12-9-2016

A Genome-Wide Characterization of Differentially Expressed Genes Encoding mRNAs and miRNAs and Methylation Analysis of Phytochrome Genes in a Cotton Phytochrome A1 RNAi line

Qing Miao

A genome-wide characterization of differentially expressed genes encoding mRNAs and miRNAs and methylation analysis of phytochrome genes in a cotton phytochrome A1 RNAi line

By

Qing Miao

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Molecular Biology
in the Department of Biochemistry, Molecular Biology, Entomology & Plant Pathology

Mississippi State, Mississippi

December 2016

A genome-wide characterization of differentially expressed genes encoding mRNAs and miRNAs and methylation analysis of phytochrome genes in a cotton phytochrome A1 RNAi line

By

Qing Miao

Approved:

_____

Din-Pow Ma
(Director of Dissertation)

_____

Johnie N. Jenkins
(Committee Member)

_____

Sukumar Saha
(Committee Member)

_____

Jiaxu Li
(Committee Member)

_____

Kenneth Willeford
(Committee Member/Graduate Coordinator)

_____

George M. Hopper
Dean
College of Agriculture and Life Sciences

Name: Qing Miao

Date of Degree: December 9, 2016

Institution: Mississippi State University

Major Field: Molecular Biology

Director of Dissertation: Din-Pow Ma

Title of Study: A genome-wide characterization of differentially expressed genes encoding mRNAs and miRNAs and methylation analysis of phytochrome genes in a cotton phytochrome A1 RNAi line

Pages in Study 131

Candidate for Degree of Doctor of Philosophy

Silencing phytochrome A1 gene (*PHYA1*) by RNA interference in upland cotton (*Gossypium hirsutum* L. cv. Coker 312) had generated *PHYA1* RNAi lines with increased fiber length, strength and low micronair (finer fiber). In order to identify and characterize mRNAs and miRNAs that are differentially expressed in the RNAi plants, transcriptome and miRNAome analyses via high-throughput RNA sequencing were performed. Total RNA isolated from 10-DPA (days post anthesis) fibers and small RNAs isolated from 5-, 10-, and 15-DPA fibers of RNAi and Coker 312 lines were used to construct 6 RNA libraries and 18 small RNA libraries, respectively, which were sequenced using the Illumina HiSeq system. A total of 142 differentially expressed genes (DEGs) were identified in *PHYA1* RNAi compared to Coker 312. GO analysis showed that these DEGs were mainly involved in metabolic pathways, binding and regulating enzymes (hydrolase, transferase, and oxidoreductase activities), and cell structures which were reported to play important roles in fiber development. Twenty-eight KEGG pathways were mapped for 142 DEGs, and the pathways related to glycolysis/gluconeogenesis and pyruvate metabolism were the most abundant, followed by cytochrome P450-involved pathways.

Sixty-one conserved miRNA families and thirty-five novel miRNAs were identified in upland cotton. The targets of 6 conserved miRNAs, which expressed differentially in the RNAi line, were reported to participate in primary cell wall synthesis and phytohormone signaling pathways. The 35 novel miRNAs were identified in cotton for the first time, and their target genes were predicted. Nine novel miRNAs were identified to target cytochrome P450 TBP. Together, the results imply that miRNAs involved in fine-tune gene regulation might confer to the phenotype of the RNAi line with improved fiber quality.

Besides characterizing mRNAs and miRNAs, the CpG site methylation status within coding regions of phytochrome genes in RNAi line in leaves and 10-DPA fibers was determined using bisulfite genomic sequencing. The *PHYA1*, *PHYC* and *PHYE* in RNAi line had higher methylation levels in leaves than those in Coker 312, but *PHYB* had lower methylation levels. In fibers, the methylation levels of *PHYB* also decreased in RNAi plants. However, the methylation of other phytochrome genes showed no significant changes.

DEDICATION

To my parents, my sister, my husband, my son and my prospective baby for their endless love, continuous support and encouragement.

ACKNOWLEDGEMENTS

This research was conducted under monumental direction and support from my advisor Dr. Din-Pow Ma. I thank him for providing me tremendous help at every critical moments and giving me uttermost trust to pursue my ideas throughout this study.

I appreciate the great work of Dr. Johnie N. Jenkins, Dr. Sukumar Saha and Dr. Ibrokhim Y. Abdurakhmonov, the scientists have inventively created the *PHYA1* RNAi line and generously supplied the plants for this work.

My committee Dr. Johnie N. Jenkins, Dr. Sukumar Saha, Dr. Kenneth Willeford, and Dr. Jiaxu Li deserve great recognition for providing inspirational perspectives and for their unwavering belief in me during my whole research.

I would like to thank my husband Peng Deng for offering free labor in next-generation sequencing analysis and giving me the greatest patience ever to show me the way, without which neither could I finish my work so efficiently.

I would also like to thank Dr. Xueyan Shan for generously offering access to the Lightcycler 480 qRT-PCR instrument and the demonstration of instrument operation and data analysis with great patience.

Thanks to all members of Dr. Din-Pow Ma's lab. It was a huge pleasure to work alongside such talented and self-motivated individuals.

I am grateful to Dr. Jeffrey F.D. Dean and the Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology for the financial support to complete my degree.

Last but not least, I am fortunate to have all my friends and those persons who let me truly enjoy my stay at Mississippi State.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

A	adenine
ac	acetylation
APS	ammonium persulfate
ATP	adenosine 5´-triphosphate
bp	base pair(s)
BR	brassinosteroid
bZIP	basic leucine zipper
BZR	brassinazole-resistant
C	cytosine
CpG	5'-C-phosphate-G-3'
CRY	cryptochrome
CTAB	cetyltrimethylammonium bromide
C-terminal	carboxy-terminal
ddH_2O	double distilled water

DEGs	differentially expressed genes
DEPC	diethyl pyrocarbonate
DIECA	diethyldithiocarbamic acid
DNA	deoxyribonucleic acid
DPA	days post anthesis
dNTPs	deoxynucleotide triphosphates
ds	double strand
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EST	expressed sequence tag
eIF2A	eukaryotic initiation factor 2 subunit A
F	forward
G	guanine
GAF	cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA
Gb	*Gossypium barbadense*
Gh	*Gossypium hirsutum*

GO	gene ontology
hr	hour(s)
H3	histone 3
H4	histone 4
5-hmC	5-hydroxymethylcytosine
IPTG	isopropyl-β-D-thiogalactopyranoside
K	lysine
kDa	kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	luria broth
M	molar concentration
me	methylation
me3	Tri-methylation
min	minute(s)
mRNA	messenger RNA
MYB	myeloblastosis
N	normality

nm	nanometer
NP-40	nonidet P-40
nt	nucleotide(s)
N-terminal	amino-terminal
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pol	polymerase
PVP-40	polyvinylpyrolidone-40 (molecular weight 40,000)
QTL	quantitative trait loci
R	reverse
RNA	ribonucleic acid
RNAi	RNA interference
ROS	reactive oxygen species
rpm	revolutions per minute
RT	reverse transcription
SDS	sodium dodecyl sulfate/lauryl sulfate

sec	second(s)
seq	sequencing
Ser	serine
siRNA	small interfering RNA
SOC	super optimal broth with catabolite repression
T	thymine
TAE	Tris-acetate/EDTA
TBE	Tris-borate/EDTA
TBP	TATA-box binding protein
TE	Tris-EDTA
TEMED	N, N, N′, N′-tetramethylethylenediamine
Tris	tris (hydroxymethyl) aminomethane
UV-B	ultra violet B
V	volt(s)
WLIM1a	initials of LIN-11, ISL1, and MEC-3
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CHAPTER I

INTRODUCTION

Cotton is the largest source of textile fiber in the world. Transgenic cotton plants carrying herbicide-resistance and insecticide genes have resulted in an increase in lint yields. However, the efforts to improve the fiber quality has led to reduced fiber yields or the size or number of seeds [1]. By silencing the phytochrome A1 gene via RNA interference, Abdurakhmonov *et al.* [2] developed *PHYA1* RNAi cotton (*Gossypium hirsutum* L.) plants that increased fiber length and strength and lower micronair (finer fiber) when compared to non-transformed Coker 312. The RNAi plants also exhibited earlier flowering, boll maturing, and produced more flowers and bolls [2]. The RNAi cotton plants would provide an excellent model to study how and why the suppression of the phytochrome *PHYA1* gene could lead to better fiber quality and superior agricultural traits.

Phytochromes are a family of red and far-red light photoreceptors, and it was reported that a high ratio of far-red light to red light perceived by the photoreceptors increased fiber length [3]. The cotton phytochrome gene family consist of four members: phyA, phyB, phyC, and phyE. In contrast to the reduced expression of *PHYA1* in *PHYA1* RNAi plants, the transcript levels of other phytochrome genes including *PHYA2*, *PHY*B, *PHYC* and *PHYE* had 2-20 fold increases in RNAi plants [2]. The phytochrome A epiallele (*phyA´*) gene in *Arabidopsis thaliana* displayed hypermethylation in several

CpG sites within the gene body, and these methylations conferred a similar phenotype to a *phyA* mutant, which had long hypocotyl and unexpanded cotyledons under FR light [4, 5]. Thus, the methylation status of phytochrome genes in *PHYA1* RNAi lines is one of epigenetic mechanisms that needs to be investigated.

Phytohormones including auxin, gibberellins (GAs), bassinosteroids (BR), and ethylene (ET) have been shown to induce fiber cell initiation and positively affect fiber development [6-9], whereas abscisic acid (ABA) and cytokinin (CK) have a negative effect on fiber development [10]. Phytochromes were reported to mediate plant hormone signaling through direct interaction between their negative regulators phytochrome-interacting factors (PIFs) and certain phytohormones signalings components, such as DELLA proteins [11]. MYB transcription factors, expansions, cell wall-loosening proteins and many genes involved in carbohydrate and cutin metabolism, transportation and ROS pathways, were reported to affect the development of fiber cells. Therefore, a genome-wide transcriptome profiling is required to characterize those genes that are differentially expressed in the *PHYA1* RNAi plant.

MicroRNAs and their targets in cotton were recently identified by several research groups [12-16]. Many miRNAs were found to be differentially expressed during cotton fiber development [17, 18], and they were shown to play important roles in fiber initiation, elongation and secondary cell wall synthesis [19, 20]. In addition, many genes encoding phytochrome responsive factors and proteins related to fiber development were found to be the targets of microRNAs. In this research, miRNA sequencing was used to reveal the profiles of miRNAs and their expression in fiber cells of the *PHYA1* RNAi plant.

CpG methylation analysis of phytochrome genes, mRNA transcriptome profiling and comparative miRNAome analysis were performed in this study in order to understand the molecular mechanisms conferring superior phenotypes to *PHYA1* RNAi plants. The differences in genome-wide gene expression of mRNA and miRNA profiles between the *PHYA1* RNAi plant and Coker 312 were determined. Real-time qRT-PCR was used to validate the relative transcript levels of differentially expressed mRNA genes and miRNAs in fiber cells. The methylation status of phytochrome genes *PHYA1*, *PHYB*, *PHYC* and *PHYE* in leaf and fiber were also determined by bisulfite genomic sequencing.

In conclusion, the CpG site methylation status within coding regions of *PHYB* in both leaves and 10-DPA fibers were decreased, *PHYA1*, *PHYC* and *PHYE* had higher methylation levels in leaves, but no significant changes in 10-DPA fibers in the RNAi line. A total of 142 genes, 7 conserved miRNAs families were differentially expressed in *PHYA1* RNAi cotton compared to Coker 312. GO and KEGG analysis together with conserved miRNA target prediction suggested that improved fiber quality of RNAi cotton might be due to the regulation of some critical transcription factors and genes involved in carbohydrate metabolism and phytohormone signaling. A total of 35 novel miRNAs were identified in *G. hirsutum* fiber cells for the first time. Of these, 9 miRNAs were predicted to target Cytochrome P450 TBP. Thus cytochrome P450 involved pathways, which are apparently mediated by fine-tune gene regulation of miRNAs, might play a role in fiber improvement of RNAi plants.

The developed *PHYA1* RNAi cotton lines provided an excellent model to study how the suppression of a single phytochrome *PHYA1* gene could lead to the improvement of fiber quality and agricultural traits. The identified and characterized mRNAs and

miRNAs that were differentially expressed in *PHYA1* RNAi cotton lines provided useful information for the genetic improvement of cotton. Their identification could be useful for understanding molecular mechanisms of fiber development and developing early-maturing and productive commercial cotton cultivars with superior fiber quality by using transgenic technology. The methylation of phytochrome genes is known to regulate their expression. The determination of cytosine methylation levels in *PHYA1* and other phytochrome genes in *PHYA1* RNAi lines demonstrated that epigenetic effects could contribute to the yield and quality of cotton fiber. The long-term goals of this project are aimed to produce cotton plants with considerably increased fiber yield and better quality (longer, stronger and finer) of lint fiber, which should have a large economic impact on the cotton industry. Novel knowledge and technologies gained through this project will be potentially useful for future sustainable agriculture in cotton production.

CHAPTER II

LITERATURE REVIEW

Cotton and Cotton genome

Cotton is the largest source of textile fiber in the world. The cotton plant belongs to the genus *Gossypium* of the family *Malvaceae*. The *Gossypium* genus consists of more than fifty species distributed worldwide [21, 22]. Among them forty-five species are diploid (2n = 2x = 26) which belong to eight genome groups (Groups A-G and K) [23, 24] based on cytogenetic and genomic diversity, and the remaining species are allotetraploid (2n = 4x = 52), whose genome constitutions are designated $(AD)_1$ through $(AD)_5$ [22, 25]. Of these *Gossypium* species, four are cultivated, including two diploids *Gossypium arboreum* L. (tree cotton) and *Gossypium herbaceum* L. (levanct cotton) and two allotetraploids *Gossypium hirsutum* L. (upland cotton) and *Gossypium barbadense* L. (sea island cotton). Upland cotton comprises more than 90 percent of the world's cotton fiber yields [26]. Sea island cotton provides 8 percent of the world's cotton production, and tree cotton together with levanct cotton contributes the remaing 2 percent of the production [27]. Over years, the phylogenetic history of two cultivated allotetroploid species of *Gossypium* has been resolved. A transoceanic dispersal of an A-genome hybridized with a local D-genome led to the evolution of AD-genome alloploidy. A polyploidization event finally contributed to the evolution of AADD-genome allotetraploid. *Gossypium herbaceum* L. (A_1) and *Gossypium raimondii* (D_5) or

Gossypium gossypioides (D_6) have been shown to be the most likely progenitors of *Gossypium hirsutum* and *Gossypium barbadense* [27] based on the consistency of the genome. Cytogenetic research indicates that *Gossypium hirsutum* and *Gossypium barbadense* are most closely related to *Gossypium arboreum* (A_2) and *Gossypium herbaceum* (A_1) [28]. In 2012, the genome of *Gossypium raimondii* (D_5) was sequenced and assembled [29]. Due to the complex genome constitution of allotetraploid (A_tD_t), the genome of *Gossypium hirsutum* L. acc. TM-1 was not sequenced and assembled till 2015 by two different research groups [30, 31].

Cotton fiber

Cotton fiber is a singular cell derived from the epidermal layer of the ovule of a cottonseed. The A-genome progenitor species produce both long fibres (lint) and fuzz-like short fibres. On the contrary, the D-genome progenitor species produce very few and much shorter lint fibres than those produced by A-genome progenitor [32].

Cotton fiber development can be divided into two major processes, designated fiber cell fate determination and cell wall biogenesis [24]. At maximum only one out of four epidermal layer cells differentiates and develops into lint fibres [10]. Overall the process of fiber development can be divided into four overlapping stages: initiation, elongation (primary cell wall synthesis), secondary cell wall synthesis, and maturation [33]. During the first three stages, the fiber cells are alive and actively growing, and fibers becomes metabolically inactive during the maturation stage upon boll opening. For each fiber, the initiation stage may last only a day or so and involves the initial isodiametric expansion of the epidermal cell above the surface of the ovule [34]. The fiber initials have a balloon-like shape and they are commonly found at the first 5 or 6

days post anthesis (DPA) from several waves of fiber initiations [35]. The vacuole in the fiber initials continues expanding after differentiation until it completely occupies the whole body of the cell, leaving only a thin layer at 2 DPA, meanwhile a large spherical nucleolus is generated through fusion of small nucleoli [36].

The elongation phase encompasses the major expansion growth phase of the fiber, which may last for 25 to 30 DPA [34]. A tubular-shape cell is found during this stage and the fiber deposits a thin, expandable primary cell wall composed of a variety of carbohydrate polymers [37]. As the fiber approaches the end of elongation, the major phase of secondary wall synthesis starts [34].

The secondary cell wall synthesis lasts until the boll opens at 50 to 60 DPA. During this stage the cell wall becomes progressively thicker due to continuous deposit of cellulose [34]. Usually fibers are simultaneously elongating and depositing secondary cell wall, but large amounts of microtubules and microfilaments formation during this stage will slow down primary cell wall synthesis [38].

Fiber maturation, the last stage of fiber development, occurs at 40 to 50 DPA and is marked by boll opening. In the end of fiber development, the fiber cells are mostly occupied by crystallized cellulose, and their highly dehydration causes an irreversible influence on the fiber properties [39].

Genes related to fiber development

Cotton fibers are seed trichome cells. It has been shown that cotton genes *GhMYB2*, *GhMYB109* and *GhMYB25* encoding MYB transcription factors play a role in fiber cell fate determination and development [40]. These cotton MYB transcription factors are similar in structure and function to the *Arabidopsis* GLABRA1 MYB protein

which is a central regulator in trichrome development. Besides *MYB* genes, many phytohormones and their related genes are involved in light signaling pathways and other pathways which mediate cotton fiber development. Auxin and gibberellins (GA_3) promote early stages of fiber initiation and have a positive effect on fiber elongation [41]. However abscisic acid (ABA) inhibits fiber initiation by counteracting the promotion of auxin and gibberellins [41]. Ethylene (ET) was reported to increase the thickness of secondary cell walls by promoting the assimilates of cotton bolls [41]. In addition, ET- and brassinosteroid (BR)-related genes are up-regulated during fiber elongation [10]. Certain calmodulin-associated genes and calmodulin-binding proteins were up-regulated in fibre initials [10]. Cotton fiber contains nearly pure cellulose, thus the *CesA* gene for cellulose synthase and many genes involved in carbohydrate metabolism and transportation are up-regulated during fiber development. The enzymes and proteins for maintaining high turgor pressure are enriched in growing fiber cells, including SUS (sucrose synthase), β-1,3-gucanase, GhVIN 1 (vacuolar invertase1), and GhPIP2 (plasma membrane intrinsic protein 2) [40]. Besides carbohydrates, the rapid growth of the primary cell wall requires large amounts of cutins, thus many lipid transfer protein (LTP) [42, 43] and FDH (Fiddlehead) [44] involved in cutin synthesis also have increased expression in fiber cells. During the transition from fiber elongation to secondary cell wall synthesis, expansins and wall-loosening proteins, such as EXPA (expansin), ACT (actin), GhRDL1 (RD22-like 1) and WLIM1a [40] are highly accumulated. Genes including *GhPDF1* (protodermal factor 1), *GbML1* (meristem layer 1) and *GhCaM7* (calmodulin 7) [40] which are involved in ROS pathway also affect fiber cell elongation. Recent genome-wide analyses of gene expression via comparative transcriptome

profiling have identified many genes that are differentially expressed in fiber cells at the four developmental stages. These include genes coding for transcription factors, such as *MYB1*, *MYB3-6* [45], *KCH1* (kinesin), *GhMAPK* (MAP kinase like protein), *GhGA20ox1* (gibberellin 20-oxidase), *GhPEL* (pectatelyase), *GhPOX1* (peroxidase), *ON035N9* (fatty acid elongase), and *ON033M19* (lipid transfer protein) [45]. In addition, two miRNAs Gh-miR167 and 164 had high levels of expression in fibers [18]. In contrast, miRNAs Gh-miR159, 160, 165/166, 168, 172 and 390 are expressed at low levels in fibers. The whole genome map showed that fiber-related QTLs were clustered on chromosome 7 and 23 [46]. Transcriptome analysis showed that fiber-related genes were located in the A subgenome, which was consistent with the phenotype in the A but not in the D progenitor [31]. In addition, during the domestication of allotetraploid cotton, the A subgenome contributed to fiber improvement and the D subgenome to adaptation traits [31].

Phytochrome gene family

Plants are photosynthetic organisms and non-motile, and thus light sensing plays a critical role in plant growing and many developmental stages. Seed germination, seedling establishment, the development of photosynthetic machinery, the architecture of the vegetative plant, the timing of flowering, circadian rhythms, tuberization and bud dormancy, and the allocation of resources to root, stem, leaf, reproductive or storage structures are all controlled by the reception of environmental light signals by the photoreceptors [47-51]. Plants respond to a broad spectrum of light, ranging from UV-B to far-red light, through four photoreceptor systems—phyotochromes [52], cryptochromes (CRY) [53], photochopin [54] and UV-B receptors [55]. Of several photoreceptors, the red (r)/far-red (fr) light photoreceptors can convert between inactive

and active forms, allowing photoreversibility to occur *in vivo*. In general, the physiological and developmental responses in plants are induced by red light which converts the chromophore to the far-red-absorbing conformation, Pfr, and the responses are canceled by far-red light to convert back to the red light absorbing form, Pr. These observations lead to the suggestion that Pfr is an active form and Pr is the inactive form.

Evolution of phytochrome genes

Phytochrome receptors were first discovered in plants in 1950s [56]. Phytochromes are soluble proteins and exist as homodimers. The molecular mass of the monomer is around 120 kDa and each monomer is covalently attached to a linear tetrapyrrole (phytochrome bilin) via a thioether link at a cysteine positioned at amino-acid residue 374. Phytochrome is synthesized in its red-absorbing form (Pr), which has a major absorption peak at 660 nm, but it also absorbs over the whole visible spectrum with a secondary absorption peak around 380 nm. Pr is biologically inactive and upon absorption of red photons it is converted to Pfr, from which biological action stems. Pfr is converted back to Pr by far-red photons. The phytochromes are encoded by a small gene family. In *Arabidopsis thaliana*, there are five members—phytochrome A (phyA) to phytochrome E (phyE) [57]. Phytochromes are widely present in mosses, ferns, green algae, angiosperms and gymnosperms [58]. Phylogenetic analysis among a wide range of angiosperms reveal that the maximum sizes of phytochrome family are likely to be five. Monocotyledonous plants may not contain phyE homologues [58]. For instance, maize and rice only have three phytochrome genes (*phyA* to *phyC*) [59]. In diploid cotton, there are two paralogs of phyA (designated as *PHYA1* and *PHYA2*), and a single copy each of phyB, phyC and phyE. In allotetraploid cotton, at least four copies of phyA and two

copies of phyB, phyC and phyE exist as a consequence of genome duplication. No single copy of phyD is identified in any cotton genome [60]. Comparison sequence analysis of higher plant *PHY* sequences indicate that four major gene duplication events occurred in the evolution of the *PHY* genes [61]. The first one happened around the time of the origin of seed plants, resulting in generation of two lines of *PHYA/C* and *PHYB/D/E*. Two later duplications occurred at about the time of the origin of the flowering plants; consequently *PHYA* was separated from *PHYC* and *PHYB/D* from *PHYE*. *PHYB* and *PHYD* diverged much more recently and they therefore share ~80% amino acid identity [57]. Because light is so crucial for plants, phytochrome evolution has in turn created a family of proteins which detect identical environmental signals but place those signals in different functions. Phy are classified into two groups—type I (phyA in *Arabidopsis*) is light-liable and type II (phyB to phyE in *Arabidopsis*) is light stable [62]. PhyA is most abundant in dark-grown (etiolated) seedlings, and its level drops rapidly upon exposure to red or white light (a 50-100 fold drop in green plants) [63]. PhyB is the most abundant phytochrome in light-grown plants, whereas phyC, phyD and phyE are less abundant [64].

General modular domain architecture of phytochromes

The phytochrome molecule possesses two structural domains: a globular N-terminal photosensory half (~ 70 kDa) connected to a more linear C-terminal regulatory half (~ 550 kDa) by a flexible hinge region. The N-terminal half contains four subdomains: N-terminal extension domain (NTE), Per-Arnt-Sim (PAS/P2), GAF/P3, and PHY/P4, while the C-terminal half contains two subdomains: the PAS-related domain (PRD) and the histidine kinase-related domain (HKRD). The NTE (P1) domain known to

inhibit dark reversion [60] is uniquely present in plant phytochromes. Besides NTE, the PRD domain is also exclusive to plant phys, which can be used either as platforms for protein-protein interaction or as response modules for small ligands or changes in light conditions, oxygen levels, and redox potentials [52]. The rest of the domains are found in all phytochrome-like proteins of various organisms [65], such as Cph1 (cyanobacterial phytochrome 1), BphP (bacterialphytochrome) and Fph (fungal phytochrome) [60]. Interestingly, plant, bacterial, and fungal apophytochromes are all able to self-ligate to appropriate bilins *in vitro* in the absence of other proteins or cofactors. The intrinsic bilin lyase activity of phytochromes had been mapped to their GAF/P3 domains by truncation analysis [66]. PAS/P2 and PHY/P4 domains play an important role in tuning the spectroscopic properties of the bound bilins [66]. It is now believed that the N-terminal 450-amino acid fragment (including NTE, PAS, and GAF) is composed of the core signaling domain of phytochrome [67].

Function of phytochromes in plant

Individual phytochromes play both unique and overlapping roles throughout the life cycle of plants [68]. Three modes of actions have been identified for phytochromes, including low-fluence responses (LFRs), very-low-fluence responses (VLFRs) and high-irradiance responses (HIRs) [69]. The 'type I phytochrome' phyA which is highly sensitive and displays non-reversible functions in response to low qualities of light is termed VLFRs. Photoreversible responses mediated by 'type II phytochrome' displayed relative stability in the Pfr form (phyB-phyE) are termed LFRs. Continuous exposure to high light intensities primarily responsible for seedling de-etiolation under all wavelengths of lights is termed HIRs.

The VLFR mediated by phyA

The induction of seed germination is a response to the fluence of the light pulse. The first stage is VLFR, and followed by the second LFR stage. VLFR is absent in the *phyA* mutant and therefore VLFR is mediated by phyA [70]. Only a small portion of phyA molecules is needed to be in its active form (Pfr) to trigger the VLFR [71].

The LFR mediated by type II phytochrome

The transformation of a large portion of the Pr form into Pfr is required by the LFR. Far-red light can cancel LFR by reducing the levels of Pfr established by red light. Thus the LFR is reversible [71] and it is mediated mainly by phyB.

The HIR mediated by phyA

The dark-grown plants are characterized by an elongated hypocotyl, unexpanded cotyledons and the absence of chlorophyll. Continuous or prolonged exposure to light is required to inhibit the elongation of hypocotyl. Continuous red light and far-red light are very effective, acting via phyB and phyA, respectively, for the inhibition of hypocotyl elongation. Continuous red light can be replaced by pulses of red-light and cancelled by far-red light. The continuous red light, therefore, can be interpreted as a repeated LFR of phyB [71]. HIR triggered by continuous far-red light requires *PHYA*-specific sequences primarily located at the NTE domain and secondarily at the C-terminal domain [72].

Function of phytochromes at various plant developmental stages

Germination

PhyA regulates germination via VFLR in red and far-red light, while phyB plays a predominant role in red light via the LFR mode [73]. In addition, phyA also promotes

germination in continuous far-red light via the HIR mode [74]. PhyD plays a small role in light-induced germination, however, phyE itself may operate as a far-red photoreceptor in regulation of germination [74].

Seedling de-etiolation

PhyA plays a unique role in mediating de-etiolation in far-red light [75-77]. In addition, phyA regulates de-etiolation in continuous blue light via HIR [78, 79]. PhyB is the predominant receptor in regulating de-etiolation in white and red lights [76, 80]. Redundancy between phyA and phyB caused red light mediated hypocotyl inhibition and cotyledon expansion [79, 80]. Despite high sequence similarity between *PHYB* and *PHYD*, the role of phyD in de-etiolation appears minor [81]. PhyE has a minor role in de-etiolation in red light [82]. Although phyC has a higher sequence similarity to phyA than phyB, D or E, no role for phyC in mediating de-etiolation in far-red light has been ever reported [83].

Shade avoidance

When growing in close proximity to neighboring vegetation, plants experience a reduction in red: far-red ratio and initiate a suite of developmental changes termed shade-avoidance syndrome, which is characterized by elongation of stems and petioles and increased apical dominance. The suppression of shade avoidance responses in high red: far-red ratio is mediated mainly by phyB [84, 85]. PhyD and phyE play a redundant role to phyB in suppression of shade avoidance responses [82].

Other aspects

PhyB regulates stomatal development, and mutants deficient in phyB displayed reduced stomatal index in white and red lights [86]. The period patterns of some genes related to circadian clock were changed in mutants deficient in phyA or phyB or both [87, 88]. Small additional roles of phyD and phyE in regulating period length were also observed [88]. In addition, phyA was shown to perceive long photoperiods, which is of importance in regulating flowering of long-day plants, such as Arabidopsis [88]. PhyB acted redundantly with phyD and phyE to repress flowering in high red: far-red conditions [89]. The *PHYC* locus attributes considerable latitudinal variation in flowering time, thereby explaining the role of phyC in regulating flowering in Columbia background but not in the Wassilewskija background of Arabidopsis [90].

Phytochrome signaling

Light induced subcellular localization of phytochromes

Using Arabidopsis as a model, light-regulated nuclear localization of all five phytochromes is reported to be a pivotal step in phytochrome signaling. The patterns of each phytochromes, however, show distinct characteristics. Previous studies showed that continuous red and blue light were effective in inducing the nuclear translocation of phyB-phyE, and this process was reversible by far red light [91]. Very recent research reported that phyB was also rapidly translocated into the nucleus during transition from dark to far-red light [92]. A detailed domain-function analysis revealed that a putative nuclear localization signal (NLS) which resides in the proline-rich PRD domain (amino acid 594-917) led to the translocation of phyB [93].

On the contrary, the phyA nuclear import can be effectively induced by continuous illumination or pulses of all light wavelengths (red, far-red, and blue lights). PhyA lacks a known NLS, and its nuclear input requires two homologous chaperone proteins—the primary FHY1 (far-red elongated hypocotyl1) and the secondary FHL (FHY1 like) [94, 95]. PhyA utilizes the NLS of FHY1/FHL for its nuclear translocation through their direct physical interaction [94, 95]. In addition, FHY3 (far-red elongated hypocotyl 3) and FAR1 (far-red-impaired response 1), and two homologous transposase-derived transcription factors, can directly activate *FHY1/FHL* expression and thus indirectly control phyA nuclear accumulation [96].

Nuclear bodies (NBs)

Upon import into the nucleus, both phyA and phyB form discrete subnuclear foci, termed nuclear bodies (NBs) or speckles [91, 97]. Two types of phytochrome NBs have been defined: small and transient NBs and larger and stable NBs formed under prolonged light exposure [98, 99]. Interestingly, a group of phytochrome-interacting factors (PIFs) were found to be required for the formation of the early transient NBs of both phyA and phyB, suggesting that these early NBs are the sites for phy-PIF interaction and their subsequent degradation [98, 100]. Thus NBs may serve as sites of proteolysis for phytochromes [67, 101].

Phytochrome signaling intermediates

Negative regulators

Phytochrome-interacting factors (PIFs)

PIFs belong to a family of basic helix loop helix (bHLH) transcription factors. They bind to DNA, which induces their ubiquitination and subsequent degradation in the proteasome. However, PIFs's DNA binding ability is reduced when they are bound by phytochromes. In this way, phytochromes release the repression of phytomorphogenesis imposed by PIFs. Six PIFs including PIF1 [102], PIF3 [103], PIF4 [104], PIF5 [105], PIF6 [105] and PIF7 [106] have been identified. PIF3 and PIF1 are bound by phyA through the APA (active phyA binding) domain. PIF4 and PIF5 are the targets of phyB by binding to an APB (active phyB binding) motif. PhyA indirectly affects PIF4 and 5 by inducing the expression of a bHLH transcription factor HFR1 (long hypocotyl in far-red) [107]. The target genes of FHY1/FHL-phyA complex include PKS1 (phytochrome kinase substrate 1), NDPK2 (nucleoside diphosphate kinase 2), CRY1 and 2, AUX/IAA, COP1 (constitutive photogenic 1), HY5 (long hypocotyl 5), LAF1 (long after far-red light 1) and HFR1 [65, 108].

COP1 and SPA1 (suppressor of PHYA-105)

COP1 is a RING finger E3 ubiquitin ligase and targets some photomorphogenesis-promoting proteins for degradation, including HY5, HYH (HY5 homologue), LAF1, HFR1 and phytochromes [67]. SPA1 is a repressor of phyA, and three SPA1-like proteins including SPA2, 3 and 4 have been identified. SPA proteins physically interact with COP1, forming a heterogeneous complex in the control of photomorphogenesis [67].

EID1 (Empfindlicher im Dunkelroten Licht 1)

EID1 is a F-box protein that targets positive acting components in phyA signal transduction to ubiquitin-dependent proteolysis [67].

Positive regulators

HY5 and HYH

HY5 is a basic leucine zipper (bZIP) transcription factor in promoting photomorphogenesis. It can directly induce both upregulation and downregulation of light related genes. HY5 is a target of COP1. HYH is a homologue of HY5 and is also a target of COP1 [67].

HFR1 and LAF1

HFR1 is a (bHLH transcription factor [109], and LAF1 [110], a myb factor, acts like HY5. They are two positive regulators in phytochrome A (phyA) signaling and they work independently [111]. In addition, they are the substrates of the COP1 [112] as well as HY5. HY5 but not its homolog, HYH, can interact with HFR1 and LAF1 to modulate the FHY1/FHL signal pathway [111].

Phosphorylation and dephosphorylation of phytochromes

The oat phyA contains two sites Ser8 and Ser18 within the NTE region, and they can be autophosphorylated *in vitro* [113]. Another site Ser599, however, is not autophosphorylated. So far, no phytochrome related protein kinase has been ever reported. Until now there is no conclusion about the intrinsic kinase activity of phytochromes. However, several protein phosphatases that dephosphorylated phytochromes, including FyPP (2A) [114], PAPP5 (phytochrome-associated protein

phosphatase 5) and PAPP2C (phytochrome-associated protein phosphatase 2c) were identified. Autophophorylation in NTE of phyA controls its stability. Unphosphorylated phyA is protected by FHY1 and FHY3 from its binding by the COP1/SPA1 complex [67, 115].

Epigenetic modification of phytochromes

Histone acetylation and methylation and DNA methylation are involved in gene expression regulation of phytochromes genes. Alteration in many histones marks by HDA19 (histone deacetylase 19) had been reported to regulate the transcription of *phyA*, including H3K9/14ac, H3K27ac, H3K4me3, H4K5ac, H4K8ac, H4K12ac, and H4K16ac [115]. Similarly, the light-induced chromatin compaction had been shown to be controlled by phyB and HDA6 (histone deacetylase 6) [116]. In addition, it was reported that a phytochrome A epiallele (*phyA´*) gene in *Arabidopsis thaliana* displayed hypermethylation in several CpG sites within gene body, and these methylations conferred a similar phenotype to a phyA mutant, which had long hypocotyl and unexpanded cotyledons under FR light [4, 5].

Crosstalk between phytochromes and phytohormones

The DELLA protein, a pivotal GA (gibberellins) signaling component, was reported to physically interact with PIF3 and PIF4. These two PIF factors yet are negative regulators and function in the downstream of phyA signaling pathway (as previously described). Binding with phyA triggers PIFs' degradation. GA promotes hypocotyl elongation. In the presence of GA, the DELLA protein was ubiquitinated and degraded, releasing PIFs from binding to target promoters and thus promoting growth [11, 117,

118]. In response to BR signaling, the transcription factor BZR1 was dephosphorylated and activated. This activated BZR1 preferred to interact with DELLA and in turn deactivated it. Inactivation of PIFs and BZR1 can be counteracted by GA by inducing the degradation of DELLA. In response to light, hypocotyl elongation is inhibited due to the induced interaction between PIF and phyA, which confers the degradation of PIFs [11, 119].

Specific functions of phytochromes in cotton

Two paralogues of the *PHYA* gene and a single copy of *PHYB*, *PHYC*, and *PHYE* genes were identified in diploid cotton by using degenerate PCR [60]. Physiological experiments suggest that phytochromes play important roles in many stages of cotton growth and development, including seed germination, morphology architecture, photoperiodic flowering, fiber elongation, and responses to drought resistance [3, 120-122]. Although the majority of cultivated cottons including *Gossypium hirsutum* L. and *Gossipium barbadense* L. exhibit photoperiod independent flowering [123], it should be noted that a high ratio of far-red light to red light perceived by the phytochrome photoreceptors increased the fiber length [124]. This observation implicates the potential effects of phytochrome on fiber development. Recently *PHYA1* RNAi transgenic cotton lines developed by Abdurakhmonov *et al*. (2014) were found to have improved fiber quality (longer, stronger and finer fiber) and increased numbers of the bolls [2]. In addition, the RNAi plants exhibited earlier flowering, produced more flowers and showed earlier boll maturing phenotypes [2]. In comparison to the down regulation of *PHYA1*, the other phytochrome genes (*PHYA2, PHYB, PHYC* and *PHYE*) in RNAi lines, however,

were up-regulated [2]. These findings indicate that phytochrome genes may play a significant role in cotton fiber development.

DNA methylation

Definition of DNA methylation

DNA methylation is one of the most important epigenetic modifications naturally occurring in both prokaryotic and eukaryotic organisms. In prokaryotes DNA methylation is used to protect host DNA from digestion by restriction endonucleases that are destined to eliminate foreign DNA. In contrast, DNA methylation in higher eukaryotes plays a pivotal role in the regulation of gene expression at transcriptional and post-transcriptional levels [125]. DNA methylation is a covalent modification in which the 5′ position of cytosine is methylated [126]. DNA methylation is found in CpG-rich sequences. In mammalian cells, only CpG methylation is detected [127]. However, in plant cells, DNA methylation occurs in the contexts of CG, CHG, and CHH (H = A, C, or T) [128].

The roles of DNA methylation

The methylation of promoter regions commonly leads to gene inactivation that is inherited throughout cell divisions. This allows the daughter cells to retain the same expression pattern as the precursor cells, which is important for many cellular processes including imprinting, X chromosome inactivation, heterochromatin maintenance, developmental controls, and tissue specific expression [129]. The methylated DNA is often associated with inactive chromatin marks, for example deacetylated histones H3 and H4, histone H3 lysine 9 methylation (H3K9me), and histone H3 lysine 27

methylation (H3K27me) [129]. Furthermore, the methylation of gene bodies may prevent false initiation of gene transcription [127, 130].

DNA methyltransferase and demethylation-associated enzymes

DNA methylation is catalyzed by DNA methyltransferases (DNMTs). There are two classes of DNMTs: de novo DNMTs, which establish methylation patterns during early development, and the maintenance DNMTs that preserve methylation patterns during cell division [131]. Five DNA methyltransferases including Dnmt1, Dnmt2, Dnmt3A, Dnmt3B and Dnmt3L have been identified in mammalian cells. Among them, Dnmt3A, Dnmt3B, and Dnmt3L are the de novo methyltransferases. Dnmt3A and Dnmt3B are active methyltransferases, and Dnmt3L is catalytically inactive and functions as a regulator to modulate the catalytic activities of Dnmt3A and Dnmt3B when it physically associates with the two methyltransferases. Dnmt1 is the maintenance methyltransferase, which prefers hemimethylated DNA over unmethylated DNA for methylation and is localized at the replication foci during S phase. In addition, Dnmt2 also displays weak DNA methyltransferase catalytic activity [131-133]. The process of DNA demethylation is also found in mammalian cells, thus the DNA methylation might be reversible. In plants, CG methylation is mediated by MET1 (Methyltransferase 1), a homologue of Dnmt1. Non-CG methylation is maintained by CMT3 (Chromomethylase 3) and DRM2 (Domains rearranged methyltransferase 2). DRM2 activity is regulated by RNA-directed DNA methylation (RdDM) [134]. In mammals, three enzyme families involved in active DNA demethylation have been identified. These enzymes firstly modify the methylated cytosine (5mC), which in turn leads to the replacement of 5mC by DNA repair [135]. The three enzyme families are: 1.The ten-eleven translocation (TET)

family, which modifies 5mC by hydroxylation and then oxidation; 2. The AID/APOEBC family, which deaminates 5mC or 5hmC; and 3. The base excision repair glycosylases, which mediate DNA repair [135]. In plants, DNA demethylation is characterized by 5-methylcytosine glycosylases, which includes ROS1 (Repressor of silencing 1), DME (Demeter), DML2 (DME-like 2) and DML3 [136]. It was reported that cotton fibers harvested during May to June were significantly longer than those harvested in Feb to March. This phenomenon was due to the changes in expression levels of DNA demethylases and methyltransferases in cotton fibers [137]. In addition, the regulation of cotton fiber growth by the genes *ERF6* (ethylene response factor 6), *SUR4* (suppression of RVS 161 delta 4) and *KCS13* (3-ketoacyl-CoA synthase 13) was due to the modifications on certain CHH sites in their promoter regions [137].

DNA methylation detection methods

Methylation profiles of epigenome analysis can be global or gene-specific depending on downstream applications. The methyl-cytosine levels in genomic DNA could be quantified by three methods: chemical hydrolysis following by high-performance liquid chromatography (HPLC); methylation analysis by fluorescent labeling; or the methyl-acceptor assay using *Sss*I DNA methyltransferase [138]. Currently, the most commonly used techniques for DNA methylation study were based on bisulfite treatment. Downstream methylation detection methods include end-point PCR, real-time PCR, primer extension, single-stranded conformational polymorphism assays, blotting, microarray analysis and sequencing. More recently, a global mapping of cytosine methylation at single-base resolution has become possible via Next-generation

sequencing [138]. For example, a genome-wide methylome map of *Arabidopsis* has been constructed [139, 140].

MicroRNAs (miRNAs)

MicroRNAs are the second most abundant class of plant small RNAs. To date, 337 and 713 miRNAs have been identified in *Arabidopsis* and rice, respectively (miRBase version 20) [141]. They are classified as a family of small (21-25 nt) noncoding RNA molecules that regulate eukaryotic gene expression post-transcriptionally in a sequence-specific manner. The silencing of target gene expression by miRNAs may occur through translational inhibition (e.g. miRNA 172 regulated APETALA2 expression [142]) or accelerated mRNA decay (miR-125b and let-7 induced accelerated deadenylation, leading to rapid mRNA decay [143]), or slicing within miRNA-mRNA base pairing [144, 145].

Biogenesis of miRNAs

MicroRNA genes (MIR) are mainly transcribed by RNA polymerase II (pol II). The MIR gene is first transcribed into a large primary fold-back transcript (pri-miRNA), which has a 5′-cap and a 3′-poly(A) tail as the unique characteristics of pol II products [146]. In addition, the promoters of many MIR genes have been reported to contain the TATA box— a typical pol II element [147]. It was also reported that a few miRNA genes might be transcribed by pol III [148]. The pri-miRNA was further processed into a ~60-70 nt stem-loop precursor (pre-miRNA). This initial cleavage was facilitated by Drosha enzymes in animals and by Dicer enzymes in plants (plant Drosha homologues have not been identified yet). The pre-miRNA was then exported into the cytoplasm by Exp5

(Exportin 5) in animals and further cleaved by a Dicer into a small, dsRNA duplex. One of the RNA strand's in the duplex with the weakest 5′ end base pairing was then selected as the mature miRNA. The other RNA strand, called miRNA*, was then degraded [149]. In plants, the final cleavage by Dicer (DCL1) assisted by other factors, such as HYL1 (a two dsRBD-containing nuclear protein) and HEN1 (a protein with a dsRBA and a methyltransferase domain) [150] occurred within the nucleus rather than in the cytoplasm. After the cleavage, the miRNA:miRNA* duplex was exported into cytoplasm by HASTY in plants [151]. One strand of the mature miRNA is loaded to RISC (RNA-induced silencing complex). The miRNA in the miRNA-RISC complex is guided by ARGONAUTE proteins (as a part of RISC) and targets homologous mRNA complementary sequences and triggers mRNA cleavage or translational inhibition [152]. There are more differences that exist between plant miRNAs and animal RNAs. It was proposed that methylation at the 2′ hydroxyl groups on the 3′ most nucleotides of miRNA:miRNA* duplex played a crucial role in its biogenesis [153]. In addition, plants miRNAs usually have an exact match to their mRNA targets and cause the cleavage and degradation of the target transcript. Although the miRNA pathway is involved in the plant-virus interaction for protection of the host plant by degrading virus mRNA, it is more considered as a mechanism for the plant to control its own gene expression [154-156].

Cotton miRNAs with fiber development

It was reported that phenotypic variation in *Arabidopsis* allotetraploids could be mediated, at least to some extent, by miRNA. This variation was partially related to the strength of promoter and modifications in chromatin and putative inhibitors present in

allotetraploids [40]. As previously reviewed, the most commonly grown cotton—upland cotton (*Gossypium hirsutum*) is an allotetraploid, it is thus interesting to determine if the same mechanism of miRNAs causing phenotypic variation is also present in the cotton genome. By using computational prediction and small RNA sequencing, approximately 80 miRNAs have been identified in cotton [157]. Based on small RNA blotting, it was shown that many miRNAs were differentially expressed during fiber development. More importantly, many genes related to fiber development, such as MYB transcription factors, phytochrome receptors, and hormone related factors have been reported to be the targets of miRNAs. These observations imply that miRNAs might play crucial roles in regulating fiber development, either as negative regulators if the targets have a positive role or as positive regulators otherwise [18, 40].

miRNAs and trans-acting siRNAs (ta-siRNAs)

When the target of miRNA is a *TAS* locus, the cleavage triggered by miRNA produces a new category of small RNAs called trans-acting siRNAs (ta-siRNAs) The ta-siRNAs, a class of endogenous short interfering RNAs, have been identified in both plants and animals and play a role in gene expression. During the biogenesis of ta-siRNA, the cleaved mRNA does not end up with pre-miRNA, instead it serves as a template for RDR6 (RNA-dependent RNA polymerase 6) to generate a dsRNA (pre-ta-siRNA), which is further processed by DCL4 into a mature ta-siRNA. In cotton, the *GhMYB2* gene promotes fiber development and its transcript is targeted by GhmiR828 and GhmiR858. In addition to the cleavage of *GhMYB2* transcripts, GhmiR828 also participates in transcript processing to produce ta-siRNAs [157].

Criteria for annotation of plant miRNA and its target prediction

Precise biogenesis of a stem-loop pre-miRNA is the primary criterion for miRNA prediction [158]. The stem-loop structure can be predicted by RNA secondary structure prediction software using genome sequence or ESTs as inputs [158]. Auxiliary criteria for miRNA identification include evolutionary conservation, having proper targets, and DCL1-dependence, etc. However, when controversy happens, the primary criterion is the one and the only one to adhere to [158]. The determination of the miRNA function is largely dependent on its target identification. Bioinformatics predictions of miRNA targets is more powerful and reliable for plant miRNAs due to their perfect or nearly perfect complementary to their targets [159]. Basically, 2-8 nt from the 5′-end of each miRNA designates the seed region, and a single mismatch within this region will induce off-target silencing [160].

CHAPTER III

MATERIAL AND METHODS

Plant materials

Gossypium hirsutum L. cv. Coker 312 and Phytochrome A1 (*PHYA1*) RNAi transgenic plants were grown in the greenhouse at Agricultural Research Service (ARS) of USDA and in the field at the North Farm R. R. Foil Plant Science Research Center of Mississippi State University. Flowers were tagged on the day of anthesis (0 DPA). Leaves and bolls (5, 10, 15, and 20 DPA) were harvested with three biological replications. Fibers were carefully dissected from cotton bolls and immediately frozen in liquid nitrogen for the preparation of genomic DNA, mRNAs and miRNAs.

DNA methylation status analysis

Isolation of genomic DNA from leaves and 10 DPA cotton fibers

Genomic DNAs were isolated from cotton leaves and 10 DPA fibers using a modified method as described by Paterson *et al* (1993) [161]. Frozen cotton tissues were ground into fine powder with a pre-chilled mortar and pestle in liquid nitrogen. One gram of the powder was weighed out and transferred into a pre-chilled 15-ml centrifuge tube. The sample was resuspended with 10-ml of ice-cold DNA extraction buffer (0.35 M glucose, 0.1 M Tris-HCl, pH 8.0, 5 mM EDTA, pH 8.0, 2% (w/v) PVP-40, 0.1% (w/v) DIECA, 0.1%(w/v) ascorbic acid, 0.2% (v/v) β-mercaptoethanol) by vortexing and then centrifuged at 2,700 x g for 20 min at 4°C. The supernatant was decanted and the nuclear

pellet was resuspended with 4 ml of nuclei lysis buffer (0.1M Tris-HCl, pH 8.0, 1.4 M NaCl, 20 mM EDTA, pH 8.0, 2% (w/v) CTAB, 0.1% (w/v) DIECA, 0.1% (w/v) ascorbic acid, 0.1% (v/v) β-mercaptoethanol) and then incubated in a 65°C water bath for 30 min. An equal volume of chloroform-isoamyl alcohol (24:1, v/v) was added to the lysate and the sample was mixed by vortexing vigorously for 15 sec, and then centrifugd at 2,700 x g for 5 min. After centrifugation, DNA in the upper aqueous phase was transferred into a new 15-ml centrifuge tube and precipitated by adding 0.6 volume of isopropanol and incubation at room temperature for 10 min. The sample was centrifuged for 10 min at 10,000 x g for 10 min. The supernatant was decanted, and the pellet was washed with 1 ml of 70% ethanol by vortexing and centrifugation at 10,000 x g for 5 min. The ethanol was decanted, and the DNA pellet was air dried for 10 min and dissolved 100 µl of ddH_2O or TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

Detection and Quantitation of DNA Methylation

Bisulfite genomic sequencing

Bisulfite treatment of genomic DNA

Genomic DNA samples from *PHYA1* RNAi and Coker 312 were treated with sodium bisulfite using an EZ DNA Methylation Kit (ZYMO Research, Irvine, California). Five µl of M-Dilution Buffer were added to 0.5 µg DNA sample, and the sample volume was then adjusted to 50 µl with ddH_2O. The sample was mixed by pipetting up and down and then incubated in a 37°C water bath for 15 min. After incubation, 100 µl of the prepared CT Conversion Reagent were added to the sample and mixed well. The sample was incubated in the dark (covered by alumni foil) at 50°C for 16 hr and followed by incubation on ice for 10 min. Four hundreds µl of M-Binding

Buffer was added into a Zymo-Spin™ IC Column which was placed into a 1.5 ml collection tube. The treated DNA sample was loaded into the Zymo-Spin™ IC Column containing 400 µl of M-binding Buffer. The sample was mixed by inverting the column in the collection tube several times and centrifuged at 13,200 rpm for 30 sec. The flow-through was discarded, and 100 µl of M-Wash Buffer were added to the column. After centrifugation at 13,200 rpm for 30 sec, 200 µl of M-Desulphonation Buffer were added to the column and followed by incubation at room temperature for 20 min. After incubation, the sample was centrifuged at 13,200 rpm for 30 sec. The DNA sample was washed two times by adding 200 µl of M-Wash Buffer to the column and followed by centrifugation at 13,200 rpm for 30 sec. The column was then placed into a new 1.5 ml microcentrifuge tube and 10 µl of M-Elution Buffer was added to the center of the column matrix. The sample was centrifuged for 30 sec at 13,200 rpm to elute the DNA. The DNA was stored at -20°C for later use.

Touch-down PCR

The bisulfite-converted genomic DNA was used as a template for touch-down PCR amplification. The PCR reaction mixture was prepared in a final volume of 50 µl containing 5 µl of 10 x Taq buffer solution (100 mM Tris HCl pH 9, 500 mM KCl, 15 mM $MgCl_2$, 1% Triton X-100), 1 µl of 10 mM dNTPs stock, 1 µl of forward primer (10 µM), 1 µl of reverse primer (10 µM), 1 µl bisulfite-converted DNA, 0.5 µl of Taq polymerase (5 units μl^{-1}) (GenScript, Piscataway Township, New Jersey) and 40.5 µl of sterilized ddH_2O. Bisulfite primers of four phytochrome genes including *PHYA1* (HM143735), *PHYB* (HM143750), *PHYC* (HM143755), and *PHYE* (HM143759) and a standard gene *eIF2A* (Q922565) [137] were designed using Bisulfite Primer Seeker

(http://www.zymoresearch.com/tools/bisulfite-primer-seeker). The sequences of the primers are listed in Table 3.1:

Table 3.1 The primers used in bisulfite genomic sequencing

Primer ID	Sequences (5′→3′)
PHYA1-F	TGGTTGTTATYGTTAATGATGGAGATGAAGAAG
PHYA1-R	AACCTAAAAATAATATTATAACATACTACCAAACCCC
PHYB-F	ATGTTTGTTTTGATATTTTGTTTATGTTGTTTTAATTG
PHYB-R	ACCACCCAATAAACTTTTCCATAACAATATTC
PHYC-F	GTTGTTGAGAGATTTTTTTGTAGGAATTG
PHYC-R	CAAAATCATACTTTACACCACCCC
PHYE-F	GYGTTAGGATGATTTGTGATTGTTATG
PHYE-R	ACRAAATATAATAACAAACAACCAATCCCCAAAAC
eIF2A-F	GTYGTGAGTGTTTATTTTTTAAGTTTTAATTTAG
eIF2A-R	AAAAAAACACCRCTATTCCATTATCCCATAACC

The assembled PCR reaction components in a 200 µl thin wall tube were subject to touch down PCR amplification using a cycling program with initial denaturation at 95°C for 3 min, and then followed by 25 cycles of denaturation at 95°C for 30 sec, annealing with a temperature starting at 65.5°C and gradually reduced (0.5°C/every second cycle) to 53°C for 30 sec, and extension at 72°C for 1 min. The amplification was continued for another 30 cycles at the annealing temperature of 53°C and a final extension at 72°C for 5 min. The PCR product was electrophoresed on a 1% (w/v) agarose gel (20 mM Tris-HCl, pH 7.5, 7.5 mM sodium acetate, 0.5 mM EDTA, and 1 µg μl^{-1} ethidium bromide) at 76 V for 1hr in a 1 x TAE running buffer (20 mM Tris-HCl, pH 7.5, 7.5 mM NaOAc and 0.5 mM EDTA). The correct amplicons were selected based on their size and then purified by using a QIAquick PCR Purification Kit (Qiagen, Valencia, California). The purified PCR products were ligated into pGEM-T easy vector using a pGEM-T Easy kit (Promega, Madison, Wisconsin). The T4 ligase-catalyzed ligation

reaction was carried out at room temperature for 3-4 hr, and the ligation mixture was then used for transformation into *E. coli* XL1-Blue cells.

Preparation of E. coli XL1-Blue competent cells

A single colony of *E. coli* XL1-Blue was inoculated into 5 ml of LB medium (1% (w/v) Bacto-tryptone, 0.5% (w/v) Bacto-yeast extract, and 1% (w/v) NaCl, pH 7.0) in a 15 ml sterilized tube and cultured at 37°C overnight. Two ml of this fresh overnight-cultured cells were then subcultured into 250 ml of LB medium in a one-liter flask at 37°C with agitation (~220 rpm) until the value of OD_{600} reached to 0.6. The cells were immediately chilled on ice and collected by centrifugation at 4,200 rpm at 4°C for 8 min. The supernatant was discarded and the cell pellet was resuspended with 60 ml of ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 min. The cells were recollected by centrifugation as before and were resuspended with 10 ml ice-cold 0.1 M $CaCl_2$ and 2.5 ml of 80% (v/v) glycerol. The prepared competent cells were aliquoted into 1.5 ml microfuge tubes (100 µl per tube) and stored at -80°C until use.

Chemical transformation into E. coli XL1-Blue competent cells

The ligation mixture (10 µl) was added into 100 µl of *E. coli* XL1-Blue competent cells and incubated on ice for 30 min. After heat shock at 42°C for 2 min, the cells were immediately transferred into ice and incubated for another 2 min. A SOC medium (0.5% (w/v) Bacto-yeast extract, 2% (w/v) Bacto-tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM Glucose) of 500 µl was added, and the cells were incubated at 37°C with agitation (~220 rpm) for 45 min to 1 hr. The

transformed cells were spread on LB/ampicillin (50 μg ml^{-1}) /IPTG (0.5 mM)/ X-Gal (80 μg ml^{-1}) plates and incubated at 37°C overnight.

Preparation of DNase-free RNase A

The boiling method [164] was used to prepare a DNase-free RNase A stock solution for plasmid isolation. Powdered RNase A (Sigma, St. Louis, Missouri) was dissolved in ddH_2O at a concentration of 10 mg ml^{-1}. The solution was heated to 100°C for 15 min, allowed to cool to room temperature, and then centrifuged at full speed (13,200 rpm) to remove any precipitates. The stock solutions were stored in frozen aliquots (-20°C) until use.

Isolation of recombinant pGEM-T Easy/GhPHY plasmids

For each phytochrome gene, a total of 20-40 white colonies were picked up, inoculated individually into 2 ml of LB/ampicillin (50 μg ml^{-1}) medium, and cultured at 37°C overnight. The recombinant plasmids were isolated by the alkaline lysis mini preparation method. The bacterial cells (1.5 ml) were collected by centrifugation at full speed (13,200 rpm) for 1 min and then re-suspended in 100 μl of suspension solution (25 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, 50 mM glucose, and 100 μg ml^{-1} DNase-free RNase A). Two hundred μl of lysis buffer 0.2 N NaOH and 1% (w/v) SDS), freshly prepared before plasmid isolation, were added to the cell suspension solution and mixed by inverting 4-6 times. Next, 150 μl of neutralization solution (3 M NaOAc, pH 5.2) were added to the mixture and inverted for 4-6 times immediately. The mixture was kept in ice for 10 min and white flocculent precipitate was formed. The tube was centrifuged at full speed (13,200 rpm) for 7 min, and the supernatant containing the plasmid was then

transferred into a new 1.5 ml micro-centrifuge tube. The plasmid DNA was precipitated with 2 X volumes of 96% ethanol and kept at -20°C for 30 min. The plasmid was collected by centrifugation at full speed (13,200 rpm) for 10 min. The supernatant was decanted and the plasmid pellet was washed with 1 ml of 70% ethanol. The sample was vortexed for 1 min and centrifuged for another 10 min at full speed to recollect the DNA pellet. The ethanol was gently poured off. The sample was re-centrifuged for 1 min, and the remaining ethanol was removed by using a 100 µl pipet. The plasmid was air dried for 10 min and then resuspended in 50 µl of ddH_2O.

DNA cycle sequencing

A total of 20-40 subcloned DNA samples for each phytochrome gene were sequenced using the Big Dye terminator v1.1 Cycle Sequencing Kit on ABI PRISM 310 Genetic Analyzer (ABI, Foster City, California). The sequencing reaction mixture contained 2 µl of BigDye Terminator v3.1 Ready Reaction Mix, 3 µl of 5 x reaction buffer, 1 µl of plasmid, 1 µl of M13 forward primer (5′-GTTTTCCCAGTCACGAC-3′, 10 µM) or M13 reverse primer (5′-CAGGAAACAGCTATGAC-3′, 10 µM) and 13 µl of ddH_2O. The cycle sequencing reaction was carried out with preheating at 95°C for 3 min and followed by 30 cycles of 3-step PCR including denaturation at 95°C for 15 sec, primer annealing at 50°C for 10 sec and extension at 60°C for 4 min in MultiGene II Thermal Cycler (Labnet, Edison, New Jersey). The sequencing DNA product was precipitated with 2 µl of 3 M NaOAc (pH 4.8 to 5.2) and 50 µl of 96% ethanol. The mixture was incubated at room temperature for 15 min and then centrifuged at full speed (13,200 rpm) for 10 min. The DNA pellet was washed with 500 µl of 70% ethanol by vortexing for 1 min and re-centrifuged for 10 min to recollect the DNA pellet. The

ethanol wash was then completely removed by using a pipette, and the DNA pellet was air dried for 10 min and then resuspended in 20 µl of Hi-Di formamide (ABI, Warrington, UK). The DNA samples were denatured at 95°C for 2 min, chilled on ice immediately, and loaded onto the ABI PRISM 310 DNA Genetic Analyzer for sequence analysis.

Bisulfite sequencing data analysis

The nucleotide sequence alignment of individual clones was performed by DNASTAR, and those identical sequence reads were removed from further analysis. The sequences were exported as multi-FASTA files and compiled in one zip file by WinZip. The sequences were uploaded to, analyzed and finally displayed by BISMA (Bisulfite Sequencing DNA Methylation Analysis) (http://services.ibc.uni-stuttgart.de/BDPC/index.php).

Identification of Specifically or Differentially Expressed Transcripts

Messenger RNA Transcriptome Sequencing

Isolation of total RNA from 10 DPA cotton fibers

Total RNA was isolated from 10 DPA fibers using a modified hot borate method [162]. Frozen fiber tissues were ground into fine powder with a pre-chilled mortar and pestle in liquid nitrogen. About 0.3 gram of the powder was weighed out and transferred into 2 ml of pre-heated (80°C) borate exaction buffer (200 mM sodium tetraborate decahydrate, pH 9.0, 30 mM EGTA, 1% SDS, 1% sodium deoxycholate, 2% PVP-40, 0.5% NP-40 and 10 mM DTT) in a 15 ml centrifuge tube. Fifty µl of Proteinase K (20 mg ml^{-1}) were added and mixed well by vortexing, and the mixture was incubated at

42°C for 1.5 hr with agitation at 100 rpm. The homogenate was adjusted to a final concentration of 160 mM KCl by adding 178 µl of 2 M KCl with gentle mixing. After incubation on ice for 1 hr, the sample was centrifuged at 4°C for 20 min at 12,000 x g. Subsequently, the supernatant was transferred into a new 15 ml centrifuge tube, and the RNA was precipitated on ice overnight in the presence of 2 M LiCl (adding 967 µl of 8M LiCl to the mixture). After centrifugation at 4°C for 20 min at 12,000 x g, the supernatant was discarded, and the RNA pellet was washed with 1 ml of 2 M LiCl by gentle vortexing and centrifuged again at 4°C for 20 min at 12,000 x g. The wash steps were repeated 2-3 times until the supernatant becomes colorless. The RNA pellet was dissolved in 1 ml of 10 mM Tris-HCl, pH 7.5, and the RNA solution was adjusted to 200 mM KAc by adding 71.4 µl of 3 M KAc (pH 5.5). The mixture was incubated on ice for 15 min and then centrifuged at 4°C for 20 min at 12,000 g. The supernatant was transferred into a new 15 ml centrifuge tube, and the RNA precipitated by adding 2.5 x volumes of 96% ethanol. The precipitation step could be performed at -20°C overnight or at -80°C for at least 30 min. After centrifugation at 4°C for 15 min at 12,000 x g, the supernatant was discarded and the RNA pellet was washed with 1 ml of 70% ethanol (prepared with DEPC-treated water) by vortexing for 1 min. The sample was then centrifuged for 10 min at 12,000 x g to harvest the RNA pellet. The ethanol was gently poured off, and the RNA pellet was air dried for 10 min and resuspended in 50 to 100 µl of RNase-free water.

RNA-Seq library construction and Illumina sequencing

Genomic DNA contamination in the RNA samples was removed with DNase I (Promega, Madison, Wisconsin). Briefly, 10-50 μg of total RNA were treated with 5 μl of DNase I (1 unit μl^{-1}) at 37°C for 20 min in a 50 μl reaction according to the manufacturer's instructions. After DNase treatment, the RNA sample was cleaned up with a Qiagen RNasey Mini Kit (Qiagen, Valencia, California). The quality of the RNA sample was pre-checked on a 1-1.2% agarose gel by electrophoresis, and the RNA concentration was determined with Nanodrop 2000c and a Qubit® fluorimeter by using a Qubit RNA BR Assay Kit (Invitrogen, Carlsbad, California). The quality of total RNA was finally estimated using the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, California).

Two μg of total RNA with three biological replicates were used to construct strand-specific mRNA seq libraries using the NEB Next1 Ultra™ Directional RNA Library Prep Kit (NEB, Ipswich, Massachusetts) according to the manufacturer's instructions. Index codes were added to the corresponding RNA samples. Briefly, a Ribo-Zero rRNA Removal Kit (Epicentre Biotechnolgies, Madison, WI) was used to treat the total RNA, and fragmentation was carried out using divalent cations under elevated temperature in the NEB Next First Strand Synthesis Reaction Module buffer (NEB, Ipswich, Massachusetts). First strand cDNA was synthesized using random hexamer primers and M-MLV Reverse Transcriptase (NEB, Ipswich, Massachusetts). Second strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H (NEB, Ipswich, Massachusetts). The resultant library fragments were purified with AMPure XP system (Beckman Coulter, Danvers, MA) to select the preferred cDNA

fragment. The USER enzyme (3 U, NEB, Ipswich, Massachusetts), a mixture of uracil DNA glycosylase (UDP) and the DNA glycosylase-lyase endonuclease VIII, was used to treat size-selected, adaptor-ligated cDNA at 37°C for 15 min and followed by 5 min at 95°C before PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, universal PCR primers and Index Primer. The PCR products were purified by AMPure XP system and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System (Illumina, USA) using TruSeq PE Cluster Kit v3-cBot-HS (Illumina, USA) according to the manufacturer's instructions. The prepared libraries were sequenced on Illumina Hiseq 2500 platform (Illumina, USA) and 2 x 50 bp paired-end reads were generated from the whole transcriptome.

Data processing and analysis

FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) was used to test data quality. Trimmomatic [163] was used to trim the low-quality reads and adaptor sequences. *Gossypium hirsutum* L. acc. TM-1 genome [164] was used as reference for short reads mapping. ArrayStar (Version 12, DNASTAR Inc., USA) was used for RNA-Seq analysis. The expression of genes was normalized by Reads Per Kilobase of transcript per Million mapped reads (RPKM) method [165]. Blast2GO [166] and InterProScan 5 [167] were used for identifying orthologous genes and adding additional Gene Ontology (GO) terms, if necessary. GO terms were mapped to the GO-slim [168] to get a broad functional overview of the transcriptome. The Fisher's exact test [169] method was used for gene set enrichment analysis [170, 171]. KEGG database [172] was used for additional protein functions analysis.

MicroRNA Sequencing

Isolation of small/microRNAs from different developmental stages of cotton fibers (5, 10, 15 DPA)

For isolation and analysis of miRNAs, fiber samples were collected with three biological replications. MicroRNAs were isolated from 5, 10, and 15 DPA cotton fibers using the mirPremier™ microRNA Isolation Kit (Sigma, St. Louis, Missouri). The frozen fibers were ground three times into fine powder in liquid nitrogen using a pre-chilled mortar and pestle. It should be noted that insufficient grinding of the tissue would result in reducing RNA yields. After the evaporation of liquid nitrogen, approximately 100 mg (90-110 mg) of the tissue powder were quickly weighed out in a 1.5 ml microcentrifuge tube, which was pre-chilled in liquid nitrogen. The weighed sample was kept on dry ice or stored at -80°C before microRNA purification. A Lysis Mix for microRNA purification was prepared by adding 650 µl of microRNA NA Lysis Buffer (M1070), 300 µl of Binding Solution (L8042) and 50 µl of 96% ethanol in a 1.5 ml microfuge tube and mixed well. The solution was then supplemented with 10 µl of 2-mercaptoethanol per ml of the solution. The Lysis Mix was prepared just prior to the start of the purification procedure. To lyse the samples, 750 µl of the Lysis Mix were added to the tissue powder and vortexed immediately and thoroughly for 30 sec. The mixture was incubated in a 55°C water bath for 5 min and then centrifuged at 16,000 x g for 5 min to remove cellular debris, genomic DNA, and large RNAs. The lysate supernatant was then pipetted into a Filtration Column (blue retainer ring) and centrifuged at 16,000 x g for 1 min to remove residual debris. The clarified flow-through lysate was saved and its volume was measured with a pipette. Ninety-six percent ethanol (1.1 volume) was added into the clarified lysate, and the sample was mixed immediately and thoroughly by inversion. The mixture

of 700 µl was transferred onto a Binding Column (red retainer ring) by pipetting. The mixture was centrifuged at full speed (16,000 x g) for 30 sec and the flow-through liquid was decanted. The remaining mixture was then transferred to the binding column to repeat the binding step. The column was sequentially washed with 700 µl of 96% ethanol and 500 µl of binding solution by centrifugation at 16,000 x g for 30 sec and 1 min, respectively. The column was then further washed twice with 500 µl of the Ethanol-diluted Wash Solution 2 by centrifugation. After the four washes, the column was further centrifuged at full speed (16,000 x g) for 1 min to remove residual wash solution. Finally, the column was transferred onto a new 2 ml collection tube, and 50 µl of Elution Solution were added directly onto the center of the filter inside the column. The tube was let sit for 1 min and then centrifuged at 16,000 x g for 1 min. The elution step was repeated once in order to increase the yield of the RNA. Purified RNA in the flow-through was ready for library construction or stored at -20°C (short term) or -80°C (long term).

Construction and Sequencing of miRNA Libraries

The NEBNext Multiplex Small RNA Library Prep Set (NEB, Ipswich, Massachusetts) for Illumina was used to convert purified miRNAs into barcoded cDNA libraries for high throughput next generation DNA sequencing. Briefly, about 500 ng (6 µl) of small RNAs were ligated to a 3′ SR Adaptor (a 5′-adenylated, 3′ blocked oligodeoxynucleotide) using T4 RNA ligase in the absence of ATP. A SR RT primer was then used to hybridize to the excess of 3′ SR adaptor to form dsDNA molecule which would prevent adaptor-dimer formation. After the hybridization, the 3′ SR Adaptor-ligated small/miRNA samples was then ligated to a freshly denatured 5′ SR Adaptor with T4 RNA ligase in the presence of ATP. The 5′ and 3′ ligated small RNAs were then

converted into cDNAs by reverse transcription with ProtoScript II reverse transcriptase and a SR Primer in the presence of murine RNase inhibitor. The cDNAs were then subjected to PCR amplification using the SR primer and an index primer with a different index sequence. The sequences of all the adaptors and primers used in the construction of miRNA libraries were listed in Table 3.2.

Table 3.2 Primers and adaptors used in the construction of miRNA libraries

Primer ID	Sequences
3′ SR Adaptor	5′-rAppAGATCGGAAGAGCACACGTCT-NH2-3′
SR RT primer	5′-AGACGTGTGCTCTTCCGATCT-3′
5′ SR Adaptor	5′-rGrUrUrCrArGrArGrUrUrCrUrA rCrArGrUrCrCrGrArCrGrArUrC-3′
SR Primer	5′-AATGATACGGCGACCACCGAGATC TACACGTTCAGAGTTCTACAGTCCG-3′
Index primer (Index* was italicized)	5′-CAAGCAGAAGACGGCATACGAGAT*CGTGAT*G TGACTGGAGTTCAGACGTGTGCTCTTCCGATC-3′

*All index sequences were listed in Table 4.3.

The PCR amplification reaction was conducted under the following conditions: initial denaturation at 94°C for 30 sec, 13 cycles of denaturation at 94°C for 15 sec, annealing at 62°C for 30 sec, extension at 70°C for 15 sec, and a final extension at 70°C for 5 min. The PCR amplified cDNA construct was purified by using a QiAQuick PCR Purification Kit (Qiagen, Valencia, California) according to the manufacturer's instructions except that the lid of the spin column was open during the 5 min centrifugation step in order to completely remove any residual ethanol.

Size selection of miRNA library (~147 bp) using AMPure XP Beads

The amplified DNA was eluted in 27.5 µl of Nuclease-free Water. One µl of the purified PCR product was loaded on the Bioanalyzer using a DNA 1000 chip according to the manufacturer's instructions. To the purified PCR product (25 µl), 32.5 µl (1.3 x) of resuspended AMPure XP beads were added and mixed well by pipetting up and down at least 10 times. After incubation for 5 min at room temperature, the tube was placed on a magnetic stand to separate beads from the supernatant. After the solution was clear (about 5 min), the supernatant (57.5 µl) was carefully transferred into a new tube. The beads that contained the large DNA fragments was discarded. An aliquot of 92.5 µl (3.7 x) of resuspended AMPure XP beads was added into the supernatant (57.5 µl), mixed well and incubated for 5 min at room temperature. The tube was then placed on a magnetic stand to separate beads from the supernatant. After the solution was clear (about 5 min), the supernatant was carefully removed and discarded. The beads that contain DNA targets were washed with 200 µl of freshly prepared 80% ethanol while the tube was in the magnetic stand. The sample was incubated at room temperature for 30 sec, and the supernatant was then carefully removed and discarded. This wash step was repeated one more time. After the second ethanol wash, the tube was placed back in the magnetic stand and the residual ethanol was completely removed. The beads were air dried for 10 min while the tube was on the magnetic stand with lid open. The DNA target on the bead was eluted with 13 µl of nuclease-free water. After mixing well by pipetting up and down, the tube was placed back in the magnetic stand until the solution was clear (about 1 min). The supernatant was transferred into a clean PCR tube. One µl of this product was analyzed on the Bioanalyzer High Sensitivity chip to check the peak distribution and

molarity of the small RNA library. If too many large DNA bands were still present in the library, the whole bead size selection procedure was repeated one more time to get more pure small RNA libraries. The microRNA cDNA libraries were subsequently sequenced by using the HiSeq next-generation sequencing system (Illumina, San Diego, California) with the single end (1 x 50) method. The Illumina sequencing platform has the capability to pool samples in a single lane to reduce the cost of sequencing. Sample pooling, or multiplexing, was accomplished with a barcode sequence incorporated into DNA fragment during library construction that was used to sort reads post-run. The Illumina indexing system allows the pooling of up to 48 samples of small RNA libraries.

Size selection of small RNA library (~147 bp) by using polyacrylamide gel

A 6% denatured polyacrylamide gel containing 7.5 M urea was prepared by mixing 1.5 ml of 40% acrylamide/Bis 19:1, 2 ml of 5 x TBE buffer (0.445 M Tris base, 0.445 M boric acid, 0.01 M EDTA) and 4.5 g urea. After all the reagents were dissolved, 125 µl of 10% APS and 10 µl of TEMED were added to catalyze the polymerization reaction, and the gel solution was immediately injected into a glass cassette with a pipette. A comb (1mm thick) was immediately inserted into the gel solution before it solidified in 10 min. The gel could be also made one-two days before use and stored in 4°C to get a better sample resolution. Before electrophoresis, the gel was pre-run for at least 15 min to remove the extra salts settled in the slots. Ten µl of the purified PCR product mixed with 10 µl of 2 x loading dye were loaded into one well (one sample was loaded into three wells). The pBR322 plasmid digested with *Msp*I was served as a DNA marker for size comparison. The electrophoresis was carried out in 1 x TBE buffer (5 x TBE: 1.1M Tris-HCl; 900 mM Borate; 25 mM EDTA; pH 8.3) at 180 V for about 1 hr

until the bromophenol blue dye reached the bottom of the gel before it migrated off the gel. The gel was stained with Diamond nucleic acid dye (Promega, Madison, Wisconsin) in a clean container for 20 min on an agitator, and the gel was then viewed and photographed with a blue light trans-illuminator. The 140 nt DNA band represented adapter-ligated constructs derived from the 21-24 nt miRNA fragments. These bands were cut from the gel and three gel slices from the same sample were placed in a 1.5 ml microfuge tube. The gel slices were crushed with the RNase-free Disposable Pellet Pestles and soaked in 250 µl of DNA Gel Elution buffer (1 x). The sample was rotated end-to-end for at least 2 hr at room temperature, and the elute and the gel debris together were then transferred to a Costar Spin-X ® Centrifuge tube filter column (Sigma, St. Louis, Missouri). The column was centrifuged at full speed (13,200 rpm) for 2 min. The DNA in the elute was precipitated by adding 1 µl of Linear Acrylamide, 25 µl of 3 M sodium acetate, pH 5.5 and 750 µl of 96% ethanol and incubating at -80°C for at least 30 min. The DNA pellet was collected by centrifugation at full speed for 30 min at 4°C. The supernatant was discarded and the pellet was washed with 80% ethanol by a vortex mixer. The sample was centrifuged at full speed for another 30 min at 4°C. The DNA pellet was then air dried for up to 10 min at room temperature to remove residual ethanol. The DNA sample was finally suspended in 13 µl of RNase free water. One µl of the size selected purified library was analyzed on a 2100 Bioanalyzer using a DNA High Sensitivity DNA chip according to the manufacturer's instructions.

miRNA identification and target prediction

Total raw sequences generated from Illumina instrument were trimmed to remove low-quality reads and adaptor sequences using Trimmomatic [163]; FastQC

(http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) was used to test data quality. The miRBase database release 21 (http://www.mirbase.org/) and the *Gossypium hirsutum* L. acc. TM-1 genome sequence [164] were used as references for known conserved miRNA mapping and novel miRNA identification, respectively. SeqMan NGen and ArrayStar (DNASTAR, Inc.) were used for mapping of clean short reads (18-44 nt) to the reference and expression level analysis. The miRNA expression was normalized by the Reads Per Kilobase per Million mapped reads (RPKM) method [165].

Conserved miRNAs were identified by comparing to miRBase with the criterion that tags were similar to their homologues within two mismatches and without gaps [173]. Potential novel miRNAs were firstly BLAST verified against non-coding rRNA, tRNA, small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA) in GenBank, tRNA database (http://gtrnadb.ucsc.edu) and Rfam database [174]. Sequences with no hits to known miRNAs or other cellular RNAs were mapped to the *Gossypium hirsutum* L. acc. TM-1 genome. The mireap 0.2 (https://sourceforge.net/projects/mireap/) was used to analyze the pre-miRNAs [175]. The psRNATarget was used for target prediction [176]. Sequence folding prediction was made by MFOLD [177]. The miRNAs with negative hairpins folding energy from -25.7 to -230.5 kcal mol^{-1} [178] were selected for further analysis.

Quantitative RT-PCR

Real time RT-PCR was performed to validate the relative transcript levels of differentially expressed mRNAs and miRNAs determined from RNA Seq.

qRT-PCR analysis for mRNA Seq

Fiber 10-DPA total RNA was used as template for first-strand cDNA synthesis using M-MLV reverse transcriptase (Promega, Madison, Wisconsin) according to the manufacturer's instructions. Two µg (volume was dependent on the concentration of each RNA sample) of fiber RNA, 2 µl (0.5 µg μl^{-1}) of oligo d(T)18 mRNA primer (NEB, Ipswich, Massachusetts) and ddH_2O were transferred to a 0.6 ml microfuge tube at a final volume of 17 µl and then mixed, The samples were denatured at 70°C for 5 min to melt the RNA secondary structure and cooled on ice immediately. Then 5 µl of M-MLV 5 x reaction buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT), 1.25 µl of 10 mM dNTPs, 0.625 µl (25 units) of Murine RNase Inhibitor (NEB, Ipswich, Massachusetts), and 1 µl (200 units) of M-MLV RT were added to the tube to a final volume of 25 µl. The reverse transcription reaction was performed by incubating the tube at 42°C for 1 hr , and the reverse transcriptase was then inactivated by heating at 70°C for 10 min. The qRT-PCR reaction was carried out on lightcycler 480 (Roche, Basel, Switzerland) with the following cycling profile: 95°C for 5 min, following by 45 cycles of 10 sec at 95°C, 15 sec at 60°C and 15 sec at 72°C. Each qRT-PCR reaction mixture contained 10 µl SYBR Green I Master (Roche, Basel, Switzerland), 1 µl of forward primer (10 µM), 1 µl of reverse primer (10 µM), 2 µl of cDNA (4 x fold dilution), 6 µl of ddH_2O. Cotton Ubiquitin 7 gene (*GhUBQ7*) was used to normalize the expression levels of RT-PCR products [179]. All reactions were performed with three replicates. The $2^{-\Delta\Delta Ct}$ method [ΔCt = Ct (differentially expressed gene) – Ct (UBQ7), $\Delta\Delta Ct = \Delta Ct$ (RNAi) – ΔCt (Coker 312), $2^{-\Delta\Delta C}t$ = Relative Expression] was used to calculate relative expression of differentially expressed genes [180]. All primers were designed by using IDT online

tool (https://www.idtdna.com/scitools/Applications/RealTimePCR/) and the sequences of the primers are listed in Table 3.3.

qRT-PCR for miRNA Seq

Cotton fiber small RNAs were polyadenylated and reverse transcribed to form first-strand cDNAs using a Mir-X miRNA First-Strand Synthesis Kit (Clontech, Mountain view, California). A 10 µl of reaction mixture containing 5 µl of 2 x mRQ buffer, 3.75 µl (500 ng) of small RNA and 1.25 µl of mRQ enzyme (Clontech, Mountain view, California) was incubated at 37°C for 1 hr, and the reaction was then terminated at 85°C for 5 min. The synthesized cDNA was then used as the template for quantitative real-time PCR on Lightcycler 480 (Roche, Basel, Switzerland) by using SYBR Advantage qPCR Premix, mRQ 3´ Primer (Clontech, Mountain View, California) and a miRNA-specific 5' primer. The qRT-PCR reaction mixture (20 µl) was prepared by mixing 10 µl of 2 x SYBR Advantage Premix, 0.5 µl (0.25 µM) of miRNA-specific primer, 0.5 µl (0.25 µM) of mRQ 3´ primer, 2 µl of cDNA (1/10 dilution) and 7 µl of ddH_2O. The qRT-PCR was carried out by pre-denaturation at 95°C for 15 sec, following by 45 cycles of 3-step PCR with denaturation at 95°C for 5 sec, annealing at 55-60°C for 15 sec, and extension at 72°C for 15 sec. The cDNA was also used for qRT-PCR amplification of U6 snRNA with U6 forward and U6 reverse primers under the identical conditions. All qRT-PCR were conducted with three biological replicates and included no template controls. Gene expression levels were presented as fold-change and calculated using the comparative threshold cycle (CT) method as described [180] with U6 snRNA as the internal reference. All the primers used for miRNA qRT-PCR were listed in Table 3.3.

Table 3.3 Primers used for Quantitative RT-PCR

Primer Name	Primer Sequences (5′→3′)
Primers for mRNA qRT-PCR	
Gh_D01G0906F	GCGAGGGATGTTTGATTTCTTG
Gh_D01G0906R	CCAGGAGACTGTGTTTCTTCTC
Gh_D08G1232F	TGTGCCAGAGTCTCCTCTAA
Gh_D08G1232R	CCACCTAGGCATCATCTTTCTC
Gh_A05G1317F	GAAAGCAAGCCCGTTTACATTAG
Gh_A05G1317R	TACCTTTGGAGCGAGAAAGAAG
Gh_D10G2595F	CCATAACCGCGACCTTAACT
Gh_D10G2595R	CGGCAACTTCTTCACTGTTTC
Gh_D02G0992F	TCGAGTTCCTCGGGAATTTAAG
Gh_D02G0992R	CAGTTGGGTCAAACCAGAGATA
Gh_D02G0733F	GCAGAAGGATGAAGTCCGTATTA
Gh_D02G0733R	GACCTAGTATGCGAGGAAACAG
Gh_D07G1319F	GGTATTCTGCCTACTTGCTTCT
Gh_D07G1319R	CACCCTCCTTCCCATCAAAT
Gh_D08G1639F	GGTGTCCCTGTGTGATGATT
Gh_D08G1639R	TGCCCAACATCCTCTCTTTAC
Gh_D05G3509F	GAGGTCCTTATGTGCGGATTAC
Gh_D05G3509R	CTGCTCGGATAGTTCCATCAAA
Gh_A05G0931F	TCCGTCCGGGTCTCTATTATC
Gh_A05G0931R	TTGGCCTCTTCCTCTTCAATATC
Gh_D09G0953F	GAGGAGGCAAGGGATTGATTT
Gh_D09G0953R	CCCAACTCACCGACTCATTAC
Gh_D05G0148F	CTCCTGTGCTTGTGTCTCTAAG
Gh_D05G0148R	CGTTCTCCTTGACCACCTTT
Gh_D11G2932F	CTTTGGCACTTCACGAACAAG
Gh_D11G2932R	AACTCACGCCTCTGGTAAAC
Gh_D13G0889F	CTTTGGCACTTCACGAACAAG
Gh_D13G0889R	AACTCACGCCTCTGGTAAAC
Gh_Sca016160G01F	AGTGCTTTACAACCCGAAGG
Gh_Sca016160G01R	ACACGGTCCAGACTCCTAC
Gh_D05G0805F	GTGGACCAAACCAGCTATACA
Gh_D05G0805R	AGGGCATGAAGTGGAGAAAG
Gh_UBQ7F	GAAGGCATTCCACCTGACCAAC
Gh_UBQ7R	CTTGACCTTCTTCTTCTTGTGCTTG
Gh_A12G1250F	GGGCGTTTCCTTTCATGATTT
Gh_A12G1250R	CTCTCAGAAGCGCGGATAAA
Gh_Sca142710G01F	CTCAGAACTGGTACAGACAAAGG
Gh_Sca142710G01R	CAGAGCACTGAGCAGAAATCA

Table 3.3 (Continued)

Gh_A05G2828F	GGCCTCATAATCACTAGCCAAT
Gh_A05G2828R	TGGATTGGTGTTGTTCCTCTAC
Gh_A07G0477F	GACGAACAGTCGACTCAGAAC
Gh_A07G0477R	GAGCACTGGGCAGAAATCA
Gh_Sca006071G01F	CTTCGTATTCTCCGCTCATACC
Gh_Sca006071G01R	GGATCGAAGCTTGAACGATACT
Primers for miRNA qRT-PCR	
MIR2950	TGGTGTGCAGGGGGTGGAATA
MIR169b	CAGCCAAGGATGATTTGCCGG
MIR160	TATGAGGAGCCATGCATGTAT
MIR399c	TGCCAAAGGAGAGTTGGCCTT
MIR399d	TGCCAAAGGAGATTTGCCCTG
Novel-miR-22	TGTGTCAAATCGGCGGCTACATCT
Novel-miR-2	CCGACTGTTTAATTAAAACAAAGT
Novel-miR-3	GCGGCAAAATAGCTCGACGCCAGGAT
Novel-miR-4	GGCTCAGCCGGAGGTAGGGTCCAG
Novel-miR-5	CCGACCTTAGCTCAGTTGGTAGA
Novel-miR-6	AAGAATTTGGGCTTTTGTGACTCG
Novel-miR-7	GGCCAAGATCAATAGACAGGCGTG
Novel-miR-8	TTCCACAGCTTTCTTGAACTT
MIR172	AGAATCCTGATGATGCTGCAG
MIR390a/b/c	AAGCTCAGGAGGGATAGCGCC
MIR166b	TCGGACCAGGCTTCATTCCCC
MIR167a/b	TGAAGCTGCCAGCATGATCTA
MIR164	TGGAGAAGCAGGGCACGTGCA
MIR396a/b	TTCCACAGCTTTCTTGAACTG
MIR162a	TCGATAAACCTCTGCATCCAG

CHAPTER IV

RESULTS

Comparison of CpG methylation status within gene bodies of phytochromes genes between *PHYA1* RNAi transgenic plants and Coker 312 by using bisulphite genomic sequencing

The different CpG site methylation status within coding regions of phytochrome genes in both leaves and 10 DPA fibers from *PHYA1* RNAi transgenic plants and their genetic standard Coker 312 had been determined using bisulfite genomic sequencing. The full sizes of four *PHY* amplicons in this study are shown in Figure 4.1.

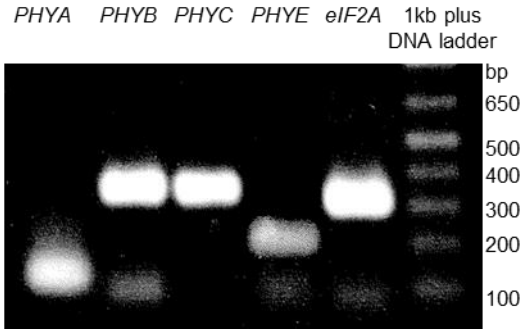

Figure 4.1 Amplicons of cotton phytochromes genes for CpG methylation studies.

In leaves, partial coding regions (gene bodies) of four phytochrome genes *PHYA1*, *PHYB*, *PHYC* and *PHYE* with the sizes of 61, 103, 160, and 228 bp containing two, four, three and six CpG sites, respectively, were amplified and analyzed by BISMA (Table 4.1).

Table 4.1 The sizes of *PHY* amplicons for methylation analysis and their numbers of CpG sites

	PHYA1	*PHYB*	*PHYC*	*PHYE*
Leaves	61 | 2	103 | 4	160 | 3	228 | 6
10 DPA Fibers	61 | 2	279 | 6	160 | 3	170 | 5

For *PHYA1*, the CpG sites at the position 50 were unmethylated in both RNAi line and Coker 312. The other CpG site at the position 11 had 46.7% (7 out of 15) methylation in Coker 312; however, the methylation level increased to 100% in the RNAi line. For *PHYB*, all four CpG sites were highly methylated in Coker 312; there was 80% (12 out of 15) methylation levels in the first two CpG sites and 100% in the latter two sites. In the RNAi line, the methylation level at the first two sites dropped to 46.7% (7 out of 15), and only one out of the fifteen subclones was unmethylated at the third site and completely methylated at the last one. For *PHYC*, the second and third CpG sites showed no differences on methylation between the RNAi line and Coker 312, however methylation at the first CpG site in RNAi line was 100%, compared to 86.7% in Coker 312. For *PHYE*, the methylation level of the RNAi line was slightly increased compared to Coker 312 in general. The second, third and fourth CpG sites were completely methylated in both RNAi and Coker 312 lines. However, at the first, fifth and sixth (the

last) CpG sites, the methylation levels of the RNAi line were 43.8%, 18.8% and 100%, respectively. At the relevant sites in Coker 312, the methylation levels were 20%, 26.7% and 73.3%, respectively (Figure 4.2 A).

In fibers, 61-, 279-, 160-, and 170-bp amplicons of four phytochrome genes *PHYA1*, *PHYB*, *PHYC* and *PHYE* containing two, six, three and five CpG sites, respectively, were identified and analyzed by BISMA (Table 4.1). For *PHYA1* and *PHYC*, their CpG site methylation levels at each sites were exactly the same in RNAi lines and Coker312. For *PHYE*, the methylation levels of three CpG sites in the middle were also identical in both RNAi lines and Coker 312. While, the methylation level at the first and the last CpG sites were determined to be 40% and 20%, respectively, in the RNAi lines; however at the relevant sites in Coker 312, the methylation levels were 21.4% and 28.6%, respectively. More interesting findings were that the CpG methylation levels of *PHYB* of *PHYA1* RNAi line was decreased by ~ 50% within coding regions compared to those of Coker 312 (Figure 4.2 B).

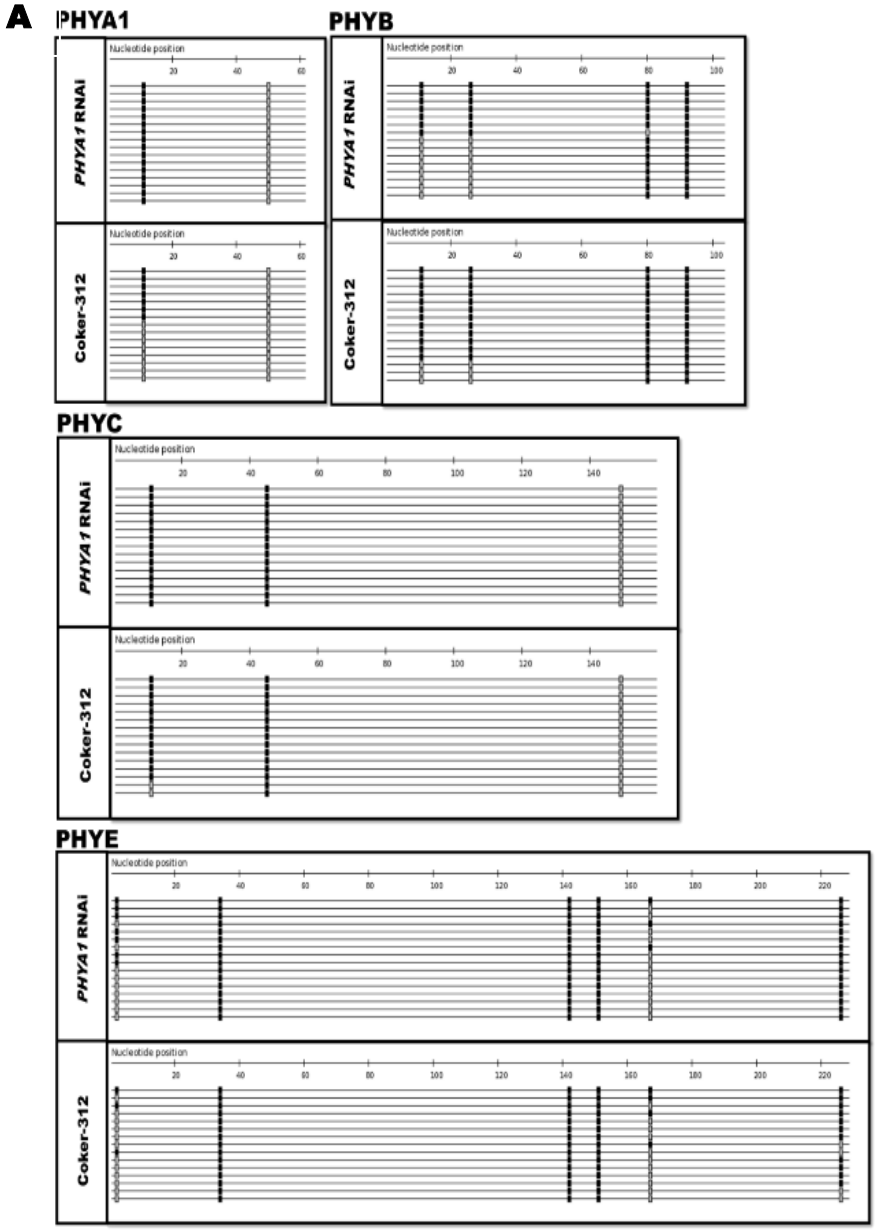

Figure 4.2 Comparison of CpG methylation patterns of *PHYA1*, *PHYB*, *PHYC* and *PHYE* gene bodies in leaves (A) and 10 DPA fibers (B) between *PHYA1* RNAi line and Coker 312.

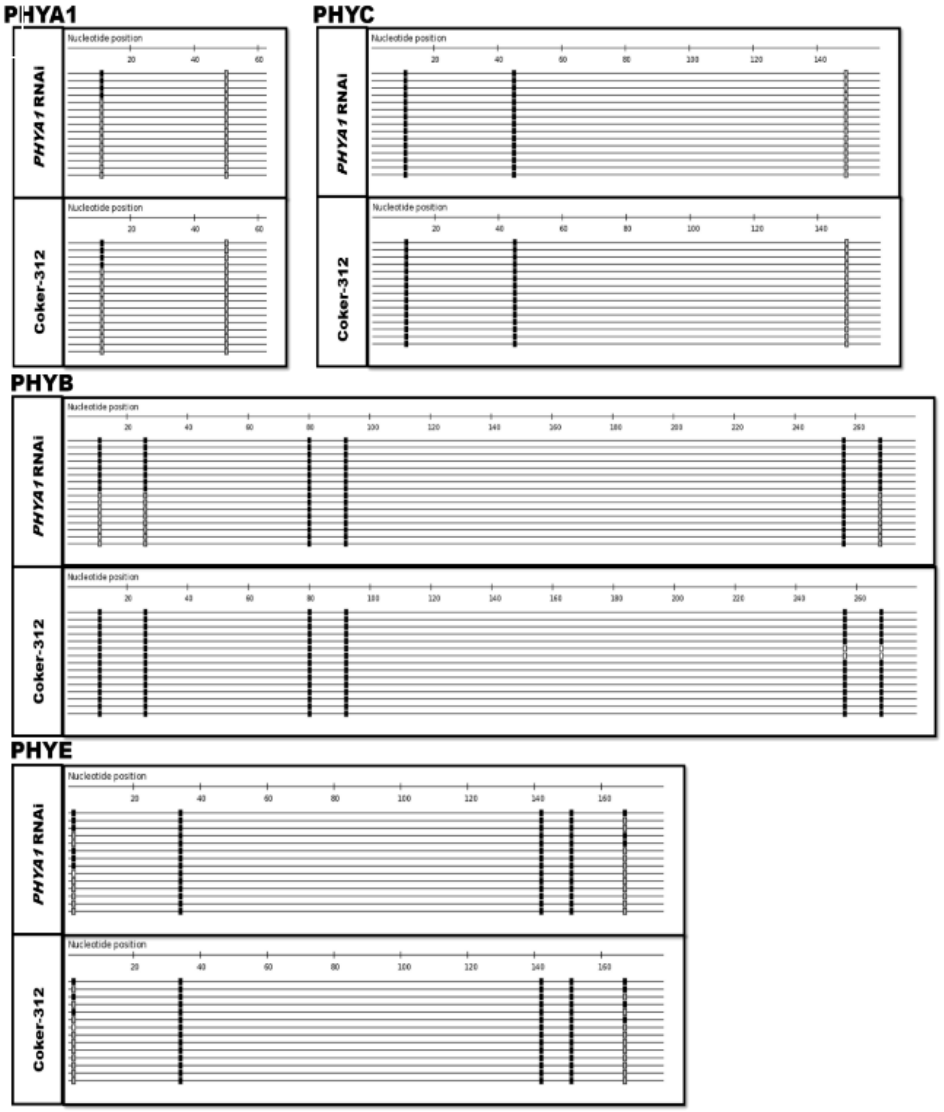

Figure 4.2 (Continued)

The *PHY* fragments of interest were amplified by sequence-specific primers using bisulfite-treated genomic DNA as the template. Fifteen or sixteen unique "non-sister" individual clones from independent PCR reactions were selected for sequencing. Each line represents a unique "non-sister" individual bisulfite sequencing result. The length of each line is proportional to the size of the fragment. The black boxes represent methylated sites, grey boxes denote unmethylated sites and blank boxes indicate the undermined ones.

The methylation studies indicated that the gene bodies of *PHYA1*, *PHYC* and *PHYE* in leaves of the *PHYA1* RNAi line had higher methylation levels than those in Coker 312. The RNAi line, however, had lower methylation levels in *PHYB* compared to Coker 312. In fibers, the CpG methylation levels of *PHYB* was decreased by ~ 50% within coding regions of *PHYA1* RNAi line compared to those of Coker 312. The other *PHY* genes showed no significant changes otherwise. The promoter region of housekeeping gene transcription initiation factor 2 (*eIF2A*) was selected as a control to evaluate the accuracy of the bisulfite sequencing data (Figure 4.3). There was a very low level of DNA methylation in the *eIF2A* upstream promoter region in both RNAi line and Coker 312 as previously reported [137].

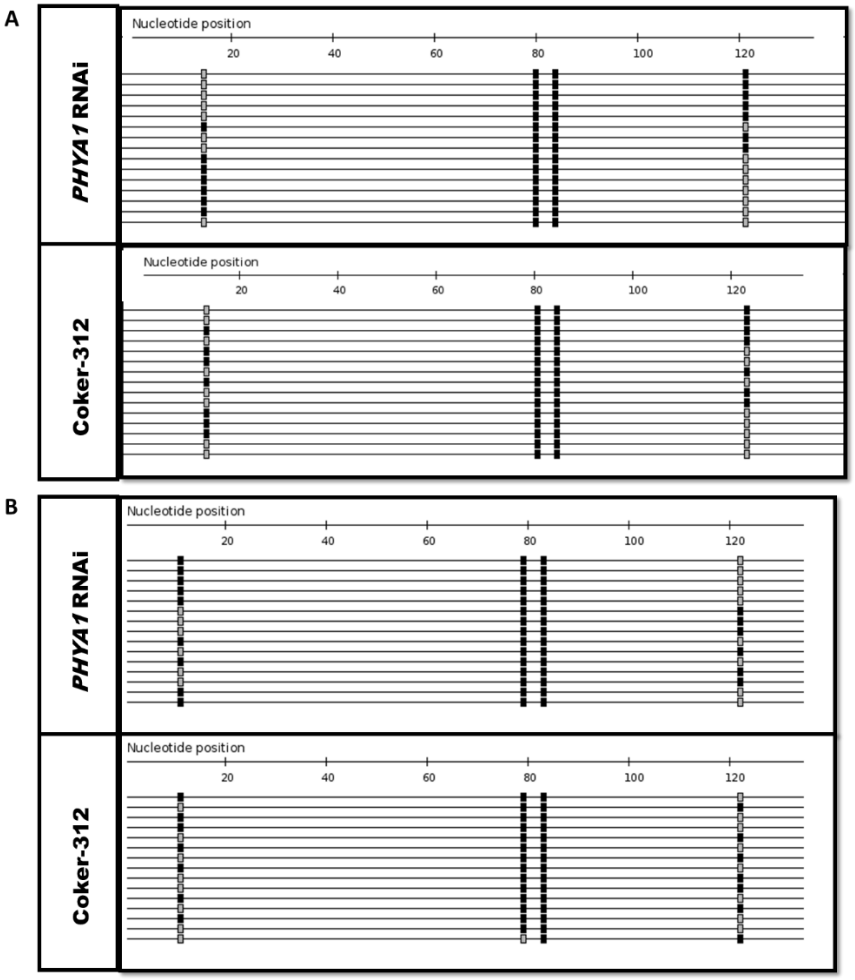

Figure 4.3 Comparison of CpG methylation patterns of eIF2A gene promoter regions in leaves (A) and 10-DPA fibers (B) between *PHYA1* RNAi line and Coker 312.

Transcriptome analysis of 10-DPA fiber of *PHYA1* RNAi line by Illumina RNA Sequencing

Sequence statistics of six 10-DPA fiber libraries

To determine global transcriptome changes in 10-DPA fibers between *PHYA1* RNAi and Coker 312 lines, the 10-DPA fiber RNA samples from three individual plants of the two lines were used to generate 6 independent libraries that were sequenced using the Illumina HiSeq 2500 sequencing platform. Approximately 565.8 million reads of 50-bp long were generated from all 6 libraries (Table 4.2). Low quality reads and adaptor sequences were then trimmed off with Trimmomatic and the clean reads were mapped to 70,478 transcripts from the *Gossypium hirsutum* L. acc. TM-1 genome sequence [164].

Table 4.2 Total reads statistics of RNA-Seq libraries

Sample	CF1	CF2	CF3	RF1	RF2	RF3
Total Sample Reads	94,720,967	93,523,005	94,121,986	95,064,738	94,462,679	93,860,621
Total Reads Mapped to Reference	94,694,073	92,987,578	93,840,825	94,718,530	94,205,949	93,693,368
Total Unique Reads	28,996,162	27,493,684	28,244,922	28,662,405	27,759,879	26,857,354

Note: CF and RF denote the fibers from Coker 312 and *PHYA1* RNAi plants, respectively. The RNA samples from Coker 312 and PHYA1 RNAi lines have 3 biological replicates (replicates 1-3).

Global transcriptome changes in *PHYA1* RNAi plant during fiber development

The mapped sequenced reads of all identified transcripts were used for differential expression analysis with the ArrayStar program (Version 12, DNASTAR Inc., USA) with a FDR of <0.05. The overall expression patterns of the mapped genes between the RNAi line and Coker 312 were similar with a linear correlation R^2=0.9747 (Figure 4.4). A total of 142 genes (Fold Change ≥ 2) were either up or down regulated in *PHYA1* RNAi plant compared to Coker 312. Of these genes, 138 genes were down-regulated and 4 genes were up-regulated. The 142 genes differentially expressed in *PHYA1* 10-DPA fibers were listed in Table 4.3.

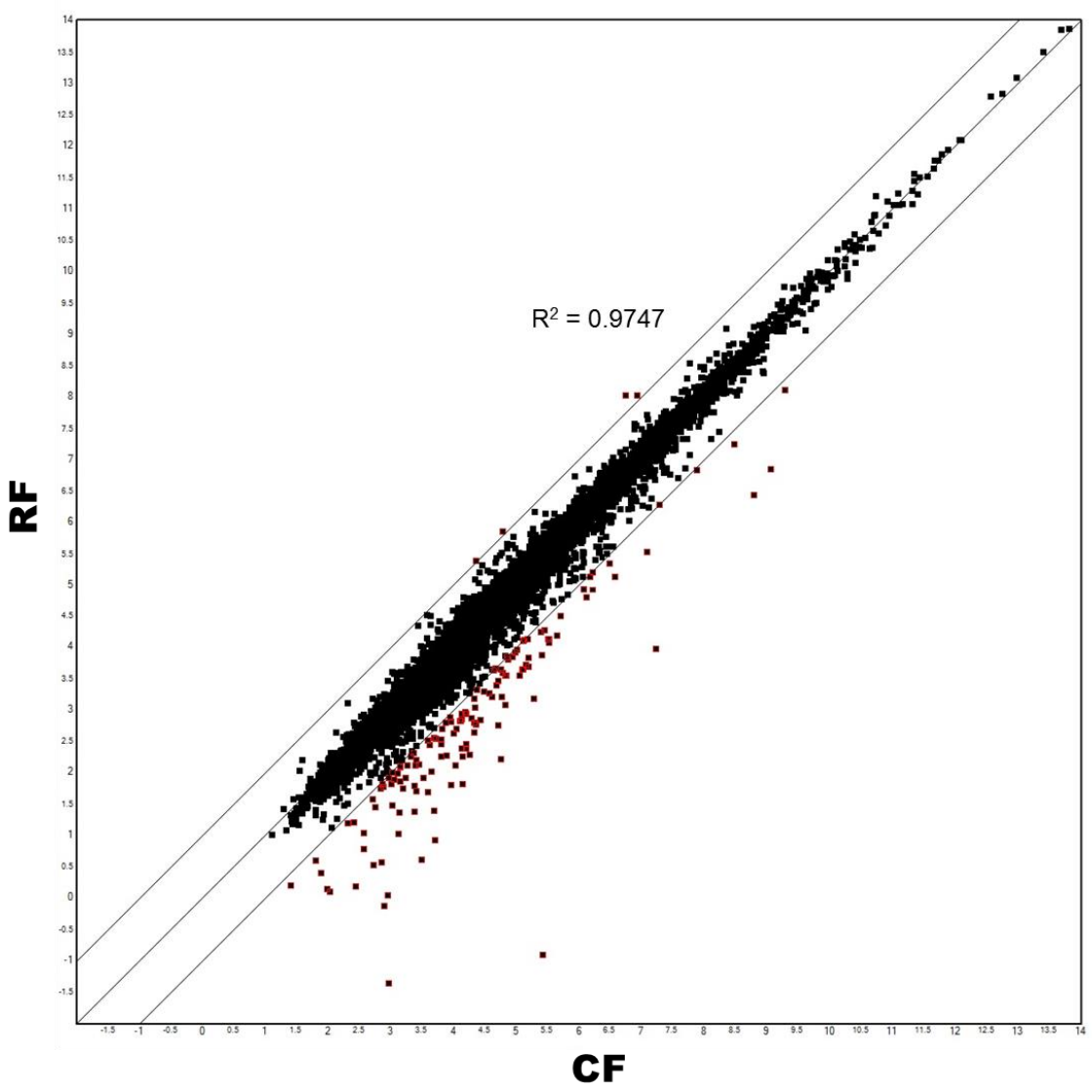

Figure 4.4 Transcriptome scatter plot of PHYA1 RNAi plant (RF) compared to Coker 312 (CF).

Genes were normalized by the RPKM method. Red data points indicate genes that show 2-fold or greater changes in expression with a FDR adjusted $p < 0.05$, $R^2 = 0.9747$.

Table 4.3 The list of 142 differentially expressed genes (fold change ⩾ 2) in *PHYA1* RNAi line

Gene Accession Number (NBI)	Gene Annotation	Fold change
Gh_D01G0906	phosphoprotein ecpp44-like	0.0125
Gh_D08G1232	probable wrky transcription factor 41	0.0500
Gh_A05G1317	pyruvate decarboxylase 1	0.1048
Gh_D10G2595	---NA---	0.1233
Gh_A08G2417	probable wrky transcription factor 41	0.1330
Gh_D04G1318	probable wrky transcription factor 33	0.1359
Gh_D02G0992	xyloglucan endotransglucosylase hydrolase family protein	0.1463
Gh_D04G1245	bahd acyltransferase dcr	0.1735
Gh_D02G0733	alcohol dehydrogenase	0.1948
Gh_D07G1319	uncharacterized calcium-binding protein at1g02270-like isoform x1	0.1988
Gh_D08G1639	abscisic acid 8 -hydroxylase 1-like	0.2031
Gh_D12G1525	lov domain-containing protein	0.2059
Gh_D05G3509	scarecrow-like protein 21	0.2105
Gh_A02G0689	alcohol dehydrogenase	0.2159
Gh_A05G0931	probable protein phosphatase 2c 25	0.2180
Gh_D08G2557	phosphatase 2c family isoform 1	0.2244
Gh_A08G1344	abscisic acid 8 -hydroxylase 1-like	0.2330
Gh_D02G0732	alcohol dehydrogenase class-p-like	0.2338
Gh_D09G0953	probable ccr4-associated factor 1 homolog 11	0.2504
Gh_D05G0148	ein3-binding f-box protein 1-like	0.2542
Gh_D08G2556	nuclease harbi1	0.2573
Gh_A05G0839	exocyst complex component exo70b1-like	0.2622
Gh_A08G2389	oxalate-- ligase	0.2650
Gh_D04G0293	nac domain-containing protein 2-like	0.2686
Gh_A09G0923	probable ccr4-associated factor 1 homolog 11	0.2710
Gh_A05G2138	serine threonine-protein kinase-like protein ccr4	0.2800
Gh_D06G1403	ethylene-responsive transcription factor 4-like	0.2846
Gh_D07G0283	PREDICTED: uncharacterized protein LOC105769700	0.2886
Gh_A10G0504	linoleate 13s-lipoxygenase 3- chloroplastic	0.2941
Gh_A08G0965	nadp-specific glutamate	0.2982
Gh_D05G3030	xyloglucan endotransglucosylase hydrolase family protein	0.2990
Gh_D06G1513	endoplasmic reticulum oxidoreductin-1-like	0.3001
Gh_D08G2484	ein3-binding f-box protein 1-like	0.3076
Gh_D02G1480	probable ribose-5-phosphate isomerase 2	0.3100
Gh_A05G2724	xyloglucan endotransglucosylase hydrolase family protein	0.3247

Table 4.3 (Continued)

Gh_A08G2191	nuclease harbi1	0.3272
Gh_D07G2014	nam protein	0.3300
Gh_A08G2112	ein3-binding f-box protein 1-like	0.3311
Gh_A07G1811	nam protein	0.3327
Gh_D01G1679	---NA---	0.3346
Gh_D05G1208	tetratricopeptide repeat-like superfamily	0.3389
Gh_D08G1346	protein rolling stone-like isoform x2	0.3394
Gh_D03G1189	scarecrow-like protein 34	0.3401
Gh_A03G0210	probable wrky transcription factor 25	0.3409
Gh_A13G1709	nuclease harbi1	0.3460
Gh_D01G1033	zinc finger protein zat10-like	0.3467
Gh_D10G2455	zinc finger ccch domain-containing protein 29-like	0.3469
Gh_A03G2061	protein yls9-like	0.3497
Gh_D05G0600	wrky transcription factor 40 -like protein	0.3524
Gh_A03G0889	phosphate transporter pho1 homolog 3-like	0.3552
Gh_D08G0665	auxin-induced protein 22b	0.3636
Gh_D12G0236	maternal effect embryo arrest 14 isoform 1	0.3643
Gh_A10G1443	protein srg1	0.3655
Gh_D01G0450	ethylene-responsive transcription factor 4-like	0.3663
Gh_D02G0832	ethylene-responsive transcription factor rap2-4-like	0.3667
Gh_A09G1902	non-specific lipid-transfer protein at2g13820	0.3691
Gh_D02G0417	ethylene receptor 2	0.3707
Gh_A03G2168	---NA---	0.3749
Gh_Sca006021G02	calcium-dependent lipid-binding family	0.3758
Gh_A04G0305	ethylene-responsive transcription factor 4-like	0.3810
Gh_A08G1031	sucrose synthase 1	0.3839
Gh_A06G1144	ethylene-responsive transcription factor 4-like	0.3890
Gh_A10G2134	zinc finger ccch domain-containing protein 29-like	0.3900
Gh_D06G1078	wrky transcription factor 40 -like protein	0.3922
Gh_D03G1371	probable wrky transcription factor 25	0.3947
Gh_A02G0671	u-box domain-containing protein 28-like	0.3974
Gh_D12G0679	maternal effect embryo arrest isoform partial	0.4007
Gh_A07G0226	hypothetical protein F383_30923	0.4008
Gh_A12G2296	heavy metal-associated domain protein	0.4038
Gh_A11G1993	probable indole-3-acetic acid-amido synthetase	0.4044
Gh_A01G0875	prolyl endopeptidase	0.4047
Gh_D08G2140	probable protein phosphatase 2c 25	0.4066

Table 4.3 (Continued)

Gh_A12G0132	nitronate monooxygenase	0.4074
Gh_A10G0255	ethylene-responsive transcription factor 4-like	0.4080
Gh_D02G1817	---NA---	0.4085
Gh_D06G0817	ethylene-responsive transcription factor erf017-like	0.4092
Gh_A05G0434	ein3-binding f-box protein 1	0.4098
Gh_A06G1910	nac domain-containing protein 2	0.4106
Gh_D03G1188	scarecrow-like protein 30	0.4110
Gh_A03G1618	probable protein phosphatase 2c 63	0.4134
Gh_A05G3992	xyloglucan endotransglucosylase hydrolase family protein	0.4146
Gh_A06G1923	probable wrky transcription factor 40	0.4189
Gh_D08G1949	4-coumarate-- ligase-like 10	0.4230
Gh_A09G0534	stem-specific protein tsjt1	0.4249
Gh_A05G0483	wrky transcription factor 40 -like protein	0.4298
Gh_Sca013634G01	two-component response regulator-like aprr5 isoform x2	0.4309
Gh_D05G3348	ethylene-responsive transcription factor 4-like	0.4321
Gh_D11G3494	disease resistance protein rga4 isoform x1	0.4323
Gh_A12G1612	disease resistance protein rga3	0.4323
Gh_A05G2211	respiratory burst oxidase homolog protein d-like	0.4323
Gh_A11G0519	udp-glucuronate:xylan alpha-glucuronosyltransferase 1-like	0.4360
Gh_D02G2032	probable protein phosphatase 2c 63	0.4368
Gh_A01G1439	expansin-like a2	0.4384
Gh_D10G1768	acetate butyrate-- ligase peroxisomal-like	0.4392
Gh_A03G1070	probable ribose-5-phosphate isomerase 2	0.4397
Gh_A01G1726	protein yls9-like	0.4404
Gh_D07G1307	suppressor of disruption of tfiis	0.4435
Gh_D13G0440	classical arabinogalactan protein 9	0.4437
Gh_A04G0196	scarecrow-like protein 21	0.4467
Gh_A06G0917	wrky transcription factor 40 -like protein	0.4499
Gh_D04G1573	glutamyl-trna amidotransferase subunit a	0.4514
Gh_A05G1814	lob domain-containing protein 39-like	0.4515
Gh_A05G3754	wat1-related protein at4g08300-like	0.4520
Gh_A08G1791	probable protein phosphatase 2c 25	0.4529
Gh_D03G1808	late embryogenesis abundant hydroxyproline-rich glycoprotein isoform 1	0.4549
Gh_A13G1311	9-cis-epoxycarotenoid dioxygenase chloroplastic-like	0.4563
Gh_D02G0837	protein exordium-like	0.4565
Gh_D06G1046	ein3-binding f-box protein 1-like	0.4572
Gh_A05G3796	cyclin family isoform 1	0.4592

Table 4.3 (Continued)

Gh_A11G0944	endonuclease or glycosyl hydrolase with c2h2-type zinc finger isoform 1	0.4615
Gh_D05G0562	ein3-binding f-box protein 1-like	0.4656
Gh_A07G1599	probable sodium-coupled neutral amino acid transporter 6	0.4673
Gh_D06G2235	scarecrow-like protein 13	0.4675
Gh_A11G0020	purple acid phosphatase 17-like	0.4688
Gh_A06G0709	ethylene-responsive transcription factor erf017-like	0.4720
Gh_D05G1276	inorganic pyrophosphatase 1-like	0.4750
Gh_D11G3015	soybean gene regulated by cold-	0.4753
Gh_D12G2716	two-component response regulator-like aprr5 isoform x1	0.4763
Gh_D02G1271	phosphate transporter pho1 homolog 3-like	0.4765
Gh_A10G1517	acetate butyrate-- ligase peroxisomal-like	0.4783
Gh_D05G2829	gata transcription factor 8-like	0.4803
Gh_A09G0414	scarecrow-like protein 8	0.4814
Gh_A10G2086	uncharacterized acetyltransferase at3g50280-like	0.4815
Gh_A07G0017	probable wrky transcription factor 17	0.4832
Gh_A08G0581	calcium-transporting atpase plasma membrane-type	0.4834
Gh_A03G0639	udp-glycosyltransferase 89b1-like	0.4841
Gh_A08G1509	homeobox-leucine zipper protein hat4-like	0.4847
Gh_A10G2166	senescence associated gene	0.4847
Gh_A02G1410	phosphoenolpyruvate carboxykinase	0.4873
Gh_D11G1089	endonuclease or glycosyl hydrolase with c2h2-type zinc finger isoform 1	0.4874
Gh_D04G0606	e3 ubiquitin-protein ligase ring1-like	0.4914
Gh_A06G1742	scarecrow-like protein 13	0.4954
Gh_D01G2344	cytochrome P450 87a3-like	0.4956
Gh_D08G1309	sucrose synthase 1	0.4972
Gh_D09G1773	nac domain-containing 78 -like protein	0.4977
Gh_A01G1640	calmodulin-binding protein 25-like	0.4985
Gh_D10G0255	ethylene-responsive transcription factor 4-like	0.4985
Gh_D10G0773	glycerophosphodiester phosphodiesterase gde1-like	0.4989
Gh_D11G2932	pectinesterase inhibitor-like	2.0540
Gh_D13G0889	cytosolic sulfotransferase 12-like	2.1120
Gh_Sca016160G01	uncharacterised protein	2.1153
Gh_D05G0805	pollen ole e 1 allergen and extensin family	2.4296

A heatmap depicting the differential expression profiles of *PHYA1* RNAi line was constructed using Blast2Go and shown in Figure 4.5. Many transcription factor families were significantly affected. For example, 11 genes in the WRKY gene family (Gh_D08G1232, Gh_A08G2417, Gh_D04G1318, Gh_A03G0210, Gh_D05G0600, Gh_D06G1078, Gh_D03G1371, Gh_A06G1923, Gh_A05G0483, Gh_A06G0917 and Gh_A07G0017) encoding WRKY transcription factors or like proteins were down-regulated in 10-DPA fibers of the *PHYA1* RNAi line compared to Coker 312. Transcriptome profile analysis of WRKY gene family had revealed that many WRKY genes were involved in multiple fiber developmental processes [181]. Two genes encoding ABA 8′-hydroxylase (Gh_A08G1344 and Gh_D08G1639), which is a cytochrome P450 monooxygenase and catalyzes the first step in the oxidative degradation of ABA were down-regulated [182, 183]. ABA was reported to play a negative role in fiber development [10]. Certain proteins and enzymes which are involved in cell wall biogenesis and organization process were also affected by silencing the *PHYA1* gene. For example, the expression of 4 genes encoding xyloglucan endotransglucosylase hydrolase family protein (Gh_D02G0992, Gh_A05G2724, Gh_D05G3030 and Gh_A05G3992) and Gene Gh_A11G0519 for udp-glucuronate:xylan alpha-glucuronosyltransferase 1-like, a cell wall-modifying enzyme-like, were down-regulated. In addition, sucrose synthase and proteins related to lipid metabolism were also affected.

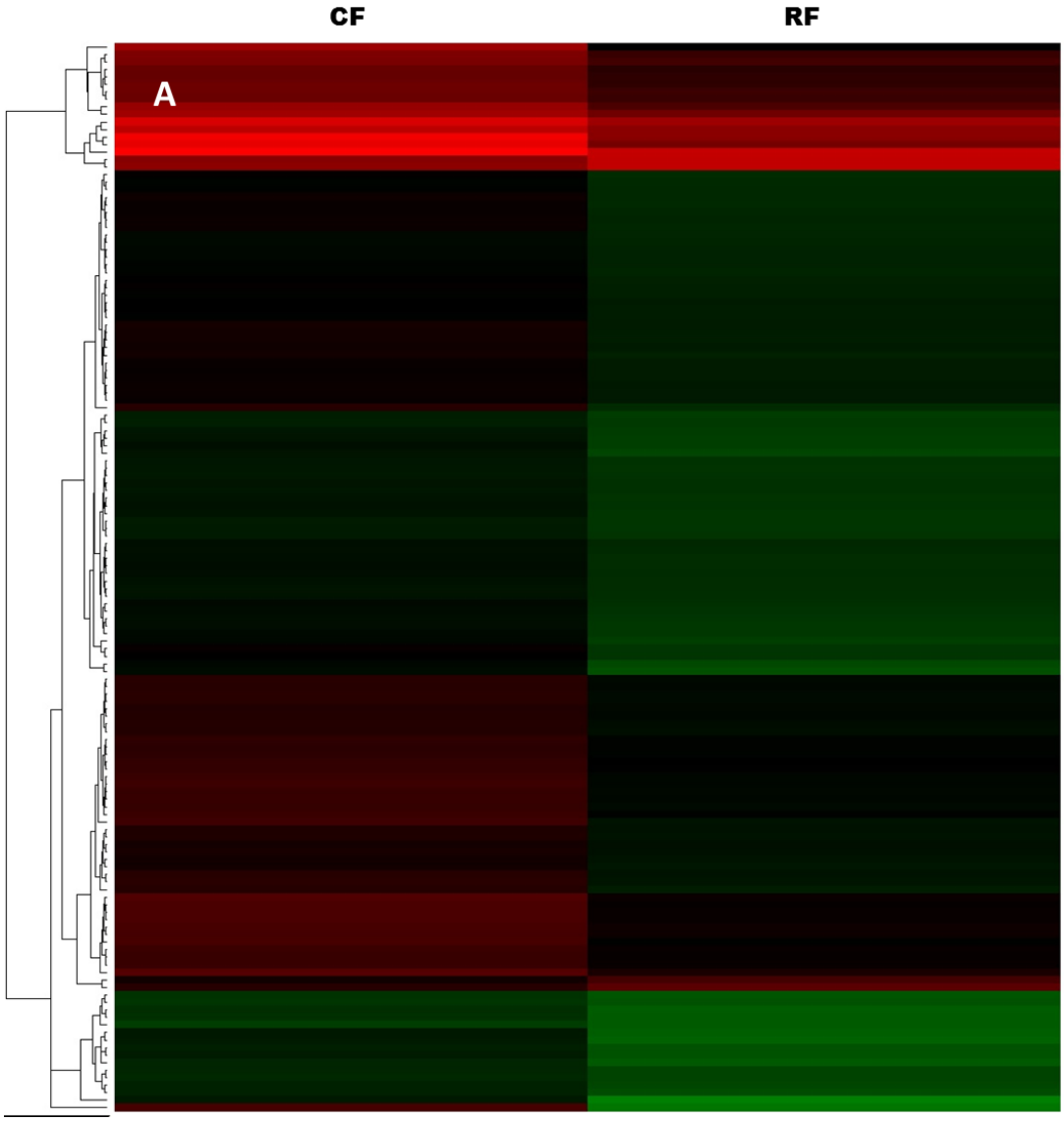

Figure 4.5 Hierarchical cluster analyses and heatmap visualization for differentially expressed genes (DEGs) in *PHYA1* RNAi line (RF) vs Coker 312 (CF) during fiber development.

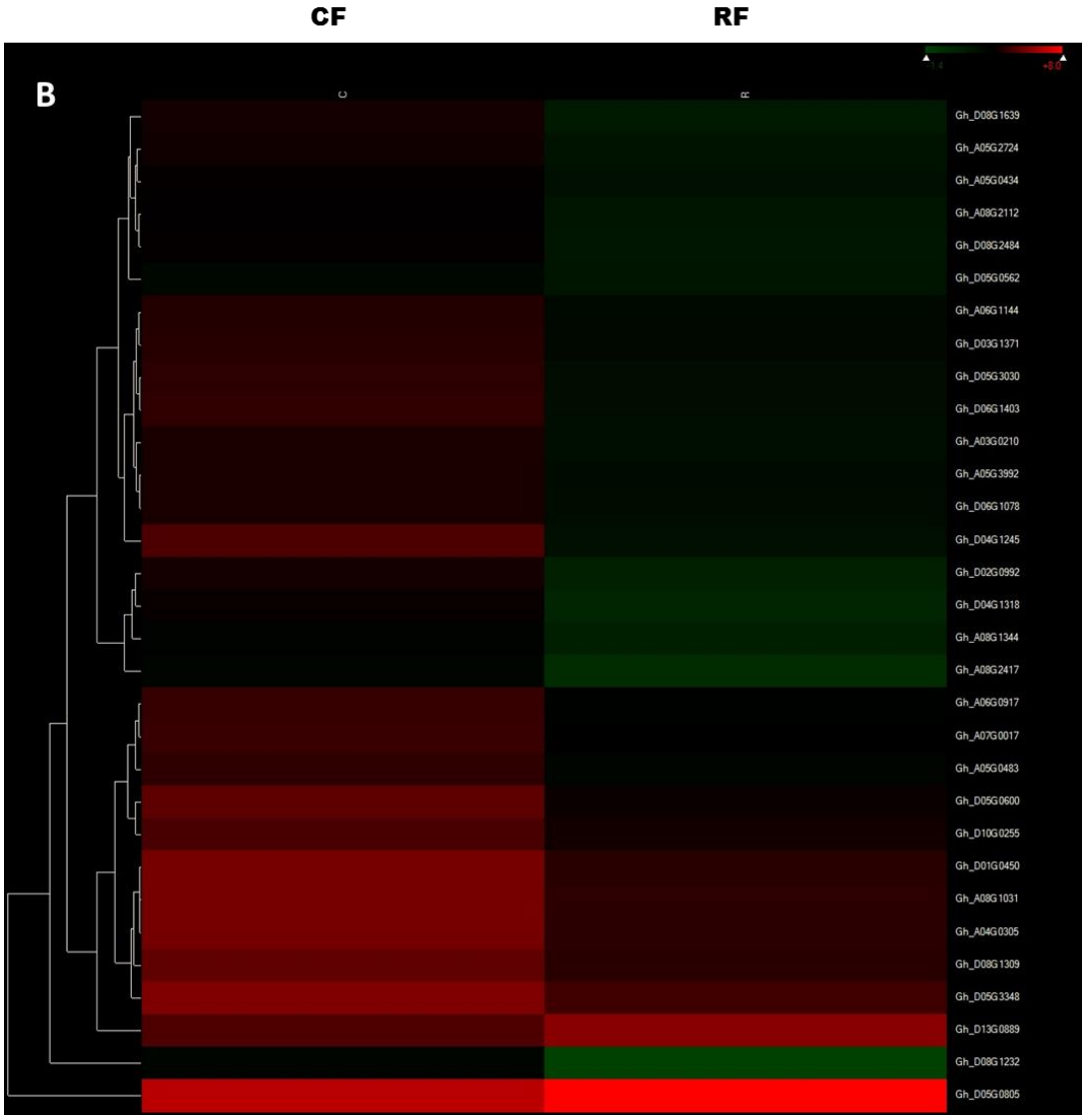

Figure 4.5 (Continued)

A) 142 DEGs B) DEGs were mentioned or discussed in this study. Red color represents up-regulated, and green color represents down-regulated.

These differentially expressed genes (DEGs) were distributed across all 26 chromosomes in the tetraploid progenitor A-genome and D-genome (Figure 4.6). Of these 142 genes, 68 genes were originated from A-genome, and 71 genes were from D-genome. In A-genome, the DEGs were most abundantly located in chromosome 7 with 12 genes, which had 18% of the DEGs distributed in A-genome. Following the chromosome 5 was the chromosome 8 with 10 DEGs, composed of 15% of the group. For D-genome, similar distribution patterns were found, chromosome 19 and 24 had the highest numbers of DEGs, but chromosome 14 also had identical numbers of DEGs as 19 and 24. All three chromosomes 19, 24 and 14 had 10 DEGs, each having 14% DEGs in D-genome (Figure 4.6).

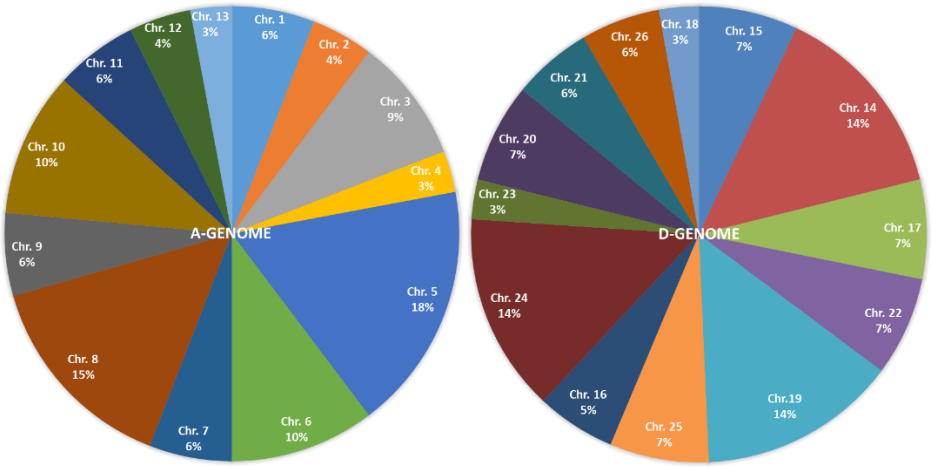

Figure 4.6 Chromosomal distribution of 142 differentially expressed genes.

Distribution of differentially expressed transcripts was determined by the alignment of *PHYA1* RNAi RNA-Seq reads to the draft genome of the allotetraploid cotton *Gossypium hirsutum* L. acc. TM-1.

Functional annotation of differentially expressed transcripts and Gene ontology

The 142 differentially expressed genes were annotated by Blast2GO [166], and additional GO terms were added to DEGs by InterProScan 5 [167] as necessary. All the DEGs were subjected to gene ontology (GO) enrichment analysis and grouped into three subsections: molecular function, biological process, and cellular components by Blast2GO (Figure 4.7). The results from the enrichment analysis indicated that most of the identified DEGs were involved in cellular metabolic process, heterocyclic/organic cyclic compound binding and multiple enzyme (hydrolase, transferase, and oxidoreductase) activities, and cell structures which have been reported to play roles in fiber development (Figure 4.7).

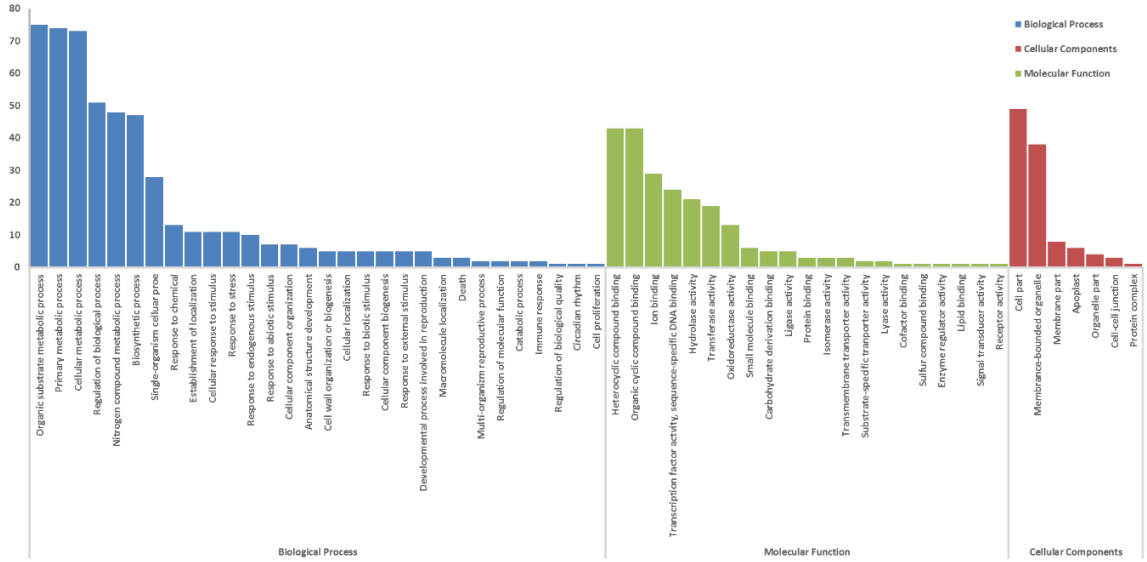

Figure 4.7 Gene ontology categories of differentially expressed genes (Fold change ⩾ 2) in *PHYA1* RNAi line compared to Coker 312.

The number of differentially expressed genes in each categories is shown in the vertical axis.

KEGG pathway analysis

To identify metabolic pathways in RNAi line during the fiber elongation stage, the 142 DEGs were mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) database. As a result, 28 pathways were mapped in the KEGG database and they were mostly related to glycolysis/gluconeogenesis, pyruvate metabolism, and metabolism by cytochrome P450 (Figure 4.8). Plant cytochrome P450s were shown to play an important role in response to environmental stimuli and participate in the synthesis of a variety of biomolecules including fatty acid conjugates, plant hormones, secondary metabolites, lignin, and defensive compounds [184]. It was reported that *PHYA1* RNAi plant had a better tolerance to abiotic stresses, including salinity, heat and drought stresses [185]. Thus the improvement of fiber quality and stress tolerance of *PHYA1* RNAi plant might be mediated through the regulation of proteins involved in cytochrome P450 pathways.

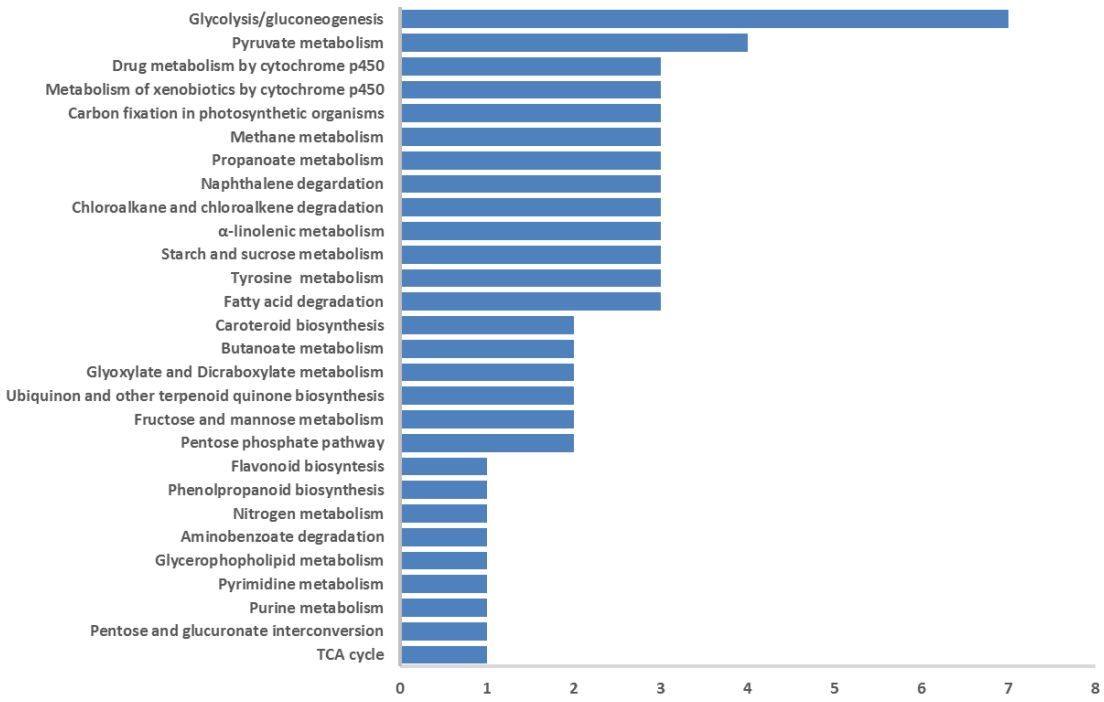

Figure 4.8 Kyoto Encyclopedia of Genes and Genomes pathway classifications of differentially expressed genes.

The number of differentially expressed transcripts in each pathways is shown.

Validation of RNA sequencing results by quantitative real-time PCR analysis (qRT-PCR)

To test whether the RNA-seq gene expression results were reliable, RNA transcripts of 15 DEGs in *PHYA1* RNAi plant were selected for qRT-PCR analysis. A comparison of the expression of each transcript by qRT-PCR and RNA-seq was made and shown in Figure 4.9. All the 15 transcripts were successfully detected by qRT-PCR, and the expression patterns of these transcripts showed general agreement with RNA-seq results. These 15 transcripts include four up-regulated genes (Figure 4.9). Three out of the four genes were successfully annotated and they were pollen ole e1 allergen and extension (Gh_D05G0805), cytosolic sulfotransferase 12-like protein (Gh_D13G0889) and pectinesterase inhibitor-like protein (Gh_D11G2932). The gene with the accession number Gh_Sca016160G01 was uncharacterized (Table 4.3). The rest of the 11 genes were down-regulated, and four of them had more than 8-fold down-regulation in RNAi plant. The expression levels of three out of the four genes were consistent with the results from RNA-seq, and they encoded WRKY transcription factor 41 (Gh_D08G1232), pyruvate decarboxylase 1 (Gh_A05G1317) and an unknown gene (Gh_D10G2595). A phosphoprotein ecpp44-like protein (Gh_D01G0906) gene was about 80-fold down-regulated in RNAi plant via RNA sequencing, however, it only showed approximately 4-fold down-regulation by the qRT-PCR method (Table 4.3 and Figure 4.9). This discrepancy between RNA-seq and qRT-PCR results could have been caused by very low counts of transcripts detected in RNAi line, which were usually unreliable for the down-stream analysis. The other seven genes include Gh_D09G0953 (probable ccr4-associated factor 1 homolog 11), Gh_D05G3509 (scarecrow-like protein 21), Gh_D05G0148 (ein3-binding f-box protein 1-like), Gh_D08G1639 (abscisic acid 8 -hydroxylase 1-like),

Gh_A05G0931 (probable protein phosphatase 2c 25), Gh_D02G0733 (alcohol dehydrogenase), Gh_D02G0992 (xyloglucan endotransglucosylase hydrolase family protein) were all down-regulated in RNAi line, which were consistent with the results from RNA-Seq (Table 4.3 and Figure 4.9). In summary, the majority (except for Gh_D01G0906) of the selected transcripts tested by qRT-PCR showed similar expression patterns as determined by the statistical analysis of RNA-Seq data.

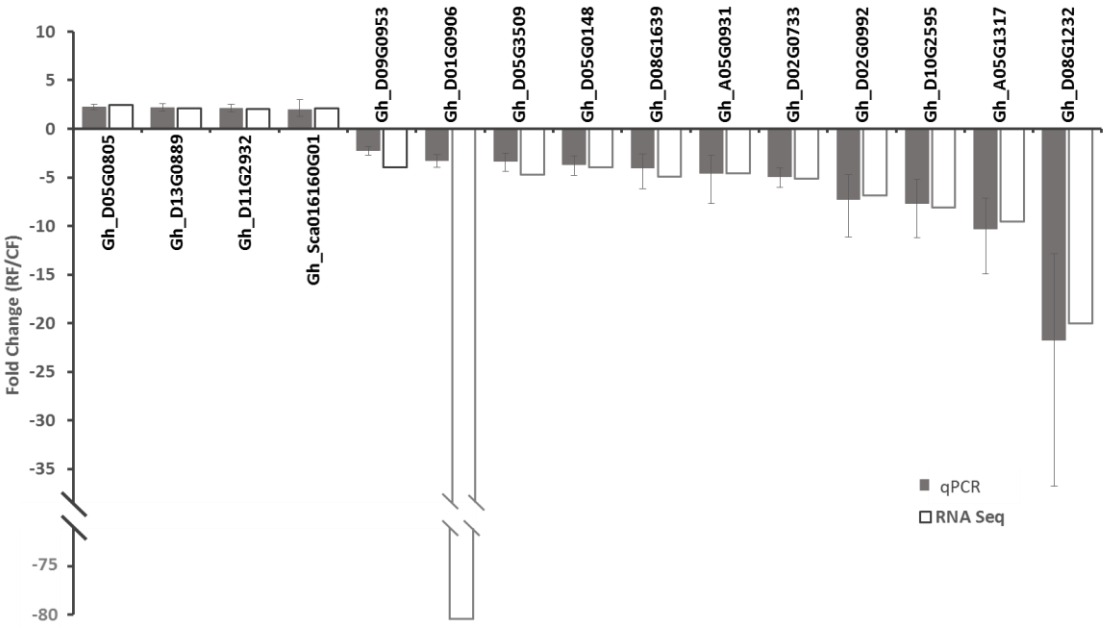

Figure 4.9 Validation of the expression patterns of differentially expressed genes in 10 DPA fiber of *PHYA1* RNAi line by qRT-PCR.

Gene expression was represented as relative fold change $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_{T\ RNAi} - \Delta C_{T\ Coker\ 312}$). The expression levels of genes were normalized by using *GhUBQ7* as a reference. Three biological replicates and three technical replicates were included for qRT-PCR experiments. The error bar represent the confidence limits.

Global analysis of the *PHYA1* RNAi line miRNAome in different fiber developmental stages (5-, 10-, and 15-DPA) fibers by miRNA-seq

Construction of small RNA sequencing libraries and quality control

To characterize small RNAs specifically or differentially expressed in fibers of the *PHYA1* RNAi plants, 18 barcoded small RNA libraries were constructed by using miRNAs extracted from cotton fibers at three different developmental stages of 5-DPA, 10-DPA, and 15-DPA, respectively, from both *PHYA1* RNAi line and the parent line Coker 312 (Table 4.4).

Table 4.4 Basic information of miRNA libraries

Genotype	Tissue type	Barcode	Index Seq	Qubit (ng μl^{-1})	Fragment Size (bp)	Concentration (nM)
Coker 312	5DPA Fiber	Adapter 1	ATCACG	5.58	147	62.38421746
Coker 312	5DPA Fiber	Adapter 2	CGATGT	3.83	148	42.53046258
Coker 312	5DPA Fiber	Adapter 3	TTAGGC	5.26	148	58.40998255
PHYA1 RNAi 3 copes	5DPA Fiber	Adapter 4	TGACCA	4.41	146	49.64075179
PHYA1 RNAi 3 copies	5DPA Fiber	Adapter 5	ACAGTG	2.57	148	28.53871771
PHYA1 RNAi 3 copies	5DPA Fiber	Adapter 6	GCCAAT	6.13	149	67.61489292
Coker 312-plant 1	10DPA Fiber	Adapter 11	GGCTAC	2.87	149	31.65656488
Coker 312-plant 2	10DPA Fiber	Adapter 10	TAGCTT	7.51	148	83.39524125
Coker 312-plant 3	10DPA Fiber	Adapter 12	CTTGTA	1.89	147	21.13013817
PHYA1 RNAi-plant 1	10DPA Fiber	Adapter 7	CAGATC	2.38	150	26.07707639
PHYA1 RNAi-plant 2	10DPA Fiber	Adapter 8	ACTTGA	2.52	147	28.17351756
PHYA1 RNAi-plant 3	10DPA Fiber	Adapter 9	GATCAG	8.60	148	95.49921102
Coker 312	15DPA Fiber	Adapter 13	AGTCAA	3.22	147	35.99949467
Coker 312	15DPA Fiber	Adapter 14	AGTTCC	5.51	150	60.37171886
Coker 312	15DPA Fiber	Adapter 15	ATGTCA	7.29	151	79.34667968
PHYA1 RNAi 3 copies	15DPA Fiber	Adapter 16	CCGTCC	3.21	148	35.64563574
PHYA1 RNAi 3 copies	15DPA Fiber	Adapter 17	GTAGAG	4.09	150	44.81312707
PHYA1 RNAi 3 copies	15DPA Fiber	Adapter 18	GTCCGC	2.51	148	27.87244415

For each sample, three biological replications were used for the experiments. The concentration and quality of these miRNA libraries were determined and analyzed by Qubit and Bioanalyzer (Figures 4.10, 4.11 and 4.12 and Table 4.4).

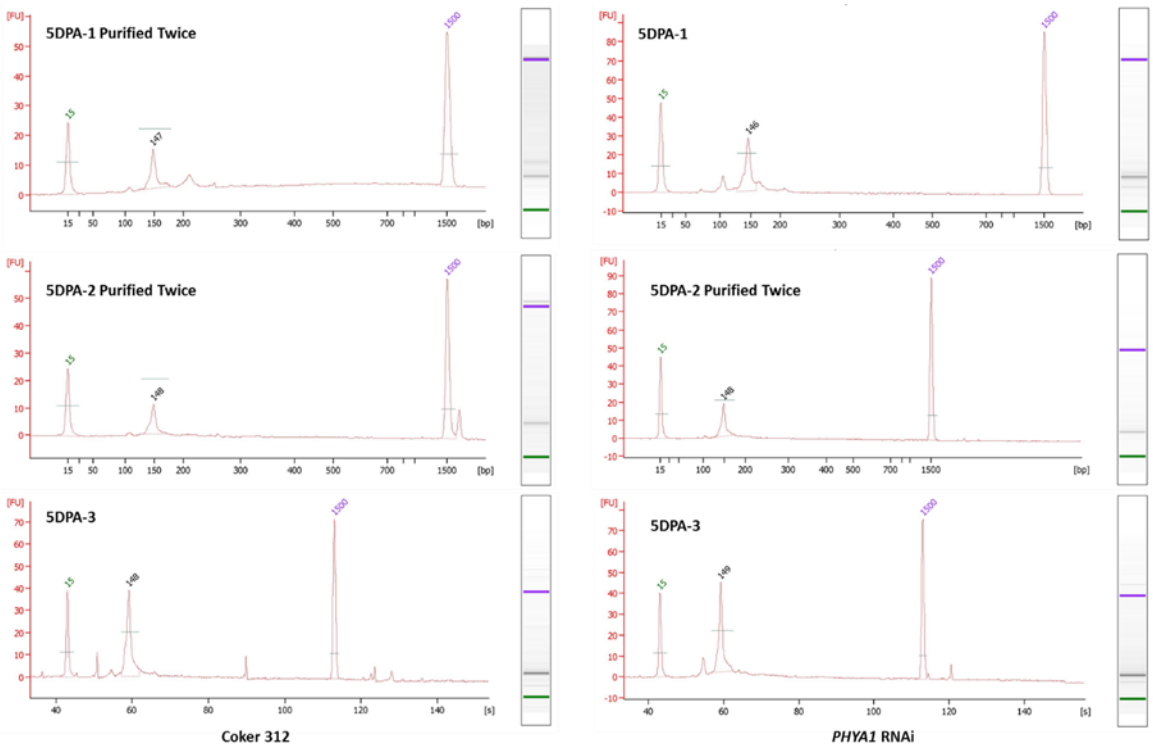

Figure 4.10 Electropherogram traces of bead size selected purified miRNA libraries from 5-DPA fibers.

The numbers 15 and 1500 represent the size (nt) of the lower and upper markers, respectively. The peak in the middle was the desired miRNA bands and their estimated sizes (146-149 nt) are labelled on the top. The Coker 312 5DPA-1 and 5DPA-2 and the RNAi 5DPA-2 library samples were purified twice with the beads.

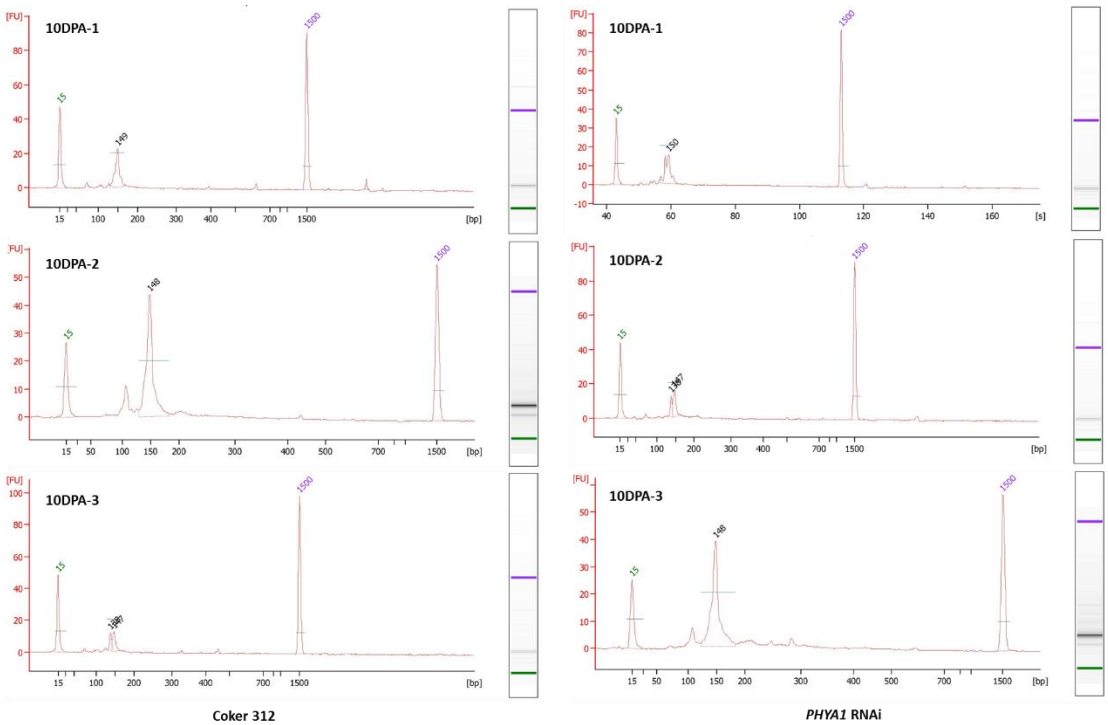

Figure 4.11 Electropherogram traces of bead size selected purified miRNA libraries from 10-DPA fibers.

The numbers 15 and 1500 represent the size (nt) of the lower and upper markers, respectively. The peaks in the middle are desired miRNA bands and their estimated sizes (147-150 nt) are labelled on the top.

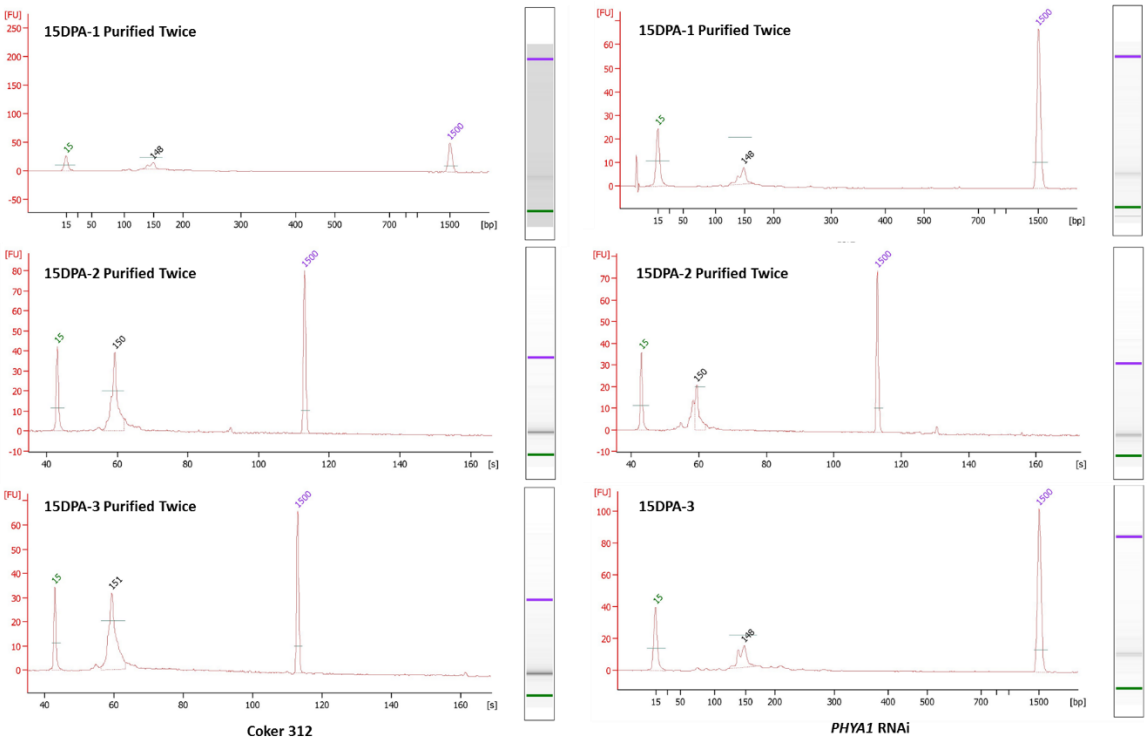

Figure 4.12 Electropherogram traces of bead size selected purified miRNA libraries from 15-DPA fibers.

The numbers 15 and 1500 represent the size of the lower and upper markers, respectively. The peaks in the middle are desired miRNA bands and their estimated sizes (148-151 nt) are labelled on the top. The Coker 312 15DPA-1, 15DPA-2 and 15DPA-3 and RNAi 15DPA-1 and 15DPA-2 samples were purified twice with the beads.

Size selection is one of the major steps during the miRNA library construction, and two different methods for size selection were performed in this study. Size selection using polyacrylamide gel electrophoresis was preferred at the very beginning with the goal to get better purity. Based on the bioanalyzer results, this method worked perfectly for 10-DPA samples (Figure 4.13), of which miRNAs were the most abundant small RNAs (Figure 4.13 A).

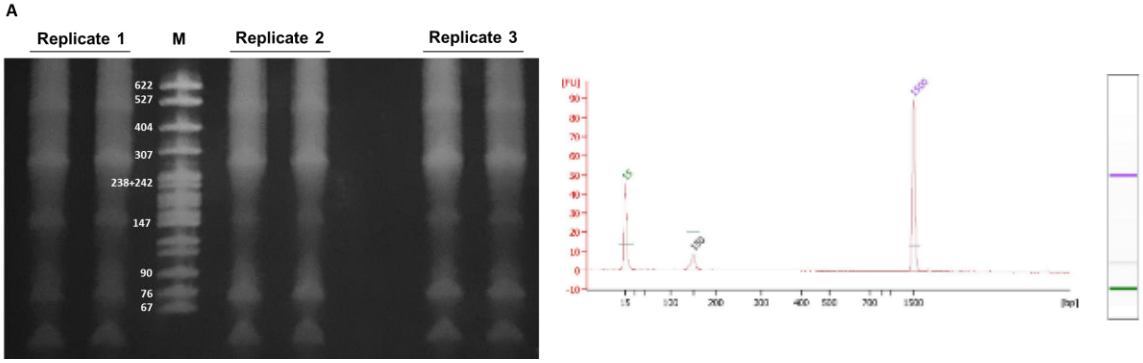

Figure 4.13 Size selection of 10-DPA fiber small RNA libraries by electrophoresis on a 6% polyacrylamide gel (A) and quality control check of the libraries by Agilent 2100 Bioanalyzer (B).

A) The 147 bp band corresponding to miRNAs (21 nt) was selected for further sequencing. The size marker (M) was *Msp*I digested pBR322, and the sizes of individual DNA bands were labeled; B) The size of miRNA library recovered from the gel was further analyzed by Bioanalyzer and confirmed to be 150 nt based on electropherogram traces. The numbers 15 and 1500 represent the size (bp) of the lower and upper markers, respectively. The peak in the middle was the desired miRNA band and its estimated size is labelled on the top. The miRNAs were extracted from 10-DPA fiber.

However, for 5-DPA and 15-DPA samples, miRNAs were not the most abundant RNA samples. The 15-DPA samples analyzed by polyacrylamide gel electrophoresis were shown in Figure 4.14, and 5-DPA samples have the same gel pattern as 15-DPA samples (Data not shown).

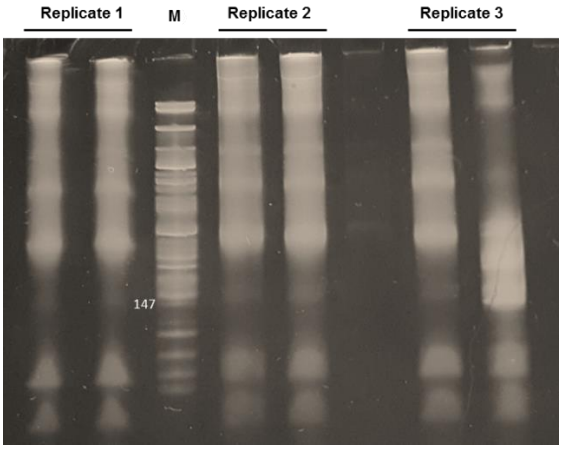

Figure 4.14 Size selection of miRNA libraries of 15 DPA fiber by electrophoresis on a 6% polyacrylamide gel.

The 15-DPA small RNA samples with three biological replicates (replicates 1-3) were used for library construction (each replicate sample was loaded onto two wells). The M lane contained DNA size marker of *Msp*I digested pBR322, and the 147 bp DNA fragment was marked.

Because enriching miRNAs by the gel method for 5-DPA and 15-DPA samples resulted in low recoveries, a second method using AMPure XP beads for size selection was thus utilized. Small fragments such as adaptor dimer (127 bp) or excess primers (70-80 bp) could not be effectively removed by XP beads, and therefore three modifications were incorporated when using XP beads for size selection: 1) Construction of miRNA libraries was done within one week in order to prevent old or thawed adaptors from dimerization and degradation. More importantly, before ligation of miRNAs to the 5′ SR

adaptor, hybridization to a reverse transcription SR RT primer was performed to remove the excess of the 3´ SR adaptor and convert the single stranded DNA adaptor into a double stranded DNA molecule. 2) The libraries were purified with beads twice to remove large size DNA contaminants (Figure 4.15). 3) Low-quality reads and junk data were removed during data analysis. In the end, 18 high quality miRNA libraries with a unique barcode sequence incorporated to each library were pooled together in one lane and loaded to the Illumina next-generation sequencing machine. The barcodes were used to sort reads post-run (Table 4.4).

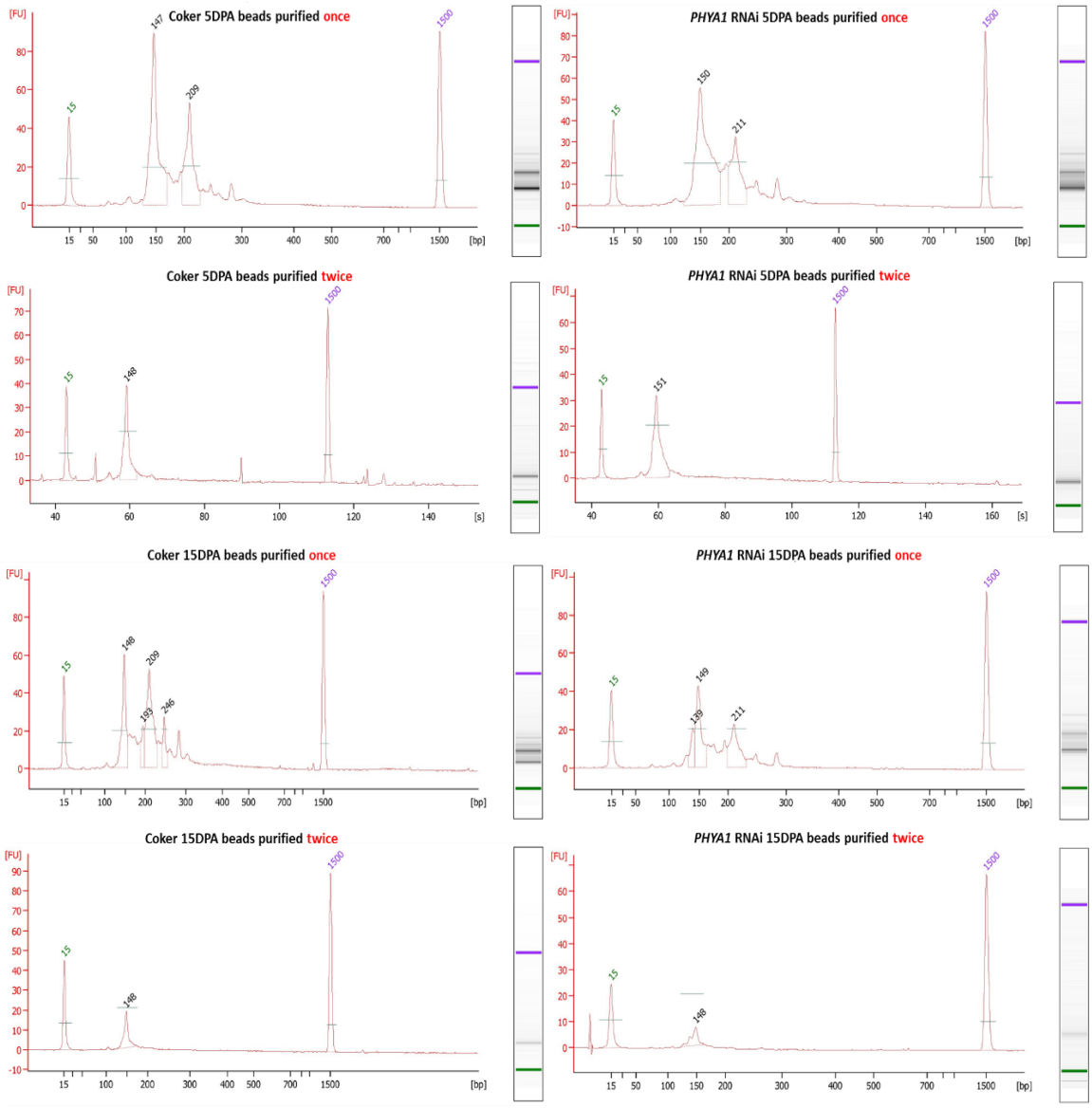

Figure 4.15 Comparison of electropherogram traces of miRNA libraries prepared through size selection once or twice using XP beads.

DNA contaminants with large sizes were successfully removed by the second size selection. The numbers 15 and 1500 represent the size (nt) of the lower and upper markers, respectively, The sizes of miRNA and other DNA samples migrated between the two markers are labeled on the top.

High-throughput sequencing of 5 DPA-, 10 DPA- and 15 DPA-fiber small RNA libraries of Coker 312 and *PHYA1* RNAi plants

A total of 18 high quality small RNA libraries with three biological replicates were subject to next generation sequencing on HiSeq 2500. Sequence data analysis indicated that each library had a minimum 8,052,792 and a maximum 12,947,430 short reads extracted from raw data (Table 4.5). Adaptor sequences and low-quality reads composed of 39.54% ~ 65.63% of raw reads were removed, and the short reads average length after trim was from 22.5 to 24.2 nt (Table 4.5).

Table 4.5 Trim reads statistics from raw data image file

Sample ID	Raw reads	Trimmed reads	Trimmed percentage (%)	Average length (nt) after trim
CF5-1	8,052,792	3,867,782	48.03	22.8
CF5-2	11,192,002	7,345,455	65.63	22.7
CF5-3	10,445,749	6,503,668	62.26	22.6
CF10-1	10,496,469	6,475,658	61.69	22.9
CF10-2	10,807,969	7,164,085	46.45	22.6
CF10-3	10,343,483	4,804,740	46.45	22.7
CF15-1	9,721,703	4,083,368	42	22.5
CF15-2	11,898,328	5,846,551	49.14	22.7
CF15-3	12,947,430	6,272,361	48.44	23.0
RF5-1	8,684,895	5,396,278	62.13	23.3
RF5-2	12,917,341	7,671,999	59.39	23.4
RF5-3	12,783,126	8,366,374	65.45	23.3
RF10-1	10,924,895	5,315,759	48.66	23.7
RF10-2	10,578,419	5,777,184	62.69	24.2
RF10-3	10,347,993	6,487,061	62.69	24.2
RF15-1	10,939,628	5,511,637	50.38	23.4
RF15-2	11,395,084	5,859,236	51.42	23.2
RF15-3	10,194,345	4,040,911	39.54	23.7

Note: RF and CF denote fibers from RNAi and Coker 312 lines, respectively. The numbers 5, 10, 15 represent the days after anthesis. Each of CF and RF samples had 3 biological replicates (1-3).

Within the trimmed reads, 8.9%, 8.9%, 9.6%, 10.3%, 10.7% and 10.4% were annotated for 5-DPA (CF5), 10-DPA (CF10) and 15-DPA (CF15) fiber from Coker 312 and 5-DPA (RF5), 10-DPA (RF10) and 15-DPA (RF15) fiber from the *PHYA1* RNAi line, respectively (Table 4.6). Among all the reads, 6,566,588, 6,588,884, 5,025,362, 8,286,014, 6,426,397, 4,741,820 unique reads were annotated as small RNAs, respectively (Table 4.6). The majority of these reads were clustered into rRNA, tRNA, snRNA, snoRNA and scRNA and removed before miRNA analysis. Among the rest of the annotated reads, 7,004, 7,291, 5,704, 8,698, 6,942, 5,559 reads, respectively, were matched to miRBase database (Table 4.6).

Table 4.6 Distribution of mapped sequence reads

Category	CF5	CF10	CF15	RF5	RF10	RF15
Total reads	17,716,905	18,444,483	16,202,280	21,434,651	17,580,004	15,411,784
Annotated	1,583,027(8.9%)	1,645,910(8.9%)	1,554,366(9.6%)	2,214,010(10.3%)	1,883,895(10.7%)	1,607,358(10.4%)
Total Identified Small RNAs	6,566,588	6,588,884	5,025,362	8,286,014	6,426,397	4,741,820
Annotated Small RNAs	373,566(5.7%)	354,203(5.4%)	308,890(6.1%)	506,395(6.1%)	390,173(6.1%)	298,707(6.3%)
Small RNAs Match miRBase	7,004	7,291	5,704	8,698	6,942	5,559
Small RNAs Match *Gossypium hirsutum*	366,562	346,912	303,186	497,697	383,224	293,148

Note: RF and CF denote fibers from RNAi and Coker 312 lines, respectively. The numbers 5, 10, 15 represent the days post anthesis.

The size distribution of small RNAs ranged from 15 to 30 nt, and the most abundant length of them was 24 nt in both Coker 312 and RNAi lines (Figure 4.16).

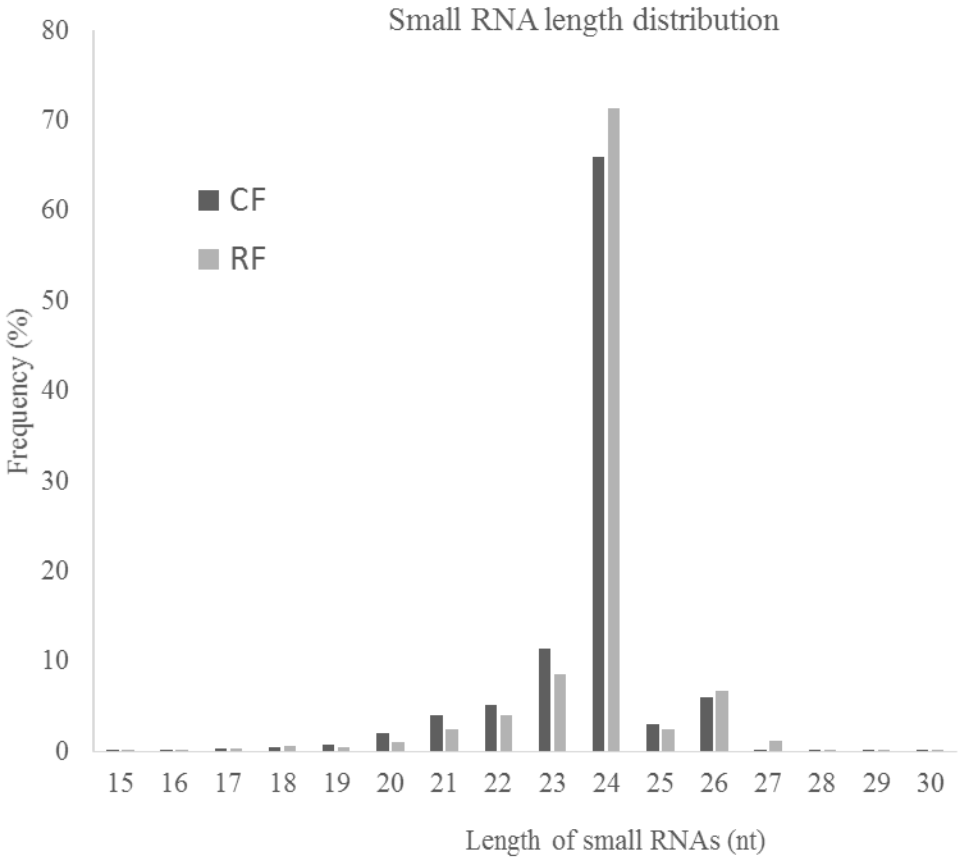

Figure 4.16 The small RNA length distribution in cotton fibers from both *PHYA1* RNAi line (RF) and Coker 312 (CF).

Identification of differentially expressed conserved miRNAs in *PHYA1* RNAi cotton fibers.

Known and conserved miRNAs were identified by mapping RNA sequence to the database of miRBase (release 21), and 77 known miRNAs belonging to 61 miRNA families were found in the two cotton lines. Of these, 34 miRNA families were cotton-specific, and 7 miRNAs were differentially expressed (≥ 2-fold change) in RNAi cotton. For majority of the miRNA families, only one member was identified. However, multiple members were identified in some miRNA families, which include miR7495, 390, 2949, 167, 7486, 394, 7484, 396, 399, 7492, 156, 827, and 7496 (Figure 4.17).

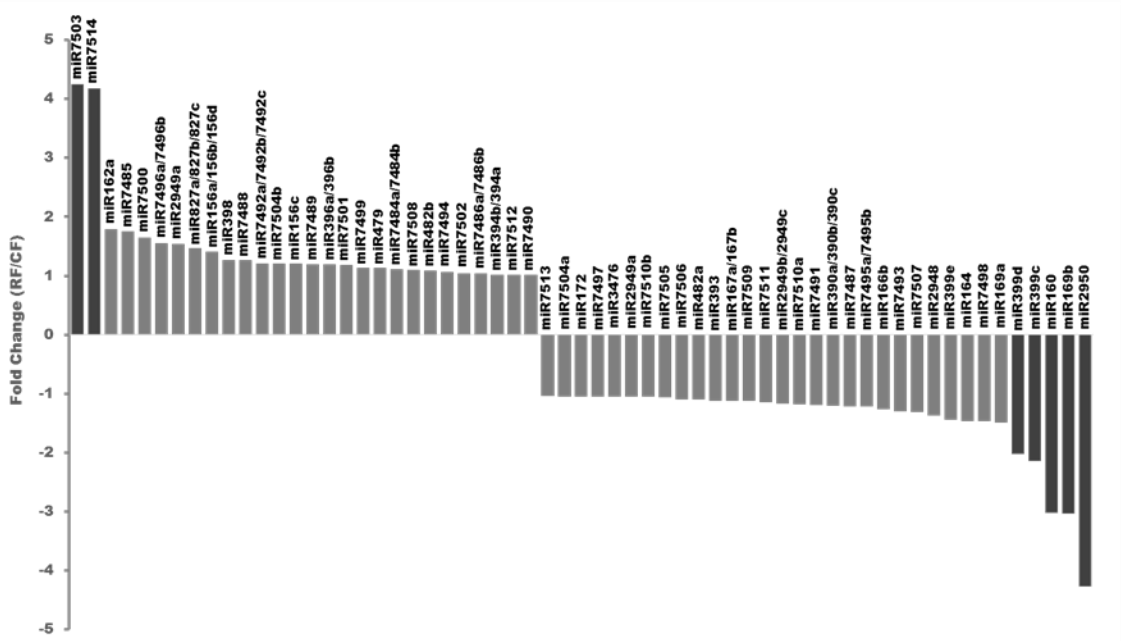

Figure 4.17 Differentially expressed known miRNA families in *PHYA1* RNAi cotton (RF) relative to the non-transformed Coker 312 (CF) in fibers.

The miRNAs with a fold change greater than 2 are displayed by darker columns.

Among 7 differentially expressed miRNAs, 5 miRNAs were down-regulated in *PHYA1* RNAi compared to Coker 312, and the other 2 miRNAs were up-regulated in the RNAi line (Figure 4.17). These results implied that miRNA-mediated gene regulation might confer to the phenotype of the RNAi line with improved fiber quality.

Identification of novel miRNAs in PHYA1 RNAi cotton fibers

In addition to successful identification of 77 conserved miRNAs in *PHYA1* RNAi line, novel miRNAs in fibers were also identified in the two cotton lines. After removing known miRNAs and other small RNAs, the stem-loop secondary structures of novel pre-miRNAs were predicted by MFOLD. The predicted hairpin structures of four novel pre-miRNAs are shown in Figure 4.18

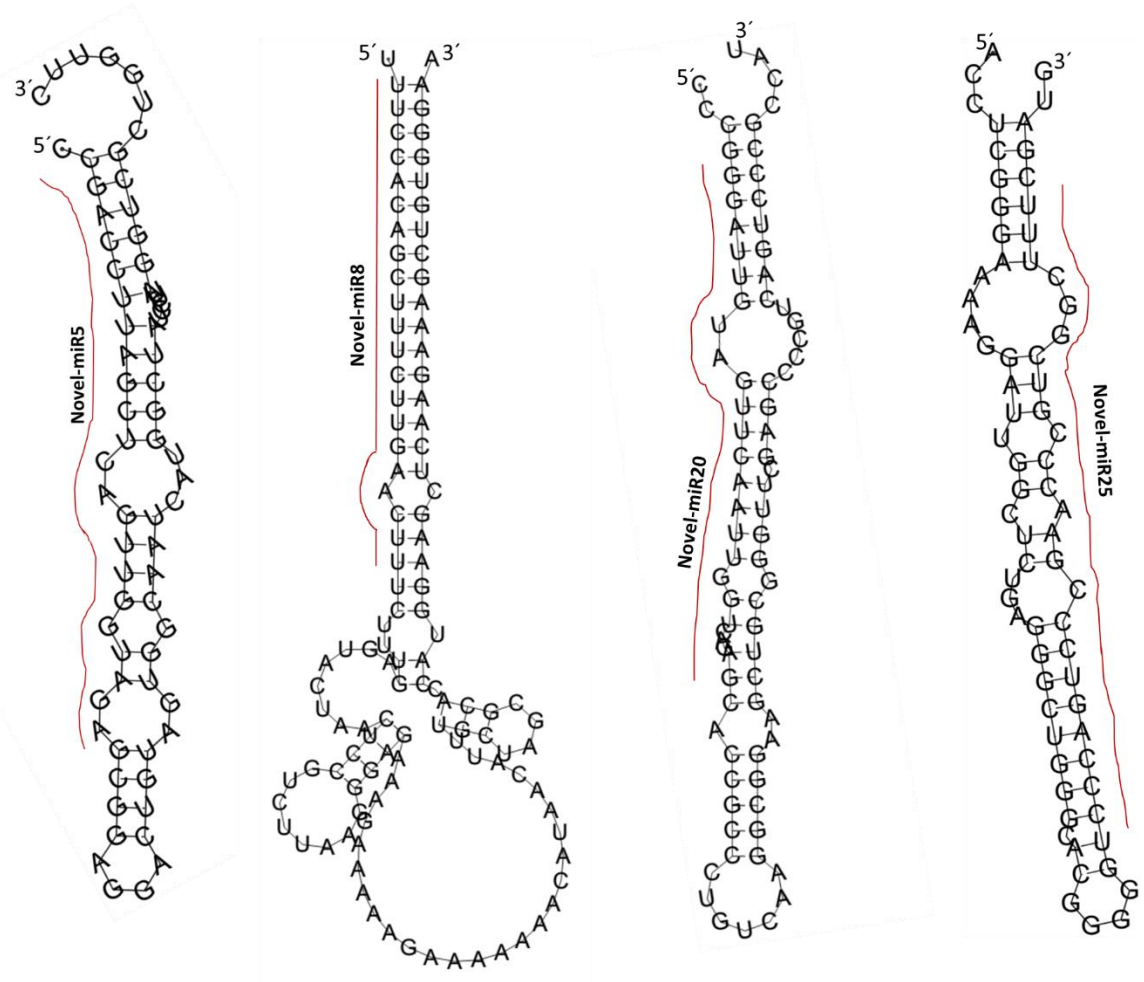

Figure 4.18 The predicted hairpin structures of four Novel pre-miRNAs.

Mature miRNAs are marked by red lines.

The criterion to test the stability of a RNA secondary structure is the minimal folding free energy (MFE). The lower the MFE, the more stable the RNA structure. The only criteria to annotate a novel miRNA is that it was generated by precise excision from the stem of a stem-loop precursor [158]. A total of 35 novel miRNAs were identified l from the 18 fiber libraries (named as Novel-miR1 to 35) (Table 4.7). The sizes of these novel miRNAs ranged from 21 to 24 nt. Twenty-seven novel miRNAs containing 24 nt were the most abundant in these newly discovered miRNAs. Five, two and one novel miRNAs were found to have the sizes of 21, 22 and 23 nt, respectively. The stem-loop precursors of these novel miRNAs predicted by MFOLD have negative free folding energies from -45.7 to -230.1 kcal mol^{-1}, which are lower than those of tRNAs (-27.5 kcal mol^{-1}) or rRNA (-33 kcal mol^{-1}) (Table 4.7).

Table 4.7 Novel miRNAs identified in all fiber libraries

Name	Sequence	Length (nt)	MFE (kcal mol^{-1})	Accession number (NBI)	Annotation	Inhibition	Cleavage/Translation inhibition site (nt)
Novel-miR1	CCGGAGACGTCGGCGGGGGCCTCG	24	-215.5	Gh_Sca142710G01	RRNA promoter binding protein	Cleavage	497
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	945
				Gh_Sca014836G01	atp synthase subunit beta	Cleavage	408
				Gh_D11G1394	chaperonin cpn60- mitochondrial	Cleavage	993
				Gh_A05G2828	senescence-associated protein	Cleavage	293
Novel-miR2	CCGACTGTTTAATTAAAACAAAGT	24	-158.2	Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	916
				Gh_Sca009741G01	Cytochrome P450 like_TBP	Cleavage	1223
				Gh_D08G0858	Cytochrome P450 like_TBP	Cleavage	530
				Gh_Sca006071G01	elongation factor 1-alpha	Cleavage	1198
Novel-miR3	CGGCAAAATAGCTCGACGCCAGGA	24	-72.3	Gh_D04G2001	ORF107c	Cleavage	429
Novel-miR4	GGCTCAGCCGGAGGTAGGGTCCAG	24	-230.1	Gh_A05G2828	senescence-associated protein	Cleavage	214
Novel-miR5	CCGACCTTAGCTCAGTTGGTAGA	23	-218.3	Gh_A02G0253	PE-PGRS FAMILY PROTEIN	Cleavage	240
Novel-miR6	AAGAATTTGGGCTTTTGTGACTCG	24	-190.7	Gh_D04G0695	N/A	Cleavage	362
				Gh_A01G0034	N/A	Cleavage	665
Novel-miR7	CCAAGATCAATAGACAGGCGTG	22	-203.2	Gh_D06G0864	alpha-aminoadipic semialdehyde synthase	Cleavage	308
				Gh_D12G1259	BZIP transcription factor-like protein	Translation	146
Novel-miR8	TTCCACAGCTTTCTTGAACTT	21	-95.1	Gh_D07G0477	chaperone protein chloroplastic	Cleavage	18
				Gh_D02G2222	calcium-binding ef-hand family protein	Cleavage	346
				Gh_D08G0162	serine threonine-protein kinase nek6	Cleavage	582
				Gh_D05G3256	bacterial-induced peroxidase precursor	Cleavage	1166
Novel-miR9	TGGATGGGCGATTGGGTGCGGTGC	24	-192.5	Gh_D07G1749	14-3-3-like protein	Cleavage	230

Table 4.7 (Continued)

Novel-miR10	ATTATCCGATGAGGCACTGGGTGC	24	-141.3	Gh_A08G2090	eukaryotic translation initiation factor isoform 4g-1-like isoform x1	Cleavage	677
				Gh_A10G1238	acetolactate synthase chloroplastic-like	Cleavage	591
				Gh_D11G2202	TINY-like protein	Cleavage	493
Novel-miR11	ATTTGTGGATCTATTTGACAGT	22	-174.9	Gh_A04G0528	gdsl esterase lipase at5g03610-like	Cleavage	353
				Gh_A04G0245	cbs domain-containing protein cbsx5-like	Cleavage	823
				Gh_Sca027989G01	Chloroplast 30S ribosomal protein S16	Cleavage	356
				Gh_A06G2111	endo- -beta-glucanase	Cleavage	903
				Gh_D08G2188	gtp-binding protein sar1a	Translation	1076
				Gh_D01G0099	global transcription factor group isoform 1	Translation	142
Novel-miR12	GTTCGGGAGTCGGTTACGCGCGAG	24	-203.3	Gh_A02G0261	Oleosin isoform	Cleavage	501
				Gh_D11G3242	Emb|CAB10291.1	Cleavage	124
Novel-miR13	TTTGTACTTTAGATGTCTCTC	21	-62.3	Gh_D11G1863	Membrane protein-like protein	Cleavage	235
				Gh_D12G1850	Receptor-like protein kinase	Translation	239
				Gh_D02G1964	3-hydroxy-3-methylglutaryl-coenzyme A reductase	Cleavage	294
				Gh_A05G3942	40s ribosomal protein s3-3	Cleavage	37
				Gh_D10G2028	Mitochondrial carrier protein, expressed	Cleavage	454
Novel-miR14	GACGGACTGGGAACGGCTCCC	21	-164.3	N/A	N/A	N/A	N/A
Novel-miR15	AATCGGTTGGCATGTGGCACAATT	24	-224	Gh_D05G0643	probable xyloglucan glycosyltransferase 5	Cleavage	1336
Novel-miR16	AAGGATTGGCTCTGAGGGCTGGGT	24	-157.7	Gh_A05G2834	Cytochrome P450 like_TBP	Cleavage	893
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	661
Novel-miR17	GCGGACCAAGATCGAGTTGCCAAG	24	-220.5	Gh_D11G0625	N/A	Cleavage	107
				Gh_D05G2413	FRO1 and FRO2-like protein	Translation	227
Novel-miR18	TGAAGCTGCCAGCATGATCTC	21	-68.8	Gh_A13G0162	lim domain-containing protein wlim2b	Cleavage	513
				Gh_A06G0984	tubulin alpha-3 chain	Cleavage	89
				Gh_A02G1655	protein mid1-complementing activity 1 isoform x1	Translation	124

Table 4.7 (Continued)

Novel-miR19	TCTGAGGGCTGGGCACGGGGGTCT	24	-148	Gh_A05G2834	Cytochrome P450 like_TBP	Cleavage	883
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	651
Novel-miR20	GGATTGTAGTTCAATTGGTCAGAG	24	-86.5	Gh_A09G1967	Bet1-like SNARE 1-2	Cleavage	177
					Predicted protein	Cleavage	610
Novel-miR21	ATTTGACGGTACCAATAATTTTGG	24	-219.2	Gh_A10G1317	N/A	Cleavage	374
				Gh_A06G2074	Octicosapeptide/Phox/Bem1p	Cleavage	769
Novel-miR22	TGTGTCAAATCGGCGGCTACATCT	24	-76.3	N/A	N/A	N/A	N/A
Novel-miR23	AATCGTCTCAATCGGACAACCGAG	24	-115.2	Gh_D03G1527	N/A	Cleavage	1938
Novel-miR24	CGGCTGCCGGCGCACTGCTCGAGT	24	-146.4	N/A	N/A	N/A	N/A
Novel-miR25	CCCAGTCCCGAACCCGTCGGCTTT	24	-149.6	Gh_A03G1313	N/A	Cleavage	738
				Gh_A05G0906	receptor-like serine threonine-protein kinase nerk isoform x3	Cleavage	457
Novel-miR26	TAGTCCGACTTTGTGAAATGACTT	24	-45.7	Gh_Sca009741G01	Cytochrome P450 like_TBP	Cleavage	343
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	346
				Gh_Sca006071G01	elongation factor 1-alpha	Cleavage	526
				Gh_A05G4003	Cytochrome P450 like_TBP	Cleavage	1026
Novel-miR27	TAAACGGCGGGAGTAACTATGACT	24	-107.1	Gh_Sca009741G01	Cytochrome P450 like_TBP	Cleavage	525
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	528
				Gh_A05G2834	Cytochrome P450 like_TBP	Cleavage	588
				Gh_D08G0858	Cytochrome P450 like_TBP	Cleavage	736
				Gh_Sca006071G01	elongation factor 1-alpha	Cleavage	1086
				Gh_A05G4003	Cytochrome P450 like_TBP	Cleavage	1208
Novel-miR28	CAGGTCTCCAAGGTGAACAGCCTC	24	-151.1	Gh_Sca142710G01	RRNA promoter binding protein	Cleavage	296
				Gh_A05G2834	Cytochrome P450 like_TBP	Cleavage	976

Table 4.7 (Continued)

				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	744
				Gh_Sca014836G01	atp synthase subunit beta	Cleavage	207
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	1193
Novel-miR29	GTTGGTCGATTAAGACAGCAGGAC	24	-89.8	Gh_Sca014836G01	atp synthase subunit beta	Cleavage	287
				Gh_D11G1394	chaperonin cpn60- mitochondrial	Cleavage	1426
Novel-miR30	CACGGAATCGAGAGCTCCAAGTGG	24	-93.1	Gh_Sca014836G01	atp synthase subunit beta	Cleavage	638
				Gh_D11G1394	chaperonin cpn60- mitochondrial	Cleavage	1543
Novel-miR31	ATGTAGGCAAGGGAAGTCGGC	21	-156.3	Gh_A05G2828	senescence-associated protein	Cleavage	23
				Gh_A05G2834	Cytochrome P450 like_TBP	Cleavage	941
				Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	709
				Gh_Sca014836G01	atp synthase subunit beta	Cleavage	172
Novel-miR32	AACAGTCGACTCAGAACTGGTACG	24	-151.4	Gh_D08G0862	Cytochrome P450 like_TBP	Cleavage	676
				Gh_A05G2834	Cytochrome P450 like_TBP	Cleavage	735
Novel-miR33	ATTTCCGTAAGACATTTTCCGTGC	24	-133.4	Gh_D11G2331	Os01g0692600 protein	Cleavage	1344
				Gh_A08G1457	boi-related e3 ubiquitin-protein ligase 1-like	Cleavage	375
				Gh_D07G1808	Zinc finger, N-recognin; WD40-like	Cleavage	148
				Gh_A01G2124	60s ribosomal protein l17-2	Cleavage	162
				Gh_A07G1464	protein sulfur deficiency-induced 1-like	Cleavage	636
				Gh_A01G0147	random slug protein 5	Translation	549
				Gh_D06G1877	serine arginine repetitive matrix protein 2	Cleavage	880
				Gh_A03G1542	2-oxoglutarate/malate translocator	Translation	1181
Novel-miR34	GTGTGACTCAAATTCTAAGAGATT	24	-188.1	Gh_D02G0923	Os09g0462400 protein	Translation	181
				Gh_D02G1201	Far-red impaired response protein	Cleavage	486
				Gh_D08G0976	Arsenical pump membrane protein	Cleavage	345

Table 4.7 (Continued)

				Gh_D02G0487	Protein FBL4	Cleavage	29
Novel-miR35	AGTGGAACCTCTCATTTAAGACGT	24	-210.5	Gh_A09G2497	Gag-pol	Cleavage	42
				Gh_A01G1980	At3g51780/ORF3	Translation	651
				Gh_D10G0595	Urophorphyrin III methylase	Cleavage	9

All miRNAs* of these novel miRNAs were identified in RNA sequencing. Among these novel miRNAs, none of them were significantly differentially expressed in *PHYA1* RNAi fibers (Figure 4.19).

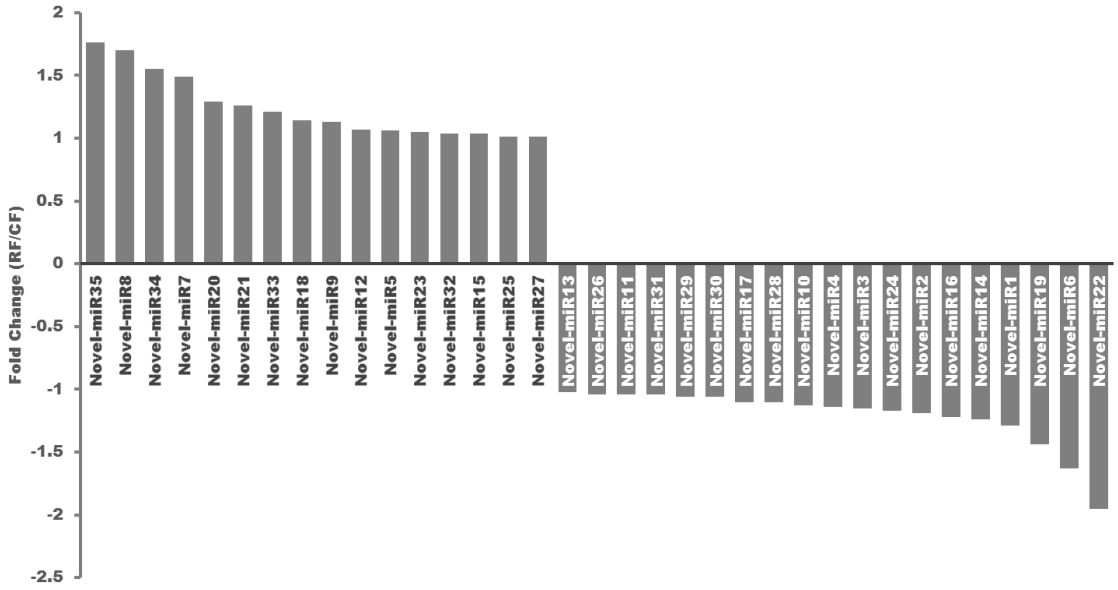

Figure 4.19 The expression pattern of novel miRNAs in *PHYA1* RNAi cotton (RF) relative to the non-transformed Coker 312 (CF) in fibers.

The 35 novel miRNAs were identified for the first time in cotton fibers, and their mature miRNA sequences were listed in Table 4.7. Some of these novel miRNAs might be specifically expressed in fiber cells, which need further analysis.

Prediction and annotation of targets for novel miRNAs

The plant miRNA has perfect or near perfect complementarity to its target, and the cleavage occurs between the 10^{th} and 11^{th} nucleotides from the 5′ end of the miRNA. These characters were used to identify the targets of novel miRNAs. By using psRNATarget, a total of 97 targets were predicted for 35 novel miRNAs. Among them,

the targets of three novel miRNAs Novel-miR14, 22, and 24 were not identified. It might be due to the reference genome used for psRNATarget prediction not being fully sequenced, or new genes were created by genome duplication. The top rated target genes for each novel miRNAs were listed in Table 4.7. The way of miRNA silencing its target through mRNA cleavage or translation inhibition and the predicted cleavage/translation inhibition sites were also presented in Table 4.7. Among these targets, the gene encoding cytochrome P450 like_TATA box binding protein (cytochrome P450 TBP) has the highest hits, being the targets of 9 different novel miRNAs, including novel-mir1, 2, 16, 19, 26, 27, 28, 31 and 32. Plant cytochrome P450s are involved in a wide range of biosynthetic reactions, including fatty acid conjugation, hormones synthesis, and generating defensive compounds. Of these 9 novel miRNAs, only Novel-miR27 was accumulated slightly higher in *PHYA1* RNAi line, and the other 8 miRNAs Novel-miR1, 2, 16, 19, 26, 28, 31, 32 were all slightly lower in RNAi cotton compared to Coker 312. These novel miRNAs were also predicted to silence RRNA promoter binding protein, membrane and ribosomal proteins. The predicted targets of these novel miRNAs also included some transcription factors and related proteins, such as a bZIP transcription factor, the global transcription factor group isoform 1 and a WD40 like transcription factor. In addition, many proteins related to translation, such as translation initiation factor and elongation factor 1α were predicted to be the targets of Novel-miR2, 10, 26 and 27. The predicted targets of these 35 novel miRNAs included many enzymes such as ATP synthase, serine threonine-protein kinase, peroxidase, acetolactate synthase, lipase, endo-β-glucanase, receptor-like protein kinase, CoA reductase, xyloglucan glycosyl-transferase and methylase, and some functional proteins, like chaperons, senescence-

associated proteins, Ca^{2+} binding proteins and 14-3-3-like protein (Table 4.7). The identification of miRNA targets provided further evidence for the presence and expression of these novel miRNAs in cotton fibers.

Validation of differentially expressed miRNAs in *PHYA1* RNAi line by qRT-PCR

Twelve known miRNAs and seven novel miRNAs expressed in 10-DPA fibers of *PHYA1* RNAi and Coker 312 lines were validated by qRT-PCR. The abundances of these miRNAs were relatively high in majority of the fiber libraries (Figures 4.17 and 4.19). The results showed that all of these 19 miRNAs were successfully detected by qRT-PCR and showed consistence with the expression profiles analyzed by small RNA sequencing (Figure 4.20). Among the known miRNAs, miR162a, 396a/b, 2950 and 160 were up-regulated, and miR172, 399, 167a/b, 390a/b/c, 164, 166b, 399c, and 169b were down-regulated (Figure 4.20).

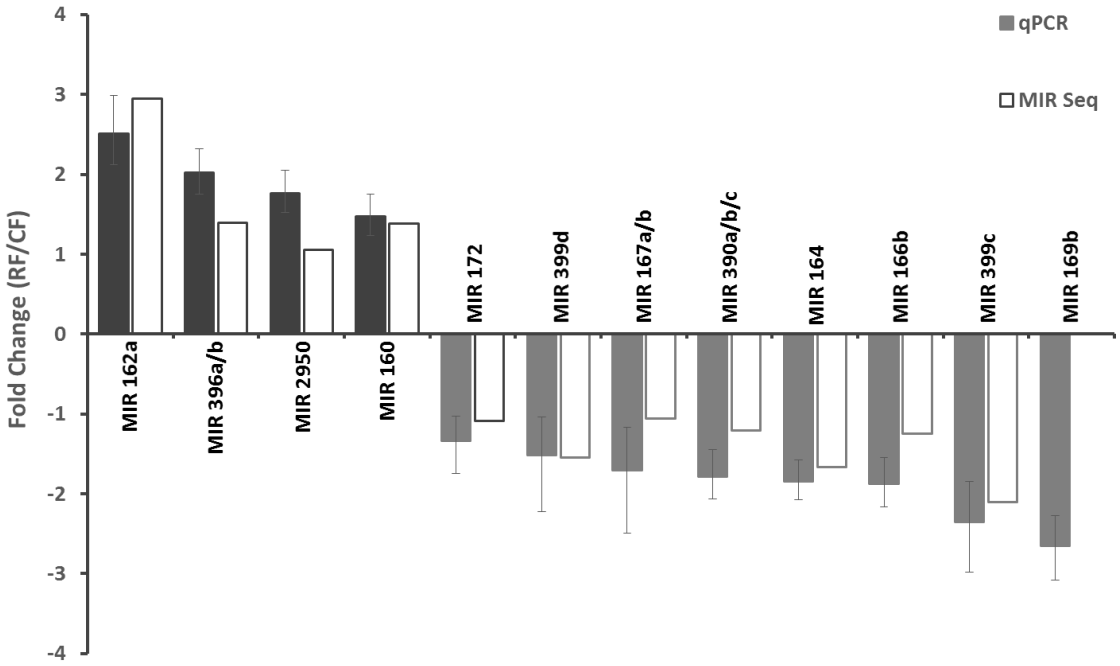

Figure 4.20 Validation of the expression patterns of differentially expressed known miRNAs in 10-DPA fiber of *PHYA1* RNAi line by qRT-PCR.

The miRNA expression was represented as relative fold change $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_{T\ RNAi} - \Delta C_{T\ Coker\ 312}$). The expression levels of miRNA were normalized by using U6 snRNA as a reference. Three biological replicates and three technical replicates were used for qRT-PCR analysis. The error bar represent the confidence limits. RF and CF denote fibers from RNAi and Coker 312 lines, respectively. MIR 169b was not detected in 10-DPA fiber by MIR Seq.

For the novel miRNAs, Novel-miR22 and Novel-miR8 were up-regulated, and Novel-miR4, 2, 6, 3, and 5 were down-regulated (Figure 4.21).

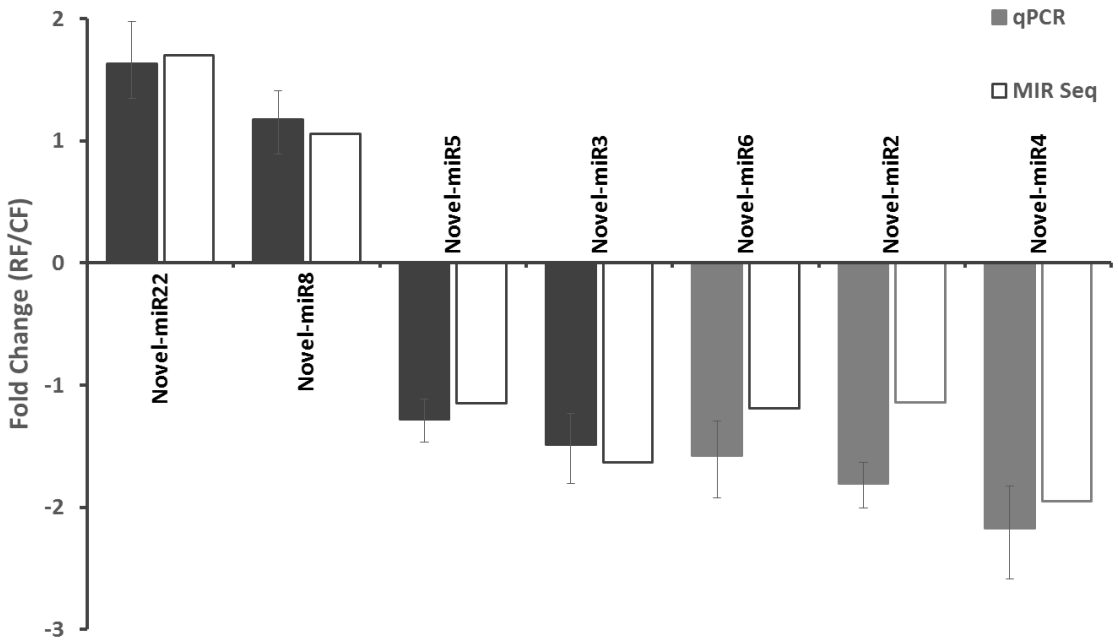

Figure 4.21 Validation of the expression patterns of differentially expressed novel miRNAs in 10-DPA fiber of *PHYA1* RNAi line by qRT-PCR.

The miRNA expression was represented as relative fold change $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_{T\ RNAi} - \Delta C_{T\ Coker\ 312}$). The expression levels of miRNA were normalized by using U6 snRNA as a reference. Three biological replicates and three technical replicates were included for qRT-PCR assays. The error bar represent the confidence limits. RF and CF denote fibers from RNAi and Coker 312 lines, respectively.

The expression patterns of miRNA2950, 169b, 160, Novel-miR22, 3, 4 and 5 in 5 DPA and 15 DPA fibers were further validated (Figures 4.22 and 4.23). Although the expression fold changes determined by qRT-PCR were different between RNAi and Coker 312 lines at different fiber developmental stages, the expression of these abundant miRNAs was detected in all three stages.

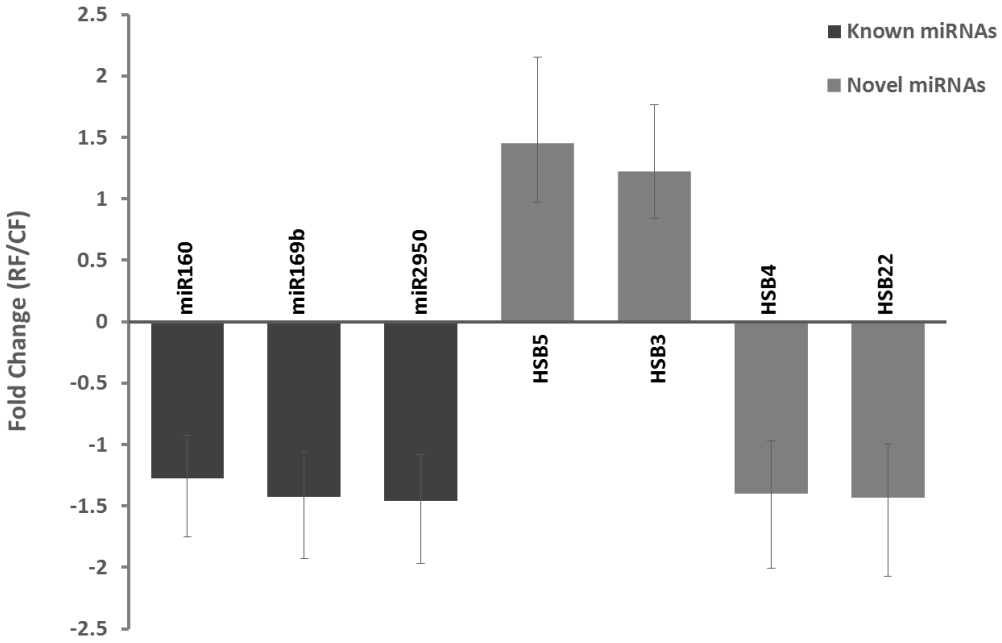

Figure 4.22 The expression levels of differentially expressed known and novel miRNAs in 5-DPA fiber of *PHYA1* RNAi line by qRT-PCR.

The miRNA expression was represented as relative fold change $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_{T\ RNAi} - \Delta C_{T\ Coker\ 312}$). The expression levels of miRNA were normalized by using U6 snRNA as a reference. Three biological replicates and three technical replicates were included for qRT-PCR assays. The error bar represent the confidence limits. RF and CF denote fibers from RNAi and Coker 312 lines, respectively.

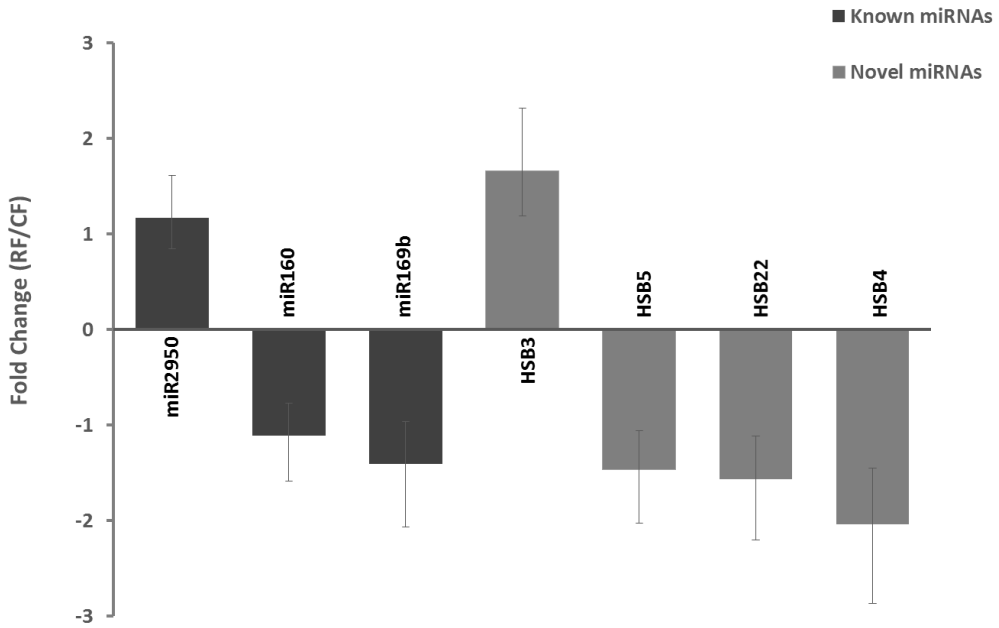

Figure 4.23 The expression levels of differentially expressed known and novel miRNAs in 15-DPA fiber of *PHYA1* RNAi line by qRT-PCR.

The miRNA expression was represented as relative fold change $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_{T\ RNAi} - \Delta C_{T\ Coker\ 312}$). The expression levels of miRNA were normalized by using U6 snRNA as a reference. Three biological replicates and three technical replicates were included for qRT-PCR assays. The error bar represent the confidence limits. RF and CF denote fibers from RNAi and Coker 312 lines, respectively.

Presence of inverse expression patterns in four pairs of miRNAs and their targets

Because miRNAs control the degradation of mRNA transcripts, the expression levels of miRNA and its target mRNA should be negatively correlated. To test this, the expression patterns of predicted targets of four novel miRNAs including Novel-miR1, 2, 4 and 8 in 10 DPA fiber were analyzed by RNA sequencing and validated by qRT-PCR (Figure 4.24 A). The results showed consistence between these two assays, and the relative expression levels of these miRNA targets were inversely correlated with the accumulation levels of corresponding miRNAs (Figures 4.24 A and B). For example, Gh_Sca142710G01 encoding a RRNA promoter binding protein was up-regulated in *PHYA1* RNAi plant (Figure 4.24 A) compared to Coker 312, and its corresponding miRNA Novel-miR1 levels was down-regulated (Figure 4.24 B). Novel-miR1 also targeted cytochrome P450_TBP (the counts were too low to count in RNA Seq), which had been shown to play crucial roles in photosynthesis and other metabolic pathways.

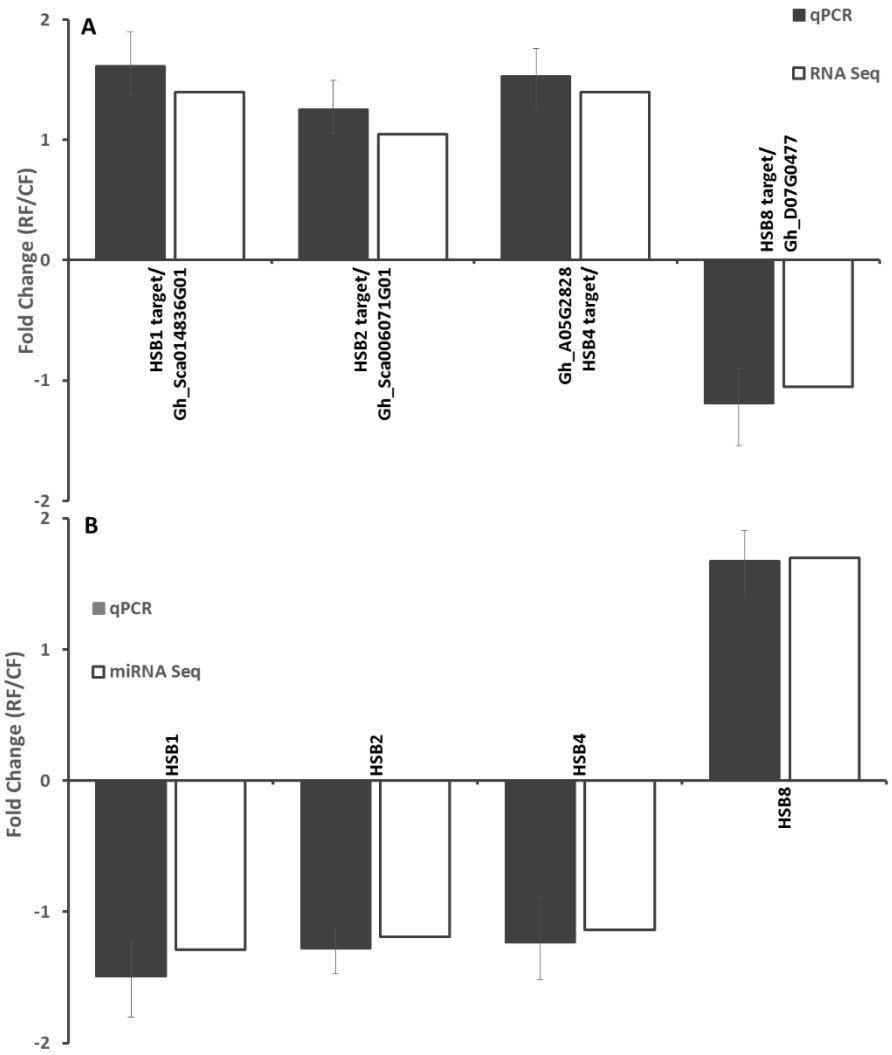

Figure 4.24 Inverse correlation between the expression of four novel miRNAs and their target genes.

A) RNA Sequencing and qRT-PCR analysis of predicted target genes of four novel miRNAs in 10 DPA fiber; B). miRNA Sequencing and qRT-PCR analysis of four novel miRNAs in 10 DPA fiber. Gene expression was represented as fold change $2^{-\Delta\Delta CT}$ ($\Delta\Delta C_T = \Delta C_{T\ RNAi} - \Delta C_{T\ Coker\ 312}$). The expression levels of the genes were normalized by *UBQ7* (Gen bank accession NO. DQ116441) and U6 snRNA as internal references for mRNAs and miRNAs, respectively. Three biological replicates and three technical replicates were included for qRT-PCR assays. The error bar represent the confidence limits.

CHAPTER V

DISCUSSION

Upland cotton (*Gossypium hirsutum* L.) produces the most commonly used textile fiber in the world. The improvement of cotton fiber quality using genetics tools has long been a key interest for cotton breeders. *PHYA1* RNAi cotton plants developed by Abdurakhmonov *et al.* had improved fiber length, strength and micronair (finer fiber) [2]. In this study, a genome-wide transcriptome and miRNAome analysis was performed to identify the differentially expressed genes for mRNAs and miRNAs, which might reveal the molecular mechanisms of the better fiber quality of RNAi cotton. In addition, determining the methylation status of CpG sites within gene body (CDS) of phytochrome genes might help to uncover potential DNA methylation effects on fiber development. In this study it was found that the methylation levels of several CpG sites within gene bodies of phytochrome genes changed in the *PHYA1* RNAi line. The RNA sequencing data showed that a total of 142 genes and 7 miRNAs were differentially expressed in *PHYA1* RNAi plants. In addition, 77 conserved miRNAs and 35 novel miRNAs were identified in this study. RNA deep sequencing has been recently used for transcription profiling in cotton functional genomics study by taking advantage of the published draft genome sequence of the allotetraploid cotton TM-1 line [164]. The whole genome sequence has allowed gene expression profiling analysis with more accuracy and less

redundancy. The TM-1 genome sequence also facilitated the identification of novel cotton-specific miRNAs.

Methylation analysis of phytochrome genes

DNA methylation has been shown to regulate of plant gene expression. DNA methylation in plants can occur in every possible context CG, CHG and CHH (H is A, T, or C), and the methylation is found in both nongenic regions (e.g. transposons) and genic regions (gene body). Nongenic methylation is usually associated with expression repression, such as transposable element silencing. In contrast, gene body methylation (CG methylation is the dominant form) generally shows a positive relationship with gene expression [134, 186, 187].

A single plant gene, present as an epigenetic allele with identical DNA sequence, often shows heritable differences in DNA methylation and gene expression. For example, an *Arabidopsis* PhyA′ epiallele mutant with high levels of CpG methylation in the *PhyA* gene body showed a phenotype similar to a *phyA* mutant, exhibiting long hypocotyls and unexpanded cotyledons [4, 5]. These observations implied that CpG-site specific methylation in the gene coding region can affect gene expression. For upland cotton, a polyploidization event occurred through hybridization of an A genome species with a D genome species, which contributed to gene duplication and the evolution of AADD-genome allotetraploid cotton [27]. It was reported that duplicated genes from the whole genome duplication had a positive relationship between the methylation of CpG gene body and gene expression in cassava [134]. Due to the limitation of sequence availability, the CpG methylation study did not cover the full coding sequences (CDS) or promoter regions of *PHY*s. Nevertheless, site-specific CpG methylation changes were still detected

in partial sequences of the *PHY* genes. In leaves, down-regulation of *PHYA1* and up-regulation of *PHYB* corresponded to slightly higher gene body methylation in *PHYA1* and lower methylation in *PHYB* gene body. The methylation levels of phytochrome genes might change if the full length of CDS were used for the studies. In addition, *PHY*s showed different CpG methylation levels between leaves and fibers, which were consistent with a previous study that various methylation patterns were found in different tissues [188].

Transcriptome profiling of 10-DPA fibers from RNAi and Coker 312 lines

The *PHYA1* RNAi cotton was reported to have a 5% increase in fiber length compared to non-transformed Coker 312 [2]. The fiber length is generated through primary wall synthesis which dominates in 10-DPA fiber. Although the components of fiber cell wall change dynamically throughout the development, the primary walls of 10-DPA cotton fibers, similar to other expanding plant cells, mainly contain cellulose, pectin, xyloglucan, lignin and hydroxyproline-rich glycoprotein [189]. Extensin, one of the glycoproteins, generates the scaffolding network of the cell wall [190]. In this study, a pollen ole e 1 allergen and extensin protein encoded by the gene Gh_D05G0805, was highly enriched in 10-DPA RNAi cotton (Table 4.3 & Figure 4.9). The ole e 1 containing proteins were first identified as major allergens and present commonly in many plants, including cotton [191]. It was expressed not only in pollen but also in other plant tissues [192-194]. An ole e 1 containing protein encoded by *AtAGP30* in *Arabidopsis* was required for root regeneration and seed germination [194]. It is reasonable to propose that extensin plays essential roles in primary cell wall architecture of cotton fiber just as in other plant cells.

Two genes Gh_08G1031 and Gh_08G1309 encoded sucrose synthase 1 (Sus 1) were down-regulated in RNAi line compared to Coker 312. It was reported that the expression level of *Sus 1* was low in 10 DPA fiber [195], which was perfectly consistent with the finding in this study. Although *GhSus3* (U73588) has been demonstrated to play an important role in ovule development and fiber cell initiation, the functional characterization of *Sus* genes however was limited [195]. The development of fiber is a complex process, and there are overlapping among each developmental stages. The transition from primary to secondary cell wall synthesis might occur in 10-DPA fibers. Sucrose synthase was reported to participate in secondary cell wall synthesis [196]. Therefore, RNAi cotton might have been lagged in primary stage much longer than Coker 312, in generating longer fiber as the result. A study on the expression of *Sus 1* at all developmental stages (-3 DPA, 0 DPA, 5 DPA, 15 DPA and 20 DPA) would provide a better understanding of the function of Sus 1 in fiber development.

In cotton, brassinosteroid (BR) positively influences fiber development [10]. It was found that the gene Gh_D13G0889 encoding a cytosolic sulfotransferase 12-like protein was up-regulated and the gene Gh_D04G1245 for a bahd acyltransferase was down-regulated in RNAi line. The *Arabidopsis* brassinosteroid sulfotransferase, a type of cytosolic sulfotransferase, was found more abundant in young seedlings and actively growing cell cultures, and its expression was induced in response to salicylic acid and methyl jasmonate and bacterial pathogens [197]. The dwarfism caused by overexpression of bahd acyltransferase can be rescued by brassinosteroid in *Arabidopsis* [198]. In addition, acylation conjugation of brassinosteroid catalyzed by bahd acyltransferase plays

a role in brassinosteroid homeostasis. Together, it suggested that fiber development was regulated by brassinosteroid metabolism and balancing in *PHYA1* RNAi cotton.

Besides brassinosteroid, the phytohormone ABA is known to inhibit fiber initiation [41]. Gh_D08G1639 and Gh_A08G1344 both encoded an abscisic acid 8′ -hydroxylase 1-like protein. The ABA 8′-hydroxylase catalyzed the first step of inactivation of ABA. Therefore the active level of ABA might be increased by decreasing the expression of ABA 8′ -hydroxylase in RNAi line. These results are consistent with the physiology data that ABA concentrations increased from the day of anthesis to 10 DPA, and then declined until 20 DPA [199]. The fiber development is a dynamic process; at the stage of 10 DPA, ABA inhibits fiber cell initiation and fiber cells are predominantly elongated in RNAi cotton.

It is well known that ethylene plays a major role in cotton fiber cell elongation [9]. In this study, five genes (Gh_D06G1403, Gh_D08G2484, Gh_A08G2112, Gh_A05G0434 and Gh_D05G0562) encoding an EIN3-binding F-box protein 1(EBF1) -like protein, were down-regulated in RNAi cotton. The reduction of EBF1 activates ethylene response by releasing EIN3 and EIL1 proteins from ubiquitination and subsequent proteasomal degradation [200]. In addition, six genes (Gh_D05G3348, Gh_D10G0255, Gh_D06G1403, Gh_D01G0450, Gh_A04G0305 and Gh_A06G1144) encoded ethylene-responsive transcription factor 4 (ERF4)-like proteins were down-regulated in RNAi line. In *Arabidopsis*, AtERF4 is a negative regulator capable of modulating ethylene and abscisic acid responses [201]. The reduction of ERF4-like proteins might promote fiber development through activating ethylene signaling pathway

mediated by EIN3. Thus, ethylene signaling pathways might also contribute to fiber elongation in RNAi cotton.

Additionally, many genes encoding WRKY or probable WRKY transcription factors were down-regulated in *PHYA1* RNAi cotton, which was consistent with the observation of Wan *et al.* (2014), in that WRKY family transcription factors were highly expressed in naked seed or fuzzless cotton mutants [202].

Identification of novel miRNAs in cotton

The utilization of allotetraploid TM-1 cotton genome as a reference is very useful for miRNA identification and target prediction. Without the complete genome sequence of *Gossypium hirsutum*, it is difficult to conduct miRNAome analysis through bioinformatics prediction. First of all, small RNA read libraries generated by previous studies might contain errors. Secondly, it is hard to evaluate the possibility of MIR gene loci duplication in allotetraploid. Thirdly, the isolation/identification of miRNAs in cotton is lagging when compared to other plant species due to the complexity of genome compensation. The major criterion for novel miRNA prediction is the hairpin structure formed by precursor miRNAs. Due to incomplete upstream and downstream sequences, it is difficult to predict the secondary structure of miRNAs [157]. The psRNATarget was used for miRNA target analysis and prediction [176]. The transcript libraries available for cotton were from *Gossypium raimondii*, and thus target prediction generated extensive redundancy. Based on our data, the precursors were predicted and isolated for all the novel miRNAs, and the redundancy of the targets was reduced by the second blasting with TM-1 as a new reference, because lots of the hits turned out to be the same gene.

Using high-throughput small RNA sequencing, 89 conserved and 8 novel miRNAs have been identified in cotton recently [157]. Based on the annotation of whole genome sequence of the cultivated *Gossypium hirsutum* TM-1 upland cotton, 602 miRNAs [203] and 301 miRNAs [204] were identified, respectively, by two research groups in the allotetraploid cotton. As a result, further identification and functional validation of cotton miRNAs are still needed, particularly for cotton specific miRNAs and fiber-development-related miRNAs. In this study, a total of 35 novel miRNAs were identified in the small RNA libraries. Of these, 16 miRNAs were located in D-genome, and the other 15 miRNAs were in A-genome. Most of miRNAs identified previously were located in D-genome due to limited sequence information of upland cotton genome [157]. An enlarged genome pool helped increase the accuracy and coverage for bioinformatics prediction and analysis. The 24 nt miRNAs were the most abundant among the 35 novel miRNAs (25), which are consistent with previous studies [205, 206]. Unlike the 21 nt miRNAs associated with AGO1, the 24 nt miRNAs were loaded onto AGO4, which has a preference for sRNAs with 5′ terminal adenine [207, 208]. As expected, most of the 24 nt miRNAs (9) had 5′ terminal adenine nucleotides. This result is consistent with the previous study [206]. The predicted hairpin precursors had negative folding free energies from -45.7 to -230.1 kcal mol^{-1}, which were in accordance with previous studies [207]. By base pairing to mRNAs, microRNAs mediate mRNA degradation or translational repression. Most of the identified novel miRNAs act to inhibit target expression through mRNA degradation, and a few of them regulate the expression of the target genes through translational repression.

Cotton fiber elongation-related miRNAs

In this study a total of 77 known miRNAs belonging to 61 miRNA families were identified in elongating fibers. Many of these conserved miRNAs were present at lower levels in RNAi cotton compared to Coker 312, including Gh-miR2950, 169b, 160, 399c and 399d. In contrast, Gh-miR7514, a cotton specific miRNA was highly enriched in RNAi cotton. Gh-miR2950, identified by Pang *et al.* in 2009, is also a cotton specific miRNA. The target gene (ES793451) of Gh-miR2950 was predicted to encode a putative gibberellin 3 hydroxylase, which accumulated to higher levels in fibers [18]. Gibberellin 3 hydroxylase was shown to control internode elongation in pea [209]. It was reported that Gh-miR2950 may affect fiber cell elongation via GA signaling [18]. The results in this study were consistent with this theory, in that down-regulation of Gh-miR2950 would promote fiber elongation in *PHYA1* RNAi cotton. miRNA169, 160, 399 are highly conserved miRNAs in plants. It was reported that miR160 and miR169 were significantly expressed at lower levels in fibers than in seedlings [210]. miRNA160 targeted three auxin response factors ARF10, 16, and 17 [18]. It is well known that auxin plays an essential role in cotton fiber development [8, 9, 211]. The results from this study suggested that miR160 can promote fiber development via the auxin signaling transduction pathway through increased expression of ARF10, 16 and 17. miR399 was predicted to target a MYB transcription factor [206]. Many MYBs, such as MYB25 and 109 were reported to promote fiber elongation, and their suppression caused shorter fibers.

As discussed above, the active levels of ABA might increase by decreasing the expression of ABA 8´-hydroxylase in RNAi line. In this study, miR7514 was

accumulated at higher levels in RNAi cotton, and it targeted the gene for Rho guanyl-nucleotide exchange factor 7 (RhoGEF7) [212]. The functional analysis of RhoGEF7 in plants has not been documented so far. However, it was reported that ABA could induce the degradation of another Rho guanyl-nucleotide exchange factor RhoGEF1 [213]. Taken together, the results in this study imply that miRNA mediated gene regulation is involved in fiber improvement of *PHYA1* RNAi cotton.

The targets of the majority of novel miRNAs (32 out of 35) were identified by psRNATarget, a plant small RNA target analysis server. Cytochrome P450 TBP had the most abundant hits, and it was the target of 9 novel miRNAs identified in this study. Of these 9 miRNAs, 8 of them had lower levels in RNAi cotton compared to Coker 312. Cytochrome P450 is involved in the biosynthesis of plant hormones (such as ABA, GA, BR) and secondary metabolites (e.g. phenylpropanoids, alkaloids, terpreoids) [214]. Cytochrome P450 enzymes also play critical roles in response to different abiotic and biotic stresses. Based on the KEGG analysis of RNA sequencing data, DEGs related to cytochrome P450-involved pathways were abundant in RNAi cotton. A stress related miRNA, miR172, was reported to target cytochrome P450 in cotton and corn [215, 216]. Taken together, although the expression of these novel miRNAs were not significantly changed in developing fibers, cytochrome P450 TBP had the highest hit as the target of 9 novel miRNAs. These observations imply that miRNA mediated gene regulation on cytochrome P450 associated pathways might play a role in improving fiber quality in *PHYA1* RNAi cotton.

The possible functions of cytochrome P450 in cotton improvement

One possible explanation to the highest hits of cytochrome P450 TBP in miRNA target prediction is the high constitution of cytochrome P450 gene (CYP) family in plant genome. It was reported that CYP genes occupied approximately 1% of the protein coding genes in six sequenced angiosperms, which include grape (*Vitis Vinifera*), papaya (*Carica papaya*), poplar (*Populus trichocarpa*), rice (*Oryza sativa*), Arabidopsis (*Arabidopsis thaliana*) and moss (*Physcomitrella patens*) and were involved in almost every aspects of the plant life [217]. Based on evolutionary analysis, the oldest CYPs are involved in biosynthesis of carotenoid and sterol. The middle branches of CYPs mediate the adaptation to environment, including abiotic stresses and biotic defenses. The newest CYPs are involved in the synthesis of structural components such as lignin, fatty acid, pigments, odorants and signaling compounds [217]. Besides with improved fiber quality, *PHYA1* RNAi plants have vigorous root and vegetative growth, and early flowering due to increased photosynthesis [2, 185]. CYPs are light-driven oxidase enzymes in electron transfer chains [218]. Thus this study support the hypothesis that increased photosynthesis of RNAi cotton might be via miRNA-mediated expression regulation on CYPs. In addition, CYPs are also involved in biosynthesis and catabolism of many plant hormones, such as ABA (CYP707), GA (CYP714) and BR (CYP724) [217]. ABA regulates osmotic stress tolerance [219], and GA plays a role in salinity adaptation [220]. As mentioned previously, CYPs also play crucial roles in environmental adaptation, which might explain the better tolerance of RNAi cotton to salinity, drought and heat stresses.

REFERENCES

1. Chen ZJ, Guan X: **Auxin boost for cotton**. *Nat Biotechnol* 2011, **29**(5):407-409.

2. Abdurakhmonov IY, Buriev ZT, Saha S, Jenkins JN, Abdukarimov A, Pepper AE: **Phytochrome RNAi enhances major fibre quality and agronomic traits of the cotton *Gossypium hirsutum* L**. *Nat Commun* 2014, **5**:3062.

3. Kasperbauer M: **Cotton fiber length is affected by far-red light impinging on developing bolls**. *Crop science* 2000, **40**(6):1673-1678.

4. Chawla R, Nicholson SJ, Folta KM, Srivastava V: **Transgene-induced silencing of *Arabidopsis* phytochrome A gene via exonic methylation**. *The Plant journal : for cell and molecular biology* 2007, **52**(6):1105-1118.

5. Rangani G, Khodakovskaya M, Alimohammadi M, Hoecker U, Srivastava V: **Site-specific methylation in gene coding region underlies transcriptional silencing of the Phytochrome A epiallele in *Arabidopsis thaliana***. *Plant molecular biology* 2012, **79**(1-2):191-202.

6. Fang L, Tian R, Li X, Chen J, Wang S, Wang P, Zhang T: **Cotton fiber elongation network revealed by expression profiling of longer fiber lines introgressed with different *Gossypium barbadense* chromosome segments**. *BMC Genomics* 2014, **15**:838.

7. Beasley C, Ting IP: **The effects of plant growth substances on in vitro fiber development from fertilized cotton ovules**. *American Journal of Botany* 1973:130-139.

8. Sun Y, Veerabomma S, Abdel-Mageed HA, Fokar M, Asami T, Yoshida S, Allen RD: **Brassinosteroid regulates fiber development on cultured cotton ovules**. *Plant and Cell Physiology* 2005, **46**(8):1384-1391.

9. Shi Y-H, Zhu S-W, Mao X-Z, Feng J-X, Qin Y-M, Zhang L, Cheng J, Wei L-P, Wang Z-Y, Zhu Y-X: **Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation**. *The Plant Cell* 2006, **18**(3):651-664.

10. Lee JJ, Woodward AW, Chen ZJ: **Gene expression changes and early events in cotton fibre development**. *Annals of Botany* 2007, **100**(7):1391-1401.

11. Wang Q, Zhu Z, Ozkardesh K, Lin C: **Phytochromes and phytohormones: the shrinking degree of separation**. *Mol Plant* 2013, **6**(1):5-7.

12. Qiu CX, Xie FL, Zhu YY, Guo K, Huang SQ, Nie L, Yang ZM: **Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags**. *Gene* 2007, **395**(1-2):49-61.

13. Zhang B, Wang Q, Wang K, Pan X, Liu F, Guo T, Cobb GP, Anderson TA: **Identification of cotton microRNAs and their targets**. *Gene* 2007, **397**(1-2):26-37.

14. Khan Barozai MY, Irfan M, Yousaf R, Ali I, Qaisar U, Maqbool A, Zahoor M, Rashid B, Hussnain T, Riazuddin S: **Identification of micro-RNAs in cotton**. *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale* 2008, **46**(8-9):739-751.

15. Ruan MB, Zhao YT, Meng ZH, Wang XJ, Yang WC: **Conserved miRNA analysis in *Gossypium hirsutum* through small RNA sequencing**. *Genomics* 2009, **94**(4):263-268.

16. Wang M, Wang Q, Wang B: **Identification and characterization of microRNAs in Asiatic cotton (*Gossypium arboreum* L.)**. *PLoS One* 2012, **7**(4):e33696.

17. Kwak PB, Wang QQ, Chen XS, Qiu CX, Yang ZM: **Enrichment of a set of microRNAs during the cotton fiber development**. *BMC Genomics* 2009, **10**:457.

18. Pang M, Woodward AW, Agarwal V, Guan X, Ha M, Ramachandran V, Chen X, Triplett BA, Stelly DM, Chen ZJ: **Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.)**. *Genome Biol* 2009, **10**(11):R122.

19. Abdurakhmonov IY, Devor EJ, Buriev ZT, Huang L, Makamov A, Shermatov SE, Bozorov T, Kushanov FN, Mavlonov GT, Abdukarimov A: **Small RNA regulation of ovule development in the cotton plant, *G. hirsutum* L**. *BMC Plant Biol* 2008, **8**:93.

20. Li Q, Jin X, Zhu YX: **Identification and analyses of miRNA genes in allotetraploid *Gossypium hirsutum* fiber cells based on the sequenced diploid G. raimondii genome**. *Journal of genetics and genomics = Yi chuan xue bao* 2012, **39**(7):351-360.

21. Fryxell PA, Craven LA, Stewart JM: **A revision of *Gossypium sect. Grandicalyx* (Malvaceae), including the description of six new species**. *Systematic Botany* 1992, **17**(1):91-114.

22. Wendel JF, Cronn RC: **Polyploidy and the evolutionary history of cotton**. *Advances in agronomy* 2003, **78**:139-186.

23. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C *et al*: **Genome sequence of the cultivated cotton *Gossypium arboreum***. *Nat Genet* 2014, **46**(6):567-572.

24. R. WFaC: **Polyploidy and the evolutionary history of cotton**. *Advances in Agronomy* 2003, **78**:139-186.

25. Gotmare V. SP, Tule BN.: **Wild and cultivated species of cotton**. *CICR TECHNICAL BULLETIN NO: 5*

26. Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD: **Toward sequencing cotton (*Gossypium*) genomes**. *Plant physiology* 2007, **145**(4):1303-1310.

27. Zhang H-B, Li Y, Wang B, Chee PW: **Recent advances in cotton genomics**. *International Journal of plant genomics* 2008, **2008**.

28. Wendel JF: **New World tetraploid cottons contain Old World cytoplasm**. *Proceedings of the National Academy of Sciences* 1989, **86**(11):4132-4136.

29. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S: **The draft genome of a diploid cotton *Gossypium raimondii***. *Nature genetics* 2012, **44**(10):1098-1103.

30. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J: **Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution**. *Nature biotechnology* 2015, **33**(5):524-530.

31. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM: **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement**. *Nature biotechnology* 2015, **33**(5):531-537.

32. Applequist WL, Cronn R, Wendel JF: **Comparative development of fiber in wild and cultivated cotton**. *Evolution & development* 2001, **3**(1):3-17.

33. Basra AS, Malik C: **Development of the cotton fiber**. *Int Rev Cytol* 1984, **89**(1):65-113.

34. Seagull RW. OV, Murphy K., Binder A., and Kothari S.: **Cotton fiber growth and development. changes in cell diameter and wall birefringence.** *The Journal of Cotton Science* 2000, **4**:97-104.

35. Stewart JM: **Fiber initiation on the cotton ovule (*Gossypium hirsutum*)**. *American Journal of Botany* 1975, **62**(7):723-730.

36. Ramsey JC, Berlin JD: **Ultrastructure of early stages of cotton fiber differentiation**. *Botanical Gazette* 1976, **137**(1):11-19.

37. Meinert MC, Delmer DP: **Changes in biochemical composition of the cell wall of the cotton fiber during development**. *Plant Physiology* 1977, **59**(6):1088-1097.

38. Quader H, Herth W, Ryser U, Schnepf E: **Cytoskeletal elements in cotton seed hair developmentin vitro: Their possible regulatory role in cell wall organization**. *Protoplasma* 1987, **137**(1):56-62.

39. Graves DA, Stewart JM: **Analysis of the protein constituency of developing cotton fibres**. *Journal of experimental botany* 1988, **39**(1):59-69.

40. Guan X, Song Q, Chen ZJ: **Polyploidy and small RNA regulation of cotton fiber development**. *Trends in plant science* 2014, **19**(8):516-528.

41. Kosmidou-Dimitropoulou K: **Hormonal influences on fiber development**. *Cotton Physiology The Cotton Foundation, Memphis, TN* 1986:361-373.

42. Ma D-P, Liu H-C, Tan H, Creech RG, Jenkins JN, Chang Y-F: **Cloning and characterization of a cotton lipid transfer protein gene specifically expressed in fiber cells**. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* 1997, **1344**(2):111-114.

43. Ma D-P, Tan H, Si Y, Creech RG, Jenkins JN: **Differential expression of a lipid transfer protein gene in cotton fiber**. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* 1995, **1257**(1):81-84.

44. Pruitt RE, Vielle-Calzada J-P, Ploense SE, Grossniklaus U, Lolle SJ: **FIDDLEHEAD, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme**. *Proceedings of the National Academy of Sciences* 2000, **97**(3):1311-1316.

45. Wang QQ, Liu F, Chen XS, Ma XJ, Zeng HQ, Yang ZM: **Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant**. *Genomics* 2010, **96**(6):369-376.

46. Wang Z, Zhang D, Wang X, Tan X, Guo H, Paterson AH: **A whole-genome DNA marker map for cotton based on the D-genome sequence of *Gossypium raimondii* L**. *G3: Genes| Genomes| Genetics* 2013, **3**(10):1759-1767.

47. Deng XW, Quail PH: **Signalling in light-controlled development**. In: *Seminars in cell & developmental biology: 1999*: Elsevier; 1999: 121-129.

48. Quail PH: **Photosensory perception and signal transduction in plants**. *Current opinion in genetics & development* 1994, **4**(5):652-661.

49. Smith H: **Physiological and ecological function within the phytochrome family**. *Annual review of plant biology* 1995, **46**(1):289-315.

50. Wang H, Deng XW: **Dissecting the phytochrome A-dependent signaling network in higher plants**. *Trends in plant science* 2003, **8**(4):172-178.

51. Jiao Y, Lau OS, Deng XW: **Light-regulated transcriptional networks in higher plants**. *Nature Reviews Genetics* 2007, **8**(3):217-230.

52. Neff MM, Fankhauser C, Chory J: **Light: an indicator of time and place**. *Genes & Development* 2000, **14**(3):257-271.

53. Ahmad M, Cashmore AR: **Seeing blue: the discovery of cryptochrome**. *Plant molecular biology* 1996, **30**(5):851-861.

54. Christie JM, Reymond P, Powell GK, Bernasconi P, Raibekas AA, Liscum E, Briggs WR: ***Arabidopsis* NPH1: a flavoprotein with the properties of a photoreceptor for phototropism**. *Science* 1998, **282**(5394):1698-1701.

55. Rizzini L, Favory J-J, Cloix C, Faggionato D, O'Hara A, Kaiserli E, Baumeister R, Schäfer E, Nagy F, Jenkins GI: **Perception of UV-B by the *Arabidopsis* UVR8 protein**. *Science* 2011, **332**(6025):103-106.

56. Butler WL, Norris KH, Siegelman HW, Hendricks SB: **Detection, assay, and preliminary purification of the pigment comtrolling photoresponsive development of plants.** *Proceedings of the National Academy of Sciences of the United States of America* 1959, **45**(12):1703-1708.

57. Clack T, Mathews S, Sharrock RA: **The phytochrome apoprotein family in *Arabidopsis* is encoded by five genes: the sequences and expression of PHYD and PHYE**. *Plant molecular biology* 1994, **25**(3):413-427.

58. Cashmore AR, Jarillo JA, Wu Y-J, Liu D: **Cryptochromes: blue light receptors for plants and animals**. *Science* 1999, **284**(5415):760-765.

59. Takano M, Inagaki N, Xie X, Kiyota S, Baba-Kasai A, Tanabata T, Shinomura T: **Phytochromes are the sole photoreceptors for perceiving red/far-red light in rice**. *Proceedings of the National Academy of Sciences* 2009, **106**(34):14705-14710.

60. Abdurakhmonov IY, Buriev ZT, Logan-Young CJ, Abdukarimov A, Pepper AE: **Duplication, divergence and persistence in the Phytochrome photoreceptor gene family of cottons (*Gossypium* spp.)**. *BMC plant biology* 2010, **10**(1):119.

61. Mathews S, Sharrock R: **Phytochrome gene diversity**. *Plant, Cell & Environment* 1997, **20**(6):666-671.

62. Hirschfeld M, Tepperman JM, Clack T, Quail PH, Sharrock RA: **Coordination of phytochrome levels in phyB mutants of *Arabidopsis* as revealed by apoprotein-specific monoclonal antibodies**. *Genetics* 1998, **149**(2):523-535.

63. Quail PH, Boylan MT, Parks BM, Short TW: **Phytochromes: photosensory perception and signal transduction**. *Science* 1995, **268**(5211):675.

64. Sharrock RA, Clack T: **Patterns of expression and normalized levels of the five *Arabidopsis* phytochromes**. *Plant Physiology* 2002, **130**(1):442-456.

65. Bae G, Choi G: **Decoding of light signals by plant phytochromes and their interacting proteins**. *Annu Rev Plant Biol* 2008, **59**:281-311.

66. Wu S-H, Lagarias JC: **Defining the bilin lyase domain: lessons from the extended phytochrome superfamily**. *Biochemistry* 2000, **39**(44):13487-13495.

67. Li J, Li G, Wang H, Wang Deng X: **Phytochrome signaling mechanisms**. *The Arabidopsis Book* 2011:e0148.

68. Fankhauser C, Chory J: **Light control of plant development**. *Annual review of cell and developmental biology* 1997, **13**(1):203-229.

69. Casal JJ, Sánchez RA, Botto JF: **Modes of action of phytochromes**. *Journal of Experimental Botany* 1998, **49**(319):127-138.

70. Botto JF, Sanchez RA, Whitelam GC, Casal JJ: **Phytochrome A mediates the promotion of seed germination by very low fluences of light and canopy shade light in *Arabidopsis***. *Plant Physiology* 1996, **110**(2):439-444.

71. Casal J, Candia A, Sellaro R: **Light perception and signalling by phytochrome A**. *Journal of experimental botany* 2014, **65**(11):2835-2845.

72. Oka Y, Ono Y, Toledo-Ortiz G, Kokaji K, Matsui M, Mochizuki N, Nagatani A: ***Arabidopsis* phytochrome A is modularly structured to integrate the multiple features that are required for a highly sensitized phytochrome**. *The Plant Cell* 2012, **24**(7):2949-2962.

73. Shinomura T, Nagatani A, Hanzawa H, Kubota M, Watanabe M, Furuya M: **Action spectra for phytochrome A-and B-specific photoinduction of seed germination in *Arabidopsis thaliana***. *Proceedings of the National Academy of Sciences* 1996, **93**(15):8129-8133.

74. Hennig L, Stoddart WM, Dieterle M, Whitelam GC, Schäfer E: **Phytochrome E controls light-induced germination of *Arabidopsis***. *Plant Physiology* 2002, **128**(1):194-200.

75. Dehesh K, Franci C, Parks BM, Seeley KA, Short TW, Tepperman JM, Quail PH: ***Arabidopsis* HY8 locus encodes phytochrome A**. *The Plant Cell* 1993, **5**(9):1081-1088.

76. Nagatani A, Reed JW, Chory J: **Isolation and initial characterization of *Arabidopsis* mutants that are deficient in phytochrome A**. *Plant Physiology* 1993, **102**(1):269-277.

77. Parks BM, Quail PH: **hy8, a new class of *Arabidopsis* long hypocotyl mutants deficient in functional phytochrome A**. *The Plant Cell* 1993, **5**(1):39-48.

78. Whitelam GC, Johnson E, Peng J, Carol P, Anderson ML, Cowl JS, Harberd NP: **Phytochrome A null mutants of *Arabidopsis* display a wild-type phenotype in white light**. *The Plant Cell* 1993, **5**(7):757-768.

79. Neff MM, Chory J: **Genetic interactions between phytochrome A, phytochrome B, and cryptochrome 1 during *Arabidopsis* development**. *Plant Physiology* 1998, **118**(1):27-35.

80. Reed JW, Nagpal P, Poole DS, Furuya M, Chory J: **Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development**. *The Plant Cell* 1993, **5**(2):147-157.

81. Aukerman MJ, Hirschfeld M, Wester L, Weaver M, Clack T, Amasino RM, Sharrock RA: **A deletion in the PHYD gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing**. *The Plant Cell* 1997, **9**(8):1317-1326.

82. Franklin KA, Praekelt U, Stoddart WM, Billingham OE, Halliday KJ, Whitelam GC: **Phytochromes B, D, and E act redundantly to control multiple physiological responses in *Arabidopsis***. *Plant Physiology* 2003, **131**(3):1340-1346.

83. Franklin KA, Davis SJ, Stoddart WM, Vierstra RD, Whitelam GC: **Mutant analyses define multiple roles for phytochrome C in *Arabidopsis* photomorphogenesis**. *The Plant Cell* 2003, **15**(9):1981-1989.

84. Franklin KA: **Shade avoidance**. *New Phytologist* 2008, **179**(4):930-944.

85. Halliday KJ, Martínez-García JF, Josse E-M: **Integration of light and auxin signaling**. *Cold Spring Harbor perspectives in biology* 2009, **1**(6):a001586.

86. Boccalandro HE, Rugnone ML, Moreno JE, Ploschuk EL, Serna L, Yanovsky MJ, Casal JJ: **Phytochrome B enhances photosynthesis at the expense of water-use efficiency in *Arabidopsis***. *Plant Physiology* 2009, **150**(2):1083-1092.

87. Somers DE, Devlin PF, Kay SA: **Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock**. *Science* 1998, **282**(5393):1488-1490.

88. Devlin PF, Kay SA: **Cryptochromes are required for phytochrome signaling to the circadian clock but not for rhythmicity**. *The Plant Cell* 2000, **12**(12):2499-2509.

89. Halliday KJ, Salter MG, Thingnaes E, Whitelam GC: **Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator FT**. *The Plant Journal* 2003, **33**(5):875-885.

90. Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D: **The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana***. *Nature genetics* 2006, **38**(6):711-715.

91. Kircher S, Gil P, Kozma-Bognár L, Fejes E, Speth V, Husselstein-Muller T, Bauer D, Ádám É, Schäfer E, Nagy F: **Nucleocytoplasmic partitioning of the plant photoreceptors phytochrome A, B, C, D, and E is regulated differentially by light and exhibits a diurnal rhythm**. *The Plant Cell* 2002, **14**(7):1541-1555.

92. Zheng X, Wu S, Zhai H, Zhou P, Song M, Su L, Xi Y, Li Z, Cai Y, Meng F: ***Arabidopsis* phytochrome B promotes SPA1 nuclear accumulation to repress photomorphogenesis under far-red light**. *The Plant Cell* 2013, **25**(1):115-133.

93. Chen M, Tao Y, Lim J, Shaw A, Chory J: **Regulation of phytochrome B nuclear localization through light-dependent unmasking of nuclear-localization signals**. *Current Biology* 2005, **15**(7):637-642.

94. Hiltbrunner A, Tscheuschler A, Viczián A, Kunkel T, Kircher S, Schäfer E: **FHY1 and FHL act together to mediate nuclear accumulation of the phytochrome A photoreceptor**. *Plant and cell physiology* 2006, **47**(8):1023-1034.

95. Hiltbrunner A, Viczián A, Bury E, Tscheuschler A, Kircher S, Tóth R, Honsberger A, Nagy F, Fankhauser C, Schäfer E: **Nuclear accumulation of the phytochrome A photoreceptor requires FHY1**. *Current biology* 2005, **15**(23):2125-2130.

96. Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H: **Transposase-derived transcription factors regulate light signaling in *Arabidopsis***. *Science* 2007, **318**(5854):1302-1305.

97. Hisada A, Hanzawa H, Weller JL, Nagatani A, Reid JB, Furuya M: **Light-induced nuclear translocation of endogenous pea phytochrome A visualized by immunocytochemical procedures**. *The Plant Cell* 2000, **12**(7):1063-1078.

98. Bauer D, Viczián A, Kircher S, Nobis T, Nitschke R, Kunkel T, Panigrahi KC, Ádám É, Fejes E, Schäfer E: **Constitutive photomorphogenesis 1 and multiple photoreceptors control degradation of phytochrome interacting factor 3, a transcription factor required for light signaling in *Arabidopsis***. *The Plant Cell* 2004, **16**(6):1433-1445.

99. Kevei E, Schafer E, Nagy F: **Light-regulated nucleo-cytoplasmic partitioning of phytochromes**. *Journal of experimental botany* 2007, **58**(12):3113-3124.

100. Chen M: **Phytochrome nuclear body: an emerging model to study interphase nuclear dynamics and signaling**. *Current opinion in plant biology* 2008, **11**(5):503-508.

101. Chen M, Galvão RM, Li M, Burger B, Bugea J, Bolado J, Chory J: ***Arabidopsis* HEMERA/pTAC12 initiates photomorphogenesis by phytochromes**. *Cell* 2010, **141**(7):1230-1240.

102. Huq E, Al-Sady B, Hudson M, Kim C, Apel K, Quail PH: **Phytochrome-interacting factor 1 is a critical bHLH regulator of chlorophyll biosynthesis**. *Science* 2004, **305**(5692):1937-1941.

103. Ni M, Tepperman JM, Quail PH: **PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein**. *Cell* 1998, **95**(5):657-667.

104. Huq E, Quail PH: **PIF4, a phytochrome‐interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in *Arabidopsis***. *The EMBO journal* 2002, **21**(10):2441-2450.

105. Khanna R, Huq E, Kikis EA, Al-Sady B, Lanzatella C, Quail PH: **A novel molecular recognition motif necessary for targeting photoactivated phytochrome signaling to specific basic helix-loop-helix transcription factors**. *The Plant Cell* 2004, **16**(11):3033-3044.

106. Leivar P, Monte E, Al-Sady B, Carle C, Storer A, Alonso JM, Ecker JR, Quail PH: **The *Arabidopsis* phytochrome-interacting factor PIF7, together with PIF3 and PIF4, regulates responses to prolonged red light by modulating phyB levels**. *The Plant Cell* 2008, **20**(2):337-352.

107. Lorrain S, Trevisan M, Pradervand S, Fankhauser C: **Phytochrome interacting factors 4 and 5 redundantly limit seedling de‐etiolation in continuous far‐red light**. *The Plant Journal* 2009, **60**(3):449-461.

108. Yang SW, Jang I-C, Henriques R, Chua N-H: **FAR-RED ELONGATED HYPOCOTYL1 and FHY1-LIKE associate with the *Arabidopsis* transcription factors LAF1 and HFR1 to transmit phytochrome A signals for inhibition of hypocotyl elongation**. *The Plant Cell* 2009, **21**(5):1341-1359.

109. Fairchild CD, Schumaker MA, Quail PH: **HFR1 encodes an atypical bHLH protein that acts in phytochrome A signal transduction**. *Genes Dev* 2000, **14**(18):2377-2391.

110. Ballesteros ML, Bolle C, Lois LM, Moore JM, Vielle-Calzada JP, Grossniklaus U, Chua NH: **LAF1, a MYB transcription activator for phytochrome A signaling**. *Genes Dev* 2001, **15**(19):2613-2625.

111. Jang I-C, Henriques R, Chua N-H: **Three transcription factors, HFR1, LAF1 and HY5, regulate largely independent signaling pathways downstream of Phytochrome A**. *Plant and Cell Physiology* 2013, **54**(6):907-916.

112. Jang IC, Yang JY, Seo HS, Chua NH: **HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling**. *Genes Dev* 2005, **19**(5):593-602.

113. Han Y-J, Kim H-S, Kim Y-M, Shin A-Y, Lee S-S, Bhoo SH, Song P-S, Kim J-I: **Functional characterization of phytochrome autophosphorylation in plant light signaling**. *Plant and cell physiology* 2010, **51**(4):596-609.

114. Kim D-H, Kang J-G, Yang S-S, Chung K-S, Song P-S, Park C-M: **A phytochrome-associated protein phosphatase 2A modulates light signals in flowering time control in *Arabidopsis***. *The Plant Cell* 2002, **14**(12):3043-3056.

115. Saijo Y, Zhu D, Li J, Rubio V, Zhou Z, Shen Y, Hoecker U, Wang H, Deng XW: ***Arabidopsis* COP1/SPA1 complex and FHY1/FHY3 associate with distinct phosphorylated forms of phytochrome A in balancing light signaling**. *Molecular cell* 2008, **31**(4):607-613.

116. Tessadori F, van Zanten M, Pavlova P, Clifton R, Pontvianne F, Snoek LB, Millenaar FF, Schulkes RK, van Driel R, Voesenek LA *et al*: **Phytochrome B and histone deacetylase 6 control light-induced chromatin compaction in *Arabidopsis thaliana***. *PLoS genetics* 2009, **5**(9):e1000638.

117. de Lucas M, Davière J-M, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S, Fankhauser C, Blázquez MA, Titarenko E, Prat S: **A molecular framework for light and gibberellin control of cell elongation**. *Nature* 2008, **451**(7177):480-484.

118. Feng S, Martinez C, Gusmaroli G, Wang Y, Zhou J, Wang F, Chen L, Yu L, Iglesias-Pedraz JM, Kircher S: **Coordinated regulation of *Arabidopsis thaliana* development by light and gibberellins**. *Nature* 2008, **451**(7177):475-479.

119. Bai M-Y, Shang J-X, Oh E, Fan M, Bai Y, Zentella R, Sun T-p, Wang Z-Y: **Brassinosteroid, gibberellin and phytochrome impinge on a common transcription module in *Arabidopsis***. *Nature Cell Biology* 2012, **14**(8):810-817.

120. Singh G, Garg O: **Effect of red, far-red radiations on germination of cotton seed**. *Plant and cell physiology* 1971, **12**(3):411-415.

121. Kasperbauer M: **Cotton plant size and fiber developmental responses to FR/R ratio reflected from the soil surface**. *Physiologia Plantarum* 1994, **91**(2):317-321.

122. Ouedraogo M, Hubac C: **Effect of far red light on drought resistance of cotton**. *Plant and Cell Physiology* 1982, **23**(7):1297-1303.

123. Brubaker CL, Bourland F, Wendel JF: **The origin and domestication of cotton**. *Cotton: Origin, history, technology, and production John Wiley & Sons, New York* 1999:3-31.

124. Kasperbauer MJ: **Cotton fibre length is affected by far-red light impingping on developing bolls**. *Crop Science* 2000, **40**(6):1673-1678.

125. Costello JF, Plass C: **Methylation matters**. *Journal of medical genetics* 2001, **38**(5):285-303.

126. Bird AP, Wolffe AP: **Methylation-induced repression--belts, braces, and chromatin**. *Cell* 1999, **99**(5):451-454.

127. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP: **A human B cell methylome at 100-base pair resolution**. *Proc Natl Acad Sci U S A* 2009, **106**(3):671-678.

128. Chan SW, Henderson IR, Jacobsen SE: **Gardening the genome: DNA methylation in *Arabidopsis thaliana***. *Nature reviews Genetics* 2005, **6**(5):351-360.

129. Miranda TB, Jones PA: **DNA methylation: the nuts and bolts of repression**. *J Cell Physiol* 2007, **213**(2):384-390.

130. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S: **Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription**. *Nat Genet* 2007, **39**(1):61-69.

131. Goll MG, Bestor TH: **Eukaryotic cytosine methyltransferases**. *Annual review of biochemistry* 2005, **74**:481-514.

132. Wang ZG, Wu JX: **DNA methyltransferases: classification, functions and research progress**. *Zhongguo yi chuan xue hui bian ji* 2009, **31**(9):903-912.

133. Jurkowska RZ, Jurkowski TP, Jeltsch A: **Structure and function of mammalian DNA methyltransferases**. *Chembiochem : a European journal of chemical biology* 2011, **12**(2):206-222.

134. Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I: **CG gene body DNA methylation changes and evolution of duplicated genes in cassava**. *Proceedings of the National Academy of Sciences* 2015, **112**(44):13729-13734.

135. Bhutani N, Burns DM, Blau HM: **DNA demethylation dynamics**. *Cell* 2011, **146**(6):866-872.

136. He X-J, Chen T, Zhu J-K: **Regulation and function of DNA methylation in plants and animals**. *Cell Research* 2011, **21**(3):442-465.

137. Jin X, Pang Y, Jia F, Xiao G, Li Q, Zhu Y: **A potential role for CHH DNA methylation in cotton fiber growth patterns**. *PloS one* 2013, **8**(4):e60547.

138. Fraga MF, Esteller M: **DNA methylation: a profile of methods and applications**. *Biotechniques* 2002, **33**(3):632-649.

139. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: **Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning**. *Nature* 2008, **452**(7184):215-219.

140. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in *Arabidopsis***. *Cell* 2008, **133**(3):523-536.

141. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence microRNAs using deep sequencing data**. *Nucleic acids research* 2014, **42**(D1):D68-D73.

142. Chen X: **A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development**. *Science* 2004, **303**(5666):2022-2025.

143. Wu L, Fan J, Belasco JG: **MicroRNAs direct rapid deadenylation of mRNA**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(11):4034-4039.

144. Zhang B, Wang Q, Pan X: **MicroRNAs and their regulatory roles in animals and plants**. *Journal of cellular physiology* 2007, **210**(2):279-289.

145. Voinnet O: **Origin, biogenesis, and activity of plant microRNAs**. *Cell* 2009, **136**(4):669-687.

146. Cai X, Hagedorn CH, Cullen BR: **Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs**. *Rna* 2004, **10**(12):1957-1966.

147. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN: **MicroRNA genes are transcribed by RNA polymerase II**. *The EMBO journal* 2004, **23**(20):4051-4060.

148. Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grässer FA, van Dyk LF, Ho CK, Shuman S, Chien M: **Identification of microRNAs of the herpesvirus family**. *Nature methods* 2005, **2**(4):269-276.

149. Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD: **A protein sensor for siRNA asymmetry**. *Science* 2004, **306**(5700):1377-1380.

150. Boutet S, Vazquez F, Liu J, Béclin C, Fagard M, Gratias A, Morel J-B, Crété P, Chen X, Vaucheret H: ***Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance**. *Current Biology* 2003, **13**(10):843-848.

151. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS: **Nuclear processing and export of microRNAs in *Arabidopsis***. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(10):3691-3696.

152. Bushati N, Cohen SM: **microRNA functions**. *Annu Rev Cell Dev Biol* 2007, **23**:175-205.

153. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, Steward R, Chen X: **Methylation as a crucial step in plant microRNA biogenesis**. *Science* 2005, **307**(5711):932-935.

154. Leung AK, Sharp PA: **microRNAs: a safeguard against turmoil?** *Cell* 2007, **130**(4):581-585.

155. Carthew RW: **Gene regulation by microRNAs**. *Curr Opin Genet Dev* 2006, **16**(2):203-208.

156. Pillai RS: **MicroRNA function: multiple mechanisms for a tiny RNA?** *Rna* 2005, **11**(12):1753-1761.

157. Xie F, Jones DC, Wang Q, Sun R, Zhang B: **Small RNA sequencing identifies miRNA roles in ovule and fibre development**. *Plant biotechnology journal* 2015, **13**(3):355-369.

158. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ: **Criteria for annotation of plant MicroRNAs**. *The Plant Cell* 2008, **20**(12):3186-3190.

159. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP: **Prediction of plant microRNA targets**. *cell* 2002, **110**(4):513-520.

160. Alemán LM, Doench J, Sharp PA: **Comparison of siRNA-induced off-target RNA and protein effects**. *Rna* 2007, **13**(3):385-395.

161. Paterson AH, Brubaker CL, Wendel JF: **A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis**. *Plant Molecular Biology Reporter* 1993, **11**(2):122-127.

162. Wan CY, Wilkins TA: **A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.)**. *Analytical biochemistry* 1994, **223**(1):7-12.

163. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.

164. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J: **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement**. 2015, **33**(5):531-537.

165. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq**. *Nat Methods* 2008, **5**(7):621-628.

166. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research**. *Bioinformatics* 2005, **21**(18):3674-3676.

167. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier**. *Nucleic Acids Research* 2005, **33**(Web Server issue):W116-W120.

168. The Gene Ontology C, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS *et al*: **Gene Ontology: tool for the unification of biology**. *Nature genetics* 2000, **25**(1):25-29.

169. Fisher RA: **Statistical methods for research workers**: Genesis Publishing Pvt Ltd; 1925.

170. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. *Proceedings of the National Academy of Sciences* 2005, **102**(43):15545-15550.

171. Auer PL, Doerge RW: **Statistical Design and Analysis of RNA Sequencing Data**. *Genetics* 2010, **185**(2):405-416.

172. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 2000, **28**(1):27-30.

173. Yang X, Wang L, Yuan D, Lindsey K, Zhang X: **Small RNA and degradome sequencing reveal complex miRNA regulation during cotton somatic embryogenesis**. *J Exp Bot* 2013, **64**(6):1521-1536.

174. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR: **Rfam: an RNA family database**. *Nucleic Acids Research* 2003, **31**(1):439-441.

175. Li Y, Zhang Z, Liu F, Vongsangnak W, Jing Q, Shen B: **Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis**. *Nucleic Acids Research* 2012, **40**(10):4298-4305.

176. Dai X, Zhao PX: **psRNATarget: a plant small RNA target analysis server**. *Nucleic Acids Res* 2011, **39**(Web Server issue):W155-159.

177. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction**. *Nucleic Acids Res* 2003, **31**(13):3406-3415.

178. Wang ZM, Xue W, Dong CJ, Jin LG, Bian SM, Wang C, Wu XY, Liu JY: **A comparative miRNAome analysis reveals seven fiber initiation-related and 36 novel miRNAs in developing cotton ovules**. *Mol Plant* 2012, **5**(4):889-900.

179. Fu W, Shen Y, Hao J, Wu J, Ke L, Wu C, Huang K, Luo B, Xu M, Cheng X: **Acyl-CoA N-acyltransferase influences fertility by regulating lipid metabolism and jasmonic acid biogenesis in cotton**. *Scientific reports* 2015, **5**.

180. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method**. *methods* 2001, **25**(4):402-408.

181. Ding M, Chen J, Jiang Y, Lin L, Cao Y, Wang M, Zhang Y, Rong J, Ye W: **Genome-wide investigation and transcriptome analysis of the WRKY gene family in *Gossypium***. *Molecular genetics and genomics : MGG* 2015, **290**(1):151-171.

182. Krochko JE, Abrams GD, Loewen MK, Abrams SR, Cutler AJ: **(+)-Abscisic Acid 8′ -Hydroxylase Is a cytochrome P450 monooxygenase**. *Plant Physiology* 1998, **118**(3):849-860.

183. Saito S, Hirai N, Matsumoto C, Ohigashi H, Ohta D, Sakata K, Mizutani M: ***Arabidopsis* CYP707As Encode (+)-Abscisic Acid 8′ -Hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid**. *Plant Physiology* 2004, **134**(4):1439-1449.

184. Narusaka M, Seki M, Umezawa T, Ishida J, Nakajima M, Enju A, Shinozaki K: **Crosstalk in the responses to abiotic and biotic stresses in *Arabidopsis*: analysis of gene expression in cytochrome P450 gene superfamily by cDNA microarray**. *Plant molecular biology* 2004, **55**(3):327-342.

185. Abdurakhmonov IY, Ayubov MS, Ubaydullaeva KA, Buriev ZT, Shermatov SE, Ruziboev HS, Shapulatov UM, Saha S, Ulloa M, Yu JZ: **RNA interference for functional genomics and improvement of cotton (*Gossypium* sp.)**. *Frontiers in plant science* 2016, **7**.

186. Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y *et al*: **Single base-resolution methylome of the silkworm reveals a sparse epigenomic map**. *Nat Biotech* 2010, **28**(5):516-520.

187. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis***. *Cell* 2006, **126**(6):1189-1201.

188. Osabe K, Clement JD, Bedon F, Pettolino FA, Ziolkowski L, Llewellyn DJ, Finnegan EJ, Wilson IW: **Genetic and DNA methylation changes in cotton (*Gossypium*) genotypes and tissues**. *PLoS one* 2014, **9**(1):e86049.

189. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J: **The dynamic plant cell wall**. 2000.

190. Lamport DT, Kieliszewski MJ, Chen Y, Cannon MC: **Role of the extensin superfamily in primary cell wall architecture**. *Plant physiology* 2011, **156**(1):11-19.

191. Jiménez-López JC, de Dios Alche J, Rodríguez-García MI: **Systematic and phylogenetic analysis of the Ole e 1 pollen protein family members in plants**: INTECH Open Access Publisher; 2011.

192. Jiang S-Y, Jasmin PXH, Ting YY, Ramachandran S: **Genome-wide identification and molecular characterization of Ole_e_I, Allerg_1 and Allerg_2 domain-containing pollen-allergen-like genes in Oryza sativa**. *DNA Research* 2005, **12**(3):167-179.

193. Bruex A, Kainkaryam RM, Wieckowski Y, Kang YH, Bernhardt C, Xia Y, Zheng X, Wang JY, Lee MM, Benfey P: **A gene regulatory network for root epidermis cell differentiation in *Arabidopsis***. *PLoS genetics* 2012, **8**(1):e1002446.

194. Van Hengel AJ, Roberts K: **AtAGP30, an arabinogalactan‐protein in the cell walls of the primary root, plays a role in root regeneration and seed germination**. *The Plant Journal* 2003, **36**(2):256-270.

195. Chen A, He S, Li F, Li Z, Ding M, Liu Q, Rong J: **Analyses of the sucrose synthase gene family in cotton: structure, phylogeny and expression patterns**. *BMC plant biology* 2012, **12**(1):85.

196. Bai W-Q, Xiao Y-H, Zhao J, Song S-Q, Hu L, Zeng J-Y, Li X-B, Hou L, Luo M, Li D-M *et al*: **Gibberellin overproduction promotes sucrose synthase expression and secondary cell wall deposition in cotton fibers**. *PLoS ONE* 2014, **9**(5):e96537.

197. Klein M, Papenbrock J: **The multi-protein family of *Arabidopsis* sulphotransferases and their relatives in other plant species**. *Journal of Experimental Botany* 2004, **55**(404):1809-1820.

198. Wang M, Liu X, Wang R, Li W, Rodermel S, Yu F: **Overexpression of a putative *Arabidopsis* BAHD acyltransferase causes dwarfism that can be rescued by brassinosteroid**. *Journal of Experimental Botany* 2012, **63**(16):5787-5801.

199. Davis LA, Addicott FT: **Abscisic acid: correlations with abscission and with development in the cotton fruit**. *Plant physiology* 1972, **49**(4):644-648.

200. Binder BM, Walker JM, Gagne JM, Emborg TJ, Hemmann G, Bleecker AB, Vierstra RD: **The *Arabidopsis* EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling**. *The Plant Cell* 2007, **19**(2):509-523.

201. Yang Z, Tian L, Latoszek-Green M, Brown D, Wu K: ***Arabidopsis* ERF4 is a transcriptional repressor capable of modulating ethylene and abscisic acid responses**. *Plant molecular biology* 2005, **58**(4):585-596.

202. Wan Q, Zhang H, Ye W, Wu H, Zhang T: **Genome-wide transcriptome profiling revealed cotton fuzz fiber development having a similar molecular model as *Arabidopsis* trichome**. *PloS one* 2014, **9**(5):e97313.

203. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J *et al*: **Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution**. *Nat Biotech* 2015, **33**(5):524-530.

204. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM *et al*: **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement**. *Nat Biotech* 2015, **33**(5):531-537.

205. Wang Z-M, Xue W, Dong C-J, Jin L-G, Bian S-M, Wang C, Wu X-Y, Liu J-Y: **A comparative miRNAome analysis reveals seven fiber initiation-related and 36 novel miRNAs in developing cotton ovules**. *Molecular plant* 2012, **5**(4):889-900.

206. Xue W, Wang Z, Du M, Liu Y, Liu J-Y: **Genome-wide analysis of small RNAs reveals eight fiber elongation-related and 257 novel microRNAs in elongating cotton fiber cells**. *BMC genomics* 2013, **14**(1):1.

207. Bonnet E, Wuyts J, Rouzé P, Van de Peer Y: **Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences**. *Bioinformatics* 2004, **20**(17):2911-2917.

208. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C: **Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5′ terminal nucleotide**. *Cell* 2008, **133**(1):116-127.

209. Lester DR, Ross JJ, Davies PJ, Reid JB: **Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase**. *The Plant Cell* 1997, **9**(8):1435-1443.

210. Xie F, Zhang B: **microRNA evolution and expression analysis in polyploidized cotton genome**. *Plant biotechnology journal* 2015, **13**(3):421-434.

211. Kim HJ, Triplett BA: **Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis**. *Plant physiology* 2001, **127**(4):1361-1366.

212. Zhang Y, Wang W, Chen J, Liu J, Xia M, Shen F: **Identification of miRNAs and their targets in cotton inoculated with *Verticillium dahliae* by high-throughput sequencing and degradome analysis**. *International Journal of Molecular Sciences* 2015, **16**(7):14749-14768.

213. Li Z, Waadt R, Schroeder JI: **Release of GTP exchange factor mediated down-regulation of abscisic acid signal transduction through ABA-induced rapid degradation of RopGEFs**. *PLoS Biol* 2016, **14**(5):e1002461.

214. Ramamoorthy R, Jiang S-Y, Ramachandran S: ***Oryza sativa* cytochrome P450 family member OsCYP96B4 reduces plant height in a transcript dosage dependent manner**. *PloS one* 2011, **6**(11):e28069.

215. Zhang L, Chia J-M, Kumari S, Stein JC, Liu Z, Narechania A, Maher CA, Guill K, McMullen MD, Ware D: **A genome-wide characterization of microRNA genes in maize**. *PLoS genetics* 2009, **5**(11):e1000716.

216. An W, Gong W, He S, Pan Z, Sun J, Du X: **MicroRNA and mRNA expression profiling analysis revealed the regulation of plant height in *Gossypium hirsutum***. *BMC Genomics* 2015, **16**:886.

217. Nelson DR, Ming R, Alam M, Schuler MA: **Comparison of cytochrome P450 genes from six plant genomes**. *Tropical Plant Biology* 2008, **1**(3-4):216-235.

218. Jensen K, Jensen PE, Møller BL: **Light-driven cytochrome P450 hydroxylations**. *ACS chemical biology* 2011, **6**(6):533-539.

219. Chinnusamy V, Schumaker K, Zhu JK: **Molecular genetic perspectives on cross‐talk and specificity in abiotic stress signalling in plants**. *Journal of experimental botany* 2004, **55**(395):225-236.

220. Zhu W, Miao Q, Sun D, Yang G, Wu C, Huang J, Zheng C: **The mitochondrial phosphate transporters modulate plant responses to salt stress via affecting ATP and gibberellin metabolism in *Arabidopsis thaliana***. *PloS one* 2012, **7**(8):e43530.