions: 1 cycle of 95 ºC for 10 minutes, 40 cycles of 95 ºC for 15 seconds and 60 ºC for one minute. Fluorescence was acquired at the 60 ºC step and a 60 ºC to 95 ºC melting cycle was then carried out. The results of parents and hybrids from different time points were averaged and normalized to values of a reference gene, AT4G34270. The expression level of this gene is unchanged between hybrids and parents in our transcriptome data (data not shown), and is unchanged in Col during early developmental stages (Czechowski et al., 2005, Dekkers et al., 2012).

Table 3.8. Primers used in ChIP-qPCR and expression level validation

Gene ID	Name	Forward Primer	Forward Primer
		ChIP-qPCR	
at5g09810	ACTIN7	CGCTGTTGTTTCTCCTCCAT	GCGAACGGATCTAGACTCA
at1g37110	TA3	AGACAGCTCTGCGTGGAAGTC	TTATCAGTCTCAGCATTACACAG
at4g18960	AGAMOUS	CCCAAAGATTTTAGTGCCTCA	GGTTCAAGTTGGGCAATCAC
		qPCR validation for RNA-Seq data	
at3g51820	G4	ACGTCAAGTACCAGGCAAGC	ATGCCATCACTTGCCCACAT
at1g44446	CAO	CGCTGAACAGGTCTTAAACGA	CCACGATCTACTGCGTTCCT
atcg00350	PSAA	CTCCGCGATTTCTTATGGGC	TAACCACGCCCGCTGAATAG
atcg00020	PSBA	CGCGAAAGCGAAAGCCTATG	AGGAGCAGCAATGAATGCGAT
at3g27690	LHCB2.3	CTCCTCAGAGCATCTGGTACG	TGTTTCTGGATCGGCTGAGAG
at3g54890	LHCA1	ATTGCATTTGTTGAGCACCAGAG	CGCAAGCCGCCCGTT
at2g30570	PSBW	GTCCGGCTCTCCTTCTCAAG	GAAACTCCAGCCCCCATGAC
at1g08380	PSAO	GTGCATCGGGAGGAAGAGTC	CGAAGAAGAGACCCGTCAGG
atcg00490	RBCL	GGCTTCAGGGGGTATTCACG	GAGCTACTCGGTTGGCTACG
at1g67090	RBCS1A	GAATTCGAGTTGGAGCACGG	TGCACTCTTCCACTTCCTTCA

3.7 Bioinformatics analyses

The workflow of bioinformatics analyses for project is described in Figure 3.2. Both ChIP-Seq and RNA-Seq data undergo quality control to filter the reads with low sequencing quality. In turn, sequencing reads are mapped to the Arabidopsis reference genome to be labelled with chromosome coordinates. The mapped reads of ChIP-Seq are then analysed to obtain the abundance information of the four histone marks in the

enriched regions in the genome, while read counts of transcripts are obtained for every Arabidopsis gene from analysing RNA-Seq data. The hybrids and parents are statistically compared based on the levels of histone marks and gene expression to identify the differential genes.

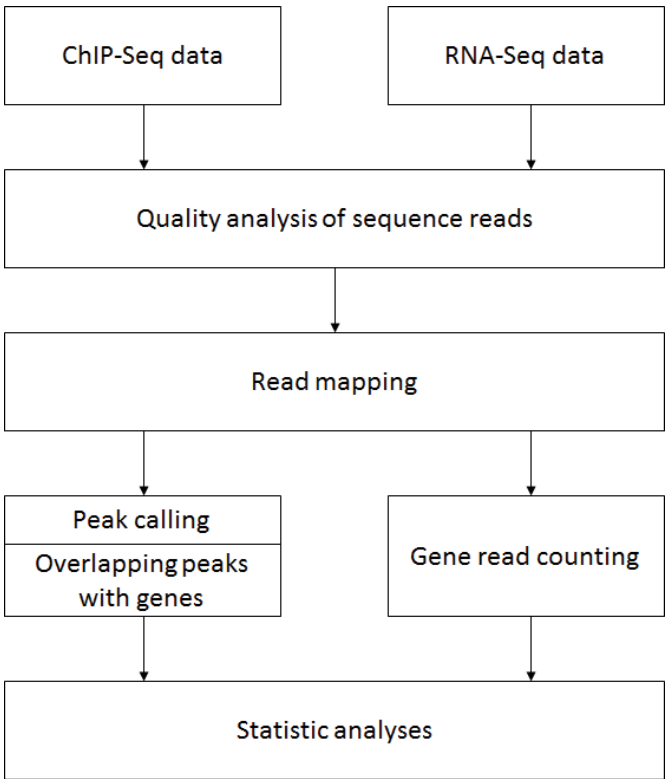


Figure 3.2. Workflow of statistical analyses for RNA-Seq and ChIP-Seq data. Firstly, the raw sequencing data undergoes quality analysis of sequence reads, which provide the preliminary results about the quality of data. Secondly, sequencing reads are mapped to reference genome

3.7.1 Read quality control and read mapping

Raw RNA-Seq and ChIP-Seq data were processed by using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) software with default settings for quality control.

The FastQC results suggested high average quality score per base in the reads (Figure 3.3

A), and high quality score per read for the majority of reads (Figure 3.3 B). The sequence content analyses suggested a consistent average percentage of sequence content per base in the reads, except for the first 10 bases, which have fluctuating percentages (Figure 3.4 A). The fluctuation could be caused by a relative low base quality score in the first several base of each read (Figure 3.3 A); this could be a minor technical issue from the deep sequencing machine. To solve it, the first 10 bases were trimmed from every read in all libraries. The percentage of GC content for the reads presents a normal distribution with the peak located around 46% (Figure 3.4 B), and the distribution curve fits the theoretical distribution curve. These two sequence content analyses suggested that there was no major sequence content bias generated in the PCR reactions in the library preparation step.

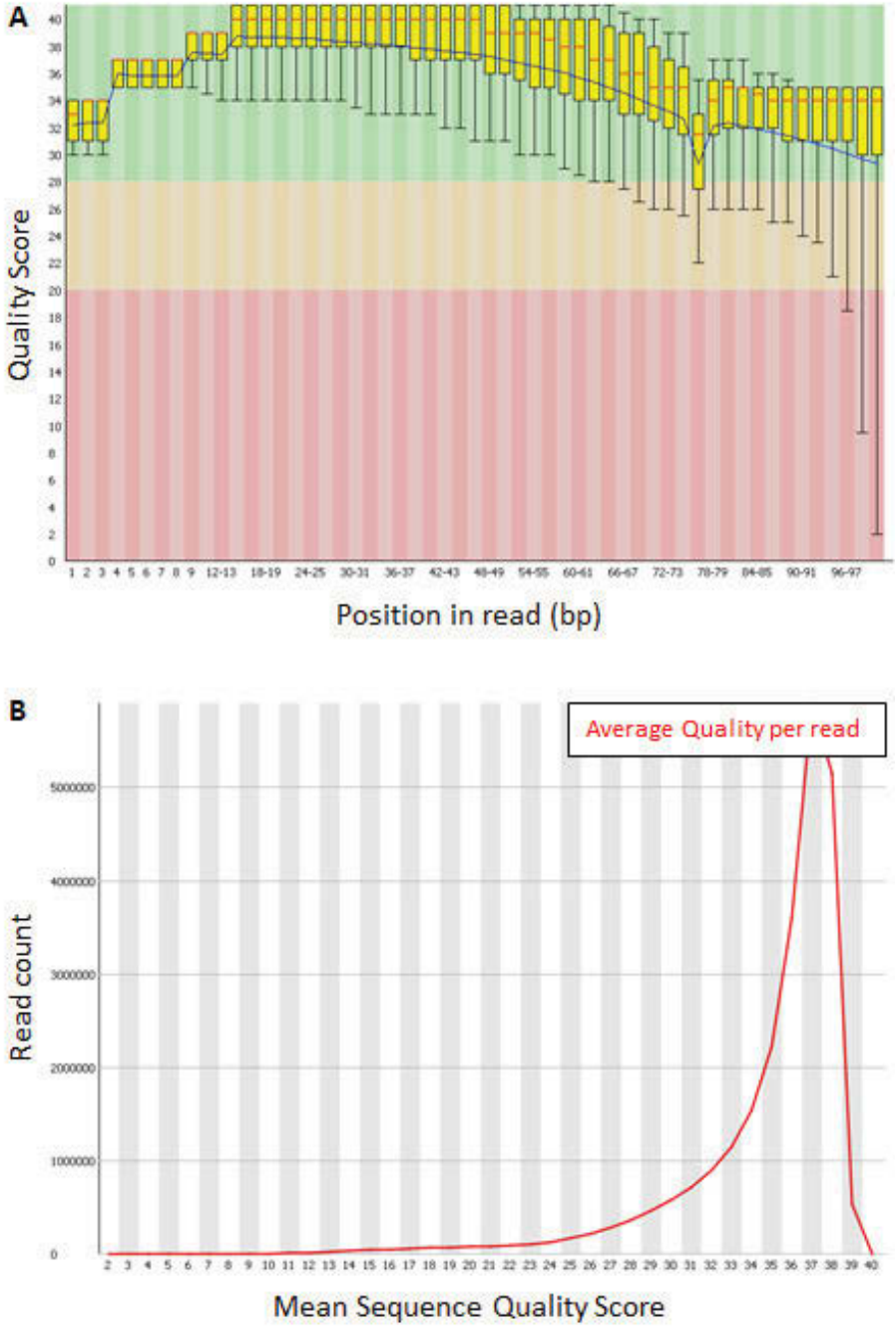


Figure 3.3. The RNA-Seq data of Ler x Ler showing high quality scores per base (A) and per read (B).

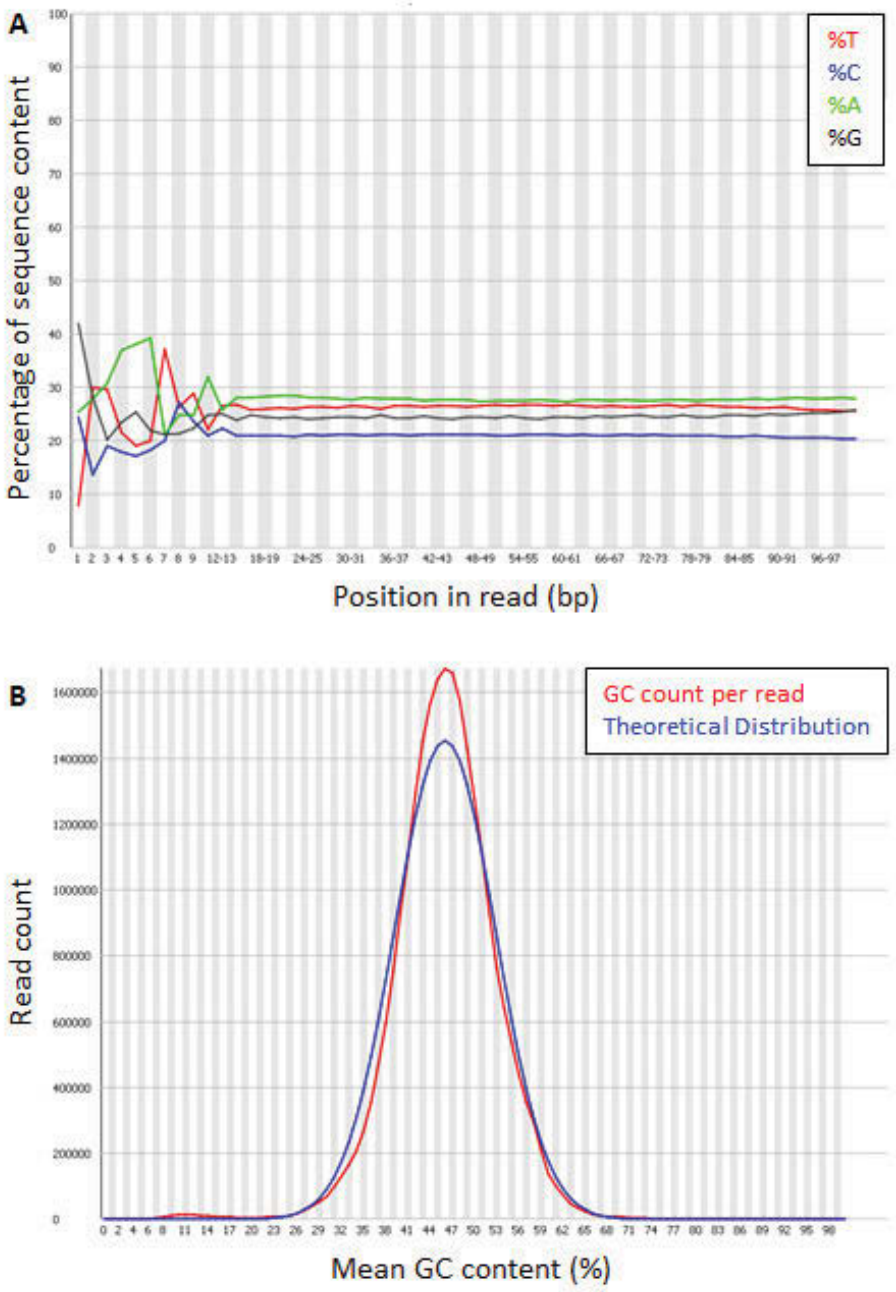


Figure 3.4. The RNA-Seq data of Ler x Ler showing constant percentage of sequence content over reads (A) and expected distribution of GC content (B).

After FastQC analyses, all the reads were mapped to the reference genome. Because of the lack of fully sequenced and annotated L*er* and C24 genomes, the reads of the four plant lines have to be mapped to the Col genome (TAIR10) by using Biokanga align (http://sourceforge.net/projects/biokanga/) with default settings and additionally applying

parameters –A25000, -M5 and –y10. ChIP-Seq reads were mapped by using Biokanga align with default settings and parameters –M5 and –y10. For ChIP-Seq data, only reads having unique sequences were considered in further analyses. Additionally, instead of using –M5, parameters -r5 and -R500 were applied in mapping ChIP-Seq reads for K9me2, because K9me2 is usually located in transposon and DNA repeat regions in genome. The Col genome has minor differences from the L*er* and C24 genomes, such as single nucleotide polymorphisms (SNPs), DNA deletions, insertions and duplications. The statistics of RNA-Seq and ChIP-Seq data are listed in Table 3.9 and 3.10, respectively. For RNA-Seq, the number of mapped reads range from 30 to 90 million, which is sufficient for further statistical analyses due to the relative small genome of Arabidopsis. For ChIP-Seq, the number of unique reads range from 5 to 16 million, which is sufficient for statistical analyses.

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

Time point	Library	Total reads	Mapped reads	% of mapped reads
0 DAS	LerxLer_rep1	98441762	85409783	86.76
	LerxLer_rep2	96156914	85488729	88.91
	C24xC24_rep1	97570538	87413631	89.59
	C24xC24_rep2	93017318	82066813	88.23
	LerxC24_rep1	91124970	82789395	90.85
	LerxC24_rep2	1.02E+08	91694282	90.13
	C24xLer_rep1	84942582	74798513	88.06
	C24xLer_rep2	84367892	75316989	89.27
3 DAS	LerxLer_rep1	61316562	55456484	90.44
	LerxLer_rep2	41476384	37215645	89.73
	C24xC24_rep1	56039214	50550467	90.21
	C24xC24_rep2	61129024	54955273	89.90
	LerxC24_rep1	60173042	54295684	90.23
	LerxC24_rep2	60325710	54629993	90.56
	C24xLer_rep1	62995656	56966347	90.43
	C24xLer_rep2	59754678	53875254	90.16
5 DAS	LerxLer_rep1	67102810	60720674	90.49
	LerxLer_rep2	34749462	31393951	90.34
	C24xC24_rep1	68284354	61630876	90.26
	C24xC24_rep2	87543680	79521234	90.84
	LerxC24_rep1	67720560	61403511	90.67
	LerxC24_rep2	70448110	63757981	90.50
	C24xLer_rep1	64277178	57925532	90.12
	C24xLer_rep2	45172974	40955545	90.66
7 DAS	LerxLer_rep1	69448718	63170118	90.96
	LerxLer_rep2	70920002	64134107	90.43
	C24xC24_rep1	50696934	45887798	90.51
	C24xC24_rep2	53866576	48569060	90.17
	LerxC24_rep1	55927726	50811505	90.85
	LerxC24_rep2	57645448	52224744	90.60
	C24xLer_rep1	60359896	54818994	90.82
	C24xLer_rep2	54228682	49057983	90.47

Table 3.10. The numbers of reads in ChIP-Seq mapped to Col genome.

Histone mark	Library	Total reads	Mapped reads	Reads with unique sequences
Total Input	LerxLer_rep1	21702128	15410926	10990356
	LerxLer_rep2	29723322	21242917	15967566
	C24xC24_rep1	20723051	14510693	10611460
	C24xC24_rep2	20367796	14044448	11204202
	LerxC24_rep1	20807873	14517991	11247100
	LerxC24_rep2	21997820	15115725	12150520
	C24xLer_rep1	22278900	15561864	12164655
	C24xLer_rep2	19520780	13438005	10809729
K4me3	LerxLer_rep1	22519892	19125612	13588096
	LerxLer_rep2	20631471	17402100	13117599
	C24xC24_rep1	21840747	18227515	13832188
	C24xC24_rep2	20575035	16992427	13026362
	LerxC24_rep1	20049541	16688105	12873211
	LerxC24_rep2	20397014	17872615	12767636
	C24xLer_rep1	21796688	18142344	13845993
	C24xLer_rep2	17097645	14288927	11292408
K9ac	LerxLer_rep1	20353166	16757110	8890002
	LerxLer_rep2	20007541	16703816	12782351
	C24xC24_rep1	19883985	16345414	9855560
	C24xC24_rep2	24626611	20174787	12217426
	LerxC24_rep1	20800522	16917582	10613739
	LerxC24_rep2	21658905	18499788	12710385
	C24xLer_rep1	23136079	18950706	11372293
	C24xLer_rep2	18970357	15587231	9432000
K27me3	LerxLer_rep1	18826747	14618875	10087398
	LerxLer_rep2	22210405	16494521	10360112
	C24xC24_rep1	19953881	15685338	9793898
	C24xC24_rep2	18145651	13884977	5924068
	LerxC24_rep1	18924776	11754877	8361808
	LerxC24_rep2	19159374	14865979	8683243
	C24xLer_rep1	21552046	16324572	11182721
	C24xLer_rep2	17761461	13699141	4726121
K9me2*	LerxLer_rep1	23135226	74121231	9176539
	LerxLer_rep2	20421937	66675222	9566308
	C24xC24_rep1	21583456	67783808	10535901
	C24xC24_rep2	18270290	59027117	7587880
	LerxC24_rep1	20899944	65271313	9285933
	LerxC24_rep2	19579014	66920226	9071781
	C24xLer_rep1	21714389	68801101	9939682
	C24xLer_rep2	18369641	64134072	7347631

* There are more mapped reads than total reads in K9me2 libraries, because parameters –r5 and –R500 were applied in mapping K9me2 reads.

To test if the deep-sequencing data are reproducible, two biological replicates of the four plant lines were analysed for their similarity of expression of every gene. In the

dendrogram of 0 DAS libraries, replicates of each of the four lines clustered together (Figure 3.5; Appendix 1 and 2), suggesting minor differences between replicates. In addition, a synthetic MPV library was artificially generated from the two parent libraries for each replicate. The two parents, L*er* x L*er* and C24 x C24 were found to be distanced from each other, while the two hybrids were relatively close to MPV and to each other in the map. This was expected due to the genetic backgrounds of the two parents being different, whereas the genetic backgrounds in hybrids are similar to each other and similar to the mix of the parental backgrounds.

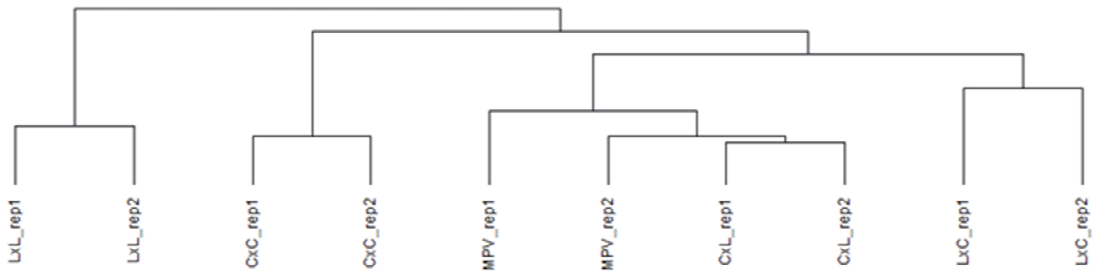


Figure 3.5. Dendrogram shows the similarity of RNA-Seq libraries in two biological replicates of the parents and the hybrids at 0 DAS. L: Ler. C:C24.

3.7.2 RNA-Seq

Mapped RNA-Seq reads were allocated to genomic features by using Biokanga map loci with default setting and additional parameters –t0 and –x1. Expression levels of each gene in Arabidopsis were determined by exon read counts. Read counts of datasets from different time points were normalized as read count per million reads. Statistical analyses were applied on read counts of genes between hybrids and parents by using DESEQ (http://bioconductor.org/packages/2.13/bioc/html/DESeq.html). Non-additive genes in hybrids were defined to have significantly differential expression levels against MPV

(fold-change ≥ 1.3, p-value ≤ 0.01). Differentially expressed genes between parents were defined to have significantly expression level against each other (fold-change ≥ 1.5, p-value ≤ 0.01).

Heat maps of gene expression were drawn using GENE-E (http://www.broadinstitute.org /cancer/software/GENE-E/) and genes were clustered based on the Pearson correlation algorithm. Venn diagrams were made by using Venny (http://bioinfogp.cnb.csic.es /tools/venny/index.html) and Venn Diagram Plotter (http://omics.pnl.gov/software/venn-diagram-plotter). Transcriptional profiles of non-additive genes over time were made using STEM (Ernst and Bar-Joseph, 2006), K-means clustering method applied. Gene Ontology (GO) analyses were performed on agriGO (Du et al., 2010) with default settings. To identify different types of non-additive genes over time, in addition to non-additive genes (fold-change from MPV ≥ 1.3, p-value ≤ 0.01), the genes with non-additive expression levels in only one reciprocal hybrid but with same direction of change (fold-change from MPV ≥ 1.2) in the other hybrid were also considered to be non-additively expressed in both hybrids. To determine changes in gene expression over time, the MPV must be above 1.2-fold greater or less than the MPV at previous time point. Motif analyses on promoters of non-additive genes were performed on MEME-ChIP function of The MEME Suite (Bailey et al., 2009) applying default settings. Only the BPGs (better-than-best-parent gene) common to both reciprocal hybrids were considered in analyses. Output common motifs were compared against known plant motifs (Franco-Zorrilla et al., 2014) by using Tomtom function of The MEME Suite.

3.7.3 ChIP-Seq

ChIP-Seq reads are enriched in localised regions of genome, named histone modification

peaks, where histone proteins are targeted by histone modifications. Peaks were called by using MACS2 callpeak (Zhang et al., 2008) with p-value less or equal to 10^{-6} for K4me3 and K9ac and with p-value less or equal to 10^{-2} for K27me3 and K9me2. Duplicated reads (reads having same nucleotide sequences) generated were removed for all histone marks by using MACS2 filterdup. Peaks of K27me3 and K9me2 were called with additional parameter –broad, as these two marks usually form broad and continuous peaks. Peaks of histone modifications from the parent and hybrid datasets were merged using BEDTools merge (Quinlan and Hall, 2010). Read counts were obtained for each peak by using BEDTools intersect. Statistical analyses were applied on read counts of genes between hybrids and parents by using DESEQ. Genes with non-additive levels of histone modifications in hybrids were identified by against MPV (fold-change ≥ 1.25, p-value ≤ 0.05). Differentially expressed genes between parents were defined to have significantly expression level against each other (fold-change ≥ 1.5, p-value ≤ 0.05). Profiles and heat-maps of histone modification levels over genes were drawn by using ngsplot (Shen et al., 2014).

3.7.4 SNP analyses

The read counts on SNP positions were called by using SAMtools (Li et al., 2009), BCFtools (Li et al., 2009) and VCFtools (Danecek et al., 2011). SNP positions were considered only when they satisfy following criteria: they have different nucleotides between L*er* and C24; read counts of SNPs must be three or more; read coverage must be present at SNP positions in all the four plant lines; they must be homologous in parent lines; they must be heterozygous in hybrid lines; they must have nucleotides consistent with previous data (1001 Arabidopsis genomes project, http://1001genomes.org/index.html). Genes and histone modification peaks were considered only when they have

one or more SPK (SNP per kb). Additionally, genes in analyses must have strong positive correlations between read counts from SNPs and read counts from whole transcript exons by comparing the parents {Correlation Factor (CF) $\leq$ 0.1, CF = Absolute[P_{Ler}(SNP) – P_{Ler}(transcript)], P_{Ler} = proportion of L*er* in parents} and by comparing hybrids to MPV {CF $\leq$ 0.1, CF = Absolute[P_H(SNP) – P_H(transcript)], P_H= proportion of hybrid in sum of hybrid and MPV}. For histone modification peaks, CF must be 0.05 or less for more stringent criteria.

Genes or peaks with changed allelic ratios from parental ratios were determined by performing a Fisher test (p-value $\leq$ 0.01; fold-change of allelic ratio from parental ratio $\geq$ 1.5). To classify parents with differential levels of gene expression and histone modifications, the fold-change must be 1.5 or greater. In analysing the change at each of the L*er* and C24 alleles in hybrids, the fold-change from the expected value (EV) must be equal to or greater than 1.5 to determine changes at alleles.

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

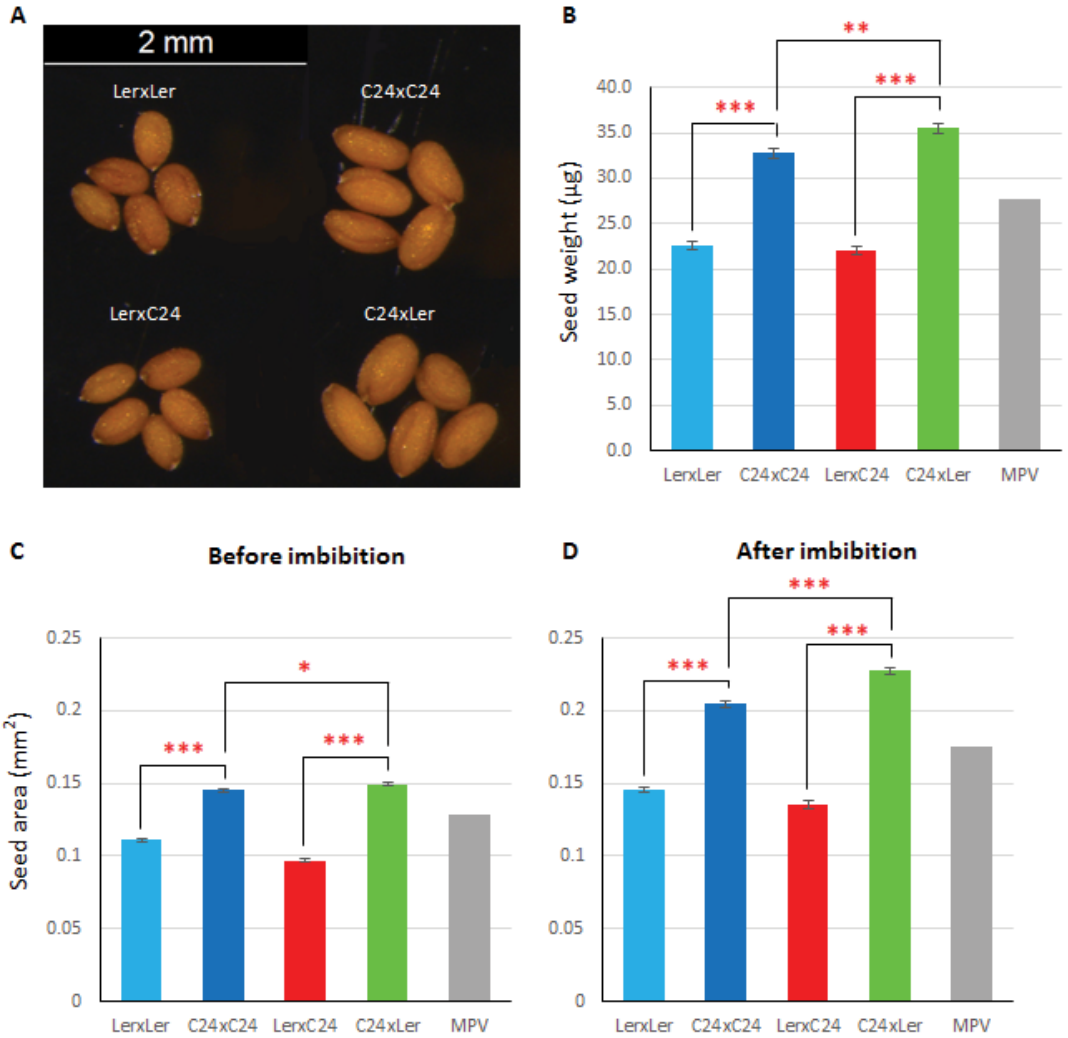
4.1 Increased seed size and weight in C24 x L*er* but not in L*er* x C24

Although increased plant size has been shown in L*er*/C24 hybrids during vegetative growth of the plants (Figure 1.1), it is necessary to know if hybrids show heterosis at the early stages of seedling development. Hybrid seeds show a maternal effect in seed weight (Figure 4.1 B). L*er* x L*er* seeds are significantly lower in weight than C24 x C24 seeds. In hybrids, C24 x L*er* has significantly heavier seeds than L*er* x C24. L*er* x C24 does not show heterosis in seed weight and has -20.2% relative heterosis level (RHL; compared to MPV). By contrast, C24 x L*er* has 28.22% RHL and is significantly heavier than the heavier parent, C24 x C24. This result suggests that, in contrast to L*er* x C24, C24 x L*er* show seed weight heterosis at the mature seed stage.

To determine if hybrids show heterosis in seed size, seed area was measured in hybrid and parental seeds. During plant growth experiments, a three-day imbibition treatment under dark conditions is included before the seeds are exposed to lights to reduce the variation in seed germination time. Seed area was measured both before and after imbibition due to the seed size being dramatically changed by absorption of water.

Like the seed weight results, L*er* x L*er* seeds are significantly smaller than C24 xC24, while L*er* x C24 has significantly smaller seeds than C24 x L*er* both before and after imbibition (Figure 4.1 A, C and D). Compared to the parents, L*er* x C24 shows negative

RHL (BI: -24.22%; AI: -22.86%), whereas C24 x L*er* shows positive RHL before imbibition (17.19%) and even higher RHL after imbibition (30.29%). C24 x L*er* is also significantly larger than the larger parent C24 x C24. This indicates that C24 x L*er* already shows seed size heterosis and maybe have better water-absorbing ability than the other three plant lines.



Relative heterosis level (%)	Seed weight	Seed area BI	Seed area AI
LerxC24	-20.20	-24.22	-22.86
C24xLer	28.22	17.19	30.29

Figure 4.1. Ler/C24 reciprocal hybrids showing significant maternally influenced seed weight and seed area. (A) Photograph shows seed phenotypes of parent and hybrid seeds. **(B)** Seed weight measurements on the parents and the hybrids. Projected area was measured on the seeds before **(C)** and after imbibition **(D)**. **(E)** Relative heterosis levels against MPV in the reciprocal hybrids. P-values were calculated using the Student's t-test. Error bars represent standard error of mean (SEM).

The two reciprocal hybrids have the same genetic background derived from the parents, but they differ significantly in seed weight and seed size. This indicates that, at seed germination stage, the phenotypes of hybrids are influenced by the phenotypes of their maternal parent. The large seed size of C24 x L*er* may be the consequence of the combination of both the maternal influence and a specific heterotic character in the C24 x L*er* hybrid regulating seed size.

4.2 Hybrids show heterosis in germination time

Germination time has not been recorded in L*er*/C24 hybrids, so a germination assay was applied to determine if hybrid seeds germinate at a different time from parental seeds. In Arabidopsis, the time required for germination is determined by physical factors, such as light, water and temperature, and biological factors, such as seed coat hardness, length of seed dormancy, hormone levels and expression levels of genes involved in regulating hormone levels (Bentsink and Koornneef, 2008). After a three-day imbibition treatment for both parent and hybrid seeds, the proportion of germinated seeds was recorded every two hours from 15 hours after sowing (HAS) for 52 hours. Germination time of each plant line was scored based on the time point when 50% of seeds were germinated.

C24 x C24 germinated four hours later than L*er* x L*er*. The average germination time of the two parental lines is about 47 HAS (Figure 4.2). The hybrids, on the other hand, are close to each other in germination time, with 38 and 39 HAS for L*er* x C24 and C24 x L*er*, respectively. In contrast to seed weight and seed size, the germination time of hybrid seeds is not influenced by maternal factors, such as seed coat hardness. Earliness in

germination time must be caused by the interactions between the two parental genomes or epigenomes in both reciprocal hybrids.

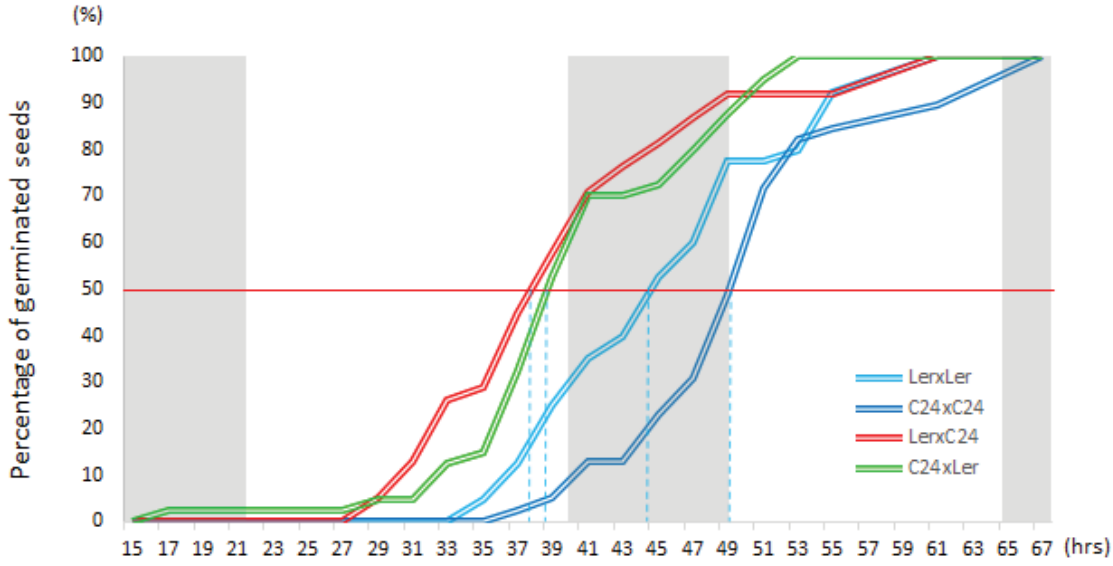


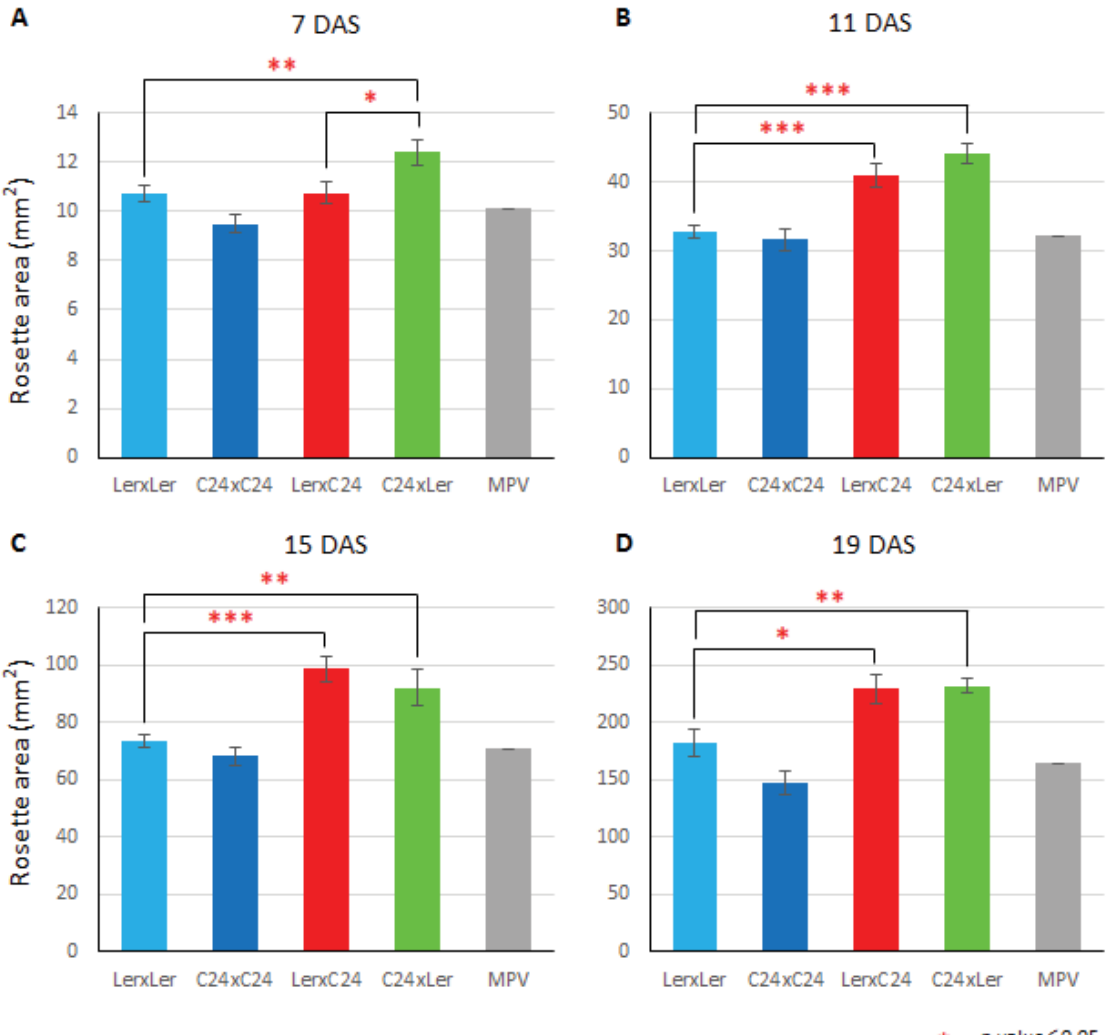
Figure 4.2. Seed germination time for the L*er* and C24 and their hybrids. Percentages of germinated seeds were tracked between 15 to 67 hours after sowing. **Grey blocks** show the dark periods of each day. The crossed points between each plant line and the **red horizontal line** (50% of seeds germinated) represent seed germination times.

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

The seed weight and seed size results suggested that the seed of L*er* x C24 does not show heterosis. However, L*er* x C24 does show plant size heterosis equal to that of C24 x L*er* at 15 DAS. To determine when hybrid seedlings start to show heterosis, rosette area was tracked in parent and hybrid seedlings until 21 DAS (L*er* x L*er* enters reproductive phase at around 21 DAS).

The rosette of L*er* x L*er* is slightly larger than that of C24 x C24 through the entire vegetative phase (Figure 4.3). In hybrids, C24 x L*er* already shows significant heterosis in rosette area with 15.74% RHL at 7 DAS, and remains at 30% - 40% RHL during 11 –

19 DAS. L*er* x C24, on the other hand, is significantly smaller than C24 x L*er* at 7 DAS, and shows 6.45% RHL compared to the average of the two parents. In the second week after sowing, L*er* x C24 shows increasing heterosis in plant size (11 DAS: 27.02% RHL; 15 DAS: 39.53% RHL), and equals C24 x L*er* in plant size at 15 DAS. After 15 DAS, the two hybrids show similar heterosis in plant size with approximately 40% RHL.



E

Relative heterosis level (%)	7 DAS	11 DAS	15 DAS	19 DAS
LerxC24	6.45	27.02	39.53	39.33
C24xLer	22.67	37.04	30.15	40.83

Figure 4.3. Ler/C24 hybrids showing heterosis in plant size at early vegetative developmental stages. (A) – D) Rosette area was measured at 7, 11, 15 and 19 DAS**. (E)** Relative heterosis levels against MPV in the reciprocal hybrids at each time point. P-values were calculated using the Student's t-test. Error bars represent standard error of mean (SEM). (Data obtained from Pei-Chuan Liu, unpublished)

The two reciprocal hybrids differ in seedling size at 7 DAS, and this difference could be influenced by the difference in seed size. Later, the hybrids overcome the maternal influence on seed size and become similar to each other in seedling size. During the first two weeks, L*er* x C24 shows a higher growth rate than C24 x L*er*, which suggests that the differences may exist in the transcriptome or metabolome between the reciprocal hybrids at early seedling stages. In addition, although L*er* x C24 is smaller than C24 x L*er* in the initial seed size, both reciprocal hybrids show similarly significant biomass heterosis in later developmental stages, implying that both reciprocal L*er*/C24 hybrids show growth rate heterosis during the early development stages following seed germination, which is consistent with previous observations in Col/C24 hybrids (Meyer et al., 2012, Fujimoto et al., 2012).

4.4 Summary

- Phenotypes of hybrid seeds reflect maternal phenotypes, including seed weight and seed size.
- Only one of the two reciprocal hybrids, C24 x L*er*, shows heterosis in seed weight and seed size.
- Both reciprocal hybrids germinate similarly earlier than the parents.
- Hybrid L*er* x C24 grows faster than C24 x L*er* at early stages until L*er* x C24 equals C24 x L*er* at about 15 DAS.
- Heterosis in plant size occurs at early developmental stages.

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

The morphological results suggest both reciprocal hybrids have heterosis in plant size at early development stages. One of the reciprocals, L*er* x C24, shows greater vigour in growth rate than the other hybrid. It is likely that the molecular basis of heterosis could be established during that short period of time. We chose four time points, 0, 3, 5 and 7 DAS to perform deep sequencing on both parent and hybrid seedling mRNA, to identify differences between parent and hybrid transcriptomes at young seedling stages.

5.1 Differentially expressed genes are identified in both reciprocal hybrids

After the quality control, statistical analyses were performed on the expression levels of all the genes in the Arabidopsis genome in order to identify genes with changed expression levels between the two parents and between the parents and the hybrids. Between the parents, there are genes identified with differential expression levels (Figure 5.1 A). At 0 DAS, the number of genes with a higher expression level in C24 is greater than those with a higher expression level in L*er*. The opposite trend was observed at 3 and 5 DAS, where L*er* has higher levels of gene expression in the majority of the genes differentially expressed between the parents. At 7 DAS, the L*er*-high and C24-high genes are equal in number. In addition, only a small proportion of the differentially expressed genes is common to the four time points (Figure 5.1 B). This suggests a dynamic pattern of the differences between L*er* and C24 ecotypes in gene expression over the period of early developmental stages.

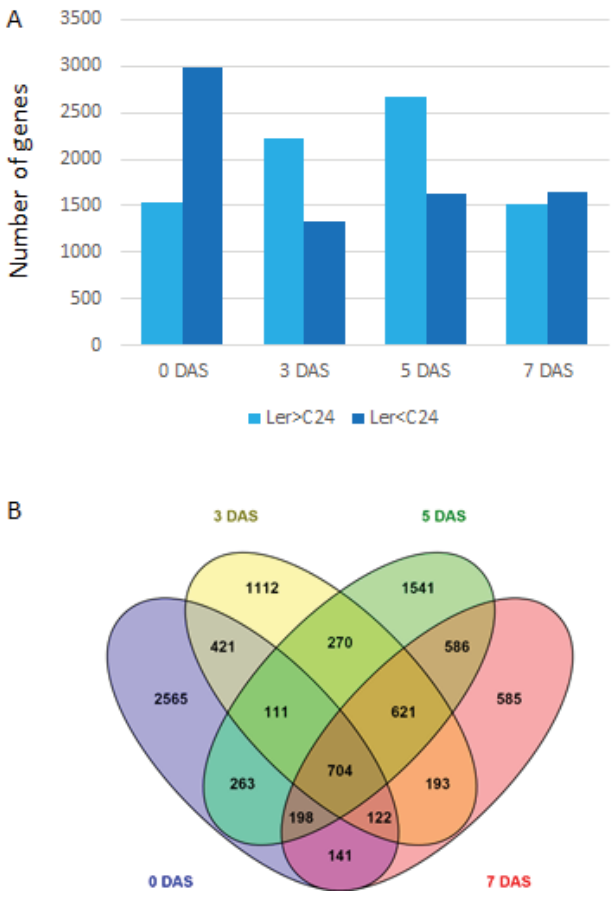


Figure 5.1. Ler and C24 showing dynamic differences in gene expression over time. (A) The numbers of differentially expressed genes between Ler and C24 at the four time points. Fold-change ≥ 1.5 and p-value ≤ 0.01. **(B)** venn-diagram showing common differentially expressed genes between parents to each time point.

To define genes with non-additive expression, the fold-change of expression levels in hybrids against MPV should equal or be above 1.3, and the p-value should equal or be below 0.01. With these criteria, non-additive genes were identified in both hybrids at each time point (Figure 5.2 A). The numbers of non-additive genes identified in the reciprocal hybrids (2,824 in L*er* x C24 and 1,516 in C24 x L*er*) at 3 DAS are higher than at other time points, suggesting there are more changes in the hybrid transcriptomes at this early time point. L*er* x C24 has more non-additive genes than the reciprocal hybrid (nearly 2 fold) at 0 DAS and 3 DAS, but not at 5 DAS and 7 DAS. This means, at these two early time points, compared to C24 x L*er*, L*er* x C24 is more different from the parents in gene

expression pattern. This result is consistent with the observation that L*er* x C24 grows faster than C24 x L*er* at early developmental stages. This indicates the differences in growth rate between parents and hybrids and between reciprocal hybrids may be caused by differences in gene expression between the four plant lines at 0 and/or 3 DAS.

Because the hybrids have a faster growth rate than the parents, it is likely that the expression levels of associated genes in the hybrids are higher or lower than those in either parent (BPG, better-than-best-parent gene). With these criteria, the expression levels of 2,179 (77.2%) and 897 (59.2%) BPGs were found in L*er* x C24 and C24 x L*er* at 3 DAS, respectively (Figure 5.2 A). However, only 20% to 45% of genes are BPGs in hybrids at other time points. These results suggest that the hybrids at 3 DAS not only have the most differentially expressed genes, but also that the majority of those genes have higher or lower expression levels than either parents.

In the 3 DAS non-additive genes, nearly half are up-regulated in both hybrids, and the other half are down-regulated. However, there is a bias between the numbers of up- and down-regulated genes at the other time points. There are slightly more up-regulated non-additive genes than down-regulated at 0 DAS, while non-additive genes tend to be down-regulated at 5 and 7 DAS (Figure 5.2 B). Assuming the existence of positive and negative regulators of heterosis, the increased activity of the positive regulators may play the major role in heterosis at the germination stage, whereas the decreased activity of negative regulators could be essential for heterosis at 5 and 7 DAS.

Because both reciprocal hybrids show strong heterosis at early stages, the hybrids probably employ similar mechanisms to establish heterosis. Therefore, non-additive

genes common to both hybrids are likely to be involved in heterosis. The non-additive genes are reciprocal-specific at 0 DAS, while the hybrids share more non-additive genes from 3 to 7 DAS (Figure 5.1 C). This indicates that the two hybrids do not undergo the same transcriptomic changes in seeds, but become similar to each other after germination.

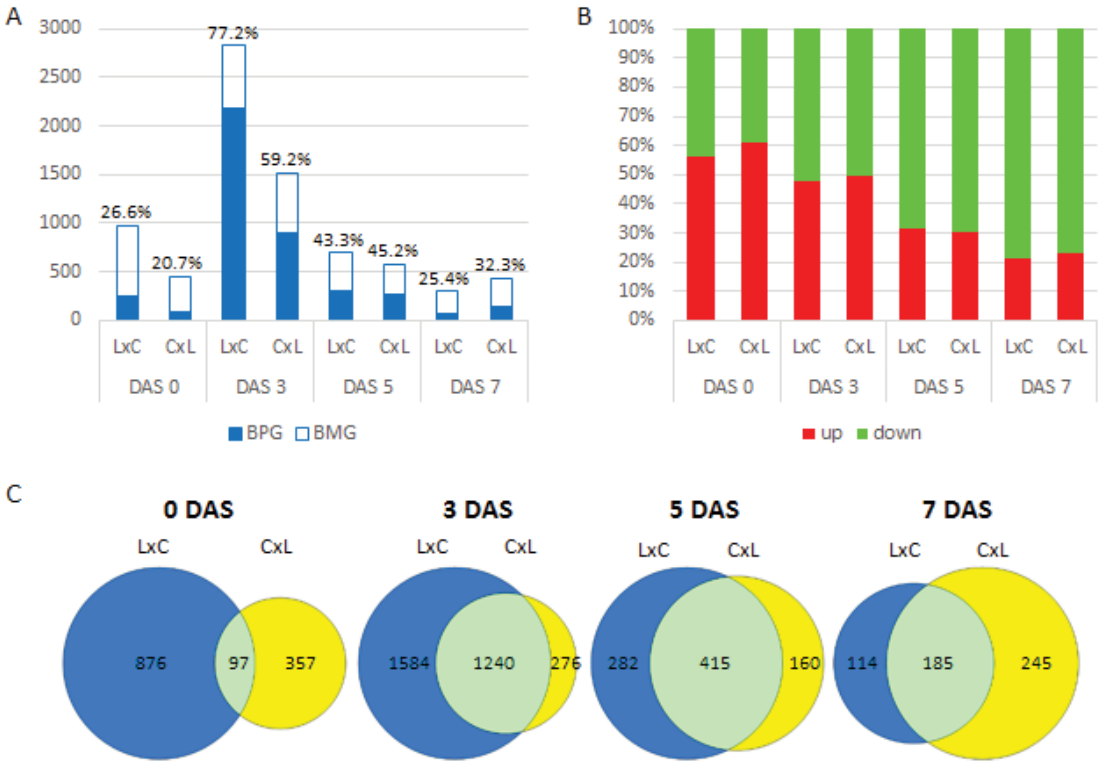


Figure 5.2. Ler/C24 hybrids showing various patterns at different time points in young seedlings. **(A)** The numbers of differentially expressed genes in reciprocal hybrids. **BPG**, better-than-best-parent gene (p-value ≤ 0.01). **BMG**, better-than-MPV-gene (fold-change from MPV ≥ 1.5 and p-value ≤ 0.01). **(B)** Distributions of up- and down-regulated genes in reciprocal hybrids. **(C)** Venn diagrams show the numbers of genes common to reciprocal hybrids.

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

To understand the expression patterns of non-additive genes over time, all non-additive genes (4,916 genes) from both reciprocal hybrids and from all four time points are

clustered together based on the similarity of gene expression. The heat map (Figure 5.3 A) shows that the differentially expressed genes from different time points do not overlap to each other, suggesting the non-additive genes are time-point-specific and heterosis is not caused by the same group of genes with continuous non-additive expression during the first week after sowing. An increasing similarity of gene expression patterns in the two reciprocal hybrids is observed at 3, 5 and 7 DAS, as suggested by previous results (Figure 5.2 C).

To visualise the expression levels of non-additive genes in both parents and hybrids at different time points, another heat map was prepared with the same order of genes (Figure 5.3 B). Almost all 4,916 differentially expressed genes have dramatically changed activity from 0 DAS to 7 DAS in all the four plant lines. 3 DAS seems to be a time point between two distinct phases of gene expression. Before 3 DAS, hybrids are different to each other due to strong maternal effects on gene expression in hybrids (Figure 5.3 B), which is expected in seeds and is consistent with maternally influenced seed size (Figure 4.1). Non-additive genes are expressed at similar levels in the two reciprocal hybrids at 5 and 7 DAS. Although both the parents and the hybrids have changed gene expression levels during the first week after sowing, the activity of these genes changes earlier in the hybrids than that in the parents; these early changes are transient and were not observed at the following time points. For example, the up-regulated genes in the hybrids at 3 DAS (Figure 5.3 B yellow box) have higher levels than that in the parents, but the expression levels of the same genes in parents are as high as those in hybrids at 5 DAS. This means that these genes changed earlier in activity in hybrids. Similar patterns occur in the down-regulated genes at 3 DAS (Figure 5.3 B blue box). All these indicate that transient non-additive expression in hybrids is caused by the transient and early changes in gene activity

in hybrids. The two reciprocal hybrids differ slightly in degree of the earliness. L*er* x C24 precedes C24 x L*er* in changing gene activity at 3 DAS, which could explain why C24 x L*er* has fewer non-additive genes than L*er* x C24 and why L*er* x C24 grows faster than C24 x L*er*.

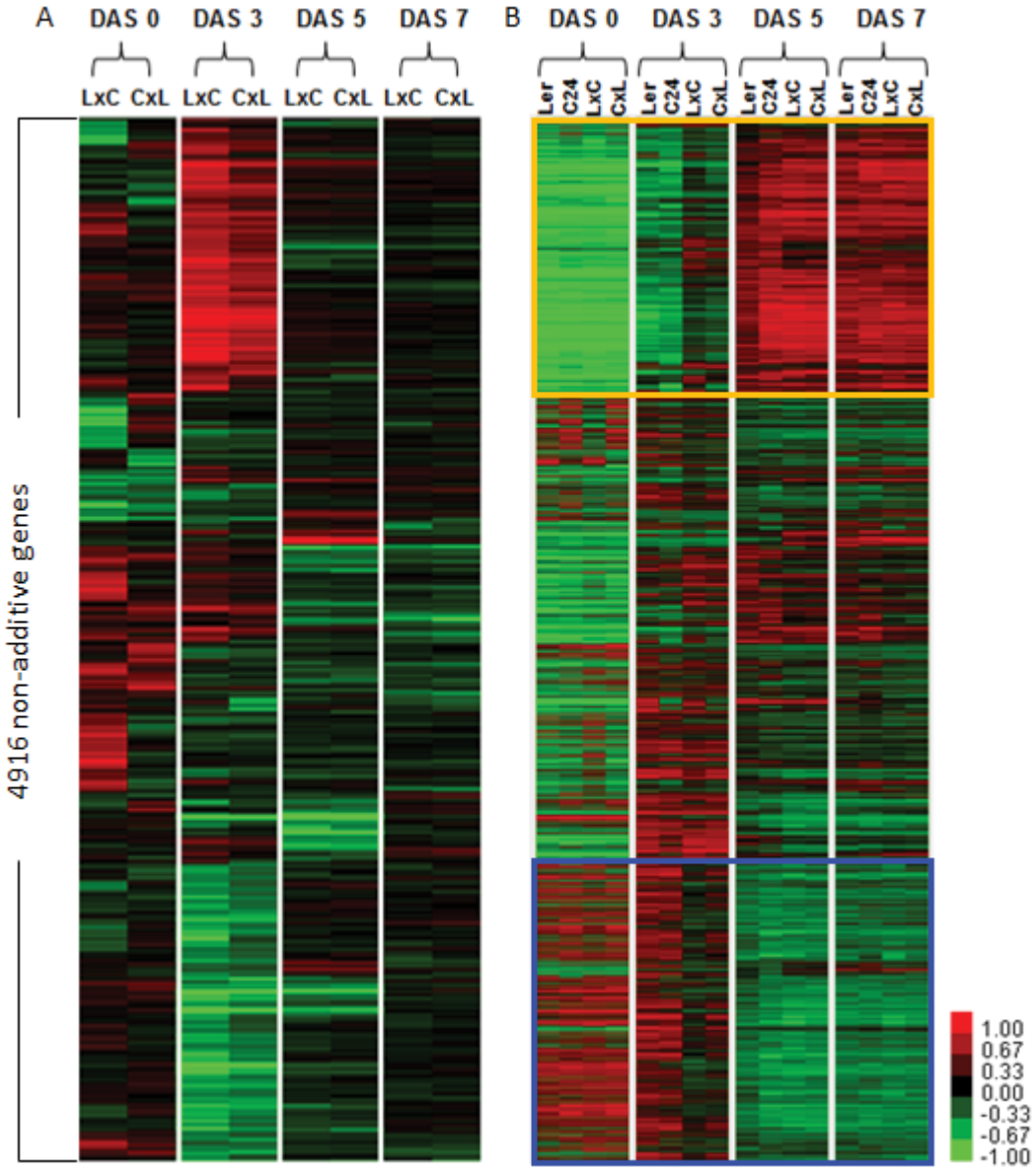


Figure 5.3. Transient and early changes in gene activity in the reciprocal hybrids at young seedling stages. (A) Heat map shows fold-change levels against MPV for all the non-additive genes at the four time points. Genes are clustered based on the similarity of expression patterns. (Log2, MPV = 0) **(B)** Heat map with the same order of gene to **(A)** shows the relative expression levels of non-additive genes against the average level of each gene across parents and hybrids and the four time points. (Log2, average level = 0)

To determine how many genes in hybrids have altered gene expression earlier than their parents, the non-additive genes were divided into five groups based on the gene expression patterns of the hybrids and the gene expression patterns of the parents at following time point (Figure 5.4 A). Because of the lack of transcriptome data after 7 DAS, only the data from the first three time points could be analysed. At 0 DAS, there are relative more Type E genes than other time points, and more Type B and Type D genes than Type A and Type C genes, respectively (Figure 5.4 B). This can be explained by the differences between transcriptomes of the reciprocal hybrids at the seed stage. Nearly 90% and 60% of non-additive genes have a Type A or Type B pattern at 3 DAS and 5 DAS, respectively. This suggests that transient and early changes in gene activity occur widely in hybrids at young seedling stages. For example, a Type A gene, *PHOTOSYSTEM I SUBUNIT O* (*PSAO*), is transiently up-regulated in the two hybrids at 3 DAS, but is additively expressed at 5 and 7 DAS. Another Type A gene, *OLEOSIN 1*, tend to decrease in gene expression over time, and at 3 DAS, the expression levels are lower in the hybrids than in the parents.

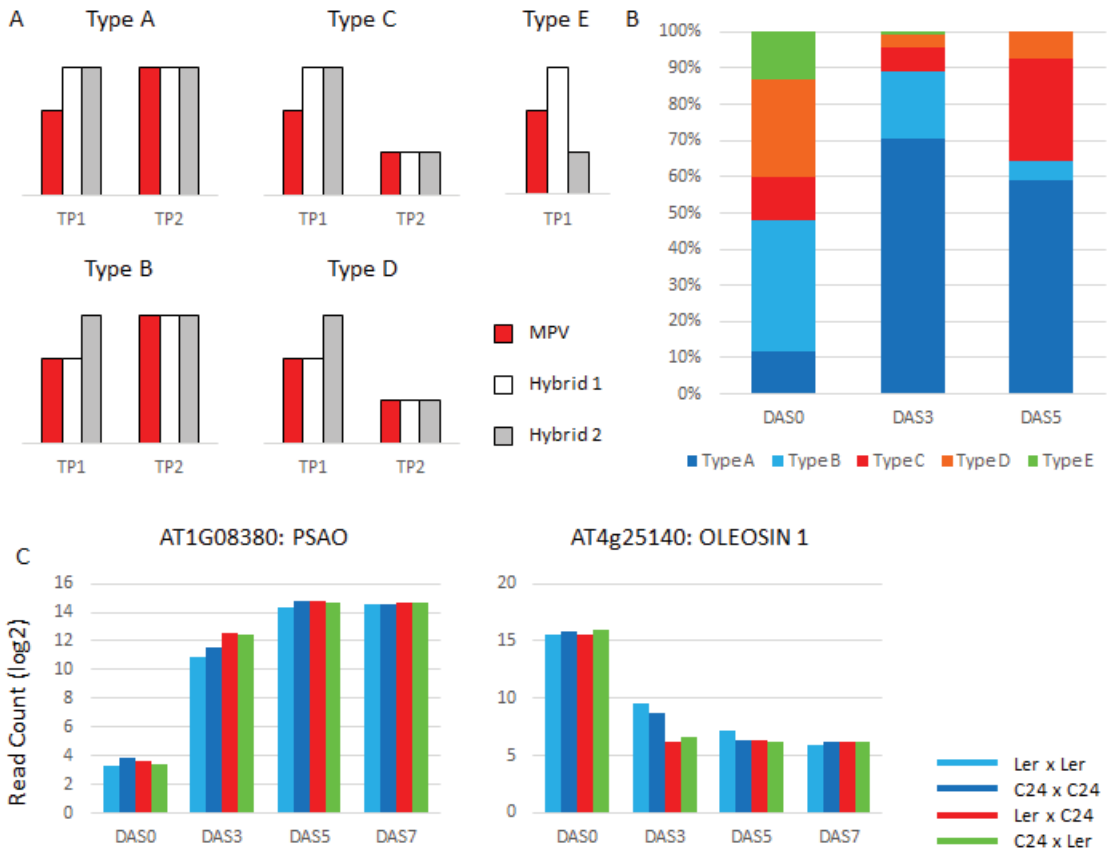


Figure 5.4. The distributions of the five types of non-additive genes at 0, 3 and 5 DAS. (A) Models of the five gene types. Type A: both reciprocal hybrids show differential expression at TP1 (first time point); MPV is changed over time at TP2 (second time point; fold-change from previous time point ≥ 1.2) with the same direction as differential gene expression at TP1. Type B: only one of the hybrids shows differential expression at TP1; MPV is changed over time (fold-change from previous time point ≥ 1.2) with the same direction as differential gene expression at TP1. Type C: both reciprocal hybrids show differential expression at TP1; MPV remains similar levels or changed with opposite direction from differential gene expression at TP1. Type D: only one of the hybrids shows differential expression at TP1; MPV remains similar levels or changed with opposite direction from differential gene expression at TP1. Type E: both reciprocal hybrids show differential expression but with opposite directions. **(B)** The distributions of each gene type in hybrids at different time points. **(C)** The read counts of two example genes in Ler and C24 and their hybrids at different time points.

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

The non-additively expressed genes in both hybrids can be separated into eight clusters based on the patterns of expression over time. No other expression pattern was specifically found in either of the hybrids, suggesting the two hybrids undergo similar

changes in their transcriptomes at these time points. To understand which pathways the non-additive genes are involved in, Gene Ontology (GO) analyses were applied to the genes in each cluster. Only the significant (False discovery rate ≤ 0.01) and common GO terms between the two hybrids were considered.

Among the GO terms, some are likely to be related to heterosis. The genes in response to high light intensity are up-regulated in hybrid seeds (Figure 5.5). High light can be harmful to plant cells (Baker et al., 2007), and the excess absorbed light energy is normally dissipated by the response systems in chloroplasts (Muller et al., 2001, Baker, 2008). The up-regulation of genes responsive to high light intensity suggests that the hybrids might have a greater tolerance to high light stimulus compared to the parents. However, this advantage does not last long, since similar expression levels of these genes were found in parents after 5 DAS.

GO terms related to energy production are also seen in hybrids at several time points. Many photosynthesis-related pathways are up-regulated in 3 DAS hybrids, including chloroplast relocation, thylakoid membrane, photosystem II assembly, chlorophyll biosynthetic process, carotenoid biosynthetic process and photosynthetic electron transport in photosystem I (Figure 5.5), indicating that the hybrids may develop photosynthetic systems earlier than their parents to produce more energy at young seedling stages.

In addition to producing more energy to achieve greater growth rate, hybrids may invest less energy and resources into pathways less important to seedling growth and development. In Arabidopsis, the main energy source to support seed germination and

early seedling growth is fatty acids, which are synthesised during embryogenesis and stored in cotyledons. After cotyledons turn green, the seedlings are able to produce energy by photosynthesis, and the majority of the stored fatty acids is hydrolysed as early as 3 DAS (Eastmond, 2006). In our results, fatty acid beta-oxidation related and lipid storage related genes are down-regulated in hybrids at 3 and 5 DAS. In addition, expression levels of genes involved in seed germination and seed dormancy are decreased at both 3 and 5 DAS. These results suggest that hybrids may shut down these pathways earlier than the parents do in order to invest more energy to the pathways relevant to seedling growth.

Hormones are key regulators of plant development and growth. We found some hormone-related genes showing non-additive expression at the time points analysed. The biosynthetic processes of salicylic acid (SA) are up-regulated leading to consistent up-regulated downstream targets of SA in hybrids at 3 DAS, whereas SA biosynthesis is down-regulated at 5 and 7 DAS (Figure 5.5). The biosynthesis pathway of another key growth factor, auxin [indole-3-acetic acid (IAA)], is also elevated at 3 DAS, with up-regulated downstream signalling pathways after two days. These results fit the theory that decreased SA and increased auxin levels cause increased cell size and cell number in hybrids (Groszmann et al., 2015). They also suggested that altered levels of SA and auxin could occur early in young seedlings of hybrids. In addition, the expression levels of genes in response to abscisic acid stimulus are decreased at 3 and 5 DAS, while the genes responding to auxin and ethylene are more active in hybrids at 5 DAS.

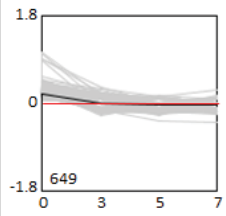
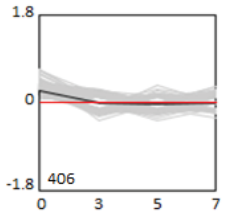
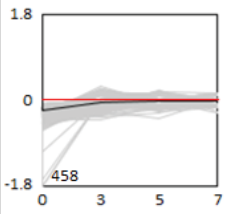
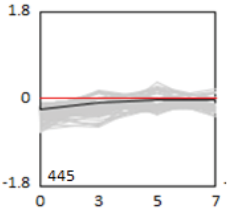
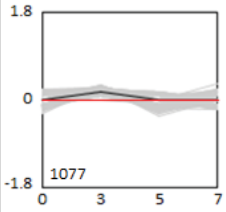
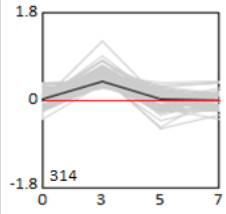
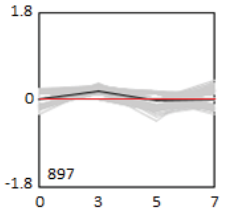
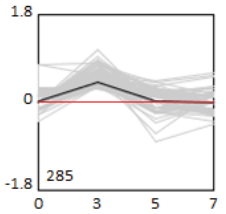
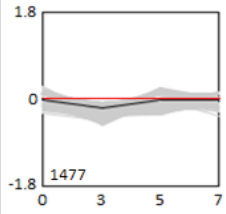
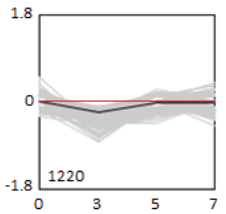
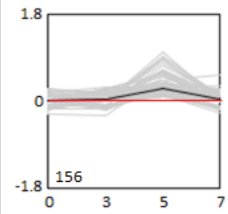
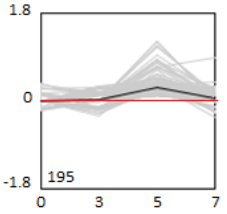
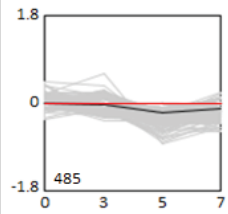
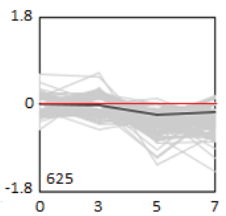
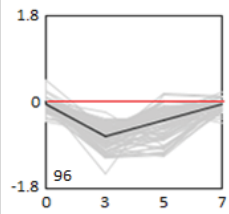
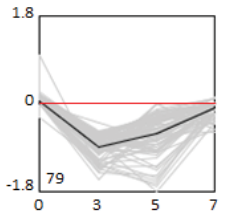
Cluster	Ler x C24	C24 x Ler	Common Significant GO terms in the two reciprocal hybrids
I			response to hydrogen peroxide response to high light intensity
II			RNA elongation
III & IV	 	 	photosystem II assembly pentose-phosphate shunt response to blue, red and far red light glucosinolate biosynthetic process rRNA processing photosynthetic electron transport in photosystem I chlorophyll biosynthetic process carotenoid biosynthetic process thylakoid membrane organization starch biosynthetic process chloroplast relocation unsaturated fatty acid biosynthetic process salicylic acid mediated signalling pathway jasmonic acid mediated signalling pathway indoleacetic acid biosynthetic process negative regulation of defence response response to high light intensity response to gibberellin stimulus salicylic acid biosynthetic process
V			pyrimidine ribonucleotide biosynthetic process seed germination fatty acid beta-oxidation protein targeting to mitochondrion seed dormancy
VI			cellular response to iron ion response to auxin stimulus cytokinin mediated signalling pathway iron ion transport lipid transport response to ethylene stimulus
VII			anthocyanin biosynthetic process response to UV-B response to chitin regulation of plant-type hypersensitive response negative regulation of programmed cell death protein targeting to membrane MAPKKK cascade salicylic acid biosynthetic process
VIII			seed dormancy lipid storage response to abscisic acid stimulus terpenoid biosynthetic process response to freezing response to water deprivation seed germination sugar mediated signaling pathway

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. Grey lines represent log2 fold-change of expression levels from MPV for each single gene in each cluster. Black lines represent the mean levels of fold-change. Red lines represent MPVs.

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

Photosynthesis is one of most important biological processes in photoautotrophic organisms. In photosynthesis, light energy is converted into chemical energy and carbon dioxide from the environment is fixed in the form of plant carbohydrates. Photosynthesis consists of two main phases, the light reactions and the Calvin cycle. The light reactions provide protons and energy to the Calvin cycle for carbon fixation. Because photosynthesis is the main energy provider in plants, it is reasonable to relate increased biomass to increased photosynthesis activity. In hybrids, the transient up-regulation of photosynthesis related genes was observed in C24/Col hybrids at early developmental stages, but those genes are additively expressed at later time points (Fujimoto et al., 2012). We find that the expression levels of photosynthetic genes are increased at 3 DAS in both reciprocal hybrids compared to their parents, but the parents have similar expression levels to the hybrids in the following days.

Pathway analyses suggest that, in 3 DAS hybrids, there are genes with increased expression levels in almost every step in both the light reactions (Figure 5.6) and the Calvin cycle (Figure 5.7), which indicates the transcriptional activities of genes in photosynthesis pathways are up-regulated in hybrids. Similar to other non-additive genes at 3 DAS, the higher expression of those genes usually disappears in hybrids relative to parents at 5 or 7 DAS. Not all the photosynthesis related genes are non-additively

expressed. This could be due to the strict criteria used to define differential regulation, and/or due to these unchanged genes being under the control of different regulators. Almost all the genes from light reactions with unchanged expression levels in hybrids are chloroplast encoded genes, and these genes also do not have a dramatic change in expression across the four time points (Figure 5.8). This suggests that the genes from the nucleus and chloroplast have different expression patterns over time, indicating the existence of two distinct systems regulating the genes involved in the light reactions.

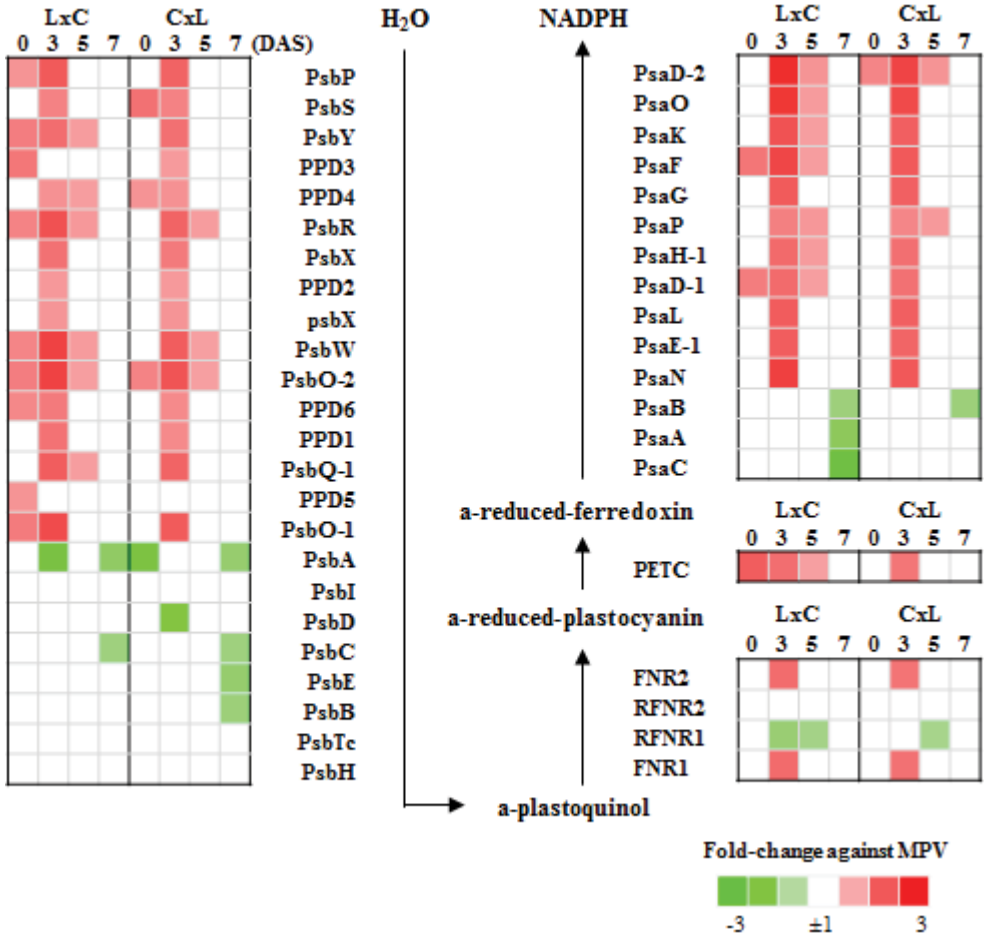


Figure 5.6. Expression fold-changes relative to the MPVs of genes involved in each step of the light reactions of photosynthesis (p-value ≤ 0.05). Numbers on top of matrixes indicate the days after sowing.

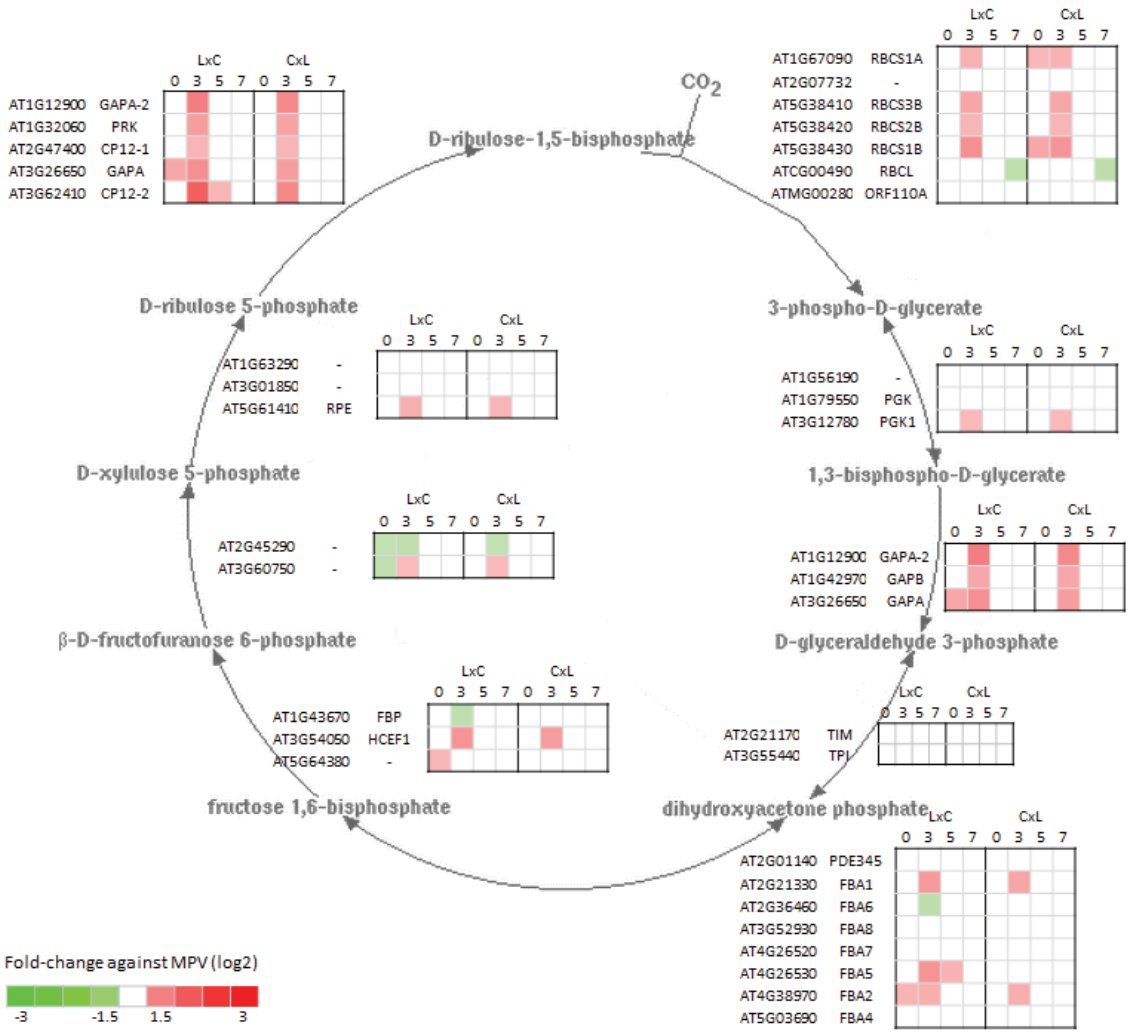


Figure 5.7. Transcriptional activities of the genes in Calvin cycle are transiently up-regulated in both reciprocal hybrids at 3 DAS. (p-value ≤ 0.05)

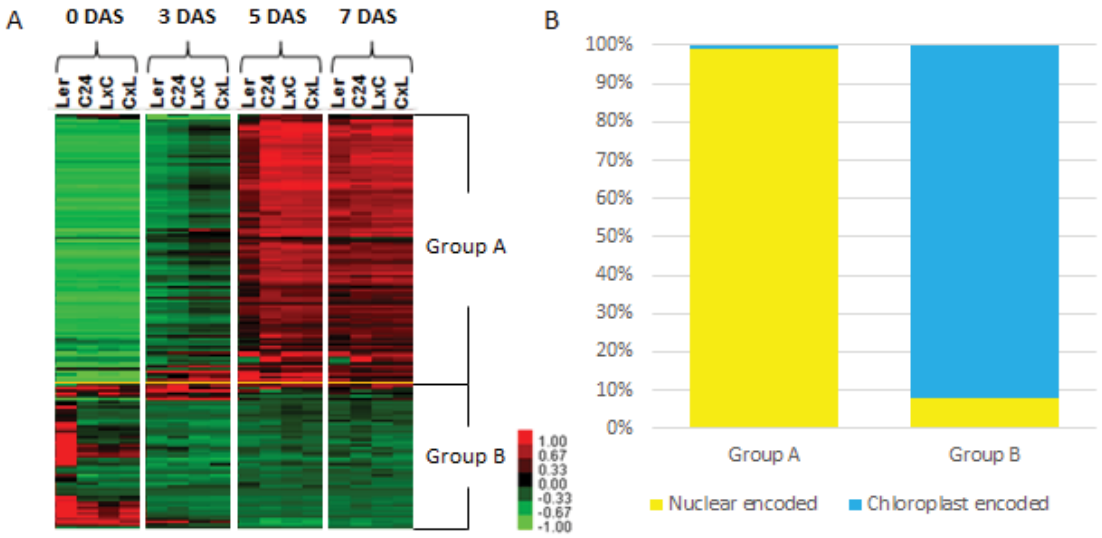


Figure 5.8. Two distinct expression patterns over time exist in the genes involved in the light reactions. (A) Heat map shows the relative expression levels of genes involved in the light reactions against the average level of each gene across parents and hybrids at the four time points. Genes in Group A show up-regulated expression levels over time; Genes in Group B show unchanged levels after 3 DAS. Expression levels are in Log2, and average level = 0 **(B)** The distributions of nucleus encoded genes and chloroplast encoded genes in Group A and Group B.

Besides photosynthesis, there are some other pathways significantly altered in hybrids at 3 DAS (Figure 5.9). The chlorophyll and carotenoid biosynthesis pathways are significantly up-regulated at 3 DAS in both hybrids. This indicates that hybrids may have more chlorophyll and carotenoid, which are important pigment molecules in light harvesting. Because of the up-regulated transcriptional activities of photosynthetic genes, the hybrids may have an enhanced capability for producing energy and carbohydrates for a short period, which leads to the increased biomass.

Triacylglycerol (TAG) is synthesised during the later stages of embryogenesis for energy storage (Mansfield and Briarty, 1992). The activity of the TAG pathway is decreased after seed germination and TAG is not the main energy source in seedlings. Our results show that the TAG biosynthesis pathway is transiently down-regulated in hybrids compared to their parents at 3 DAS (Figure 5.9).

Finally, at 3 and 5 DAS, hybrids have increased expression levels of genes involved in biosynthesis of glucosinolates (Figure 5.9), which are a group of chemicals specific to *Brassicaceae* plants and are related to the defence response to biotic stimulus (Hopkins et al., 2009). This means that the hybrids could have higher levels of glucosinolates, which might lead to greater tolerance against insects and diseases than in the parents at early development stages.

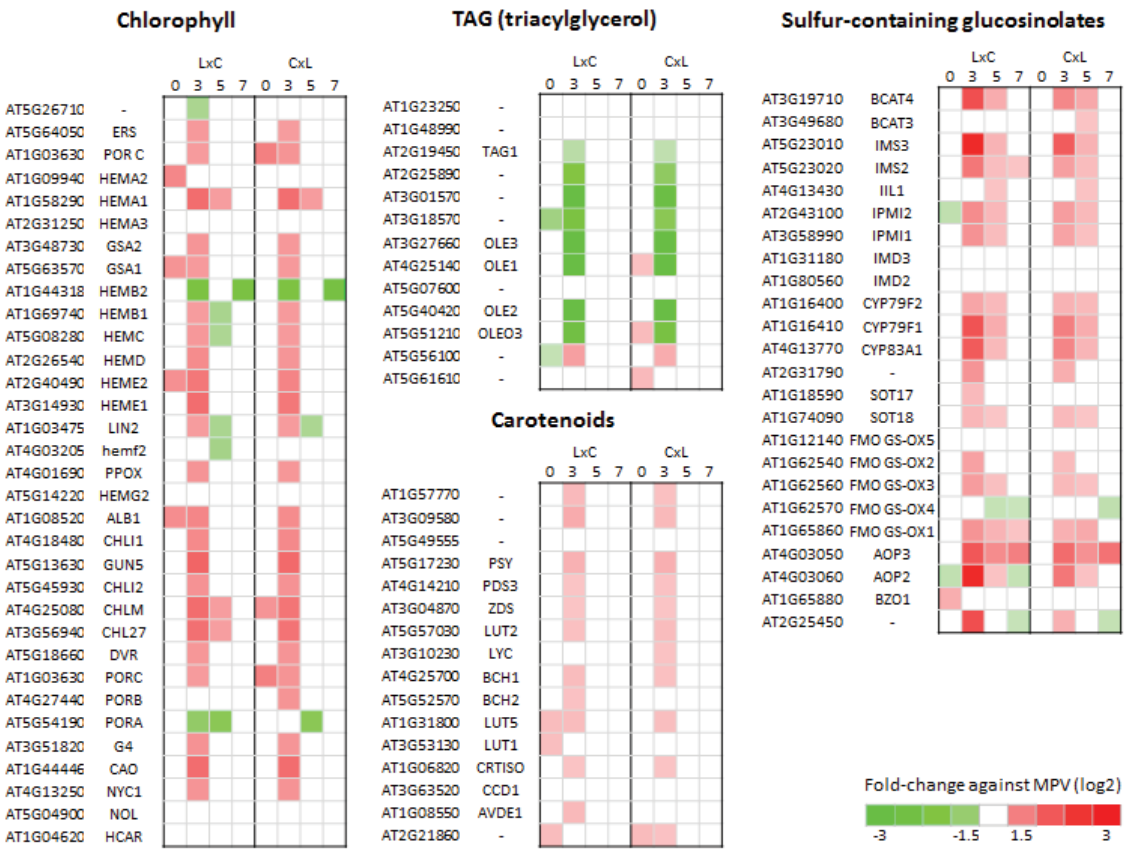


Figure 5.9. The transcriptional activity is transiently changed in young seedlings of both reciprocal hybrids in the biosynthesis pathways of chlorophyll, TAG, carotenoids and sulfur-containing glucosinolates. (p-value ≤ 0.05)

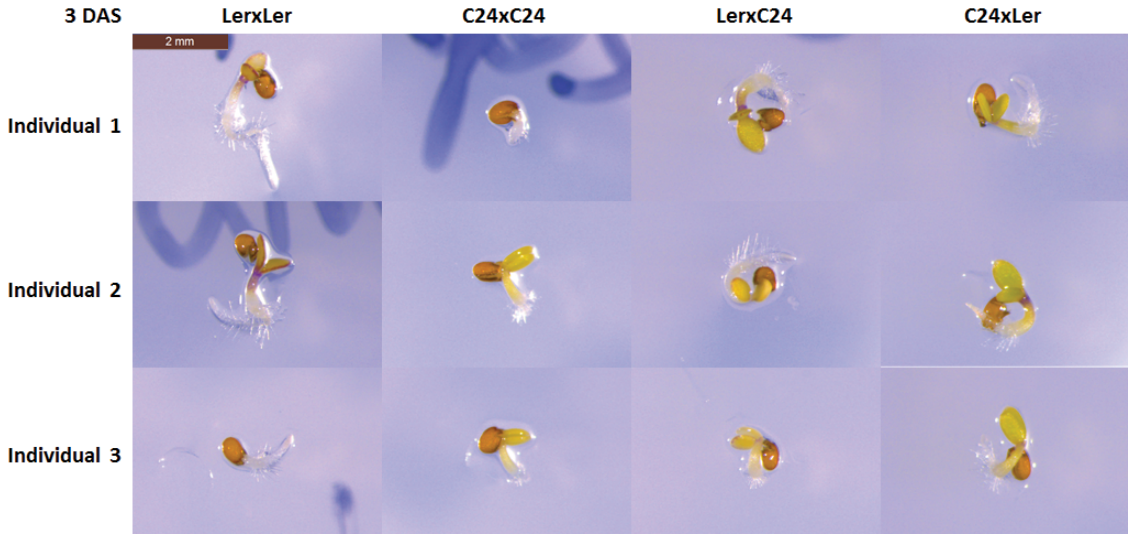
5.5 Validation of non-additive gene expression

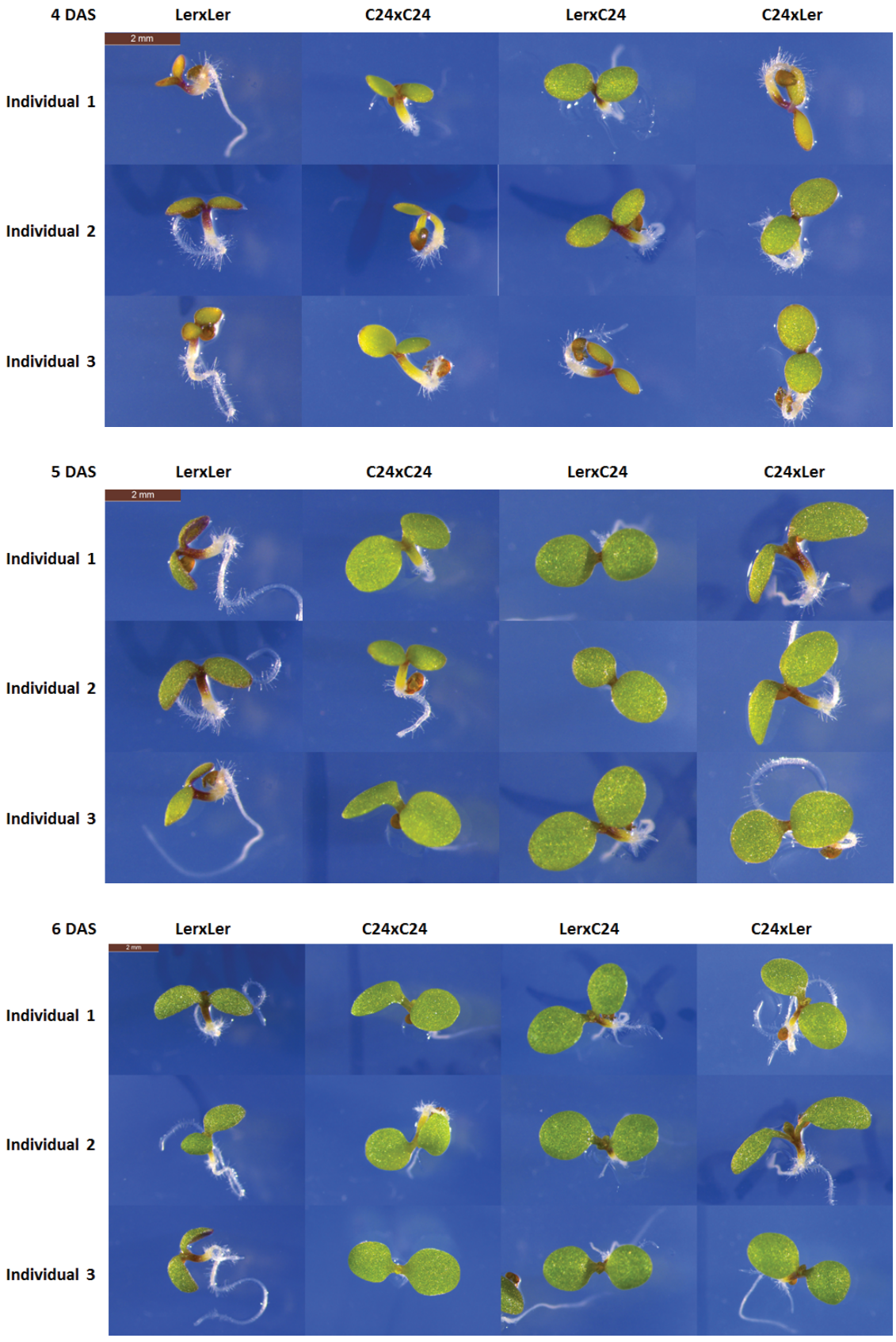
One explanation for the difference in gene activity in 3 DAS hybrids could be that the hybrids and the parents are not at exactly the same developmental stage. During photomorphogenesis of Arabidopsis seedlings, the opening and greening of young cotyledons (around 3 DAS in L*er* and C24 and the hybrids) is a key stage dividing early seedling growth into two distinct phases (Figure 5.3). In the first phase, plants are strongly dependent on stored energy in seeds and are maternally affected. In the second phase, plants develop their own energy production systems for growth and development. During the phase transition, many genes have large changes in activity (Ma et al., 2001), which is confirmed in the transcriptome data (Figure 5.3 B). Differences in developmental stage between hybrids and parents could result in large differences in gene activity, and may explain why there are many non-additive genes in hybrids at 3 DAS.

To exclude the influence of germination time on the transcriptome, the two hybrids were sown 7 hours later than the parents, based on the approximately 7-hour difference in germination time between the hybrids and the average of the parents (Figure 4.2). The hybrids and the parents are at nearly same developmental stage at 3 DAS and for the following four days, except that the hybrids develop the first pair of true leaves one day earlier than the parents (Figure 5.10).

qPCR analyses suggested that the selected nuclear encoded photosynthesis-related genes, *PHOTOSYSTEM II LIGHT HARVESTING COMPLEX GENE 2.3* (*LHCB2.3*), *PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 1* (*LHCA1*), *PHOTOSYSTEM II REACTION CENTER W* (*PSBW*), *PHOTOSYSTEM I SUBUNIT O*

(*PSAO*) and *RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1A* (*RBCS1A*), have up-regulated expression levels at 3 DAS (Figure 5.11), consistent with the transcriptome data. The chloroplast encoded genes, *PHOTOSYSTEM I REACTION CENTER PROTEIN A* (*PSAA*), *PHOTOSYSTEM II REACTION CENTER PROTEIN A* (*PSBA*) and *RUBISCO LARGE SUBUNIT* (*RBCL*), do not show up-regulated gene expression levels compared to the parents at 3 DAS (Figure 5.11), which is also reflected in the transcriptome data. These results indicate that when hybrids and parents are at similar developmental stages, the transient up-regulation of photosynthesis related genes occurs in both hybrids. However, the chlorophyll synthase genes, *G4* and *CHLOROPHYLL A OXYGENASE* (*CAO*), do not show up-regulated expression in the hybrids compared to the parents (Figure 5.11) as the transcriptome data suggest. The reason could be that the timing of transient changes in gene activity in hybrids varies for the genes from different pathways. This phenomenon might be transient, and could be missed in qPCR validation experiments.




4 DAS
LerxLer
C24xC24
LerxC24
C24xLer
2 mm
Individual 1
Individual 2
Individual 3
5 DAS
LerxLer
C24xC24
LerxC24
C24xLer
2 mm
Individual 1
Individual 2
Individual 3
6 DAS
LerxLer
C24xC24
LerxC24
C24xLer
2 mm
Individual 1
Individual 2
Individual 3

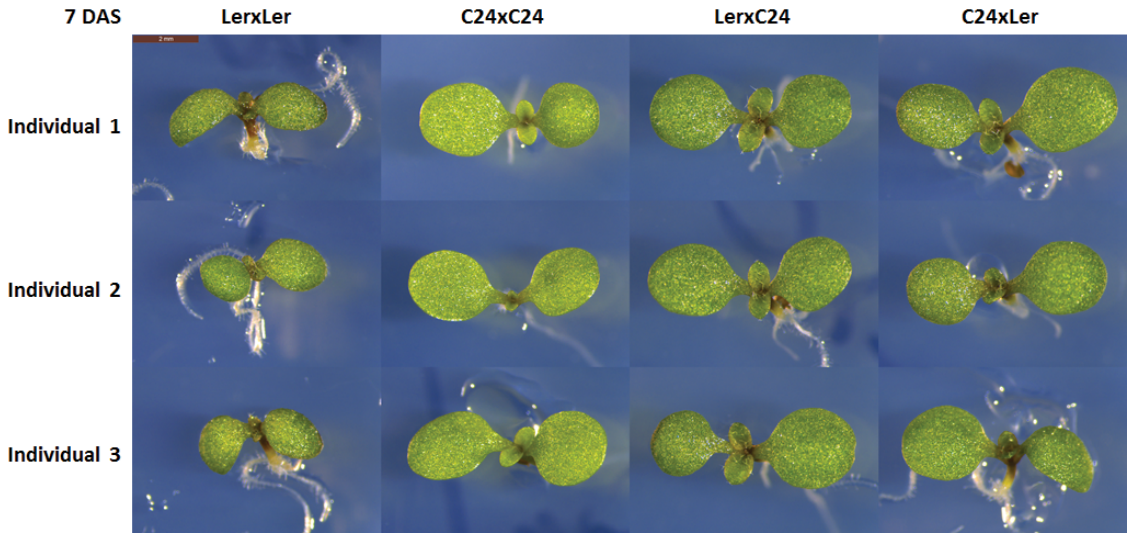
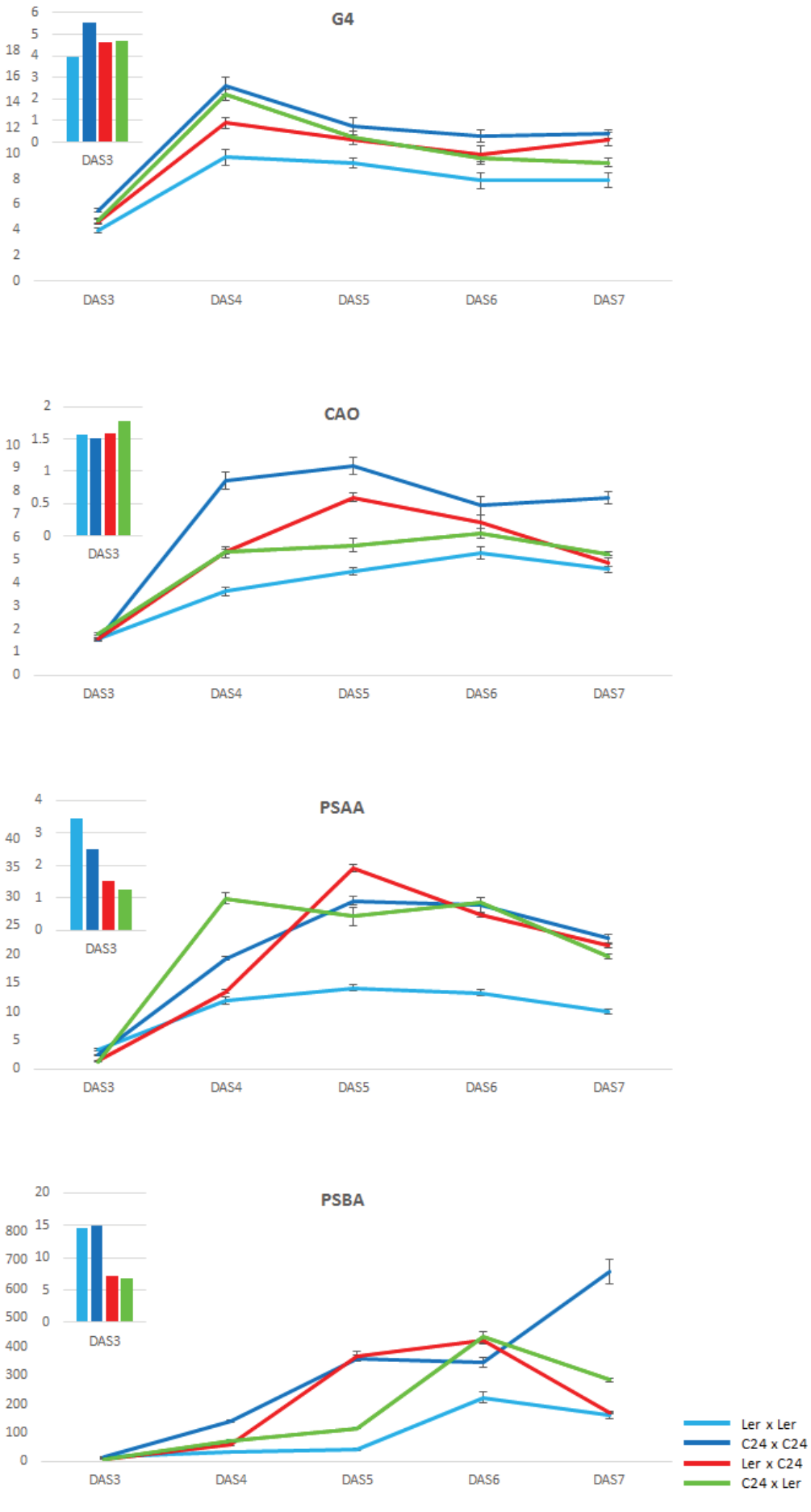
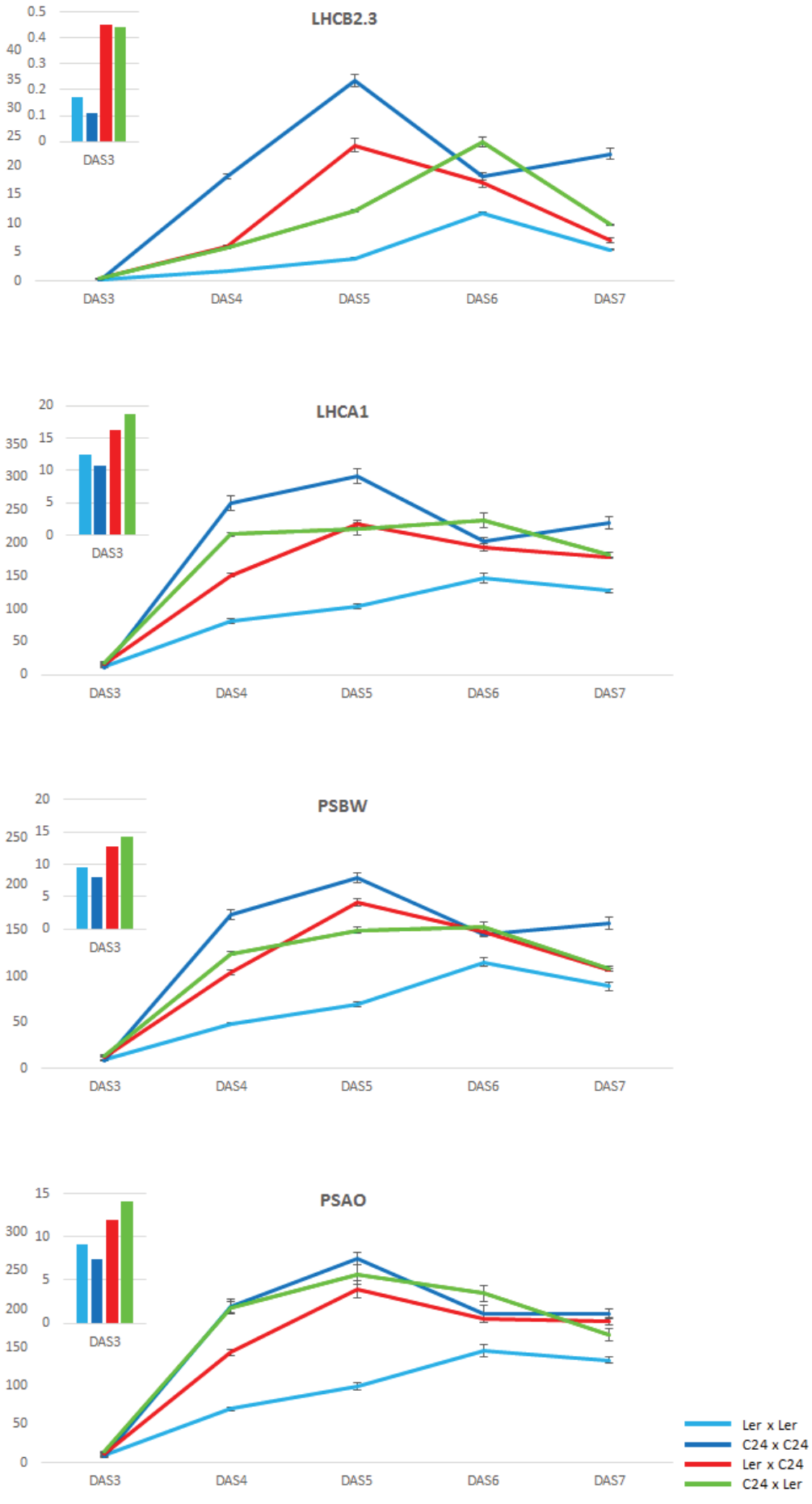


Figure 5.10. Ler/C24 hybrids sown 7 hours later are in similar developmental stages as the parents from 3 to 7 DAS.


G4
DAS3
DAS4
DAS5
DAS6
DAS7
CAO
PSAA
PSBA
Ler x Ler
C24 x C24
Ler x C24
C24 x Ler


LHCB2.3
LHCA1
PSBW
PSAO
DAS3
DAS4
DAS5
DAS6
DAS7
Ler x Ler
C24 x C24
Ler x C24
C24 x Ler

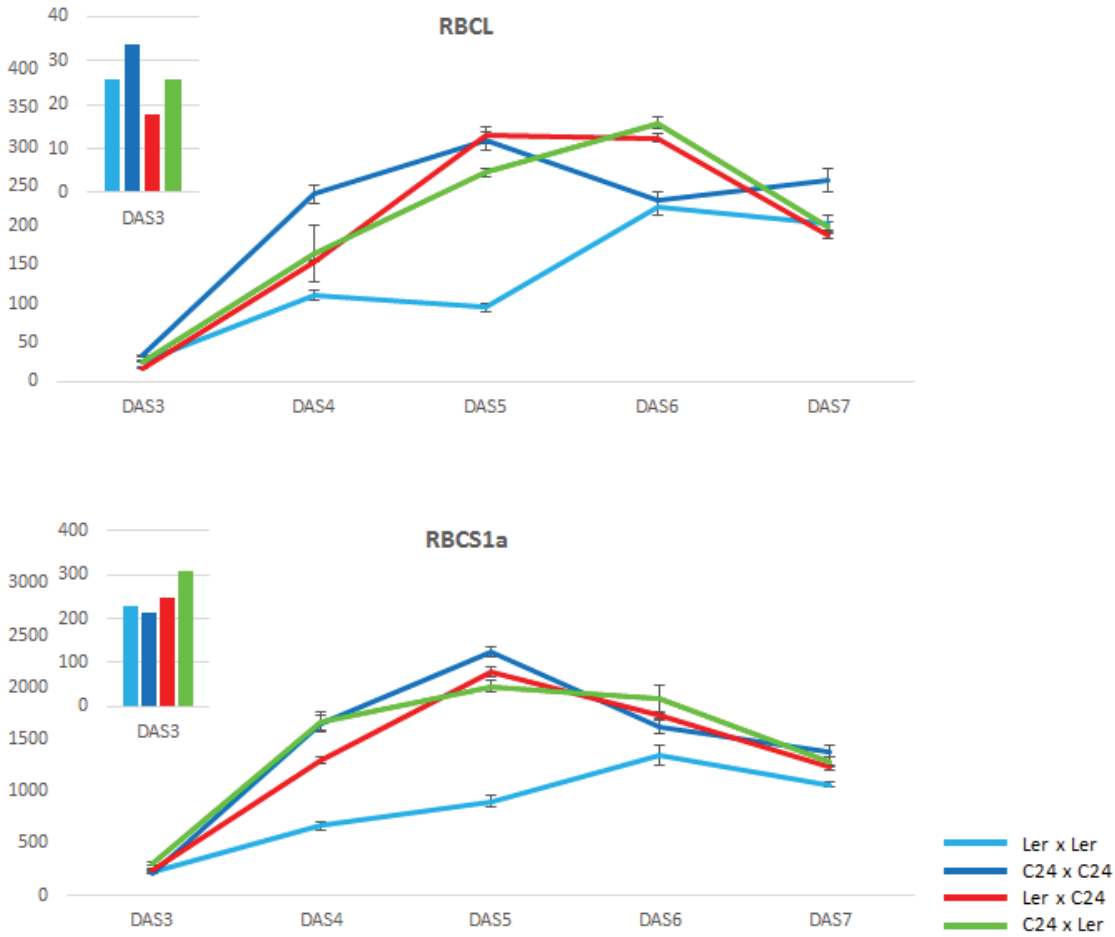
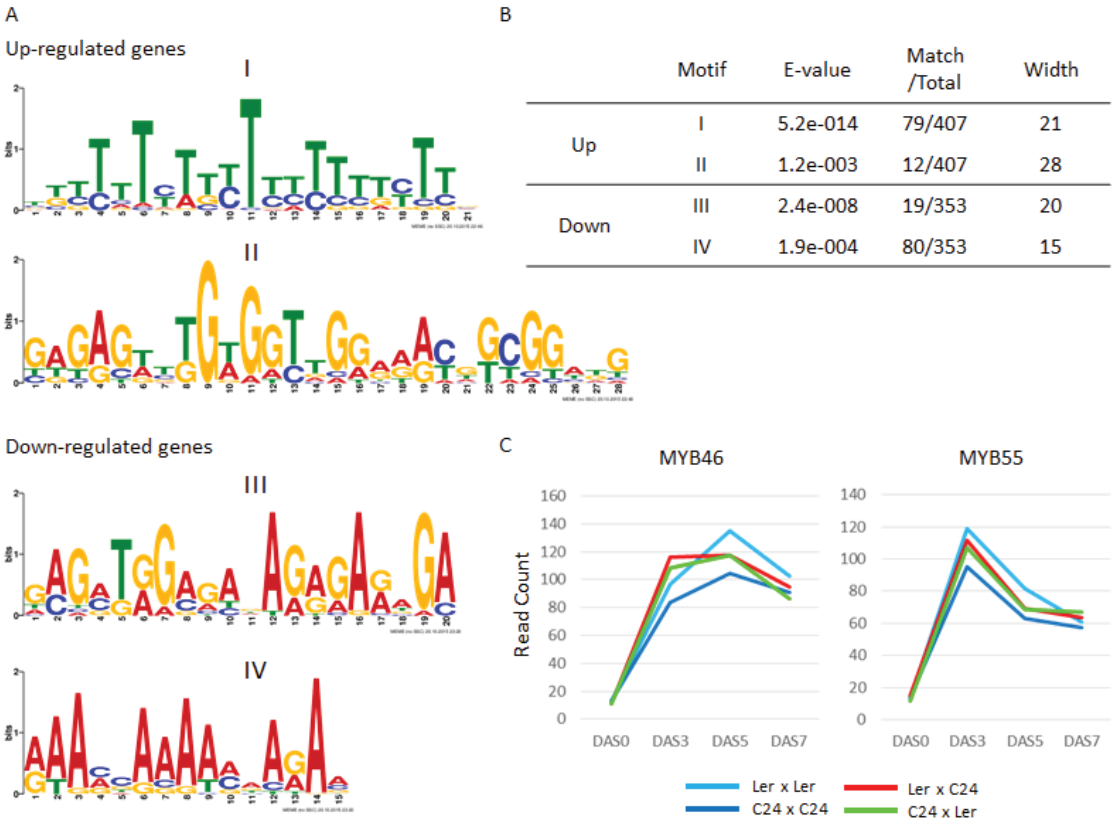


Figure 5.11. qPCR validation for the expression levels of 10 selected genes involved in the light reactions and the chlorophyll biosynthesis in parents and hybrids from 3 to 7 DAS. Two biological replicates were used, and each replicate is a pool of cDNA from ten individual plants. Expression levels are relative to AT4G34270, which has similar expression levels in the parents and hybrids (based on transcriptome data) and is stably expressed over the early stages of seedling development (Czechowski et al., 2005, Dekkers et al., 2012). Error bars represent standard error of mean (SEM).

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

To determine if non-additively expressed genes are under the control of any common transcription factor, the sequences 1kb upstream from the transcription start sites were analysed by using the MEME Suite (Bailey et al., 2009) to discover common motifs that suggest the binding positions of transcription factors. Because the two reciprocal hybrids

may use the same mechanism to achieve heterosis, only the BPGs that are common between the reciprocals are considered in the analyses. No significant common motif was identified in up- or down-regulated genes at 0, 5 and 7 DAS. Four motifs of 15 – 28 bp were significantly identified in up- or down-regulated genes at 3 DAS (Figure 5.12 A). However, except for motif II, they do not match the binding sequences of any known transcription factor. 79 up-regulated genes and 80 down-regulated genes have motif I and IV with the T- or A-rich sequences, respectively (Figure 5.12 A and B), suggesting these genes may be controlled by unknown regulatory factors specific to AT-rich regions. Motif II exists in only 12 of the up-regulated genes, and could be potentially targeted by two transcription factors, MYELOBLASTOSIS 46 (MYB46) and MYELOBLASTOSIS 55 (MYB55). The expression level of *MYB46* is transiently up-regulated in both hybrids at 3 DAS, whereas *MYB55* has an additive expression level (Figure 5.12 C). MYB46 is a transcription factor that was reported to regulate secondary wall biosynthesis, suggesting hybrids may have increased activity of cell wall biosynthesis (Zhong et al., 2007). These results indicate that non-additive gene expression is not simply regulated by one or several transcription factors, instead the cause of non-additive expression could be complex.



	Motif	E-value	Match /Total	Width
Up	I	5.2e-014	79/407	21
	II	1.2e-003	12/407	28
Down	III	2.4e-008	19/353	20
	IV	1.9e-004	80/353	15

Figure 5.12. Motif analyses on the upstream 1 kb promoter regions of non-additive genes. (A) The sequence logos of significant common motifs in up- and down-regulated BPGs common to both reciprocal hybrids at 3 DAS. **(B) E**-value, number of matched genes and sequence width of the motifs. **(C)** The expression patterns of two known transcription factors targeting motifs with sequences similar to motif II.

5.7 Summary

- Dynamic differences in gene expression between Ler and C24 ecotypes were observed over early developmental stages.
- Thousands of genes are non-additively expressed in hybrids at early developmental stages, and the largest number of differential genes is identified in 3-DAS hybrids.
- The transcriptomes of the two reciprocal hybrids are similar to each other from 3 DAS onwards.

- Non-additively expressed genes in hybrids at early development stages are mainly caused by early and transient changes in gene activity in hybrids.
- Genes in pathways, including photosynthesis, important to seedling growth have up-regulated transcriptional levels in hybrids, and could be related to biomass heterosis.
- The transient and early changes in gene activity are not caused by the difference in germination time between parents and hybrids.
- The role of known transcription factors in regulating non-additive genes in hybrids is limited at early developmental stages.

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

Many differentially expressed genes were identified in L*er*/C24 hybrids in young seedlings due to early and transient changes in gene expression. Motif analysis suggested a limited role for known transcription factors in regulating these genes. Since not all the genes in the genomes of hybrids change early in expression, non-additive gene expression in hybrids could be caused by some underlying regulatory factors. Histone modifications might be one of these factors. Hybrids may have histone modification patterns different from the parents at early time points.

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

We performed ChIP on mature seeds of parents and hybrids at 0 DAS after the 3-day imbibition. To verify the performance of antibodies used in ChIP, three genes, *ACTIN 7* (AT5G09810), *AGAMOUS* (AT4G18960) and *TA3* (AT1G37110) were chosen as control genes, as they have been shown to be K4me3- and K9ac-, K27me3- and K9me2-enriched, respectively. As expected, K4me3 and K9ac are highly enriched in *ACTIN 7*; K27me3 is enriched in *AGAMOUS*; K9me2 is enriched in the transposable element gene, *TA3* (Figure 6.1), but none of the genes has histone modification signals in the no antibody (No Ab) control, suggesting that the histone modification specific antibodies work satisfactorily. The quality and fragment length of ChIP-Seq libraries were measured, and high purity and the correct fragment length (approximately 275bp) for all the libraries were observed

(Figure 6.2).

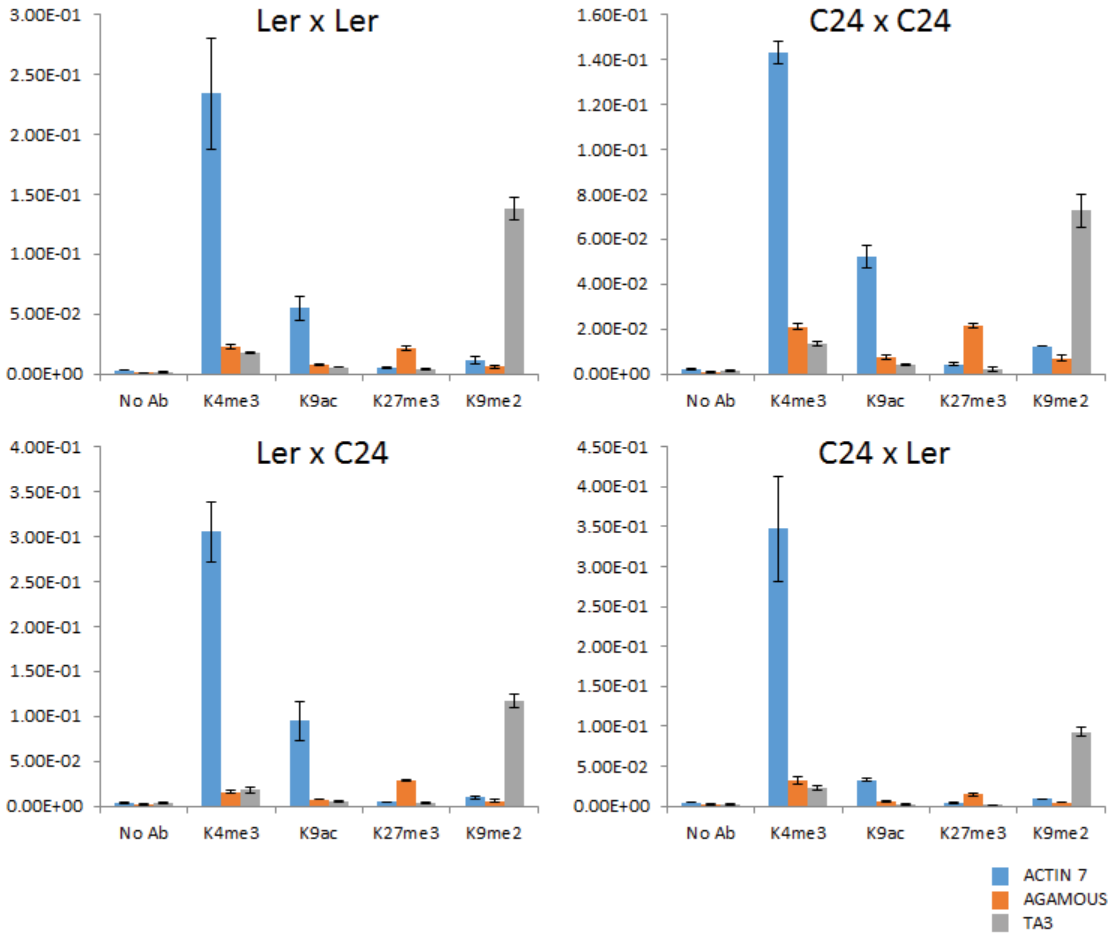


Figure 6.1. qPCR verifications for the first biological replicate of ChIP DNA libraries for deep sequencing. Two biological replicates of ChIP libraries were verified. The second replicate have similar enrichment patterns of the four histone marks in the four plant lines (data not shown). Levels of enrichment are relative to that of the total input sample. Error bars represent standard error of mean (SEM).

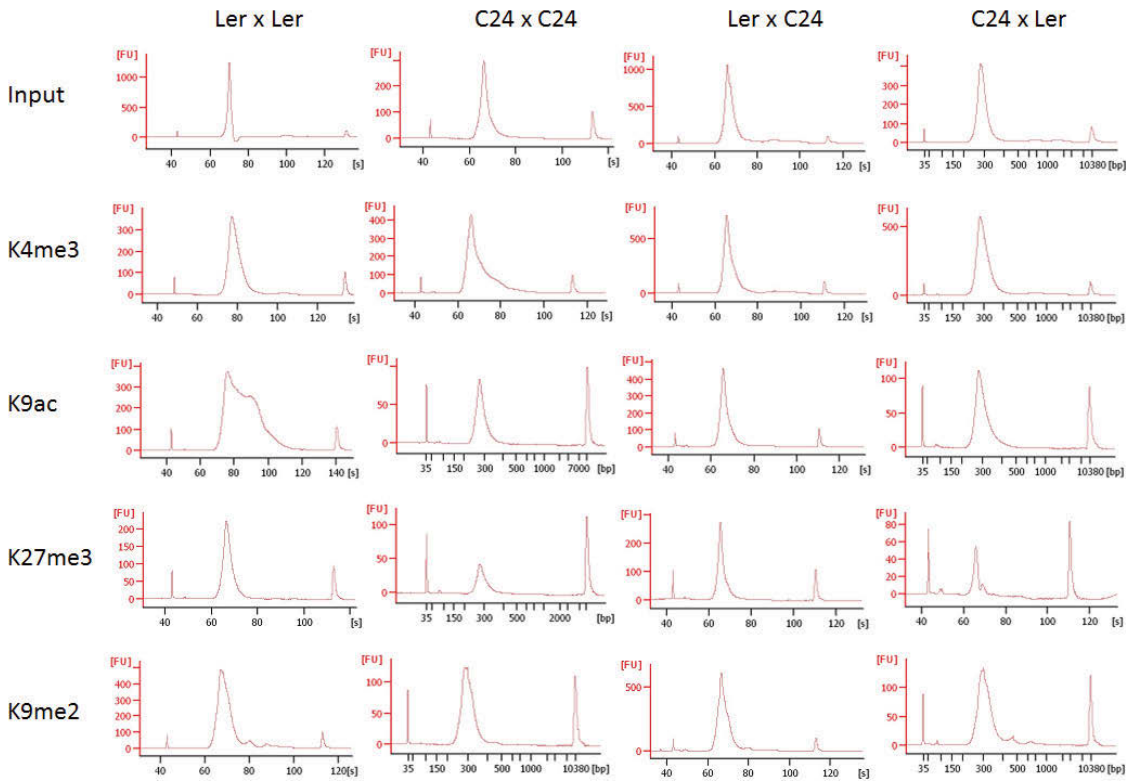


Figure 6.2. Bioanalyser results showing quality of for the first biological replicate of ChIP-Seq libraries. X axis represent fragment size. Y axis represent fluorescence units (FU). Similar results were also seen in the second biological replicate of samples.

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

Genome-wide histone modification patterns have not been determined in Arabidopsis seeds. Before comparing the epigenomes of parents and hybrids, some basic bioinformatic analyses on ChIP-Seq data are required to address the following questions: where are the different histone modifications located in the genome; where in the genes are the different histone modifications located; are histone modification levels associated with the expression levels of nearby genes; do different histone modifications occur with each other?

K4me3 and K9ac are both enriched in the same group of protein coding genes with similar enrichment levels, whereas K27me3 is enriched in a different group of genes (Figure 6.3 A). K9me2 is a transposable element (TE) specific histone mark and a high level of K9me2 is found in only a small number of genes, which have low signals of the other three histone marks. There are negative correlations between K4me3 and K9ac levels and K27me3 levels in genes. This finding is consistent with previous results in Arabidopsis seedlings (Sequeira-Mendes et al., 2014, Roudier et al., 2011), suggesting that the same relationship between histone marks is observed at the seed stage. There are still K4me3 and K9ac signals in K27me3 related genes, implying the active marks have wider distributions than the repressive marks in protein-coding genes (Figure 6.3 A). In addition to genome-wide distributions of histone marks, the profiles of enrichment of histone modifications in gene body regions and surrounding regions were prepared (Figure 6.3 B). Both K4me3 and K9ac have high enrichment levels near the transcription start sites (TSSs) of protein-coding genes, whereas the K27me3 signal is enriched in the whole gene body regions. There is poor enrichment of K9me2 in gene body regions and the level of K9me2 is generally lower than that of the other histone marks in protein-coding genes. All these results are consistent with previous findings about the profiles of histone marks over gene regions (Ha et al., 2011, Deleris et al., 2012).

In TEs, the patterns of histone marks are largely different from those in protein-coding genes. Generally, the levels of K4me3, K9ac and K27me3 are lower than K9me2 in TEs (Figure 6.4 A). The enrichment levels of K9me2 in the long TEs (≥ 1.5kb) are higher than in the short ones (Figure 6.4 A). In the long TEs, K9me2 is evenly enriched in the TE genes but has relatively low levels in the upstream and downstream regions (Figure 6.4 B). In contrast, the other three histone marks have lower levels in the TE regions than in

the surrounding regions. These results are supported by findings showing a similar distribution pattern of K9me2 in the long TEs in seedlings (Stroud et al., 2014).

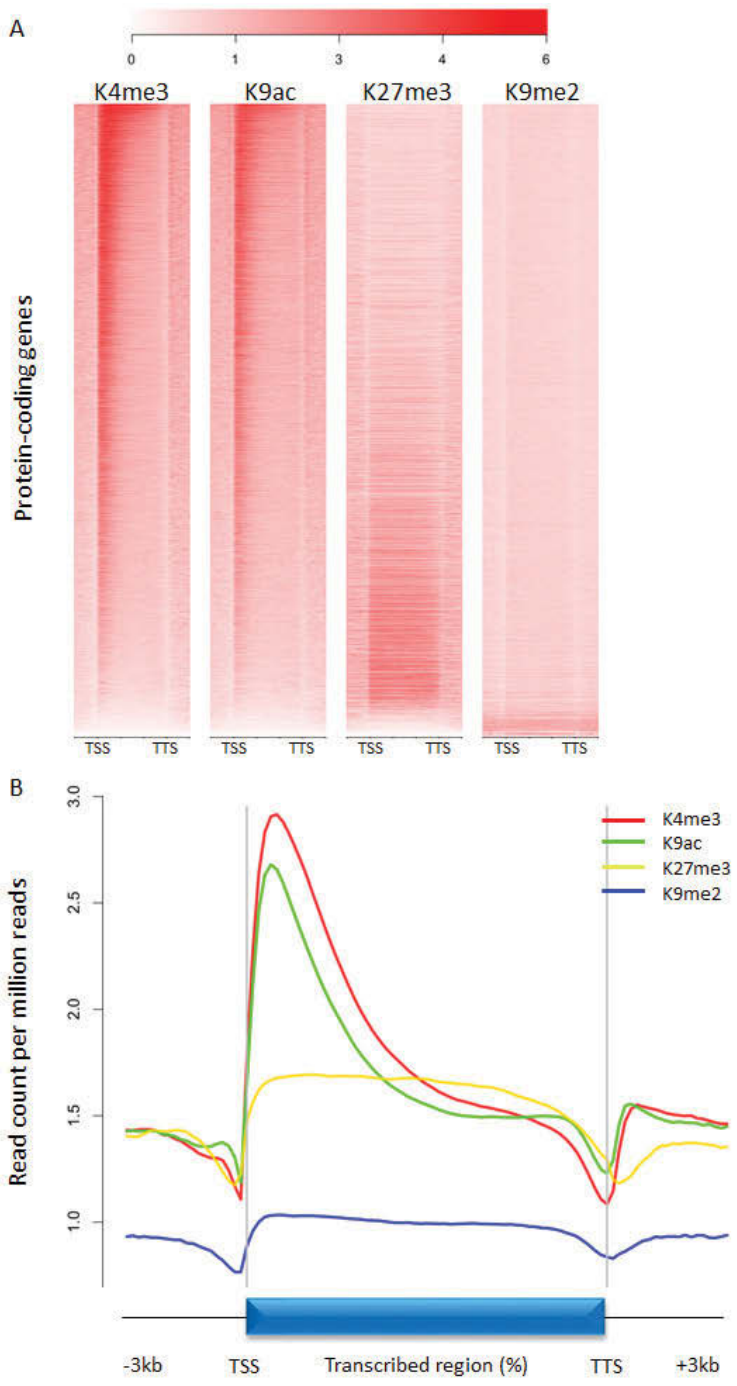


Figure 6.3. Global patterns of the four histone marks in C24 x C24 in protein-coding genes. (A) Heat maps show the enrichment levels of histone modification in all the protein-coding genes in Arabidopsis. Genes are ordered by K4me3 levels. The intensity of red represents the levels of histone marks. TSS, transcription start site; TTS, transcription termination site. **(B)** Average levels of histone modification signals over gene body and the ±3 kb regions in protein-coding genes.

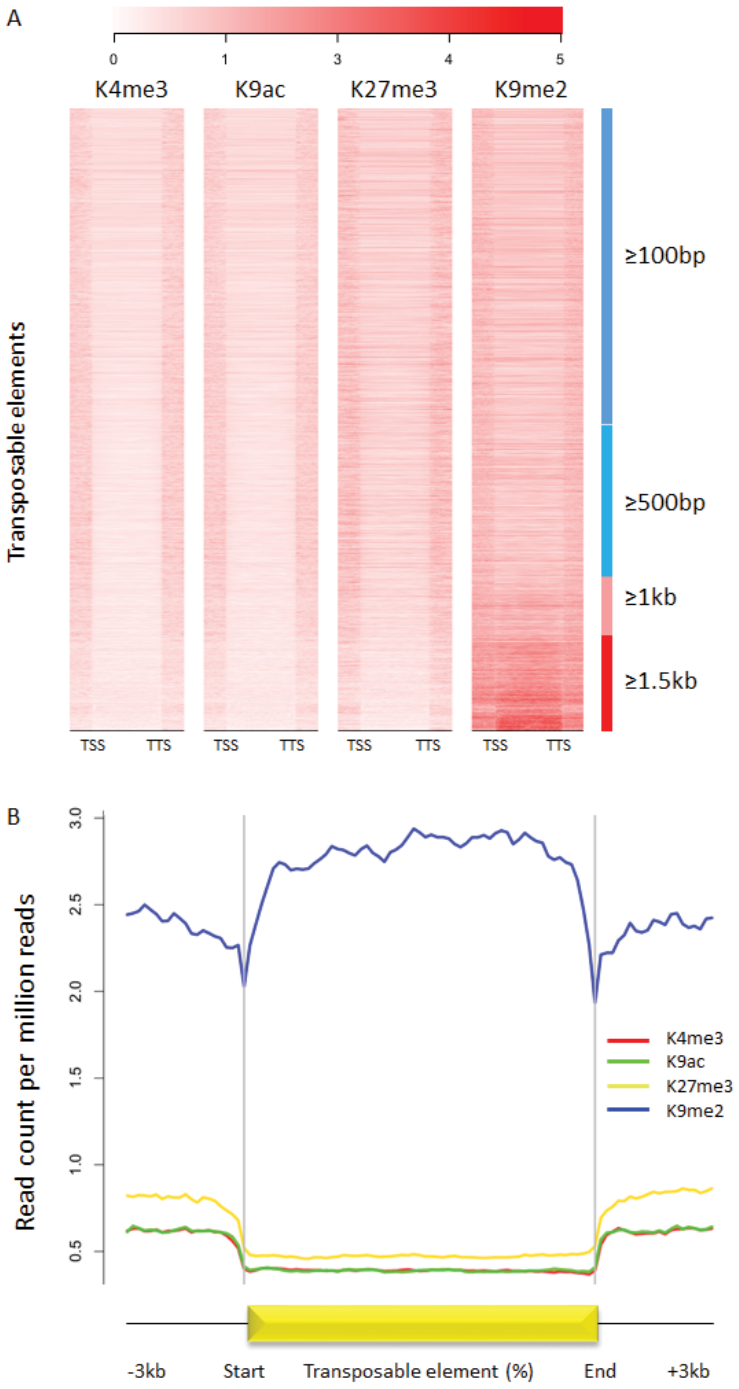


Figure 6.4. Global patterns of the four histone marks in C24 x C24 in TEs. (A) Heat maps show the enrichment levels of histone modification in all the TE in Arabidopsis. Genes are ordered by K4me3 levels. The intensity of red represents the levels of histone marks. TSS, transcription start site; TTS, transcription termination site. **(B)** Average levels of histone modification signals over TE and the ±3 kb regions in the long TE in size of above 1.5 kb.

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

Protein-coding genes were divided into 3 groups, which have high (20% of total genes in Arabidopsis), medium (60%) and low (20%) expression levels (data obtained from our transcriptome data). The histone modification enrichment patterns in each group of genes suggest positive correlations between gene expression and the active marks, K4me3 and K9ac (Figure 6.5). In contrast, the repressive marks, K9me2 and K27me3, have negative correlations with gene expression (Figure 6.5). These data suggest the active and repressive marks target different groups of genes consistent with our previous results (Figure 6.3 A). The relationship between K9me2 and gene expression is also identified in TE genes. TE genes consist of a small group of TEs and can be transcribed with poly-A tails, so that their expression levels can be detected by deep sequencing. Highly expressed TE genes have lower K9me2 levels than lowly expressed ones. Enrichment of K9me2 is found in the gene body regions, whereas it is not observed in lowly expressed TE genes, in which gene body and surrounding regions have similar K9me2 levels (Figure 6.5). The reason could be that the lowly expressed TE genes are located in heterochromatin where the K9me2 level in both genic and intergenic regions is high.

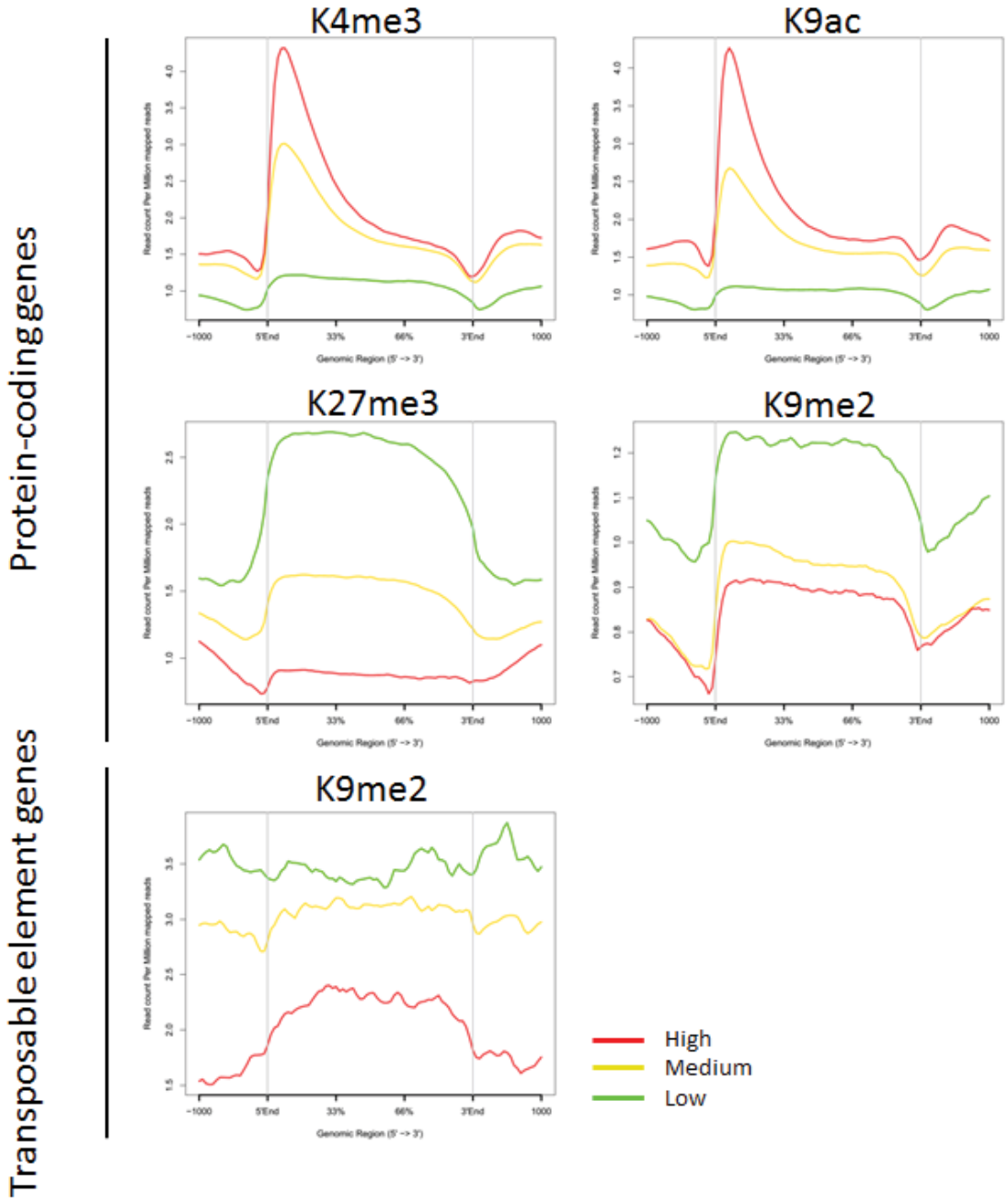


Figure 6.5. Correlations between histone modification and gene expression in C24 x C24. The top 20%, middle 60% and bottom 20% of protein-coding genes and TE genes were selected as high, medium and lowly expressed genes.

6.4 Histone modifications have the expected patterns associated with genes.

Read counts of histone modification "peaks" were compared statistically between parents

and hybrids. A histone modification peak refers to a region in the genome with significant enrichment of a histone mark. Significant histone modification peaks were identified using MACS2 software (Zhang et al., 2008). The histone modification targeted genes were identified by confirming the distances between genes and histone modification peaks. In order to obtain good correlations between gene expression and histone modification, different criteria were applied to define histone modification-targeted genes. Genes were considered to be targeted by K4me3 and K9ac when the peaks overlap the 1.5 kb regions upstream of the TSS, while genes were considered to be targeted by K9me2 when the peaks overlap the 1 kb regions upstream or downstream of the TSS. For K27me3-targeted genes, at least 50% of peak regions and 50% of gene body regions overlap with each other.

Analysis of histone modification targeted genes suggests that K4me3, K9ac and K27me3 mostly target protein-coding genes (Figure 6.6 A). By contrast, more TE genes are targeted by K9me2 than any other histone mark. There are fewer protein-coding genes targeted by the repressive marks than by the active marks (Figure 6.6 B). This is consistent with the wider distribution of the active marks compare to the repressive marks in the genome (Figure 6.3 A). In TE genes, the enrichment pattern of the four histone marks is opposite. The majority of TE genes are associated with K9me2 rather than the other three marks (Figure 6.6 C).

Histone modifications are involved in regulating gene expression. There can be more than one histone mark in or near to a gene. The analysis of the histone modification targeted genes suggests concurrent histone modifications in the genome at early developmental stages. The two active marks have a high association with each other (Figure 6.6 D). 88.98%

of K9ac-targeted genes are also targeted by K4me3, and 80.33% of K4me3-targeted genes have K9ac peaks. A poor correlation was found between the active marks and the repressive marks. K4me3 and K9ac are only found in approximately one third and one quarter of K27me3-targeted genes, respectively. However, these proportions are still higher than those in seedlings, in which active and repressive marks rarely occur in same genes (Roudier et al., 2011, Sequeira-Mendes et al., 2014), suggesting bivalent histone modifications might be widespread during embryogenesis in Arabidopsis.

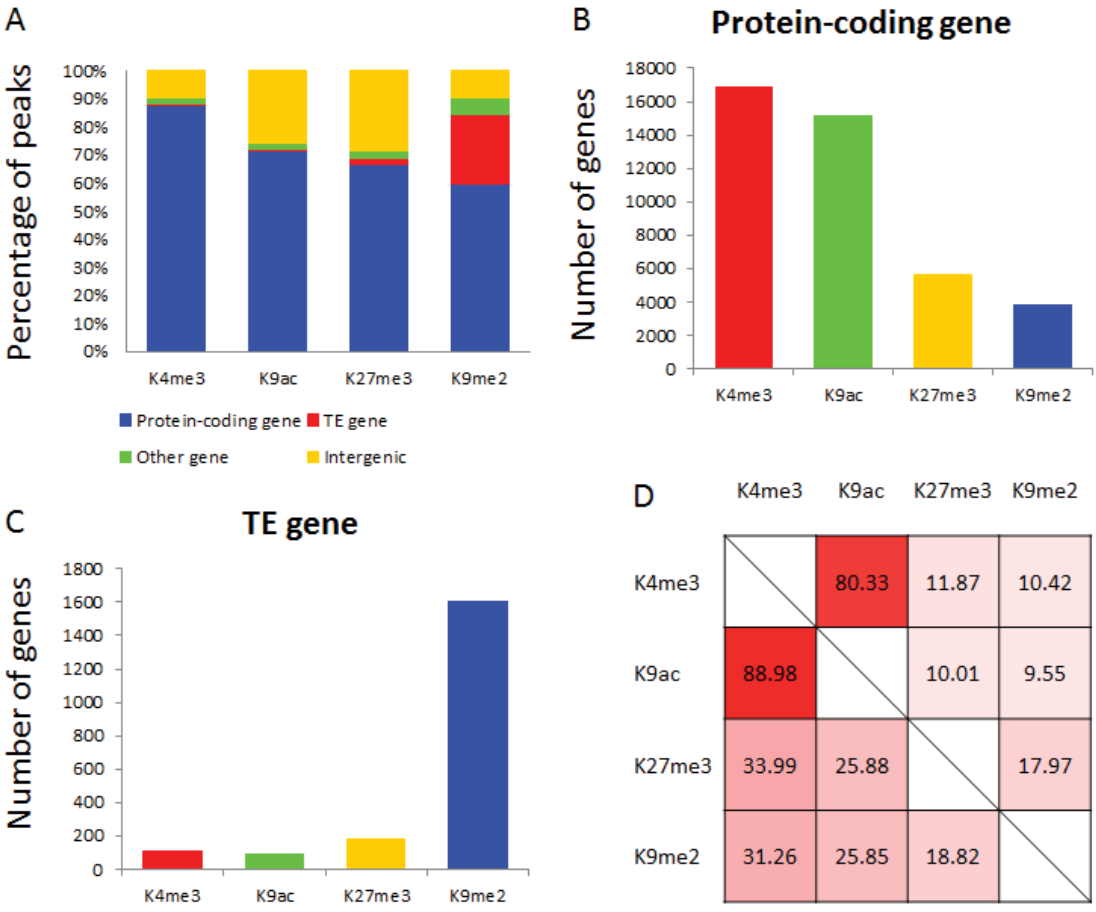
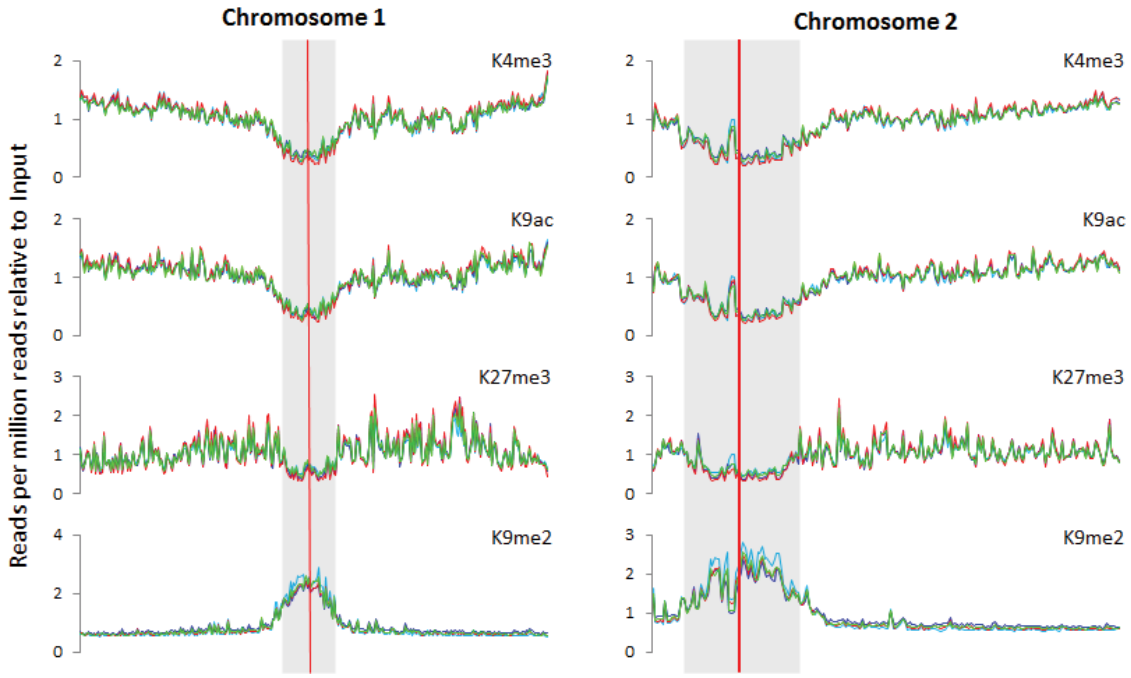


Figure 6.6. The correlation patterns between histone modification peaks and between peaks and genes in C24 x C24. (A) The percentages of histone modification peaks targeting different genomic features. **(B)** The numbers of protein-coding genes targeted by the four histone marks. **(C)** The numbers of TE genes targeted by the four histone marks. **(D)** The percentages of concurrency between the four histone marks.

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

Average levels of histone modifications in gene and TE regions were analysed in both parents and hybrids, but no significant difference were found between the four plant lines (Appendix 3 and 4). The histone modification levels of parents and hybrids were plotted in each of the five Arabidopsis chromosomes. As expected, high levels of K4me3 and K9ac were identified in chromosome arms but not in pericentric regions (Figure 6.7). In contrast, the K9me2 signal was found in pericentric regions of all five chromosomes. Patterns of histone modifications in the five chromosomes do not differ between the four plant lines. Although there are minor differences at some specific regions in the chromosomes, the general patterns of histone modification are similar between parents and hybrids (Figure 6.7). This was also shown in Arabidopsis hybrids in the seedling stage (Moghaddam et al., 2011).



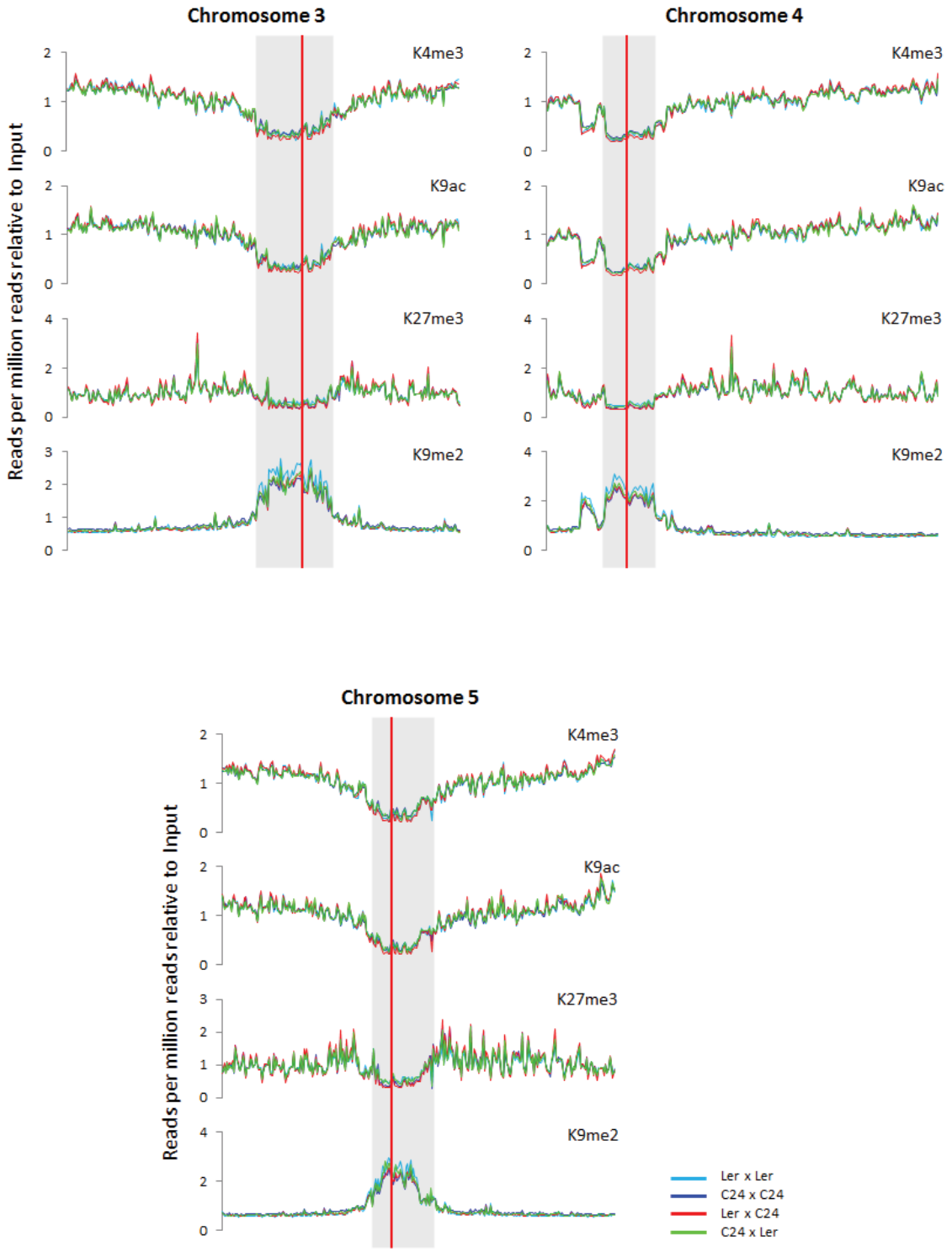


Figure 6.7. The distributions of four histone marks over the five chromosomes in Ler and C24 and their hybrids. X axis represents the locations in chromosomes. Red lines represent centromeres, grey boxes represent pericentric regions.

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

To compare histone modification levels for individual genes, statistical analyses were performed on the ChIP-Seq data of the parents and hybrids. There are 1000 to 2000 genes with altered levels of the different histone modifications (differentially histone-modification-enriched genes, DHGs) between L*er* and C24 (Figure 6.8). The majority of K4me3 and K9me2 DHGs have higher histone modification levels in L*er* than in C24, while K27me3 has the opposite pattern. Numbers of K9ac DHGs with higher and lower levels in L*er* than in C24 are equal. These results suggest that although global histone modification levels are similar between L*er* and C24, they can be different at specific genes.

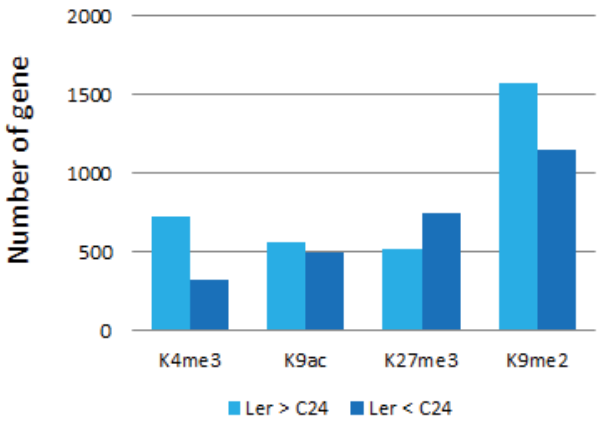


Figure 6.8. The numbers of differentially histone-modification-targeted genes (DHGs) for each histone mark between Ler and C24 in seeds. Fold-change ≥ 1.5 and p-value ≤ 0.05.

DHGs are also identified in the hybrids compared to their parents. Similar to transcriptome results at 0 DAS, the two reciprocal hybrids differ in the number of DHGs (Figure 6.9 A). L*er* x C24 have significantly more K4me3 and K27me3 DHGs and

slightly more K9ac DHGs than C24 x L*er*. In contrast, the number of K9me2 DHGs is larger in C24 x L*er* than in L*er* x C24. There is only a small group of genes with altered levels of histone modifications (0.07% - 1.7%), except for K27me3 DHGs in L*er* x C24 (16.5%). There are a small number of DHGs common to both hybrids for all the four histone marks (Figure 6.9 C).

Bias exists in directions of the changes in histone modifications levels from the MPV (Figure 6.9 B). The majority of DHGs in L*er* x C24 have increased levels of histone marks, except for K9me2. In contrast, C24 x L*er* has large proportions of DHGs with decreased levels of histone marks, except for K4me3. This implies that, in seeds, histone modification levels are differentially regulated in the two hybrids. In L*er* x C24, an unknown regulatory system could be activated to elevate K4me3 and K27me3 in many genes, but not in C24 x L*er*. The number of DHGs with higher K4me3 levels in L*er* is larger than that with higher K4me3 levels in C24 (figure in 5.6). This may cause the increased K4me3 levels only in L*er* x C24 through maternal regulatory elements. But this is not the case for K27me3, which has lower levels in L*er* but has increased levels in L*er* x C24. L*er* x C24 has a higher growth rate than C24 x L*er* at early developmental stages (Figure 3.3 and 3.4). The differences in histone modification levels at 0 DAS may be associated with changed gene expression levels, which may result in the later different growth rates between the hybrids.

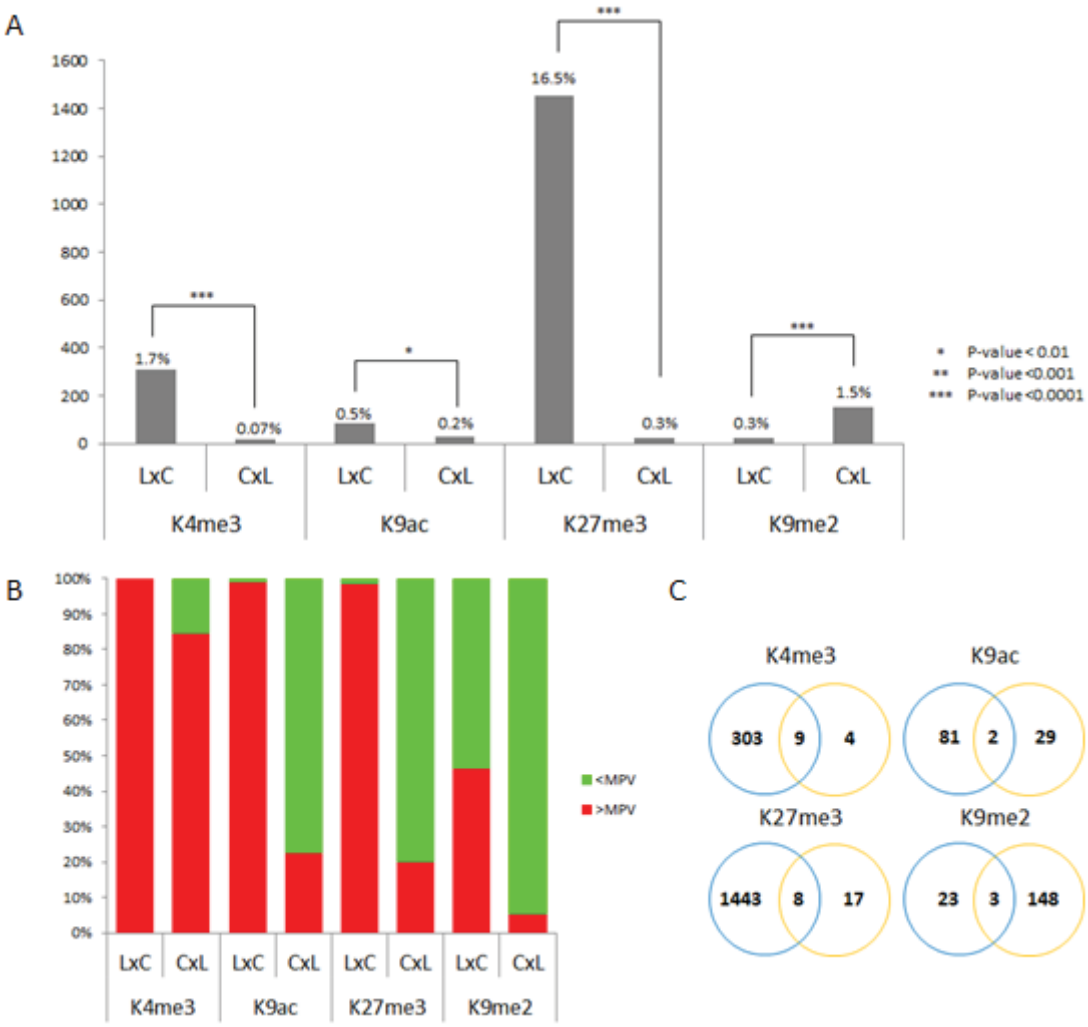


Figure 6.9. The DHGs identified in Ler/C24 hybrids in seeds. **(A)** The numbers of DHGs for each histone mark in the reciprocal hybrids. Fold-change from MPV ≥ 1.3 and p-value ≤ 0.05. **(B)** The proportions of DHGs with increased and decreased levels of histone modifications. **(C)** Venn-diagrams showing the common DHGs between the reciprocal hybrids for different histone marks.

Based on previous findings on sRNAs, the regions with largely different levels of the two epigenetic elements between parents are more likely to have changed levels in the hybrids (Groszmann et al., 2011b). To test if a similar pattern can be observed in histone modifications, all the histone modification targeted genes are plotted in the order of the degree of difference in histone modifications between parents (Figure 6.10). In C24 x L*er*, more DHGs are located in the regions with larger differences in histone modification levels between the parents. However, in L*er* x C24, the DHGs are largely distributed in

the regions where the parents have similar levels. This means that DHGs in Ler x C24 are not dependent on the histone modification levels in the parents.

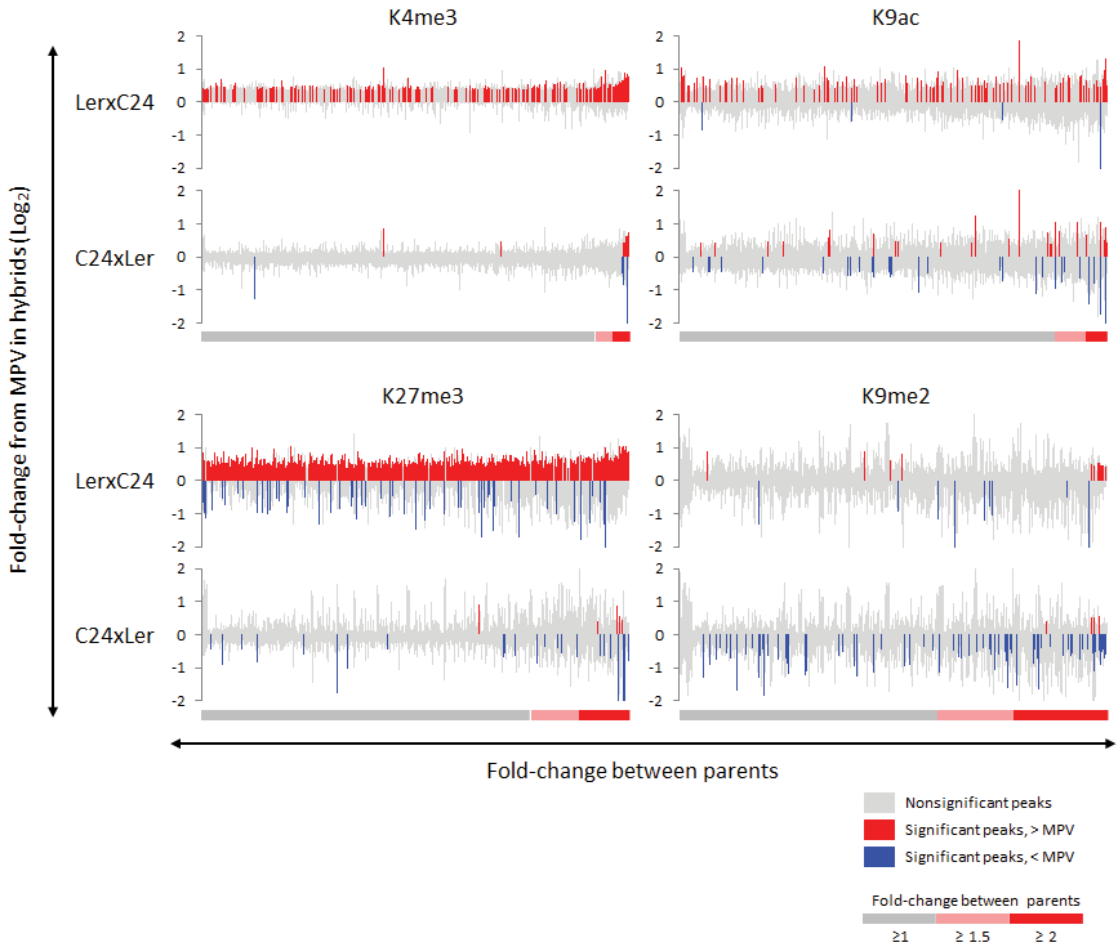


Figure 6.10. Correlations between the changes in histone modifications between parents and between parents and hybrids. To determine significant peaks, fold-change from MPV ≥ 1.3 and p-value ≤ 0.05.

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

To determine if the changed histone modification resulting in corresponding alteration in gene expression between the parents and between parents and hybrids, ChIP-Seq data was associated with RNA-Seq data at 0 DAS (Figure 6.11). Significant positive

correlations between the two active marks and gene expression occur in the DHGs with significantly changed transcript levels between L*er* and C24. The K27me3 and K9me2 DHGs do not show negatively correlated levels of gene expression between the parents. This may be because the repressive marks do not solely regulate the expression of some of the genes, and other regulatory factors could be involved in regulating expression of those genes. However, for genes with negative correlations between gene expression and histone marks, the role of the repressive marks is essential. All these results suggest that gene expression is naturally correlated with active marks but not with the repressive marks between Arabidopsis ecotypes in germinating seeds. This conclusion parallels findings in rice (He et al., 2010).

In hybrids, only a small number of genes have both significantly non-additive gene expression and significantly non-additive histone modifications, especially in C24 x L*er*. All the four histone marks are weakly correlated with the gene expression of both significantly non-additive genes and total genes (Figure 6.11). This result indicates weak correlations between non-additive gene expression and non-additive histone modifications. The changes in histone modifications do not lead to the expected alteration in gene expression in the hybrids at 0 DAS, although there are some genes showing consistent changes both in gene expression and in histone modifcaiton (Appendix 5).

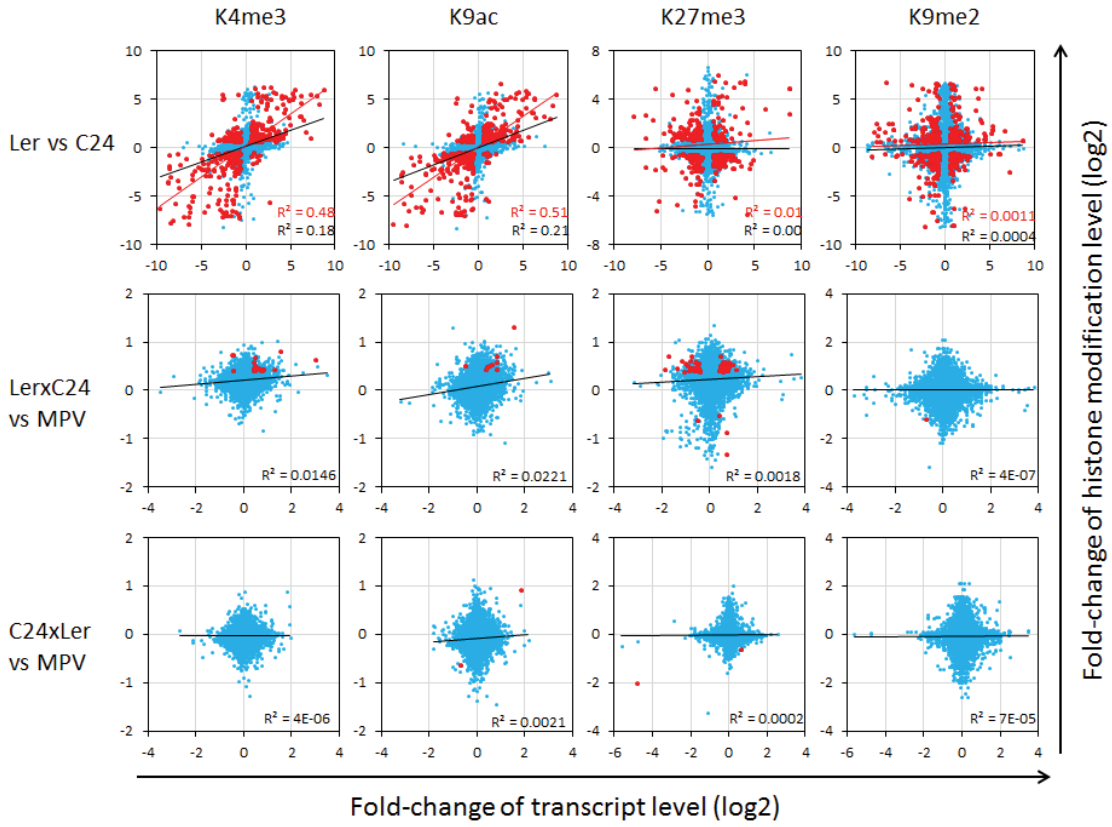


Figure 6.11. Correlations between histone modifications and gene expression in the parents and the hybrids in seeds. All the genes targeted by histone modifications were selected. **Black lines** represent trend lines of total genes. **Red lines** represent trend lines of significant genes. **Red dots** represent genes with significant changes both in histone modification and in gene expression. **Blue dots** represent non-significant genes. Significant genes for gene expression: Ler vs C24, fold-change ≥ 1.5, p-value ≤ 0.01; Hybrids vs MPV, fold-change ≥ 1.3, fold-change ≤ 0.01. Significant genes for histone modifications: Ler vs C24, fold-change ≥ 1.5, p-value ≤ 0.05; Hybrids vs MPV, fold-change ≥ 1.25, fold-change ≤ 0.05.

The average histone modification levels were analysed in genes undergoing non-additive expression in the hybrid genomes (Figure 6.12). In non-additive genes with increased or decreased expression levels, the levels of K4me3 and K27me3 are generally increased in L*er* x C24. However, they are similar as MPVs in C24 x L*er*. Furthermore, the average levels of K9me2 do not change in either up-regulated or down-regulated genes in both reciprocal hybrids. This result indicates that non-additive levels of transcripts in hybrids may not be the consequences of altered histone modification levels.

Figure 6.12. Relative average levels of gene expression and histone modifications in the non-additive genes in both hybrids at 0 DAS. Error bars represent SEM.

The changed levels of histone modifications may lead to delayed changes in gene expression in hybrids. By correlating 0 DAS ChIP-Seq data with 3 DAS RNA-Seq data, we identified a weak correlation between histone modifications and gene expression (Figure 6.13). These results imply that the changes in histone modifications in hybrid seeds could not lead to consistent changes in gene expression at a subsequent time point.

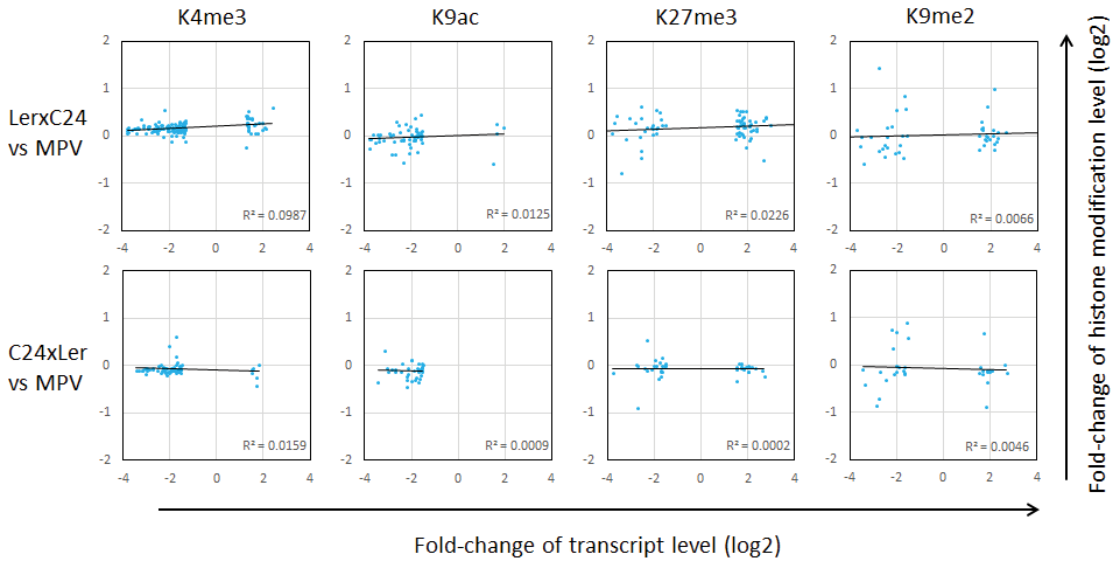


Figure 6.13. Correlations between histone modifications at 0 DAS and gene expression at 3 DAS in the hybrids. Dots represent 3 DAS non-additive genes. For gene expression, fold-change ≥ 1.3, p-value ≤ 0.01; for histone modifications, no filter applied.

6.8 Summary

- DNA from ChIP experiments were verified by qPCR.
- In Arabidopsis seeds, K4me3 and K9ac have a high enrichment in the 5' of gene body regions, and K27me3 and K9me2 are present over the entire gene body regions.
- The two active marks often occur together in the genome of Arabidopsis seeds, whereas the correspondence between active and repressive marks and between K27me3 and K9me2 is low.
- K4me3, K9ac and K27me3 target protein-coding genes, and K9me2 is correlated with TE genes.
- Globally, K4me3 and K9ac are positively associated with gene expression and negative correlations were found between K27me3 and K9me2 and gene expression.

- Although the patterns of the four histone marks in L*er*, C24 and their hybrids are similar genome-wide, there are some genes with changed levels of histone modifications in the hybrids.
- Although strong correlations were observed between gene expression and the active histone marks in the parents at 0 DAS, the changes in histone modifications in the hybrids do not lead to the expected alterations in gene expression.

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

Intra-specific hybrids have genomes of parents from the same species. Although parental genomes are similar to each other in DNA sequence, there are some differences between them, such as SNPs. By utilizing SNPs between L*er* and C24, we were able to identify the transcript and histone modification levels for each of the two alleles in the hybrid genomes. The following questions will be addressed in this chapter: Do the ratios between L*er* and C24 alleles in transcription and histone modification in hybrids follow the ratios between the two parents? Do the allelic ratios in hybrids change over time? Are the changes of allelic ratios in gene expression correlated with that in histone modification in hybrids? Global analyses of allelic expression may also provide insights into the non-additive alterations in gene expression in hybrids.

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

In SNP analyses, instead of the reads from whole transcripts, only the reads spanning the SNP identified regions were utilized. Only the genes and histone modification peaks showing SNP reads representative of transcript reads and peak reads were retained in the analyses. The numbers of retained genes and peaks are listed in Table 7.1. There are fewer genes at 0 DAS than at the following time points. This could be due to fewer genes being expressed in the seeds than in the seedlings. There are fewer histone modification peaks identified for repressive marks than those for active marks, due to the relatively long peaks of the two repressive marks.

Table 7.1. Counts of genes from the four time points and counts of peaks for the four histone marks from 0 DAS in SNP analyses.

	0 DAS	3 DAS	5 DAS	7 DAS
No. of genes	1984	5113	5008	5160
	K4me3	K9ac	K27me3	K9me2
No. of peaks	7938	6719	3121	632

In the retained genes at all four time points, the L*er* proportions (L*er*/L*er*+C24) based on SNP reads are positively correlated with the L*er* proportions based on transcript reads (Appendix 6 A). The hybrid proportions (Hybrid/Hybrid+MPV) based on SNP reads are also positively correlated with the hybrid proportions based on transcript reads (Appendix 6 B). Similar correlations were also found in the four histone marks (Appendix 7). These results mean that the patterns of gene expression and histone modifications in parents and hybrids based on SNP reads are similar to those based on overall reads of transcripts and peaks.

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

To determine the expression patterns of the two parental genomes in hybrids, the numbers of genes with significantly changed allelic ratios from parental ratios were counted in all the transcriptome datasets (Figure 7.1 A). At 0 DAS, 44% of the analysed genes analysis show allelic ratios different from the parental ratios in L*er* x C24. By contrast, only 14% to 17% of the genes have significant changes in allelic ratio from 3 to 5 DAS. Similar

patterns are also observed in C24 x L*er*. While the two parental genomes undergo large changes in gene expression in the hybrids at 0 DAS, these changes are not seen at the following three time points. The majority of the changes in allelic ratio happen in the genes that have similar expression levels between the parents, suggesting the differences in gene expression between parents may not be the reason for allelic changes in hybrids.

To understand how the two alleles change in hybrids, the genes in analysis are divided into nine groups (Figure 7.1 B), based on the expression trends of the two alleles. At 0 DAS, the majority of genes with changed allelic ratios have up-regulated levels at C24 alleles but unchanged levels of the L*er* alleles compared to the expected value (EV, which equals half of expression levels in the parents) in both reciprocal hybrids at 0 DAS. The up-regulation of C24 alleles disappears gradually from 3 to 7 DAS. This suggests that up-regulation of C24-allele-specific expression causes the changes in allelic ratio in hybrid seeds, although the up-regulation is transient.

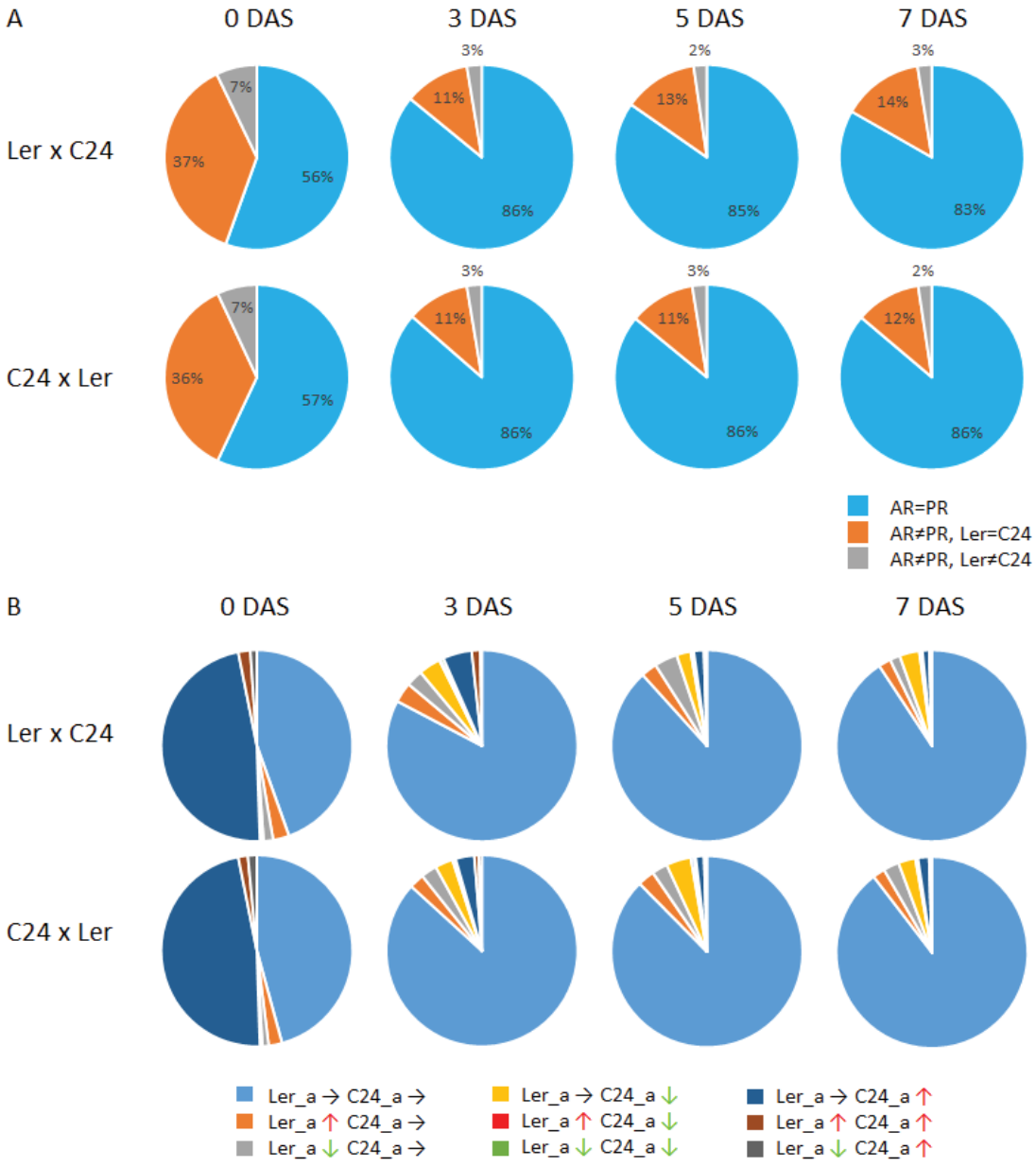


Figure 7.1. Patterns of allelic expression in hybrids at four time points. (A) Proportions of genes with and without changed allelic ratios. AR, allelic ratio; PR, parental ratio. **(B)** Proportions of the nine possible patterns of changes in gene expression at Ler and C24 alleles. The black, red and green indication arrows indicate unchanged, increased and decreased levels of the two alleles in the hybrids. To determine alteration, fold-change against the expected value is equal to or above 1.5. Ler_a and C24_a, Ler and C24 alleles.

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*

Compared to allelic gene expression, allelic histone modifications show different patterns. The majority of peaks have unchanged allelic ratios compared to the parental ratios for all the four histone marks at 0 DAS (Figure 7.2 A). The two reciprocal hybrids have similar patterns for the active marks but different patterns for the repressive marks. L*er* x C24 has more (16%) K27me3 peaks with changed allelic ratios than the C24 x L*er* hybrid (8%), and more (11%) K9me2 peaks with changed allelic ratios are found in C24 x L*er* (1%). Similar to the allelic patterns in gene expression, most peaks with changed allelic ratios in hybrids have similar histone modification levels between parents.

The two reciprocal hybrids also differ in trends of changes in the two alleles (Figure 7.2 B). Although most peaks show unchanged allelic levels relative to EVs, many peaks show increased levels of L*er* alleles and unchanged levels of C24 alleles in L*er* x C24. However, this pattern is not seen in C24 x L*er*. Instead, the levels of histone modifications are decreased at L*er* alleles in C24 x L*er*, suggesting a maternal influence. The increased levels of L*er* alleles may be associated with the increased levels of K4me3 and K27me3 in many genes in L*er* x C24 hybrid (Figure 6.10).

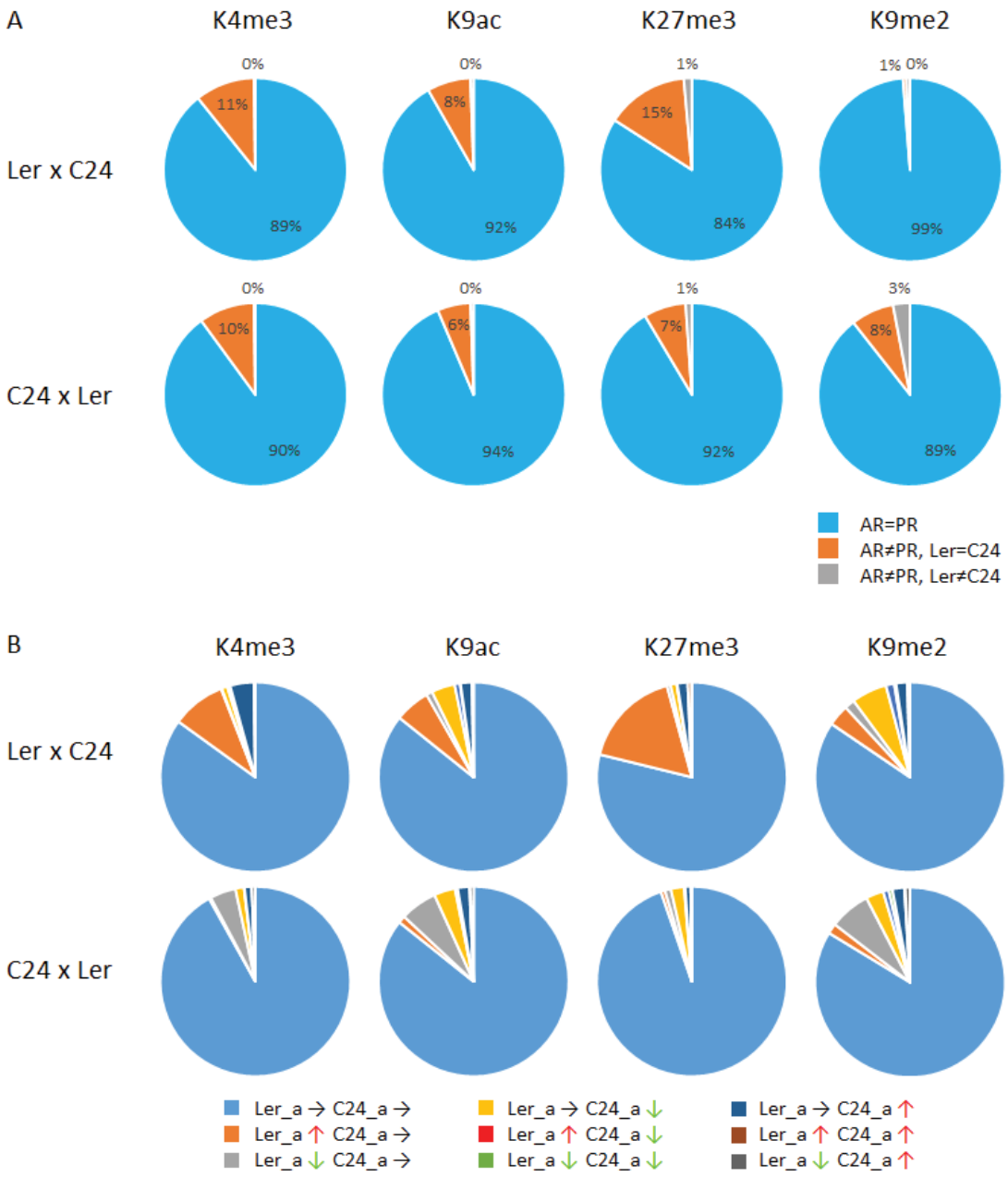


Figure 7.2. Patterns of allelic changes of the four histone marks in hybrid seeds. (A) Proportions of peaks with and without changed allelic ratios. AR, allelic ratio; PR, parental ratio. **(B)** Proportions of the nine possible patterns of changes in histone modification at Ler and C24 alleles. The black, red and green indication arrows indicate unchanged, increased and decreased levels of the two alleles in the hybrids. To determine alteration, fold-change from expected value ≥ 1.5.

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

To investigate the contributions of L*er* and C24 alleles to non-additive gene expression in hybrids, expression levels relative to EV are plotted for both alleles (Figure 7.3). In 0-DAS non-additive genes, the up-regulated expression in both hybrids is mainly due to up-regulation of the C24 alleles, whereas a similar bias is not seen in the down-regulated genes. Moreover, the bias between the two alleles in contributing to non-additive expression is not observed in either up- or down-regulated genes at the other time points. This means that changed levels in gene expression can be due to alterations occurring at either or both of the alleles. This indicates that both parental genomes play equally important roles in causing non-additive gene expression in hybrids, except at 0 DAS, where up-regulation of gene expression is mainly caused by C24 allele up-regulation in hybrids.

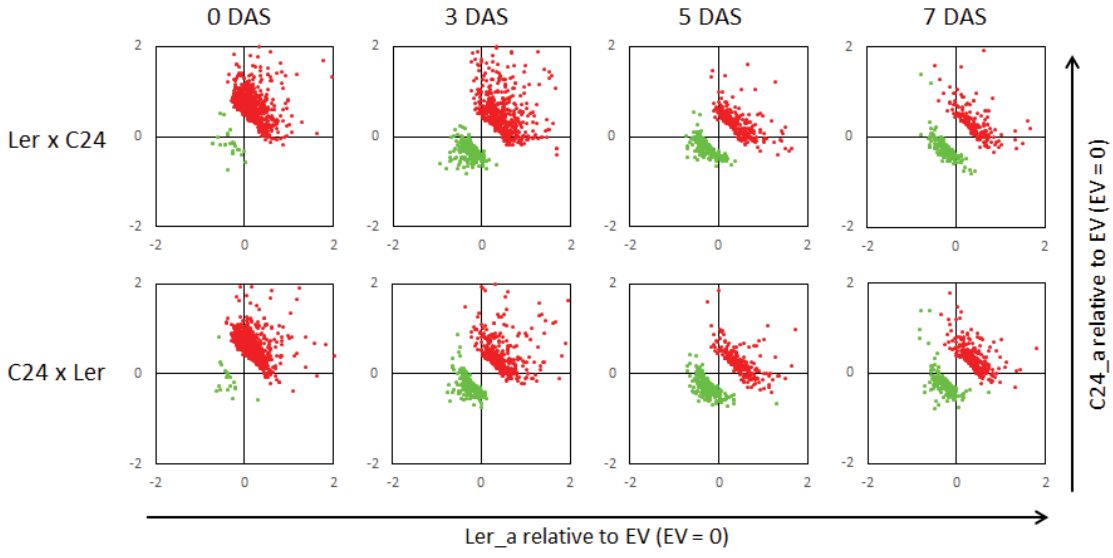


Figure 7.3. Contributions of Ler and C24 alleles to non-additive gene expression in hybrids. The up-regulated (red dots) and down regulated (green dots) non-additive genes from different time points are determined based on SNP reads. To determine non-additive genes, fold-change from MPV ≥ 1.5.

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

To identify if changes in allelic expression are regulated by changes in allelic histone modifications, only genes with overlapping histone modification peaks were included (1,257 K4me3-targeted genes; 1,518 K9ac-targeted genes; 68 K27me3-targeted genes; 21 K9me2-targeted genes). In both reciprocal hybrids, corresponding changes in histone modifications only occur at the loci with changed gene expression in a subset of the analysed genes (Figure 7.4; for C24 x Ler, see Appendix 8). However, for the other genes, histone modifications do not change consistently with gene expression at both alleles. This might be because allelic gene expression is regulated by histone modifications in only a subset of genes, or because changes in allelic gene expression and in allelic histone modification are independent of each other.

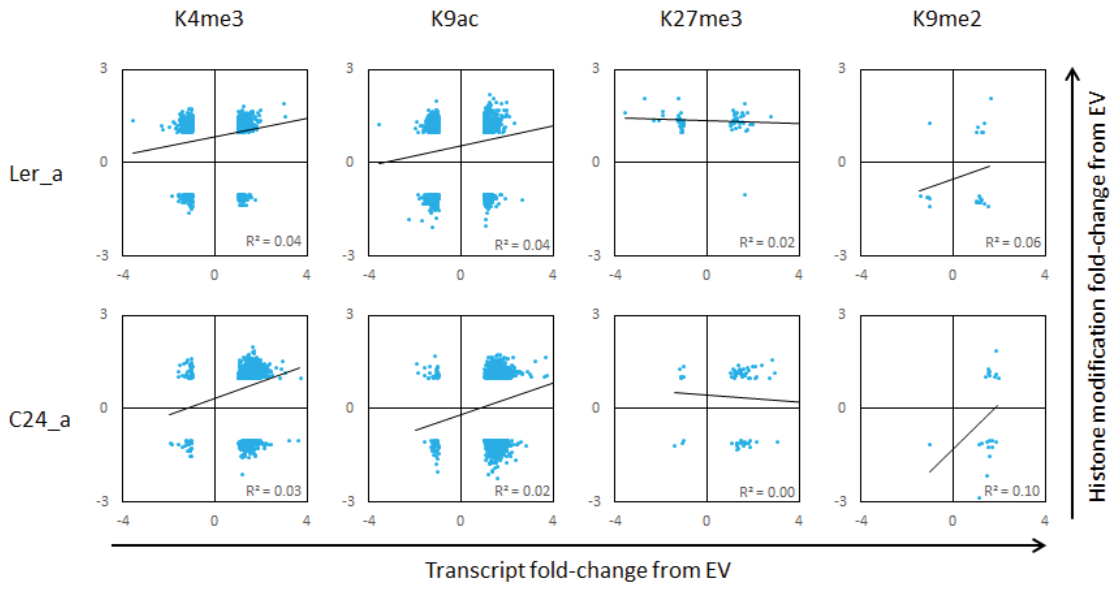


Figure 7.4. Changes of gene expression against changes of histone marks at Ler and C24 alleles in Ler x C24. Dots represent the genes targeted by histone modifications.

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

To analyse the changes of allelic expression in hybrids over time, only the genes common to all the time points were included in this analysis (1,011 genes). The analysed genes are divided into Group A (503 genes in L*er* x C24; 495 genes in C24 x L*er*; 368 genes are common to both reciprocal hybrids), in which genes have up-regulated C24 alleles (fold-change from EV ≥ 1.5), and Group B, in which genes do not have up-regulated C24 alleles (fold-change from EV < 1.5), respectively.

Group A genes show a trend of decreased L*er* proportion (L*er*/L*er*+C24) in parents at 3, 5 and 7 DAS compared to 0 DAS, whereas the parental proportions are unchanged for Group B genes (Figure 7.5 A; for C24 x L*er*, see Appendix 9 A). However, for both Group A and B genes, the L*er* allele proportions (L*er*_a/L*er*_a+C24_a) in hybrids are unchanged at 3, 5 and 7 DAS compared to 0 DAS (Figure 7.5 B; for C24 x L*er*, see Appendix 9 B). The levels of the C24 alleles in hybrids are always higher than that of the L*er* alleles throughout the early developmental stages. These data suggest that, while allelic patterns of gene expression in hybrids do not change over time, the expression levels of Group A genes are increased in the parent C24 compared to those in L*er* after seed germination, and do not change after 3 DAS.

Genes in Group A show a trend of decreased L*er* allele proportion in hybrids compared to L*er* proportions in parents at 0 DAS (Figure 7.6; for C24 x L*er*, see Appendix 10), due to up-regulation of the C24 alleles of Group A genes. The allelic proportions in hybrids at 0 DAS are generally similar to the parental proportions at 3, 5 and 7 DAS (Figure 7.6;

for C24 x L*er*, see Appendix 10). These results suggest an early 'set-up' for later high-C24 parental patterns of gene expression in hybrids at 0 DAS, or possibly at an even earlier stage. Once set up, the allelic proportions of Group A genes do not change over time in hybrids. The early 'set-up' of allelic expression only happens in a subset of genes in Arabidopsis hybrids, since Group B genes show unchanged allelic proportions in hybrids over time. A possible explanation of this phenomenon is that Group A genes are regulated by different regulatory factors in L*er* and C24. The regulatory factor of the C24 alleles could be activated later than the regulatory factor of the L*er* alleles. For an unknown reason, the factor regulating the C24 alleles is activated earlier in the hybrids, leading to up-regulation of the C24 alleles happening earlier in hybrid seeds.

To know if this early 'set-up' is related to non-additive gene expression, we analysed the genes common to Group A genes and non-additive genes from all the four time points (4,916 genes). There are only 10% (52) of genes from Group A showing non-additive expression in hybrids during early developmental stages (Figure 7.7), indicating early 'set-up' of allelic expression in hybrids may not be related to non-additive expression.

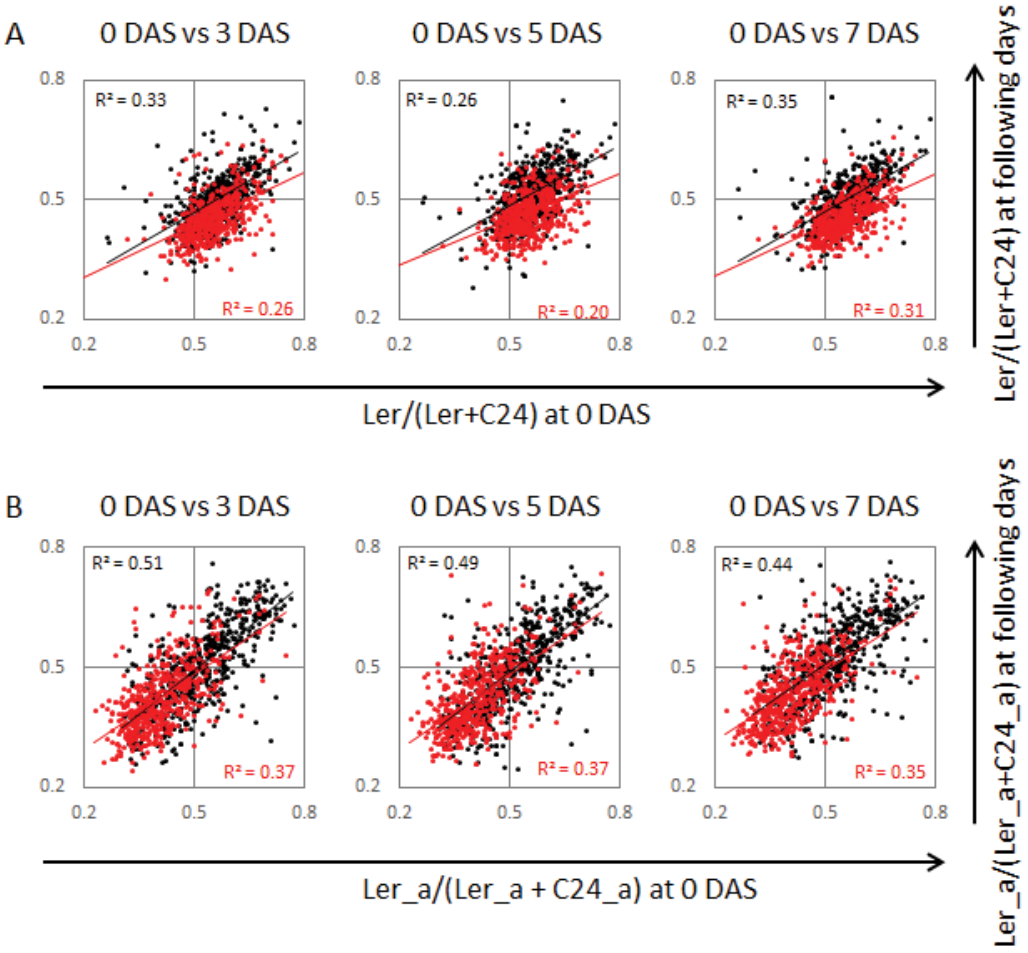


Figure 7.5. Correlations of parental (A) and allelic (B) proportions in Ler x C24 between 0 DAS and the following three time points. Red dots represent Group A genes (with up-regulated expression levels at C24 allele at 0 DAS) and black dots represent Group B genes (with unchanged and down-regulated expression levels at C24 allele at 0 DAS). To determine changed expression levels at alleles, fold-change from EV ≥ 1.5.

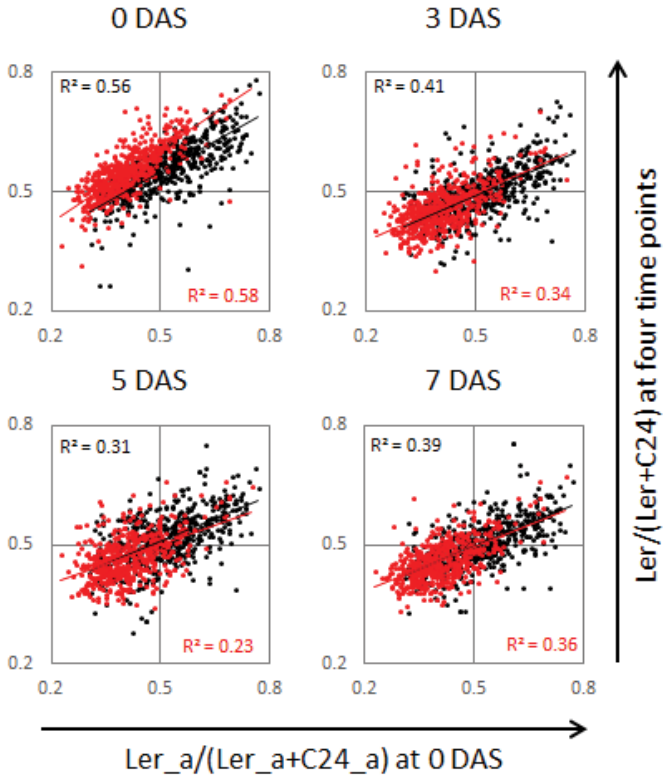


Figure 7.6. Allelic proportions against parental proportions in Ler x C24 at the four time points. Red dots represent Group A genes (with up-regulated expression levels at C24 allele at 0 DAS) and black dots represent Group B genes (with unchanged and down-regulated expression levels at C24 allele at 0 DAS). To determine changed expression levels at alleles, fold-change from EV ≥ 1.5.

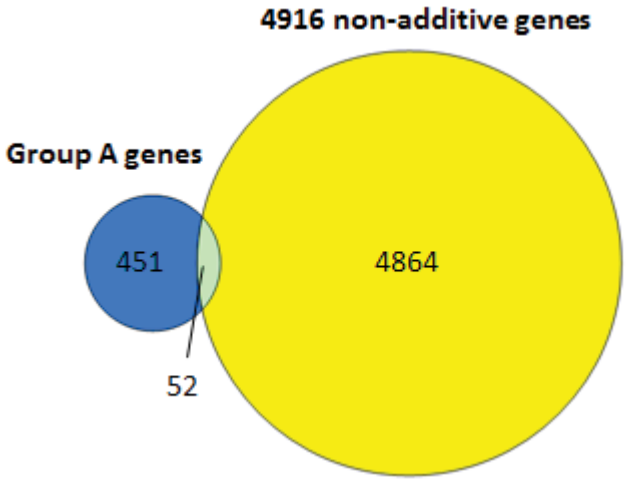


Figure 7.7. Venn-diagram showing common genes to Group A and the total non-additive genes from the four time points (4,916 genes).

7.7 Summary

- The ratio of gene expression between L*er* and C24 alleles is globally unchanged in both reciprocal hybrids at 3, 5 and 7 DAS, whereas at 0 DAS, a large number of genes have changed allelic ratios caused by up-regulation of expression of C24 alleles.
- The allelic ratios of histone modifications are generally unchanged in hybrid seeds. In the peaks with altered allelic ratios, elevated levels of K4me3 and K27me3 at the L*er* allele were identified in L*er* x C24 but not in C24 x L*er*.
- Ler and C24 alleles have equivalent roles in non-additive gene expression in hybrids at early development stages, except for 0 DAS, in which non-additive gene expression largely depends on up-regulation of the C24 allele.
- At 0 DAS, the changes in allelic expression do not completely correlate with the changes in allelic histone modification in hybrids.
- The up-regulated expression of the C24 alleles at 0 DAS is caused by the early ‘‘set-up’’ of later parental allelic patterns in hybrid seeds.

8 Chapter VII Discussion

Hybrid vigour, or heterosis, is an important topic in plant research. Understanding how hybrids perform better than their parents could be important for further improving the food industry both in quantity and in quality of crop products. Although it has been studied for nearly a hundred years, no single theory has completely explained the mechanism of heterosis. The most reasonable assumption is that changed expression levels of particular genes promote alterations in some key pathways in hybrids, such as photosynthesis (Fujimoto et al., 2012), circadian rhythm, starch production (Ni et al., 2009) and hormone biosynthesis (Groszmann et al., 2015). However, the factors driving these transcriptional changes are still not fully understood. Epigenetic factors may have roles in intra-specific heterosis, as only minor genetic differences are found between the two parental genomes, which produce considerable heterosis in hybrids. Current reports suggest that phenotypic heterosis could be established as early as the first week after sowing in Arabidopsis hybrids (Meyer et al., 2004, Meyer et al., 2012), although the timing of the beginning of heterosis is not clear at a molecular level.

In this project, we studied transcriptomes and global histone modification patterns in L*er*/C24 hybrids at early stages of seedling development. We identified thousands of genes expressed non-additively in hybrids at different time points, especially at 3 DAS in which the largest number of non-additive genes were identified. The non-additive genes at 3 DAS are involved in some important pathways of plant growth, particularly photosynthesis-related pathways that showed up-regulated transcriptional levels. By analysing gene expression patterns over time, we know that the non-additive gene

expression at early seedling stages is mainly due to earlier and transient changes in gene activity in hybrids compared to the parents. These early and transient changes happen in different groups of genes at different time points, indicating that dynamic controls are involved in regulating gene activities in hybrids. Motif analyses suggested that those non-additive genes are not commonly targeted by known transcription factors. Genome-wide analyses of four histone marks suggested that although histone modification patterns do not change globally in germinating seeds of hybrids compared to parents, they do change for particular genes. However, the changes in histone modification are not fully consistent with the changes in gene expression at the same time point, indicating that histone modifications are not the only regulatory factors of non-additive gene expression in hybrid seeds. In summary, early heterosis could result from dynamic changes of gene expression in young seedlings of hybrids, and 3 DAS could be a key time point for heterosis establishment.

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

The opening and greening of cotyledons, which happens at around 3 DAS in Arabidopsis, is an essential stage of seedling development. During this short time window, alterations in gene expression occur in a large number of genes in response to light (Ma et al., 2001). We identified approximate 3,000 genes showing transient changes in gene activity in hybrids at 3 DAS compared to those in parents; these transient changes result from earlier alteration of gene expression in hybrids relative to similar but late alteration in parents. One explanation for this is that the hybrids are at later developmental stages than their parents, which could be caused by the differences in germination time between the parents

and the hybrids (Figure 4.2). However, the qPCR verification results suggested that, at least for photosynthesis genes, the expression levels in the hybrids with delayed germination time are still higher than that in the parents (Figure 5.11). In addition, there are a large number of genes with changed expression levels in the transition phase from seeds to seedlings both in parents and in hybrids, but only a fraction of them (2,824 genes) are non-additively expressed at 3 DAS. These facts suggested that differences in germination time could not explain the changes in gene activity in hybrids at the early stages, and there must be other factors regulating non-additive gene expression.

8.2 Biomass heterosis is established in early stages of seedling development

Hybrids of various plant species have been reported to show heterosis as early as in embryogenesis. In broad beans (*Vicia faba*), size heterosis is shown in developing hybrid embryos compared to parental embryos, and this advantage remains in the mature seeds (Dieckmann and Link, 2010). Similar observations were documented in 6-day-old embryos of hybrids after pollination in maize (Jahnke et al., 2010, Meyer et al., 2007). In Arabidopsis, vigour in embryo size is only displayed in the hybrids derived from some certain species or ecotypes (Barth et al., 2003). However, there are some hybrid combinations, for example Col/C24 hybrids, showing similar seed size but greater cotyledon size compared to the parents at 3 and 4 DAS (Meyer et al., 2012). In our results, reciprocal L*er*/C24 hybrids show maternally influenced seed size. L*er* x C24 does not show the same heterosis in seed size and weight as the other reciprocal hybrid (Figure 4.1), but does show similar heterosis levels in seedling size after 11 DAS (Figure 4.3 B-D). These observations suggested that seed size heterosis is not required for generation of later heterosis in seedling size in hybrids, and heterosis in both reciprocal hybrids is likely

established during the early stages of seedling development.

8.3 Divergence in reciprocal hybrids

Reciprocal hybrids can show different heterotic phenotypes in plants. Obvious differences between reciprocal hybrids exist during embryogenesis and mature seed stages, as reported in hybrids in other Arabidopsis ecotypes (Groszmann et al., 2014), in maize hybrids (Meyer et al., 2007) and in mouse hybrids (Han et al., 2008). These differences are likely a result of maternal influences, since hybrids are in a maternal-supplying environment. In this project, we studied the differences between reciprocal hybrids at the stages of germinating seeds and young seedlings. Although reciprocal L*er*/C24 hybrids show different phenotypes in seed size and seed weight, they have similar heterosis levels relative to MPV in seedling size. Following germination, L*er* x C24 showed a greater growth rate than C24 x L*er*. Consistent observations in transcriptome analyses showed that L*er* x C24 slightly exceed C24 x L*er* in changing expression of a number of genes over the first week of early seedling growth, although the general trend of transcriptomes for the two reciprocal hybrids is that they become similar to each other after 3 DAS (Figure 5.2 C; Figure 5.3). This may suggest that the mechanisms resulting in growth vigour in hybrids could also cause greater growth vigour in L*er* x C24 compared to C24 x L*er* at the early stages of seedling growth.

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

The transcriptome results identified different groups of non-additive genes at different time points in young seedlings of hybrids. For example, the photosynthesis-related genes

are non-additively expressed at 3 DAS but have additive expression levels at subsequent time points. A similar observation in Col/C24 hybrids showed that chlorophyll biosynthesis genes are non-additively expressed only at 6 and 7 DAS (Fujimoto et al., 2012). By analysing expression patterns of non-additive genes over time, we found that the majority of non-additive expression is caused by earlier and transient changes in gene activity in hybrids At each time point, hybrids are ahead of parents in changing the expression levels of specific genes, and this causes non-additive gene expression in hybrids at early development stages (Figure 5.3; Figure 5.4). These altered patterns of gene expression do not persist for long. After each critical time point, instead of expression levels in hybrids returning back to MPV, the expression levels in parents become similar to those in hybrids. This indicates that early heterosis could result from dynamic patterns of non-additive expression over time caused by transient advantages in altering gene expression in the hybrids. Non-additive expression in mature seedlings could also result from early and transient changes in gene activity in hybrids.

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

Hybrids precede their parents not only in total transcript levels but also in allelic expression patterns. We found a large number of genes (Group A genes) showing high expression levels in L*er* relative to those in C24 but low L*er* allelic expression levels in both reciprocal hybrids at 0 DAS (Figure 7.1). By SNP analyses on transcriptome datasets from several time points, we observed allelic expression patterns in hybrid seeds are similar to those in parents at later time points (Figure 7.6). This indicated an early "set-up" of parental expression patterns from later stages, and once the ''set-up'' occurs

(perhaps during embryogenesis or even following fertilization), allelic expression patterns are not changed over time in hybrid seedlings (Figure 8.1).

By analysing allelic histone modification levels, we found that the changes of allelic expression could not be regulated by histone modifications (Figure 7.4). An explanation for this phenomenon is that the allelic expression at L*er* and C24 genomes could be controlled by different regulatory factors. The different allelic patterns in the parents at different stages are due to the activation or repression of these factors. In hybrids, for some unknown reasons, the regulation by C24 allele specific factors is activated earlier compared to in parents, resulting in the increased expression levels of the C24 alleles and unchanged levels of the L*er* alleles for a subset of genes in hybrid seeds. However, since this early ‘‘set-up’’ could not be related to non-additive gene expression (Figure 7.7), the importance of this finding to heterosis establishment is still unknown.

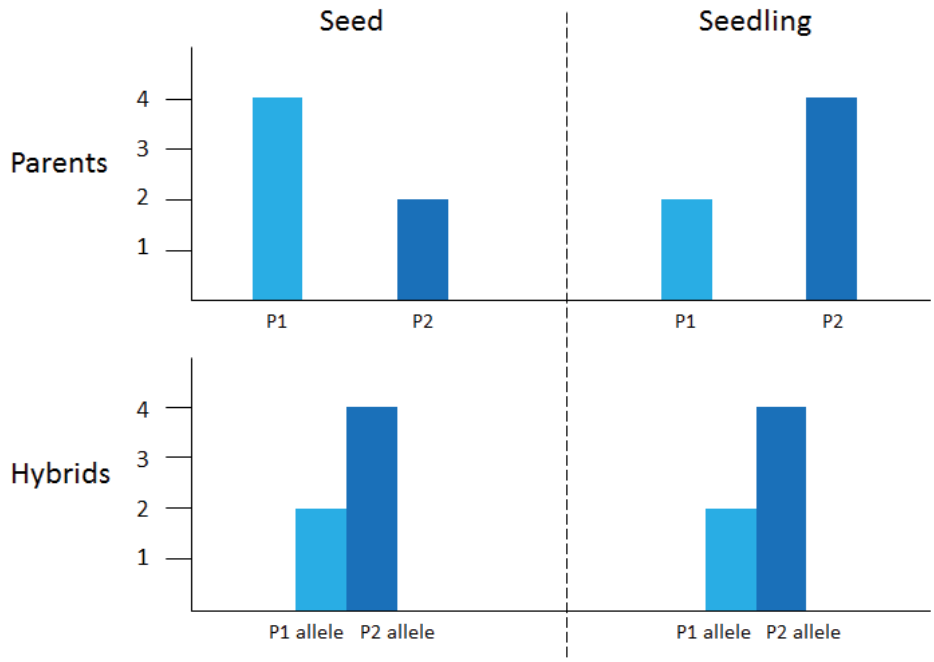


Figure 8.1. Models to describe the early “set-up” of allelic expression patterns in hybrid seeds. Ratios of gene expression levels (coloured columns) in parents can be changed from seed to seedling stages. But the seedling ratios were already established in hybrids at the seed stage, and they do not change in hybrids during the first week after sowing.

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

The regulatory factors responsive for the transcriptional advantages in hybrids at early developmental stages are still largely unknown. Our results suggested that the non-additive genes do not have common transcription factor-targeted sequences in their promoters. This indicated that non-additive gene expression is not solely regulated by a small group of transcription factors, although a transcription factor, MYB46, was found to target 12 non-additive genes and expression of *MYB46* is transiently up-regulated at 3 DAS (Figure 5.12).

We tested if non-additive gene expression could be regulated by histone modifications in hybrids. We identified strong positive correlations between gene expression and two active histone marks at 0 DAS when the two parents were compared to each other, but there is a poor correlation between them when the hybrids were compared to MPV (Figure 6.11), indicating that histone modifications may play only a limited role in regulating non-additive gene expression in hybrid seeds. This is not consistent with previous findings in rice and maize seedlings (He et al., 2010, He et al., 2013), in which histone modifications are associated with gene expression when comparing hybrids to maternal parents. This is probably because strong correlations between gene expression and histone modification can be detected only when the changes of transcript and histone modification levels are large. Our results showed that the differences between parents were usually larger than those between hybrids and MPV (Figure 5.10). In our analyses, we compare hybrids to MPV instead of comparing hybrids to the maternal parents, as our mission was to identify if histone modifications regulate non-additive gene expression. Another explanation of

the poor correlation could be that histone modifications may be involved in non-additive gene expression at later time points but not in seeds, as more genes are activated after seed germination.

Although histone modifications may not be associated with non-additive expression in seeds, we found increased levels of histone marks, especially K4me3 and K27me3, in a large number of genes in L*er* x C24 seeds, whereas the same patterns were not observed in C24 x L*er* (Figure 6.9, 6.10). These differences in histone modification patterns between reciprocal hybrids in seeds might be associated with the transcriptional differences between the reciprocal hybrids at 3 DAS, which could result in the differences between reciprocal hybrids in growth rate at later stages.

8.7 From heterotic transcriptome to heterotic metabolome

Hybrids have been shown to have a large number of genes with differential gene expression that are believed to be associated with heterosis, but whether or not hybrids show consistent changes in the proteome and in the metabolome is still largely unknown. A recent study suggested that Col/L*er* hybrids have transient decreased levels of lipid stored in seeds and transient increased levels of newly synthesized lipid compared to parents at early developmental stages of seedlings (Meyer et al., 2012). Since seed-stored lipid is the main energy source for seedlings before sugars are produced by photosynthesis in cotyledons, these findings suggested that hybrids could have higher consumption rate of the stored energy relative to their parents in young seedlings. These results also suggested that the transient up-regulated transcriptional levels of photosynthesis pathways in our results could lead to a greater ability to synthesize new lipid in hybrids

in early stages of seedling growth. As the lipid levels in parents eventually become similar to those in hybrids at later stages (Meyer et al., 2012), the transient differences of lipid levels are due to earlier changes of lipid levels in hybrids than in parents, similar to the trends we observed in hybrid transcriptomes. In addition to the differences in lipid levels, Meyer et al (Meyer et al., 2012) also observed the largest proportion of metabolites with changed levels in hybrids at 4 and 6 DAS, which may result from the changes in gene activities in hybrids at 3 DAS.

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

As phototrophic organisms, plants absorb light energy and convert it into consumable biological energy for plant growth and development. Hence, the superior biomass and growth rate in hybrids are likely due to the increase of net energy production. We identified up-regulated gene expression in photosynthesis, chlorophyll biosynthesis, thylakoid membrane organization, chloroplast relocation and starch biosynthesis in 3-DAS hybrids (Figure 5.5), implying a transient increased ability of hybrids in energy production probably due to the increased photosynthesis machinery. In addition to increased energy production, hybrids may also conserve energy by earlier cessation of some pathways, such as seed dormancy, seed germination and lipid storage, during the first week after sowing. All these observations indicate that hybrids may exceed parents in net energy available for plant growth in a short period of time. A key question is how this transient advantage of energy production in hybrids benefits the long-lasting biomass vigour at later stages.

One assumption is that the increased amount of energy around 3 DAS helps hybrids quickly establish early heterosis in cotyledon size. With increased cotyledon area, hybrids could capture more light energy to establish heterosis in true leaves at later stages. The importance of cotyledons in heterosis has been demonstrated in a recent experiment, in which Col/C24 hybrids lose biomass heterosis after blocking of photosynthesis in the cotyledons (Fujimoto et al., 2012). A similar experiment in Chinese cabbage showed that biomass heterosis disappeared when photosynthesis was blocked only in cotyledons but not in true leaves (Saeki et al., in press). It is possible that early heterosis is triggered in young cotyledons, and following this trigger hybrids develop an ongoing increased management of energy production due to larger sizes of cotyledons and leaves. The increase in total energy production in hybrids could maintain biomass heterosis at later development stages. The early heterosis in cotyledons may be the key to unlocking the long-lasting biomass heterosis at later development stages.

8.9 Future work

In this project, we showed that early heterosis in plant growth could be caused by altered expression levels in some genes. Early heterosis may be the key element to trigger long-lasting biomass heterosis at later developmental stages. To test this theory, further supporting evidence will be needed from different approaches.

Although measurements on cotyledon size have been performed to track growth vigour of L*er*/C24 hybrids from 7 DAS, accurate morphological measurements on young cotyledons of hybrids are required. Day-by-day tracking of cell size, cell number, chloroplast number and cotyledon area and thickness after seed germination could be

helpful to demonstrate the exact timing of the emergence of early phenotypic heterosis. Gene expression patterns at multiple time points could be analysed in Arabidopsis hybrids at later development stage, to test if non-additive gene expression at later development stages is also caused by early and transient changes in gene activity in hybrids. Photosynthesis assays could be performed in hybrids at around 3 DAS by measuring activity and concentrations of photosynthesis-related enzymes. To prove that hybrids have greater efficiency of energy production and consumption, starch and sugar assays could be performed in young cotyledons. Assuming that hybrids from different Arabidopsis ecotypes and species undergo similar fundamental changes at molecular levels to achieve heterosis, whether photosynthesis-related genes are up-regulated in young cotyledons of other Arabidopsis hybrids would be tested. As the mechanism of the early and transient changes in gene activity is still not clear, genome-wide histone modification patterns at 3 DAS could be investigated. It would also be worthwhile to take other known regulatory factors of gene expression into consideration, such as DNA methylation and small RNA, as these epigenetic systems could be involved in regulation of non-additive expression.

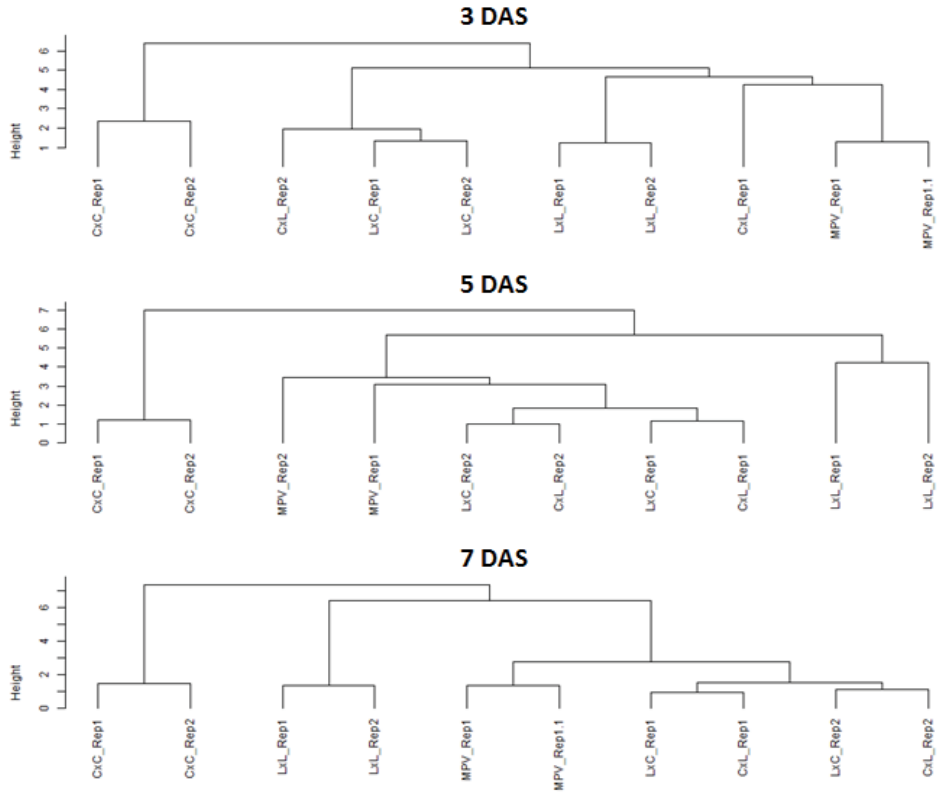
8.10 Conclusions

- Reciprocal hybrids differ from each other in gene expression and embryo size in germinating seeds, but they show increasingly similar transcription patterns and seedling size through the development of young seedlings.
- Hybrids are transiently ahead of parents in changing gene expression during early stages of plant development. These transient advantages may lead to non-additive gene expression in hybrids at early development stages.
- Poor correlations were found between altered gene expression and altered histone

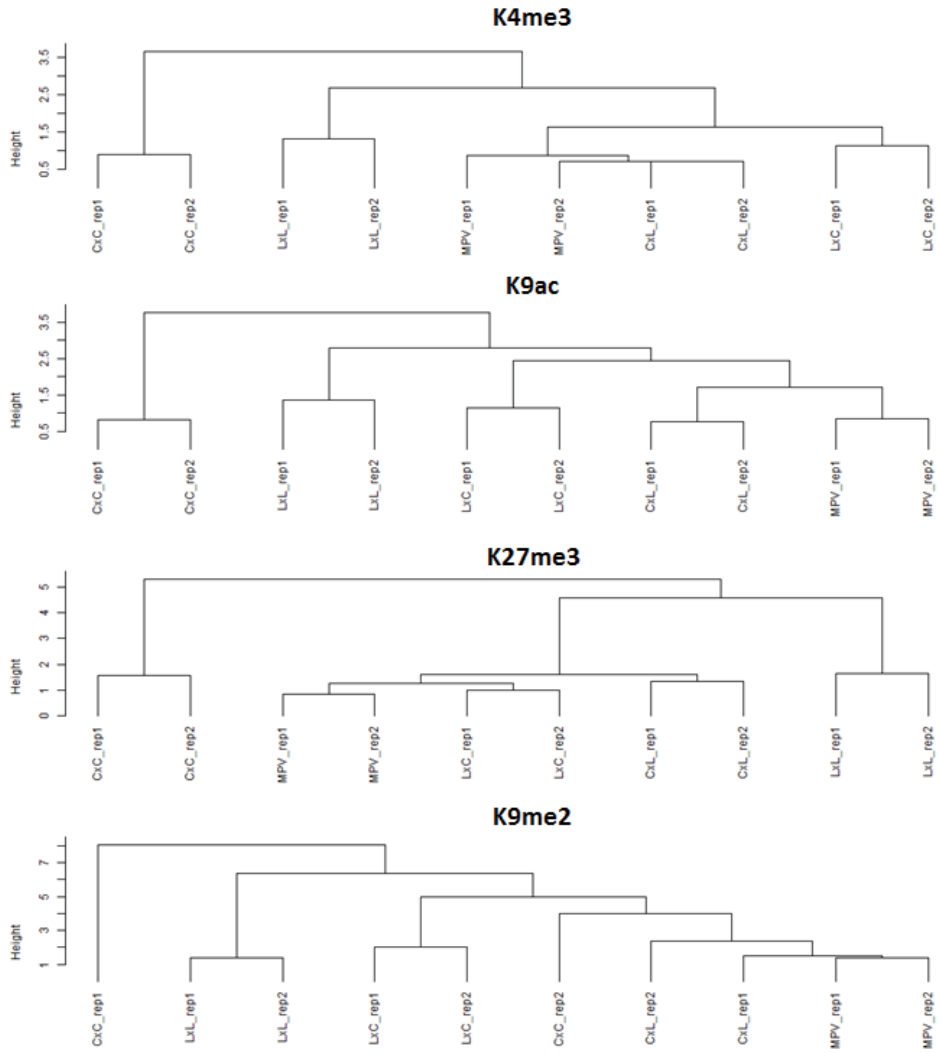
modifications in hybrid seeds. The regulatory mechanisms of non-additive gene expression at early stages is still not known.

- Energy production in young seedlings of hybrids could be elevated for a short period of time. This may help hybrids establish heterosis in cotyledons, which might be critical for long-term heterosis during the life cycle of the plant.

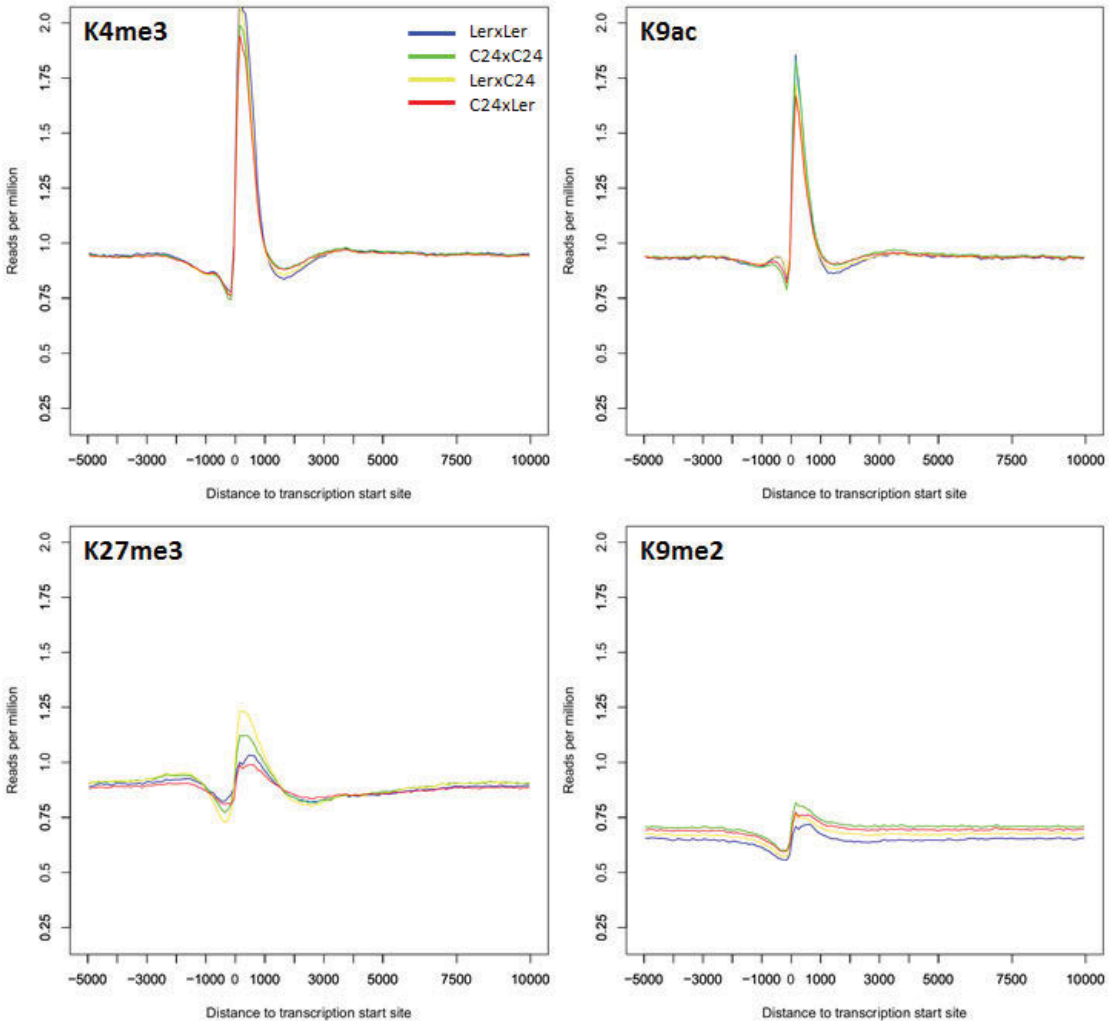
Appendices



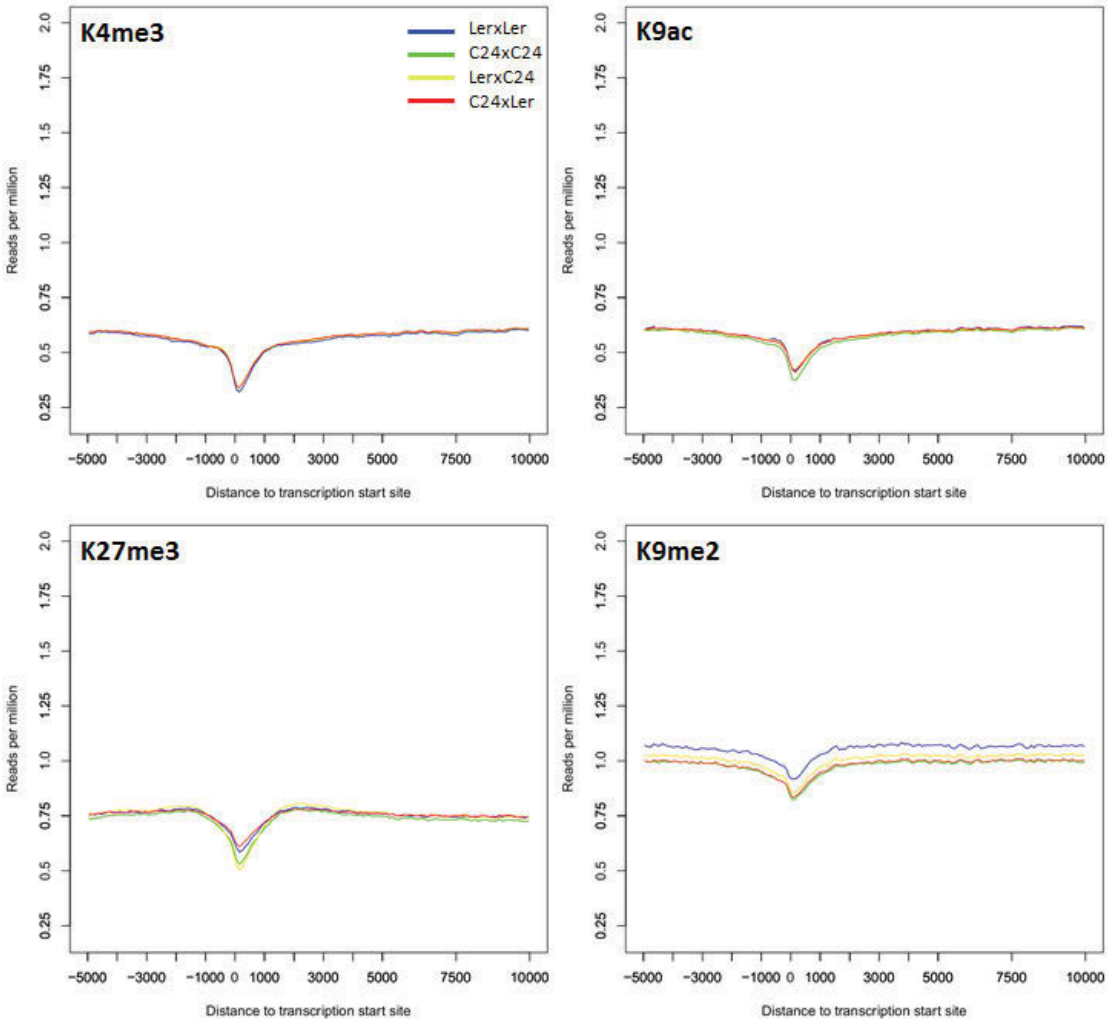
Appendix 1. Dendrogram shows the similarity of RNA-Seq libraries in two biological replicates of the parents and the hybrids at 3, 5 and 7 DAS.



Appendix 2. Dendrogram shows the similarity of ChIP-Seq libraries in two biological replicates of the parents and the hybrids at 0 DAS.



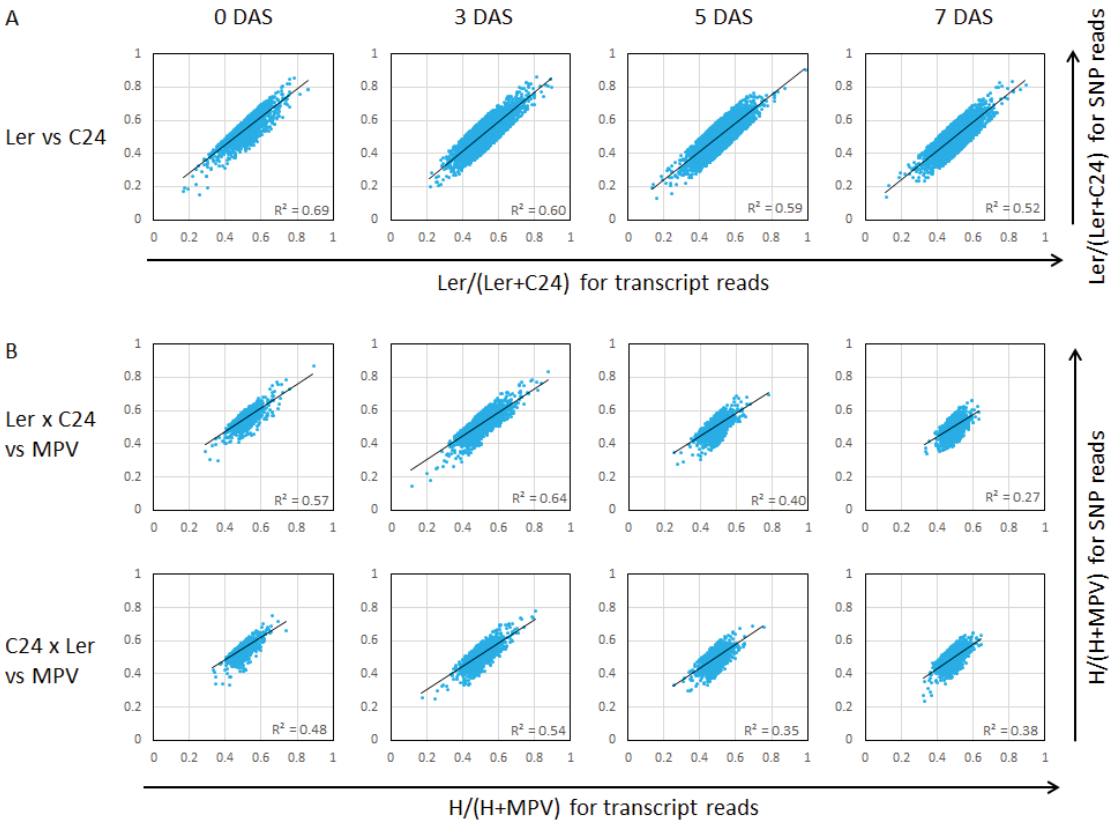
Appendix 3. Average levels of histone modification signals around TSS of protein-coding genes in the hybrids and the parents at 0 DAS.



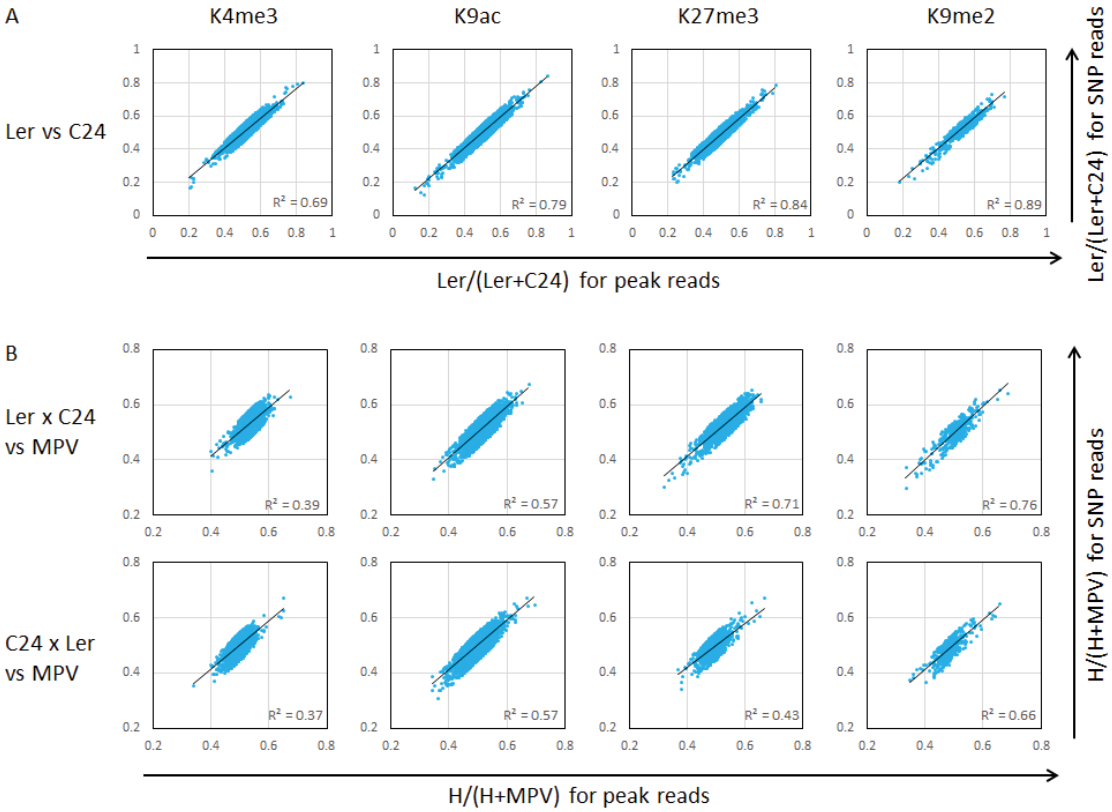
Appendix 4. Average levels of histone modification signals around 5' start sites of TEs in the hybrids and the parents at 0 DAS.

	Expression						K4me3						K9ac						K27me3						K9me2					
	Ler vs C24		LxC vs MPV		CxL vs MPV		Ler vs C24		LxC vs MPV		CxL vs MPV		Ler vs C24		LxC vs MPV		CxL vs MPV		Ler vs C24		LxC vs MPV		CxL vs MPV		Ler vs C24		LxC vs MPV		CxL vs MPV	
	L>C	L<C	>MPV	<MPV	>MPV	<MPV	L>C	L<C	>MPV	<MPV	>MPV	<MPV	L>C	L<C	>MPV	<MPV	>MPV	<MPV	L>C	L<C	>MPV	<MPV	>MPV	<MPV	L>C	L<C	>MPV	<MPV	>MPV	<MPV
AT1G01580		1		1																	1									
AT1G01640		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT1G14780			1												1										N/A	N/A	N/A	N/A	N/A	N/A
AT1G16840		1	1		1										1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT1G19394			1						1										N/A	N/A	N/A	N/A	N/A	N/A						
AT1G19396			1						1										N/A	N/A	N/A	N/A	N/A	N/A						
AT1G25550			1						1											1					N/A	N/A	N/A	N/A	N/A	N/A
AT1G27540	1		1				1		1				1		1				N/A	N/A	N/A	N/A	N/A	N/A	1					
AT1G33940				1			N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1		1									1
AT1G34580		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT1G45120			1					1	1					1					N/A	N/A	N/A	N/A	N/A	N/A	1					
AT1G49380			1						1										N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT1G51190		1		1										1					1		1				N/A	N/A	N/A	N/A	N/A	N/A
AT1G59700		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT1G66480			1						1				1												N/A	N/A	N/A	N/A	N/A	N/A
AT1G71140				1																	1									
AT1G71150		1		1			N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A			1				N/A	N/A	N/A	N/A	N/A	N/A
AT1G72570		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT1G72580		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT2G18570		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT2G19640				1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT2G23240		1		1			1														1									
AT2G30480		1	1												1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT2G34360		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT2G38940		1		1		1	N/A	N/A	N/A	N/A	N/A	N/A									1				N/A	N/A	N/A	N/A	N/A	N/A
AT2G40095	1		1				1		1						1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT3G04230		1		1																	1				N/A	N/A	N/A	N/A	N/A	N/A
AT3G19030			1						1																N/A	N/A	N/A	N/A	N/A	N/A
AT3G21600		1	1					1	1				N/A	N/A	N/A	N/A	N/A	N/A		1					N/A	N/A	N/A	N/A	N/A	N/A
AT3G26510		1	1											1	1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT3G28330	1			1			N/A	N/A	N/A	N/A	N/A	N/A	1								1				N/A	N/A	N/A	N/A	N/A	N/A
AT3G44716		1	1						1										N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT3G59190	1			1																1	1				N/A	N/A	N/A	N/A	N/A	N/A
AT3G62460	1			1			1						1								1				N/A	N/A	N/A	N/A	N/A	N/A
AT4G00895			1						1										N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT4G09590	1					1							1					1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT4G11960	1		1						1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT4G14440		1		1									N/A	N/A	N/A	N/A	N/A	N/A			1				N/A	N/A	N/A	N/A	N/A	N/A
AT4G28811				1	1								N/A	N/A	N/A	N/A	N/A	N/A			1				N/A	N/A	N/A	N/A	N/A	N/A
AT5G10140		1	1		1														1			1		1	N/A	N/A	N/A	N/A	N/A	N/A
AT5G13220	1		1						1																N/A	N/A	N/A	N/A	N/A	N/A
AT5G35960			1		1										1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT5G39640	1		1		1		1		1				1		1		1		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT5G40382		1		1									N/A	N/A	N/A	N/A	N/A	N/A			1									
AT5G45040			1						1										N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT5G48570		1	1		1										1				N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
AT5G54000		1		1									N/A	N/A	N/A	N/A	N/A	N/A			1				N/A	N/A	N/A	N/A	N/A	N/A

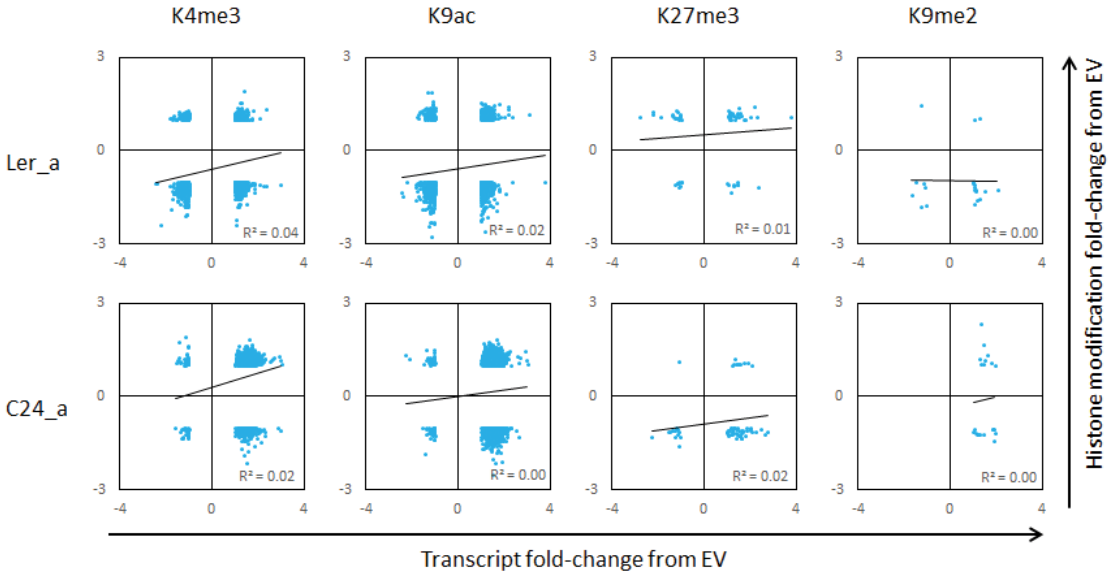
Appendix 5. Non-additively expressed genes with consistent changes in histone modifications in hybrids at 0 DAS. Significant genes were marks with '1'. N/A: no histone modification peak overlapped.



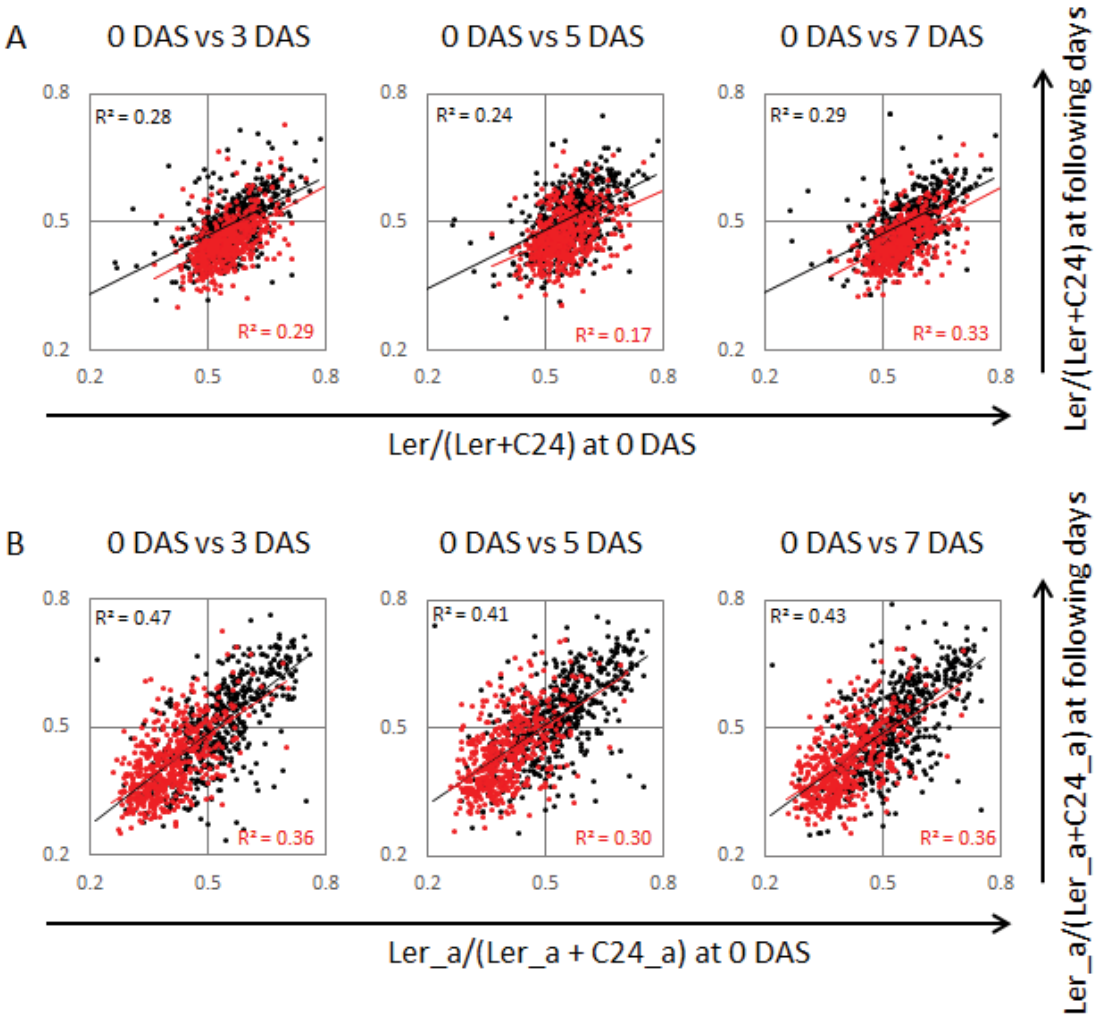
Appendix 6. Correlations between SNP reads and transcript reads at four time points. (A) Correlations between changes in parents based on transcript reads and based on SNP reads. **(B)** Correlations between changes in hybrids against MPV based on transcript reads and based on SNP reads. H, Hybrid.



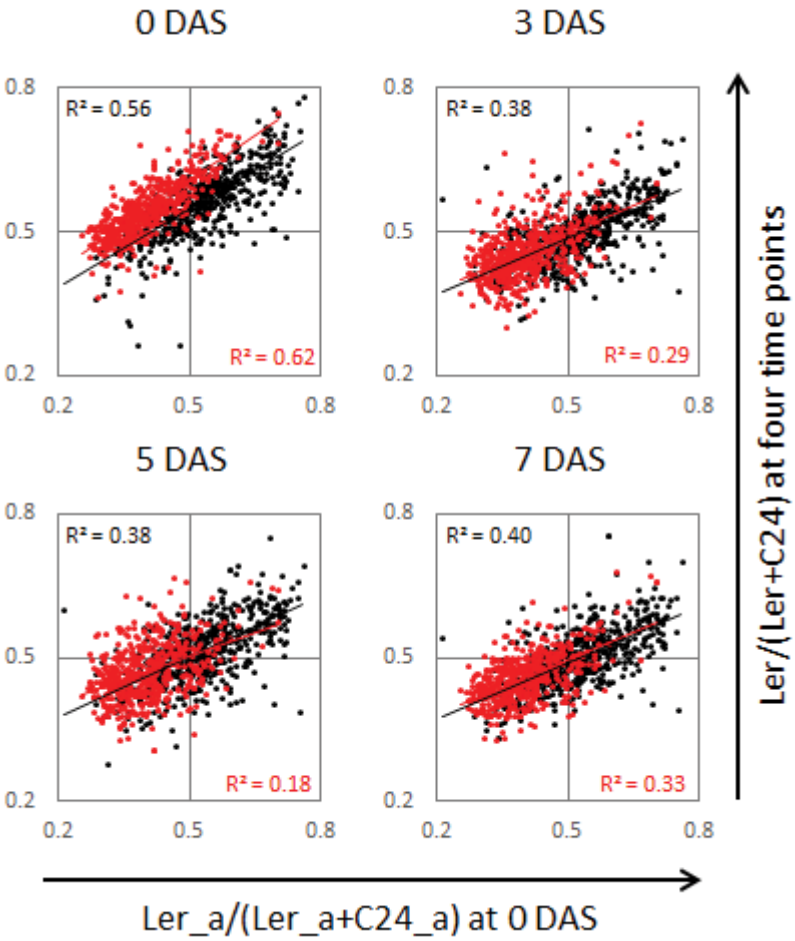
Appendix 7. Correlations between SNP reads and peak reads at four histone marks. (A) Correlations between changes in parents based on peak reads and based on SNP reads. **(B)** Correlations between changes in hybrids against MPV based on peak reads and based on SNP reads. H, Hybrid.



Appendix 8. Changes of gene expression against changes of histone marks at Ler and C24 alleles in C24 x Ler. Dots represent the genes targeted by histone modifications.



Appendix 9. Correlations of parental (A) and allelic (B) proportions in C24 x Ler between 0 DAS and the following three time points. Red dots represent Group A genes (with up-regulated expression levels at C24 allele at 0 DAS) and black dots represent Group B genes (with unchanged and down-regulated expression levels at C24 allele at 0 DAS). To determine changed expression levels at alleles, fold-change from EV ≥ 1.5.



Appendix 10. Allelic proportions against parental proportions in C24 x Ler at the four time points. Red dots represent Group A genes (with up-regulated expression levels at C24 allele at 0 DAS) and black dots represent Group B genes (with unchanged and down-regulated expression levels at C24 allele at 0 DAS). To determine changed expression levels at alleles, fold-change from EV ≥ 1.5.

Bibliography

ALBERTS, B. 2008. *Molecular biology of the cell,* New York, Garland Science.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J. Y., LI, W. W. & NOBLE, W. S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research,* 37, W202-W208.

BAKER, N. R. 2008. Chlorophyll fluorescence: A probe of photosynthesis in vivo. *Annual Review of Plant Biology,* 59, 89-113.

BAKER, N. R., HARBINSON, J. & KRAMER, D. M. 2007. Determining the limitations and regulation of photosynthetic energy transduction in leaves. *Plant Cell and Environment,* 30, 1107-1125.

BARTEE, L., MALAGNAC, F. & BENDER, J. 2001. Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes & Development,* 15, 1753-1758.

BARTH, S., BUSIMI, A. K., FRIEDRICH UTZ, H. & MELCHINGER, A. E. 2003. Heterosis for biomass yield and related traits in five hybrids of Arabidopsis thaliana L. Heynh. *Heredity (Edinb),* 91, 36-42.

BELL, A. C. & FELSENFELD, G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature,* 405, 482-5.

BENTSINK, L. & KOORNNEEF, M. 2008. Seed dormancy and germination. *Arabidopsis Book,* 6, e0119.

BERNATAVICHUTE, Y. V., ZHANG, X., COKUS, S., PELLEGRINI, M. & JACOBSEN, S. E. 2008. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS One,* 3, e3156.

BRUCE, A. B. 1910. The Mendelian Theory of Heredity and the Augmentation of Vigor. *Science,* 32, 627-8.

CAO, X. F. & JACOBSEN, S. E. 2002. Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Current Biology,* 12, 1138-1144.

CHEN, Z. J. 2010. Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci,* 15, 57-71.

CHINNUSAMY, V. & ZHU, J. K. 2009. Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol,* 12, 133-9.

COKUS, S. J., FENG, S. H., ZHANG, X. Y., CHEN, Z. G., MERRIMAN, B., HAUDENSCHILD, C. D., PRADHAN, S., NELSON, S. F., PELLEGRINI, M. & JACOBSEN, S. E. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature,* 452, 215-219.

CROW, J. F. 1998. 90 years ago: the beginning of hybrid maize. *Genetics,* 148, 923-8.

CZECHOWSKI, T., STITT, M., ALTMANN, T., UDVARDI, M. K. & SCHEIBLE, W. R. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology,* 139, 5-17.

DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C. A., BANKS, E., DEPRISTO, M. A., HANDSAKER, R. E., LUNTER, G., MARTH, G. T., SHERRY, S. T., MCVEAN, G., DURBIN, R. & GRP, G. P. A. 2011. The variant call format and VCFtools. *Bioinformatics,* 27, 2156-2158.

DAPP, M., REINDERS, J., BÉDIÉE, A., BALSERA, C., BUCHER, E., THEILER, G., GRANIER, C. & PASZKOWSKI, J. 2015. Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. *Nature Plants,* 1, 15092.

DAVENPORT, C. B. 1908. Degeneration, Albinism and Inbreeding. *Science,* 28, 454-5.

DEKKERS, B. J. W., WILLEMS, L., BASSEL, G. W., VAN BOLDEREN-VELDKAMP, R. P., LIGTERINK, W., HILHORST, H. W. M. & BENTSINK, L. 2012. Identification of Reference Genes for RT-qPCR Expression Analysis in Arabidopsis and Tomato Seeds. *Plant and Cell Physiology,* 53, 28-37.

DELERIS, A., STROUD, H., BERNATAVICHUTE, Y., JOHNSON, E., KLEIN, G., SCHUBERT, D. & JACOBSEN, S. E. 2012. Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. *Plos Genetics,* 8.

DIECKMANN, S. & LINK, W. 2010. Quantitative genetic analysis of embryo heterosis in faba bean (Vicia faba L.). *Theoretical and Applied Genetics,* 120, 261-270.

DONG, X., REIMER, J., GOBEL, U., ENGELHORN, J., HE, F., SCHOOF, H. & TURCK, F. 2012. Natural variation of H3K27me3 distribution between two Arabidopsis accessions and its association with flanking transposable elements. *Genome Biology,* 13.

DU, Z., ZHOU, X., LING, Y., ZHANG, Z. H. & SU, Z. 2010. agriGO: a GO analysis toolkit for the agricultural

community. *Nucleic Acids Research,* 38, W64-W70.
EASTMOND, P. J. 2006. SUGAR-DEPENDENT1 encodes a patatin domain triacylglycerol lipase that initiates storage oil breakdown in germinating Arabidopsis seeds. *Plant Cell,* 18, 665-675.
ENKE, R. A., DONG, Z. & BENDER, J. 2011. Small RNAs prevent transcription-coupled loss of histone H3 lysine 9 methylation in Arabidopsis thaliana. *PLoS Genet,* 7, e1002350.
ERNST, J. & BAR-JOSEPH, Z. 2006. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics,* 7, 191.
FINNEGAN, E. J., BRETTELL, R. I. & DENNIS, E. S. 1993. The role of DNA methylation in the regulation of plant gene expression. *EXS,* 64, 218-61.
FINNEGAN, E. J. & DENNIS, E. S. 1993. Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. *Nucleic Acids Res,* 21, 2383-8.
FRANCO-ZORRILLA, J. M., LOPEZ-VIDRIERO, I., CARRASCO, J. L., GODOY, M., VERA, P. & SOLANO, R. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences of the United States of America,* 111, 2367-2372.
FUJIMOTO, R., TAYLOR, J. M., SHIRASAWA, S., PEACOCK, W. J. & DENNIS, E. S. 2012. Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proceedings of the National Academy of Sciences of the United States of America,* 109, 7109-7114.
GREAVES, I. K., GROSZMANN, M., WANG, A. H., PEACOCK, W. J. & DENNIS, E. S. 2014. Inheritance of Trans Chromosomal Methylation patterns from Arabidopsis F1 hybrids. *Proceedings of the National Academy of Sciences of the United States of America,* 111, 2017-2022.
GREAVES, I. K., GROSZMANN, M., YING, H., TAYLOR, J. M., PEACOCK, W. J. & DENNIS, E. S. 2012. Trans chromosomal methylation in Arabidopsis hybrids. *Proc Natl Acad Sci U S A,* 109, 3570-5.
GROSZMANN, M., GONZALEZ-BAYON, R., GREAVES, I. K., WANG, L., HUEN, A. K., PEACOCK, W. J. & DENNIS, E. S. 2014. Intraspecific Arabidopsis Hybrids Show Different Patterns of Heterosis Despite the Close Relatedness of the Parental Genomes. *Plant Physiology,* 166, 265-+.
GROSZMANN, M., GONZALEZ-BAYON, R., LYONS, R. L., GREAVES, I. K., KAZAN, K., PEACOCK, W. J. & DENNIS, E. S. 2015. Hormone-regulated defense and stress response networks contribute to heterosis in Arabidopsis F1 hybrids. *Proc Natl Acad Sci U S A*.
GROSZMANN, M., GREAVES, I. K., ALBERT, N., FUJIMOTO, R., HELLIWELL, C. A., DENNIS, E. S. & PEACOCK, W. J. 2011a. Epigenetics in plants-vernalisation and hybrid vigour. *Biochim Biophys Acta,* 1809, 427-37.
GROSZMANN, M., GREAVES, I. K., ALBERTYN, Z. I., SCOFIELD, G. N., PEACOCK, W. J. & DENNIS, E. S. 2011b. Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci U S A,* 108, 2617-22.
HA, M., LU, J., TIAN, L., RAMACHANDRAN, V., KASSCHAU, K. D., CHAPMAN, E. J., CARRINGTON, J. C., CHEN, X., WANG, X.-J. & CHEN, Z. J. 2009. Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids. *Proceedings of the National Academy of Sciences,* 106, 17835-17840.
HA, M., NG, D. W. K., LI, W. H. & CHEN, Z. J. 2011. Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Research,* 21, 590-598.
HAN, Z., MTANGO, N. R., PATEL, B. G., SAPIENZA, C. & LATHAM, K. E. 2008. Hybrid vigor and transgenerational epigenetic effects on early mouse embryo phenotype. *Biol Reprod,* 79, 638-48.
HE, G., CHEN, B., WANG, X., LI, X., LI, J., HE, H., YANG, M., LU, L., QI, Y., WANG, X. & WANG DENG, X. 2013. Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biology,* 14, R57.
HE, G. M., ZHU, X. P., ELLING, A. A., CHEN, L. B., WANG, X. F., GUO, L., LIANG, M. Z., HE, H., ZHANG, H. Y., CHEN, F. F., QI, Y. J., CHEN, R. S. & DENG, X. W. 2010. Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *Plant Cell,* 22, 17-33.
HERRERA, C. M. & BAZAGA, P. 2010. Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet Viola cazorlensis. *New Phytol,* 187, 867-76.
HOPKINS, R. J., VAN DAM, N. M. & VAN LOON, J. J. 2009. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annu Rev Entomol,* 54, 57-83.
JAHNKE, S., SARHOLZ, B., THIEMANN, A., KUHR, V., GUTIERREZ-MARCOS, J. F., GEIGER, H. H., PIEPHO, H.

P. & SCHOLTEN, S. 2010. Heterosis in early seed development: a comparative study of F1 embryo and endosperm tissues 6 days after fertilization. *Theor Appl Genet,* 120, 389-400.

KINOSHITA, T., MIURA, A., CHOI, Y. H., KINOSHITA, Y., CAO, X. F., JACOBSEN, S. E., FISCHER, R. L. & KAKUTANI, T. 2004. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science,* 303, 521-523.

KLOC, A., ZARATIEGUI, M., NORA, E. & MARTIENSSEN, R. 2008. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr Biol,* 18, 490-5.

KRIEGER, U., LIPPMAN, Z. B. & ZAMIR, D. 2010. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat Genet,* 42, 459-63.

LAWRENCE, R. J., EARLEY, K., PONTES, O., SILVA, M., CHEN, Z. J., NEVES, N., VIEGAS, W. & PIKAARD, C. S. 2004. A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Molecular Cell,* 13, 599-609.

LEVENSON, J. M. & SWEATT, J. D. 2005. Epigenetic mechanisms in memory formation. *Nature Reviews Neuroscience,* 6, 108-118.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & PROC, G. P. D. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25, 2078-2079.

LINDROTH, A. M., CAO, X. F., JACKSON, J. P., ZILBERMAN, D., MCCALLUM, C. M., HENIKOFF, S. & JACOBSEN, S. E. 2001. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science,* 292, 2077-2080.

LIPPMAN, Z., GENDREL, A. V., BLACK, M., VAUGHN, M. W., DEDHIA, N., MCCOMBIE, W. R., LAVINE, K., MITTAL, V., MAY, B., KASSCHAU, K. D., CARRINGTON, J. C., DOERGE, R. W., COLOT, V. & MARTIENSSEN, R. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature,* 430, 471-476.

MA, L. G., LI, J. M., QU, L. J., HAGER, J., CHEN, Z. L., ZHAO, H. Y. & DENG, X. W. 2001. Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell,* 13, 2589-2607.

MANSFIELD, S. G. & BRIARTY, L. G. 1992. Cotyledon cell development in Arabidopsis thaliana during reserve deposition. *Canadian Journal of Botany,* 70, 151-164.

MEYER, R. C., KUSTERER, B., LISEC, J., STEINFATH, M., BECHER, M., SCHARR, H., MELCHINGER, A. E., SELBIG, J., SCHURR, U., WILLMITZER, L. & ALTMANN, T. 2010. QTL analysis of early stage heterosis for biomass in Arabidopsis. *Theor Appl Genet,* 120, 227-37.

MEYER, R. C., TORJEK, O., BECHER, M. & ALTMANN, T. 2004. Heterosis of biomass production in Arabidopsis. Establishment during early development. *Plant Physiol,* 134, 1813-23.

MEYER, R. C., WITUCKA-WALL, H., BECHER, M., BLACHA, A., BOUDICHEVSKAIA, A., DORMANN, P., FIEHN, O., FRIEDEL, S., VON KORFF, M., LISEC, J., MELZER, M., REPSILBER, D., SCHMIDT, R., SCHOLZ, M., SELBIG, J., WILLMITZER, L. & ALTMANN, T. 2012. Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant J,* 71, 669-83.

MEYER, S., POSPISIL, H. & SCHOLTEN, S. 2007. Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. *Plant Mol Biol,* 63, 381-91.

MILLER, M., ZHANG, C. Q. & CHEN, Z. J. 2012. Ploidy and Hybridity Effects on Growth Vigor and Gene Expression in Arabidopsis thaliana Hybrids and Their Parents. *G3-Genes Genomes Genetics,* 2, 505-513.

MOGHADDAM, A. M., ROUDIER, F., SEIFERT, M., BERARD, C., MAGNIETTE, M. L., ASHTIYANI, R. K., HOUBEN, A., COLOT, V. & METTE, M. F. 2011. Additive inheritance of histone modifications in Arabidopsis thaliana intra-specific hybrids. *Plant J,* 67, 691-700.

MULLER, P., LI, X. P. & NIYOGI, K. K. 2001. Non-photochemical quenching. A response to excess light energy. *Plant Physiology,* 125, 1558-1566.

NI, Z. F., KIM, E. D., HA, M. S., LACKEY, E., LIU, J. X., ZHANG, Y. R., SUN, Q. X. & CHEN, Z. J. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature,* 457, 327-U7.

PERRY, S. E. & WANG, H. 2003. Rapid isolation of Arabidopsis thaliana developing embryos. *Biotechniques,* 35, 278-80, 282.

POWERS, L. 1944. An Expansion of Jones's Theory for the Explanation of Heterosis. *The American Naturalist,* 78, 275-280.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26, 841-842.

RADOEV, M., BECKER, H. C. & ECKE, W. 2008. Genetic analysis of heterosis for yield and yield components in rapeseed (Brassica napus L.) by quantitative trait locus mapping. *Genetics,* 179, 1547-58.

REIK, W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature,* 447, 425-32.

ROUDIER, F., AHMED, I., BERARD, C., SARAZIN, A., MARY-HUARD, T., CORTIJO, S., BOUYER, D., CAILLIEUX, E., DUVERNOIS-BERTHET, E., AL-SHIKHLEY, L., GIRAUT, L., DESPRES, B., DREVENSEK, S., BARNECHE, F., DEROZIER, S., BRUNAUD, V., AUBOURG, S., SCHNITTGER, A., BOWLER, C., MARTIN-MAGNIETTE, M. L., ROBIN, S., CABOCHE, M. & COLOT, V. 2011. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *Embo Journal,* 30, 1928-1938.

ROUDIER, F., TEIXEIRA, F. K. & COLOT, V. 2009. Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. *Trends Genet,* 25, 511-7.

SAEKI, N., KAWANABE, T., YING, H., SHIMIZU, M., KOJIMA, M., ABE, H., OKAZAKI, K., KAJI, M., TAYLOR, J. M., SAKAKIBARA, H., PEACOCK, W. J., DENNIS, E. S. & FUJIMOTO, R. in press. Molecular and cellular characteristics of hybrid vigour in a commercial hybrid of Chinese cabbage. *BMC Plant Biology*.

SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods,* 9, 671-675.

SEGAL, E. & WIDOM, J. 2009. What controls nucleosome positions? *Trends Genet,* 25, 335-43.

SEQUEIRA-MENDES, J., ARAGUEZ, I., PEIRO, R., MENDEZ-GIRALDEZ, R., ZHANG, X. Y., JACOBSEN, S. E., BASTOLLA, U. & GUTIERREZ, C. 2014. The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *Plant Cell,* 26, 2351-2366.

SHEN, H., HE, H., LI, J., CHEN, W., WANG, X., GUO, L., PENG, Z., HE, G., ZHONG, S., QI, Y., TERZAGHI, W. & DENG, X. W. 2012. Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. *Plant Cell,* 24, 875-92.

SHEN, L., SHAO, N. Y., LIU, X. C. & NESTLER, E. 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *Bmc Genomics,* 15.

SHULL, G. H. 1908. The Composition of a Field of Maize. *Journal of Heredity,* os-4, 296-301.

SOPPE, W. J. J., JACOBSEN, S. E., ALONSO-BLANCO, C., JACKSON, J. P., KAKUTANI, T., KOORNNEEF, M. & PEETERS, A. J. M. 2000. The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell,* 6, 791-802.

STROUD, H., DO, T., DU, J. M., ZHONG, X. H., FENG, S. H., JOHNSON, L., PATEL, D. J. & JACOBSEN, S. E. 2014. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology,* 21, 64-+.

SUGIYAMA, T., CAM, H. P., SUGIYAMA, R., NOMA, K., ZOFALL, M., KOBAYASHI, R. & GREWAL, S. I. 2007. SHREC, an effector complex for heterochromatic transcriptional silencing. *Cell,* 128, 491-504.

VERDEL, A., JIA, S., GERBER, S., SUGIYAMA, T., GYGI, S., GREWAL, S. I. & MOAZED, D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science,* 303, 672-6.

WOOD, A. J. & OAKEY, R. J. 2006. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet,* 2, e147.

YUAN, L. P. 1998. Hybrid rice breeding in China. *Advances in Hybrid Rice Technology. Philippines: International Rice Research Institute*, 27-33.

ZAIDI, S. K., YOUNG, D. W., MONTECINO, M., LIAN, J. B., STEIN, J. L., VAN WIJNEN, A. J. & STEIN, G. S. 2010. Architectural Epigenetics: Mitotic Retention of Mammalian Transcriptional Regulatory Information. *Molecular and Cellular Biology,* 30, 4758-4766.

ZHANG, H. & ZHU, J. K. 2011. RNA-directed DNA methylation. *Curr Opin Plant Biol,* 14, 142-7.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology,* 9.

ZHONG, R., RICHARDSON, E. A. & YE, Z. H. 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. *Plant Cell,* 19, 2776-2792.