MOLECULAR BASIS OF QUANTITATIVE GENETICS REVEALED BY CLONING AND ANALYSIS OF 474 GENES CONTROLLING FIBER LENGTH IN COTTON

A Dissertation

by

YUN-HUA LIU

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Hongbin Zhang
Co-Chair of Committee,	C. Wayne Smith
Committee Members,	Steve Hague
	Joshua Yuan
Head of Department,	David Baltensperger

August 2014

Major Subject: Plant Breeding

ABSTRACT

Cotton (*Gossypium* spp.) is a leading textile crop in the world, generating an annual economic benefit of over hundred billion USD. However, few genes controlling fiber quality and yield traits have been cloned and characterized to date. In this study, a large number of genes controlling the upper half mean length (UHML) of fibers were cloned using a newly-developed high-throughput gene and QTL cloning system and subjected to systems analysis. Furthermore, the molecular basis and regulation mechanisms of UHML were investigated using the cloned fiber length genes.

A total of 474 *GFL* (*Gossypium* Fiber Length) genes were cloned. The effect of each *GFL* gene on UHML varied from 2.64% to 7.92%. Of 474 *GFL* genes, 88.6% decreased UHML when turned on or actively expressed in the developing fibers at the 10 days post-anthesis (10-dpa), whereas only 11.4% increased UHML. The *GFL* genes encode proteins and enzymes that are involved in a variety of biological processes and metabolic pathways. The 474 *GFL* genes interacted to form an interaction network in the 10-dpa fibers, which suggests that UHML is the consequence of interactions among the *GFL* genes. In addition, the evolution of fiber length was examined between diploid and tetraploid cottons using the *GFL* genes. The results showed that the variation of the *GFL* gene networks, including the number of genes and number of gene x gene interactions, also plays an important role in the variation of fiber length during polyploidization.

Therefore, this study has, for the first time worldwide, cloned a large number of genes controlling UHML and deciphered the underlying molecular basis and regulation

mechanisms, thus providing novel resources and knowledge for development of new toolkits for enhanced cotton fiber breeding. The UHML is determined not only by its controlling genes, *GFL*s, but also by their actions, action directions and interactions. Moreover, the results of this study have not only provided a first line of evidence that a quantitative trait is controlled by a large number of genes, but also added new molecular basis, thus forming the molecular mechanisms of quantitative genetics.

DEDICATION

To my father, mother, brother, and Solomon

ACKNOWLEDGEMENTS

I would like to thank my committee chairs, Dr. Hongbin Zhang and Dr. C. Wayne Smith, and my committee members, Dr. Steve Hague and Dr. Joshua Yuan, for their guidance and support throughout the course of this research. Especially, I would like to express my best sincere appreciation to my two major professors. Thanks to Dr. Zhang for providing me so many good opportunities and training me to be a quick learner when facing unfamiliar research areas. I was so grateful for his patience for guiding me toward the right directions in the research. Thanks to Dr. Smith for his invaluable suggestions in the field works and breeding issues. Without their helps, this exciting research project may not be on the right track.

Thanks also to my lab colleagues, Dr. Yang Zhang, Mrs. Chantel Scheuring, and Dr. Meiping Zhang for their help and support throughout the entire research and studies at Texas A&M University. Besides, I would like to thank Mrs. Dawn Deno, former graduate students, and all undergraduate assistants in the Cotton Improvement Lab for field trial experiments. I acknowledge for their assistances in the planting, harvesting, de-burring, ginning, and de-lint. I learned a lot from them while working with them as a team in the ginning lab and the field. Thank you!

TABLE OF CONTENTS

		Page
ABSTRACT		ii
DEDICATION		iv
ACKNOWLEDGEMENTS		v
TABLE OF CONTENTS		vi
LIST OF FIGURES		vii
LIST OF TABLES		x
1. INTRODUCTION		1
2. MATERIALS AND METHODS		11
	2.1 Plant materials and fiber phenotyping	11
	2.2 Transcriptome sequencing and gene digital expression profiling	13
	2.3 Data analysis	18
	2.4 Digital expression profiling of genes actively expressed in the 10-dpa fibers of diploid cottons	22
3. RESULTS		24
	3.1 UHML genetic variation in the RIL population	24
	3.2 Transcriptome sequencing and digital expression profiling	31
	3.3 Isolation and analysis of genes controlling UHML	44
	3.4 Molecular basis of UHML development	76
	3.5 *GFL* genes of diploid cottons and molecular mechanisms of fiber length evolution in the process of cotton polyploidization	124
4. DISCUSSION AND CONCLUSION		143
REFERENCES		149
APPENDIX A		156

LIST OF FIGURES

FIGURE		Page
1	UHML distributions in the TAM 94L-25 x NMSI 1331 RIL population in 2011	25
2	UHML distributions in the TAM 94L-25 x NMSI 1331 RIL population in 2010	26
3	UHML distributions in the TAM 94L-25 x NMSI 1331 RIL population in 2009	26
4	Average UHML distribution in the TAM 94L-25 x NMSI 1331 RIL population during 2009-11	27
5	Correlations of UHMLs of the TAM 94L-25 x NMSI 1331 RIL population between replicates in 2010	28
6	Correlations of UHMLs of the TAM 94L-25 x NMSI 1331 RIL population between replicates in 2011	28
7	Correlations of UHMLs of the TAM 94L-25 x NMSI 1331 RIL population between 2010 and 2011	29
8	Correlations of UHML with other HVI traits in the TAM 94L-25 x NMSI 1331 RIL population in 2011	31
9	Unigene contigs assembled with SOAPdenovo and Trinity	35
10	N50 of unigenes assembled with SOAPdenovo and Trinity	35
11	Correlation of expression levels (TPM) of genes in 10-dpa fibers between plants within a replicate (left) and between replicates (right) of the female parent, TAM 94L-25, in 2011	38
12	Correlation of expression levels (TPM) of genes in 10-dpa fibers of the female parent, TAM 94L-25, between 2010 and 2011	39
13	Distribution of numbers of 100-nuleotide (bp) clean reads derived from RNA-Seq among the 198 RILs and 2 parents	40

FIGURE Page

14 Distribution of Q20 of the RNA-Seq reads with a quality of Q20 or higher among the 198 RILs and 2 parents 40

15 Distribution of GC content of the sequence reads derived from RNA-Seq among the 198 RILs and 2 parents 41

16 Unigenes assembled with Trinity for each of the 198 RILs of the TAM 94L-25 x NMSI 1331 population 42

17 N50 of the unigenes assembled with Trinity for each of the 198 RILs of the TAM 94L-25 x NMSI 1331 population 42

18 GO items of the *GFL* genes assigned to the functional category of biological process (level 2) 58

19 GO items of the *GFL* genes assigned to the functional category of molecular function (level 2) 58

20 GO items of the *GFL* genes assigned to genes the functional category of cellular component (level 2) 72

21 Metabolism-related KEGG pathways in which the *GFL* genes are involved 76

22 The co-regulation network of the 474 *GFL* genes 77

23 Association of the variations of *GFL* gene networks with the UHML of the RILs of the TAM 94L-25 x NMSI 1331 population 89

24 Numbers of unique and shared edges among the *GFL* gene networks of different UHML RIL groups 107

25 The network of the *GFL* orthologous genes in the A-genome diploids 126

26 The network of the *GFL* orthologous genes in the D-genome diploids 126

FIGURE Page

27 Node and edge number variations of the *GFL* gene network from diploids (AA and DD) to tetraploid cottons (AADD)..... 128

28 Variation of numbers of edges among the *GFL* networks of A-, D- and AD-genome species..... 129

LIST OF TABLES

TABLE		Page
1	Fiber UHML (mm) of the RIL population during 2009-2011	24
2	Uniformity index (%) of the RIL population during 2009-2011	25
3	Pearson's correlation of UHMLs between 2009, 2010 and 2011	27
4	Pearson's correlation coefficients of HVI traits	30
5	Preliminary examination of computer programs, Trinity and SOAPdenovo, for *de novo* assembly of 100bp clean reads generated from RNA-seq of the parental lines of the TAM 94L-25 x NMSI 1331 RIL population	34
6	Preliminary examination of assembling unigene reference sequences from the parental lines of the RIL population using Trinity	36
7	Summary of RNA-Seq in 100-nucloetide paired end reads for 198 RILs and 2 parents	39
8	*GFL* genes and their effects and action direction on UHML in cotton	45
9	List of published genes controlling fiber length or trichome development	56
10	Annotation of *GFL* genes controlling UHML in 10-dpa tetraploid cotton species	59
11	KEGG pathways in which the *GFL* genes are involved	74
12	Number of edges that a single *GFL* gene has in the 474 *GFL* network of the entire RIL population	78
13	Number of edges and nodes in the gene x gene networks among groups with different UHMLs	90
14	Pairwise comparison of the nodes of the *GFL* gene networks between the RIL groups with different UHMLs	91

TABLE Page

15 Edge number variation of each gene constituting the *GFL* gene networks among the five fiber-length RIL groups, and its Pearson's correlation coefficients with that of UHML 93

16 Summary of the edges of each gene unique to the *GFL* gene networks of each fiber-length group derived from 474 *GFL* genes, and their Pearson correlation coefficients with of UHML and corresponding number of unique edges of *GFL* genes 108

17 Comparison of the networks of the *GFL* genes among A- and D-genome diploid species and AD-genome tetraploid cottons 127

18 Correlation coefficients ($P \leq 0.05$) of edges shared among the *GFL* networks of tetrapploid (AADD), A-genome diploid (AA) and D-genome diploid (DD) species 130

19 List of nodes (*GFL* genes) that have edges common among the networks of the A-, D- and AD-genome species 137

1. INTRODUCTION

Cotton (*Gossypium* spp.) is the most important natural textile crop and one of the most important oilseed crops in the world. Worldwide cotton production was estimated to be 25.4 million MT during 2013-14, and the top cotton producing countries were China (6.96 million MT), India (6.30 million MT), and United States (USA) (2.87 million MT) (USDA 2014). Cotton yarn and related products generate up to one hundred billion USD per year. In the USA, the annual business revenue, including cotton products and farm related activities, generated from cotton production was 27.62 billion USD in 2011 (NCC 2012). The USA has shifted from a primary consumer of domestic production to a primary exporter. In order for the American cotton farmers to remain competitive in this global market, superior fiber quality will be required, including longer, finer, stronger and more uniform fibers.

The cotton genus, *Gossypium* L., contains 50 species, including 45 diploid and five allotetraploid species. Diploid species consists of A, B, C, D, E, F, G and K genome groups, whereas tetraploid species have genomes of $(AD)_1$ through $(AD)_5$. Wendel and Cronn (2003) estimated that the polyploidization event that led to the birth of the tetraploids from the diploids occurred approximately one to two million years ago (MYA), in which two diploid ancestor genomes, AA and DD, merged into a single nucleus (Wendel 1989; Wendel et al. 1989). Two of the resulting tetraploid species were domesticated, including *G. hirsutum* $(AD)_1$ and *G. barbadense* $(AD)_2$, in North America and two of the A-genome diploids, *G. herbaceum* (A_1) and *G. arboreum* (A_2), were

domesticated in Africa-Asia. The two polyploid cultivated cottons, *G. hirsutum* (Upland cotton) and *G. barbadense* (Pima cotton or Egypt cotton), comprise more than 95% of commercial cultivars in the modern textile industry.

Cotton is also a good model for studying the genome evolution and polyploidization. Two tetraploid cultivated cottons used in this study, *G. hirsutum* and *G. barbadense*, are the classic examples of allopolyploidization. Changes in fiber characteristics that became important to civilized mankind were affected by this natural event. Interestingly, a majority of the loci impacting fiber yield and quality traits of tetraploid cottons was found to be derived from the non-spinnable fiber progenitor, the D-genome diploid species (Jiang et al. 1998). However, the morphological characteristics of modern tetraploid cotton fibers showed more similarities with their spinnable fiber progenitor, the A-genome diploid species. Applequist et al. (2001) used Scanning Electron Microscope (SEM) to study the morphological variations on ovules and identify the developmental differences among cultivated and wild diploid and tetraploid species. Based on their observations, fiber growth curves were similar between tetraploid species and A-genome wild types. The comparative findings of this research suggested a speculation that A-genome ancestor was morphogenetically dominant while combining D-genome in wild tetraploid cottons. This "dominance" did not conclude that genes controlling fiber development in the D-genome were silenced at the fiber elongation stage during polyploidization. Hovav et al (2008) reported that genes in tetraploid cottons were significantly biased toward A-genome or D-genome ancestor during fiber development, and the expressions of these genes was shifted strongly

toward the agronomical inferior donor, D-genome. This finding proposed a possibility that genes from the D-genome potentially contribute or provide raw materials of evolutionary innovations toward the superior fiber characteristics in modern tetraploid cultivated cottons.

Fiber quality is paramount to cotton production and textile industries. It is determined by several parameters, such as length, uniformity, micronaire, strength, elongation, color, and trash. The desired premium fiber quality in textile market requires fibers that are longer, stronger and finer, and have higher uniformity. Among the fiber properties, the fiber length is the most important to textile industries. This is because longer fibers can have a higher market price due to their ease in manufacturing. Meredith and Bridge (1972) found that fiber length and uniformity were controlled by both additive and dominant effects. Additive effects had the predominant influence on lint percentage, seed index, fiber strength, fiber elongation and fiber fineness, whereas dominant effect was responsible for boll size.

UHML is the average length of the longer half of the fibers in the sample under measurement. Of the two allopolyploid cultivated species, *G. barbadense*, also known as Extra Long Staple (ELS), produces longer and finer fibers, with an UHML of 35.05 – 36.58 mm (1.38 - 1.44 inch) (Smith and Cothren 1999), that are better suited for premium textile processes. *G. hirsutum*, known as Upland cotton, has higher fiber yield and wider environment adaptability, with an UHML of 22.86 - 29.46 mm (0.9 - 1.16 inch). Although *G. barbadense* has an outstanding fiber length, it only contributes about 8% of the world's cotton because of its lower yield and narrower adaptability. Therefore,

inter-specific crosses are often practiced between *G. hirsutum* and *G. barbadense* in cotton breeding programs to produce varieties combining high fiber quality with high fiber yield. Besides, the crosses between the two species and their derived populations have been widely used to study the molecular and genetic basis of fiber qualities in cotton (Chen et al. 2012).

Cotton fibers are seed trichomes derived from epidermal cells. They are the longest plant cells known to date, thus providing an excellent model to study cell fate, cell differentiation, and cytoskeleton dynamics. Although the highly elongated seed trichomes in cotton (fiber) are distinct from the leaf trichomes in *Arabidopsis* in cell shape and cellular components upon maturity, both of their trichomes are non-glandular hairs and have homologuous genes or transcription factors controlling the mechanisms underlying trichome initiation and development (Humphries et al. 2005). Therefore, the leaf trichomes of *Arabidopsis* have served as a model to elucidate the mechanisms of cotton fiber differentiations in the early stage of growth. It has been reported that genes that were found to regulate leaf trichome initiation in *Arabidopsis* were also found to regulate the initiation of trichome development and seed numbers bearing on epidermal trichomes in cotton (Guan et al. 2008).

Fiber development is often separated into four distinct, yet overlapping stages, i.e., initiation, elongation, secondary cell wall biosynthesis and maturation (Basra and Malik 1984). In general, the fiber initiation of tetraploid cottons starts at two days before anthesis, and ends at two days post-anthesis (dpa). After the fiber cell starts to develop from the epidermis of seed coat, the primary cell wall becomes soft and expands. During

the elongation stage, which starts approximately at three dpa and ends at approximately 16-dpa, fiber cells undergo the maximum elongation growth rate. Genes encoding cell extension-related enzymes were found to be highly or moderately highly expressed during the elongation stage, and to be down-regulated afterwards (Michailidis et al. 2009; Wang et al. 2010a; Wang and Ruan 2010; Argiriou et al. 2012; Lacape et al. 2012). Between the elongation and secondary cell wall biosynthesis stages, a short transition stage of cell wall remodeling takes place during 16- to 20-dpa. After that, most celluloses are deposited in the secondary cell wall rather than the primary cell wall during 20- to 45-dpa (Meinert and Delmer 1977), thus leading to a rapid gain of weight per fiber. At the maturation stage, the fiber cell undergoes dehydration, thus losing most of its weight. Although it takes nearly 50 days for a fiber cell to mature, the duration of each stage depends on the environmental (e.g. temperature, irrigation, etc.) and genetic factors.

Approximately 4% of a mature fiber cell is composed of primary cell wall that determines fiber length and diameter, and 96% of it is composed of secondary cell wall that determines strength and maturity. The primary cell wall is biosynthesized largely during the elongation stage that often is coupled with vigorous cell expansion, and the majority of fiber length is developed during this period. Cellulose and hemicellulose (xyloglucan) constitute 50 - 60% of dry weight of primary cell wall in elongating fiber cell, whereas pectin contributes nearly 20-30% of dry weight. Protein and phenolic acid make up the rest of elongating fiber cell. Hemicellulose and pectin that are called matrix polysaccharides are synthesized in Golgi cisternae, while cellulose is synthesized in

plasma membrane in a form of para-crystalline microfibrils (Reiter 2002). The structure network consisting of cellulose and hemicellulose that are the major load-bearing elements in the primary cell wall provides the mechanical strength while fiber cell expansion occurs.

The expansion extent of fiber cells is associated with the duration of the elongation stage, plasmodesmata closure, turgor pressure, and cell wall loosening. Given that the elongation stage is the time of the maximum speed of fiber cell expansion, the duration of this period has been found to play an important role in the determination of final length of a mature fiber. It has been documented that *G. barabsense* (Braden and Smith 2004) that often has longer fibers than *G. hirsutum* generally has a prolonged period of the elongation stage compared with short and medium UHML. A prolonged elongation stage had served an important selection factor during cotton domestication in which artificial selections were found to be made toward the prolonged elongation phase in tetraploid and diploid cottons (Applequist et al. 2001). Plasmodesmata that have been recognized to regulate of element transportation/translocation and cell-to-cell communication close at 10-dpa and re-open at 16-dpa. The genes controlling the duration of plasmodesmata closure have been reported to regulate fiber length development (Ruan et al. 2001).

Vacuoles make up 90% of the volume of an elongating fiber cell. Thus, rapid fiber cell expansion is presumed to be driven by the strong turgor pressure that is caused by H^+-ATPase and major osmoticum, including sugar, K^+, and malate (Dhindsa et al. 1975). Increased expressions of sucrose and K^+ transporters that transport soluble sugar

and K^+ from phloem in the vascular tissue to fiber cell were found to lead to an elevation of osmotic pressure (Smart et al. 1998; Ruan et al. 2001). PVP carboxylase that synthesizes malate has the highest expression level at 10- to 12-dpa, and decreases at 15-dpa. The expression level of vacuolar invertase was found to be a positive regulator of fiber length during 10-dpa (Wang et al. 2010c). Several genes that were associated with cell wall loosening, such as xyloglucan endotransferase/hydrolase (Lee et al. 2010), pectate lyase (Wang et al. 2010a), fascin-like arabinogalactan protein (Huang et al. 2013), expansin-related protein (Xu et al. 2012) and actin-binding protein (Wang et al. 2009), were found to have significant impacts on fiber length.

In addition to the inner strength of fiber cells, fiber cell extension usually accompanies changes in the orientation, organization and array of the cytoskeleton, including microfilament (actin) and microtubule (tublin) cytoskeleton. Actin, a major component of microfilament cytoskeleton, plays an important role in fiber elongation, but not in initiation (Li et al. 2005). *GhADF1*, an actin depolymerizing factor (*ADF*) that dissociates actin, was found to suppress fiber length by altering cytoskeleton structure and cellulose contents (Wang et al. 2009). Profilin, an actin-binding protein that plays essential roles in many cellular processes, appeared to play a critical role in constructing microfilament cytoskeleton during fiber elongation. Overexpression of *GhPFN2* caused a pre-terminated elongation and earlier secondary cell wall deposition, thus reducing fiber length (Wang et al. 2010b). Silencing of *KATANIN*, a microtubule protein, resulted in shorter fiber than wild-type cotton, suggesting its essential role in the organization of microtubule cytoskeleton during fiber elongation (Qu et al. 2012).

Fiber length development is regulated by phytohormones and secondary metabolites as well, including ethylene (Shi et al. 2006), gibberellin (Xiao et al. 2010), flavonoid (Tan et al. 2013), and brassinosteroid (Sun et al. 2005; Shi et al. 2006). Exogenous gibberellin (GA) and ethylene have been extensively studied in promoting fiber initiation and elongation in cotton. Recently, a GA 20-oxidase gene (*GhGA20ox1*) was found to improve fiber initiation and elongation by regulating endogenous GA levels in fiber cells (Xiao et al. 2010). The 1-Aminocyclopropane-1-Carboxylic Acid Oxidase (*ACO*) gene responsible for ethylene production was also shown to have regulatory roles in fiber initiation and elongation. Brassinosteroids (BRs), regulators of normal plant growth and development, were reported to be required for fiber initiation and elongation (Sun et al. 2005). Antisense-mediated suppression of *GhDET2* that catalyzes a major rate reduction in BR biosynthesis showed both inhibited fiber initiation and elongation (Luo et al. 2007). Flavonoids, secondary metabolites in the plant kingdom, were found to improve the fiber elongation process (Tan et al. 2013).

Fiber cell expansion during the elongation stage is mediated by several transcription factors. *GaMYB2* encoding a GL1-like MYB protein and expressing in early developing fiber cells rescued the trichome formation when transformed into the *gh1* mutant in *Arabidopsis* and induced seed trichome formation (Wang et al. 2004). Knockdown of *GhMYB109* showed depressed fiber length, providing evidence that this *MYB*-regulated gene is induced before phytohormone biosynthesis and cytoskeleton biosynthesis during fiber initiation and elongation (Pu et al. 2008). *GhMYB25* was found to regulate fiber elongation, trichome development and seed production in cotton

(Machado et al. 2009). TCP transcription factor has the same effect as *GhMYB25* (Hao et al. 2012).

At present, numerous genes differentially expressed among various development stages and between different cultivated species have been studied extensively and reported by using fuzzless-lintless mutants, ESTs (Expressed Sequence Tag), and cDNA microarray hybridizations (Ruan et al. 2001; Ji et al. 2003; Arpat et al. 2004; Shi et al. 2006; Chaudhary et al. 2008; Hovav et al. 2008; Michailidis et al. 2009; Pang et al. 2010; Rapp et al. 2010; Wang and Ruan 2010; Kim et al. 2012; Lacape et al. 2012; Padmalatha et al. 2012; Avci et al. 2013; Li et al. 2013; Liu et al. 2013; Yoo and Wendel 2014). These comparative studies revealed potential key genes and biological pathways involved across different stages in cotton fiber development and cotton genome polyploidization. However, few genes that directly control fiber length during elongation have been cloned and characterized to date. The molecular basis and mechanisms of fiber length development largely remains unknown.

In order to isolate genes controlling fiber length in a more efficient manner, a novel high-throughput gene and quantitative trait locus (QTL) cloning system to clone and characterize genes and QTLs controlling the traits of agronomic importance was developed (patent filed). The objectives of this study were [1] to identify the genes controlling fiber length from cultivated cottons and [2] to use the isolated fiber genes to study the molecular basis and mechanisms of fiber length development. This study has not only cloned 474 genes controlling cotton UHML and allowed systems studies of the molecular basis and mechanisms underlying fiber length development, but also provided

new toolkits enabling development of gene-based breeding system in cotton. This will revolutionize the efficiencies of currently-used plant breeding methods, including marker-assisted breeding.

2. MATERIALS AND METHODS

2.1 Plant materials and fiber phenotyping

2.1.1 Population selection and preparation

Four recombinant inbred line (RIL) populations at the $F_{2:8}$ generation were developed by the Cotton Improvement Laboratory, Department of Soil and Crop Sciences, Texas A&M AgriLife Research, College Station, Texas. These populations were all developed from interspecific crosses between *G. hirsutum* and *G. barbadense*, including TAM 94L-25 x SI Barbados, TAM 94L-25 x NMSI 1331, NUSI 1331 x 97M-16, and NUSI 1331 x 97M-17. To find which of the populations was better suited for this study, all of them, including parents, were planted in a completely random design (CRD) in the Texas A&M Agrilife Research Farm near College Station, Texas in 2009. All cultural practices were common for upland cotton production in central and south Texas.

All mature bolls were hand-harvested from four random plants per line. Lint quality, including UHML, micronaire, strength, uniformity index (UI), and fiber elongation (ELONG) before break in the measurement of strength, was measured by High Volume Instrument (HVI) at the Fiber and Biopolymer Research Institute, Texas Tech University, Lubbock, Texas. Since the RIL population derived from TAM 94L-25 x NMSI 1331 was shown to have the most genetic variation in lint fraction, lint yield, micronaire, UHML, UI, strength, and ELONG, it was selected for this study. To

generate sufficient seeds for field trials, the TAM 94L-25/NMSI 1331 RIL population was planted in a greenhouse (day 82 F/night 75 F) located at the Institute for Plant Genomics and Biotechnology, Norman Borlaug Center for Southern Crop Improvement, Texas A&M University in December, 2009. Seeds were harvested from each of the RILs and parents.

2.1.2 Field trials for fiber phenomics

One hundred ninety-eight RILs were randomly selected from the TAM 94L-25/NMSI 1331 population for this study. In 2010 and 2011, the 198 RILs and their parents, TAM 94L-25, and NMSI 1331, were grown in randomized complete block designs (RCBD) with three replicates at the Texas A&M Agrilife Research Farm near College Station, Texas. The field practices followed those used in standard upland cotton production in Texas. When completely matured, all cotton bolls were hand-harvested from entire plots and ginned in the Cotton Improvement Laboratory, Texas A&M AgriLife Research, College Station, Texas. Cotton seed yield, lint yield and lint percentage data were collected. Fiber quality, including UHML, micronaire, strength, UI, ELONG, was measured by HVI in the Fiber and Biopolymer Research Institute, Texas Tech University, Lubbock, Texas. Therefore, the field trials were conducted, and the RILs and parents were phenotyped in yield and fiber traits for three years, with no replication for 2009 and three replicates for 2010 and 2011. The seeds of the RIL population harvested in 2010 and 2011 were stored in a cold room (4 °C) of the Cotton Improvement Laboratory, Texas A&M AgiLife Research, College Station, Texas.

2.2 Transcriptome sequencing and gene digital expression profiling

2.2.1 RNA isolation and quality check

During cotton growth and development, 10-dpa fibers were collected from all entries of each of the RILs and parents in the 2010 field trial, and 0-dpa ovaries, 10-dpa fibers and 20-dpa fibers were collected from all entries of each of the RILs and parents in the 2011 field trial, utilizing two plants per entry. All tissues were frozen immediately in liquid nitrogen and then stored at -80 °C before use. For this study, only 10-dpa fibers collected from the first replicate of the 2011 field trial and those of the parents from different replicates in both 2010 and 2011 were analyzed. The tissues of 10-dpa fibers were analyzed because this stage of fiber development, as described above, is critical to fiber length development - the focus of this study.

Prior to RNA isolation, the 10-dpa fibers were ground into a fine powder in liquid nitrogen with mortar and pestle. RNA was isolated from each sample, following the manual of Spectrum™ Plant Total RNA Kit (Sigma Aldrich, USA). Briefly, approximately 100 mg of the fine 10-dpa fiber powder was transferred into a 1.5-ml frozen microtube, and 500 µl of the lysis buffer containing 0.1% of β-mercaptoethanol was added and mixed vigorously for 60 seconds. Then, the sample was lysed by incubation in a 56 °C water bath for five minutes. The lysate was filtered by the filtration column (Cat. No C6866, Sigma Aldrich, USA) seated in a 2-ml collection tube and supernatants collected. The filtration column was then discarded, 500 µl of the binding buffer was added to the tube containing the filtered supernatants and mixed thoroughly

by pipetting 6 −10 times. The mixture was transferred into the binding column (Cat. No C6991, Sigma Aldrich, USA) and centrifuged at room temperature and 14,000 rpm for two minutes. After all liquid flowed through the column, the binding column was washed once with 500 µl of the wash solution 1. Any contaminating DNA remaining in the RNA sample in the column was removed by digesting with DNase I (On-column DNase Digest Set, Sigma Aldrich, USA) at room temperature for 15 minutes.

After digestion, the binding column was washed again with 500 µl of wash solution 1 and centrifuged at 14,000 rpm for two minutes. Then, the binding column was washed two times with wash solution 2, 500 µl for each wash, and centrifuged at 14,000 rpm for two minutes. After all liquid flowed through the column, the binding column was transferred into a new 2-ml collection tube, 50 µl of elution buffer was added to the center of the binding column, and incubated at room temperature for two minutes. The total RNA from 10-dpa fiber samples was harvested with the flow-through liquid into the collection tube, and stored at -80°C. The quality and integrity of the isolated RNA was checked by Experion RNA StdSens chips (#700-715, Bio-Rad Laboratories, Inc., USA) and RNA quality indicator (RQI) on the microfluidics-based platforms (Experion Automated Electrophoresis System, Bio-Rad Laboratories, Inc., USA). The RNA with an RQI value of 9.0 or higher was used for further analysis.

2.2.2 Construction of RNA-seq cDNA libraries

The RNA-seq cDNA libraries were constructed from the RNAs using the Illumina TruSeq RNA Sample Prep Kit Version 2.0 - Sets A and B (Illumina, Inc., San

Diego, CA). First, 3 − 4 μg of total RNA was purified twice by using poly-T oligo-attached magnetic beads. An equal volume of completely re-suspended RNA purification beads were added to the total RNA and mixed by gentle but thorough pipetting. To denature the RNA and facilitate binding of the poly-A mRNA to the beads, the samples were incubated in a thermal cycler at 65 °C for five minutes. After cooling at room temperature for five minutes, the sample mixtures were placed in a magnetic stand (Magnetic stand-96, Life Technologies, Cat. No. AM10027) for five minutes to collect the poly-A mRNA-bound beads. The supernatant was discarded, and 67 μl of bead washing buffer was added to each sample to remove the unbound RNA by gently pipetting.

Seventeen microliters of the elution buffer was added to the samples and mixed by gentle pipetting. The mRNA and contaminant rRNA that bound non-specifically were eluted from the beads by incubating at 80 °C for two minutes and then holding at 25 °C. An equal volume of the bead binding buffer (17 μl) was added to the samples to specifically bind mRNA to the beads and reduce the non-specific binding of contaminating rRNA. After the mixture was incubated at room temperature for five minutes, it was placed on the magnetic stand for another five minutes. The supernatant was discarded, and the samples were removed from the magnetic stand and washed with the bead washing buffer again to remove residual rRNA and other contaminants.

The Elute, Prime and Fragment Mix that contains random hexamers for reverse transcription priming and serves as the first strand cDNA synthesis reaction buffer was added to the sample by gentle pipetting. The mRNA was eluted, fragmented and primed

by incubating at 94 °C for eight minutes then holding at 4 °C. After collecting clean primed RNA, the First Strand Master Mix and SuperScript II mix (1:9) was used to reverse-transcribe the mRNA into the first strand of cDNA by the following series of incubations: at 25 °C for 10 minutes, 42 °C for 50 minutes and 70 °C for 15 minutes followed by holding at 4 °C. Then, the Second Strand Master Mix was added and the reaction was incubated at 16 °C for one hour to facilitate the second strand cDNA synthesis.

To capture the double-strand cDNA, AMPure XP beads (Beckman Coulter Genomics, #A63881) were added to the reaction, mixed, and placed on the magnetic stand. After the beads bound with double-strand cDNA were washed twice with fresh 80% ethanol on the magnetic stand, the XP bead-pellets were dried at room temperature for five minutes and then gently mixed in the resuspension buffer. To make the cDNA blunt-ended, End Repair Mix was added to the sample and incubated at 30 °C for 30 minutes. As described above, AMPure XP beads were used again to purify the double-strand cDNA on a magnetic stand. Then, A-Tailing Mix was mixed with the sample and incubated at 37 °C for 30 minutes, 70 °C for five minutes and held at 4 °C to add a single ‘A’ nucleotide to the 3’-end of the end-blunted cDNA fragments.

Twenty-four specific indexing adapters were supplied with the Illumina’s TruSeq RNA Sample Prep Kit. The indexing adapters were ligated to the ends of the double-strand cDNA, with one indexing adaptor per sample, using the ligation mix at 30 °C for 10 minutes. After ligation, ligation stop buffer was added to stop the ligation reaction. AMPure XP beads were used to purify the cDNA with indexing adapters, as described

above. Before sequencing, cDNA fragments were selectively amplified by Polymorphism Chain Reaction (PCR). PCR Primer Cocktail and PCR Master Mix were added to the sample and incubated at 98 °C for 30 seconds, 15 cycles of 98 °C for 10 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds, 72 °C for 5 minutes, and 4 °C for hold. Amplified PCR products were purified with AMPure XP beads twice, as described above.

The quality and concentration of the amplified cDNA library was checked using the Experion 1K DNA Analysis Kit (Bio-Rad, cat #700-7104) on an Experion Automated Electrophoresis System (Bio-Rad, cat #700-7001). cDNA libraries were adjusted to a concentration of 10 nM using the resuspension buffer, and then were sequenced on HiSeq 2000 (Illumina, Inc., San Diego, CA) with a module of 100 PE (100 nucleotide paired ends) by BGI America (formerly Beijing Genomics Institute), Hong Kong. The quality and quantity of each sample were further validated with an Agilent 2100 Bioanalyzer (Agilent Technologies, Austin, TX) by BGI America before sequencing.

2.2.3 Multi-indexing strategy

The cDNA libraries were sequenced on HiSeq 2000 with a multi-indexing strategy using 24 indexes, with each library identified with one index. Therefore, 24 samples were multiplexed and sequenced on a flow cell lane of the HiSeq 2000 sequencer. The index sequences of the 24 adaptors supplied with the Illumina's TruSeq RNA Sample Prep Kits are ATCACG, CGATGT, TTAGGC, TGACCA, ACAGTG,

GCCAAT, CAGATC, ACTTGA, GATCAG, TAGCTT, GGCTAC, CTTGTA, AGTCAA, AGTTCC, ATGTCA, CCGTCC, GTCCGC, GTGAAA, GTGGCC, GTTTCG, CGTACG, GAGTGG, ACTGAT, and ATTCCT. Since the HiSeq 2000 sequencer can run two 8-lane flow cells each time, 384 samples can be sequenced per run.

2.3 Data analysis

2.3.1 Phenomic data in field performance

Fiber traits, including UHML, micronaire, UI, strength, ELONG, seed weight, lint yield and lint percentage, were analyzed using customized R scripts (R 3.0.1). Statistics results were all validated by JMP, a computer program for statistics developed by the JMP business unit of the SAS Institute.

2.3.2 Genotypic data from RNA-seq cDNA library sequencing

2.3.2.1 *de novo* assembly of transcriptome sequences

Library adaptors and nucleotide sequences containing more than 5% unknown nucleotides were removed from the raw sequence data. Low-quality sequencing reads, in which more than 20% of the nucleotides have reading scores of less than Q10 in a read, were also removed from the raw sequence data. Clean reads of each sample were assembled by a RNA-seq data analysis software, Trinity (version of r2013-02-25)

(Grabherr et al. 2011; Haas et al. 2013) on a server with 62 GB of RAM, allocating from eight multicores on a 3168-core IBM (iDataPlex) Linux cluster comprised of Nehalem and Westmere processors, at the Texas A&M Supercomputing Facility (http://sc.tamu.edu/), Texas A&M University, College Station, TX.

Due to lack of reference genome information for cultivated tetraploid cotton, the clean sequence reads of HiSeq 10-dpa fiber libraries of six biological replicates of the maternal parent, TAM 94L-25, harvested in 2011 were combined, assembled into a unigene set and used as a reference for gene expression profiling. A total of 85,755,752 100-nucleotide clean reads were used for the assembly. Based on the draft genome of *G. raimondii*, the full-length cDNAs of genes had an average of 2,485 bp (Wang et al. 2012). If tetraploid cottons have about 40,000 genes and approximately 60% of them are expressed in 10-dpa fibers, then, the 85,755,752 100-nucleotide clean reads would represent more than a 140X coverage of all the genes expressed in 10-dpa fibers.

2.3.2.2 Expression level estimation

The expression level of each gene in 10-dpa fibers of each RIL was estimated by the abundance of 100-nucleotide reads among the Trinity transcripts of the TAM 94L-25 reference sequences. The original HiSeq 2000 clean reads of each of the 198 RILs and the parents were aligned to the TAM 94L-25 reference sequences using the modified version of RSEM (Li and Dewey 2011) within Trinity. RSEM is a software package that is widely used to accurately and digitally quantify gene expressions from RNA-Seq data with or without a reference genome.

2.3.3 Isolation of genes controlling UHML

The genes controlling UHML was isolated using a genome-wide high-throughput gene and QTL cloning and studying system recently developed in our laboratory. The system has currently been submitted to the Texas A&M AgriLife Innovation Management Office for disclosure and filing for a patent. It is based on the hypothesis that the genetic variation of a phenotypic trait results from gene sequence mutation, variation of gene expression, variation of gene action mode, variation of gene x gene interactions, variation of gene x non-gene element interactions such as related small RNA activities, and/or gene x environment interactions. Using this system, we have successfully cloned 1,253 genes controlling maize grain yield and 606 genes controlling chickpea vernalization and flowering. This system not only determines the function(s) of the genes, as do the current gene/QTL cloning methods, but also constructs the gene networks controlling the trait, determines their molecular basis and regulation mechanisms, and develops toolkits enabling gene-based breeding. Using the system, hundreds of genes controlling different agronomic traits can be cloned within a few scientist-years. The throughput of the new system is >1,000-fold higher than those of the current gene/QTL cloning methods such as map-based cloning, gene mutagenesis and RNA inference (RNAi), thus making it possible to large-scale and rapidly clone and molecularly characterize the genes controlling UHML.

The new gene and QTL cloning system isolates genes controlling a quantitative trait through four critical steps, with a cutoff for each step, although a single such step may be sufficient for isolation of genes controlling a quantitative trait according to the

currently-used gene and QTL cloning methods such as map-based cloning, gene mutagenesis and RNA inference (RNAi). Therefore, the genes isolated using the new gene and QTL cloning system has a confidence of at least 99.999375%, which is four-fold higher than those of the currently-used gene and QTL cloning methods. The genes isolated to control UHML were identified in this study as *Gossypium* fiber length (*GFL*) gene. An Arabic numeral suffix was added to indicate the order of genes isolated, such as *GFL001*, meaning the first UHML gene isolated from *Gossypium*.

2.3.4 Systems analysis of the *GFL* genes controlling UHML

2.3.4.1 Annotation and ontology of the *GFL* genes

The unigenes that were found to control UHML described above were blasted and annotated against the NCBI non-redundant protein database (the version as of September 30, 2013) using a minimal E-value cutoff of 1.0 E-06. The unigenes that were assigned to the same Genbank number and showed overlapping and identical sequences were first merged as a single gene and defined as a *GFL* gene. The *GFL* genes were then subjected to analyses of GO (Gene Ontology, a controlled vocabulary to describe gene product characteristics and gene product annotation data), gene-encoding enzyme functions and involved metabolism pathways against the InterProScan database and KEGG using blast2GO (Conesa et al. 2005; Conesa and Gotz 2008; Gotz et al. 2008; Gotz et al. 2011).

2.3.4.2 Gene network construction

To determine whether there is any relationship in regulation and thus function among the *GFL* genes, the expression profiles of the *GFL* genes in 10-dpa fibers were subjected to correlation analysis by customized R scripts on Spearman's rank correlation using the entire RIL population. The visualization of the gene regulation interaction or network was performed by BioLayout *Expression*^{3D} (Theocharidis et al. 2009).

2.4 Digital expression profiling of genes actively expressed in the 10-dpa fibers of diploid cottons

To determine the evolution of the molecular mechanisms underlying UHML during cotton polyploidization, 10-dpa fibers were collected from diploid cotton species, including *G. herbaceum* (A_1), *G. arboreum* (A_2), *G. thurberi* (D_1), *G. davidsonii* (D_3), *G. aridum* (D_4), *G. raimondii* (D_5), and *G. gossypioides* (D_6), at the Crop Germplasm Unit of USDA/ARS, College Station, TX. The same procedures as those described above were used to isolate RNAs from the tissues and to construct and index the RNA-seq cDNA libraries. The indexed RNA-seq cDNA libraries were sequenced on HiSeq 2000 by BGI America, Hong Kong.

Clean reads of the RNA-seq sequences of diploid cottons were aligned to the *GFL* genes isolated from the tetraploid cottons using the RSEM to quantify the expression levels of the genes in diploid cottons. The diploid orthologues of the *GFL* genes were identified and the network of the orthologues was constructed, as described

above for the network construction of the *GFL* genes in tetraploid cotton using the BioLayout *Expression*^{3D}.

3. RESULTS

3.1 UHML genetic variation in the RIL population

The UHML of the TAM 94L-25 x NMSI 1331 RIL population was phenotyped by field trials over three years (Table 1). In 2009, the population had an average UHML of 28.5 mm, ranging from 23.62 mm to 33.53 mm with a variation of 41.96%. In 2010, the field trial had three replicates and the population averaged an UHML of 28.61 mm, ranging from 23.11 mm to 35.05 mm with a variation of 51.67%. In 2011, the trial also included three replicates, and the population exhibited an average UHML of 28.83 mm, with a range from 23.03 mm to 34.80 mm and a variation of 51.10%. Moreover, the fiber length of the population was quite uniform, with a UI of greater than 80% over all three years of the field trials (Table 2; Figs 1 – 4).

Table 1 Fiber UHML (mm) of the RIL population during 2009-2011

	2011				**2010**				**2009**
	Rep1	Rep2	Rep3	**Average**	Rep1	Rep2	Rep3	**Average**	**Average**
Min	21.59	23.11	23.62	**23.03**	23.11	23.11	23.62	**23.11**	**23.62**
1st Quantile	27.69	27.18	27.18	**27.43**	26.92	27.43	26.92	**27.18**	**27.18**
Median	29.21	28.19	28.70	**28.70**	28.45	28.70	28.19	**28.62**	**28.45**
3rd Quantile	30.73	30.23	29.97	**30.14**	30.23	29.97	29.97	**29.97**	**29.97**
Max	35.56	34.54	34.80	**34.80**	35.05	32.51	34.04	**35.05**	**33.53**
Range	13.97	11.43	11.18	**11.77**	11.94	9.40	10.41	**11.94**	**9.91**
Mean	29.22	28.56	28.66	**28.83**	28.57	28.68	28.43	**28.61**	**28.48**
SD	2.30	2.24	2.13	**2.02**	2.27	1.99	2.13	**2.12**	**1.95**
Skew	0.01	0.14	0.15	**0.12**	0.18	-0.33	0.18	**0.16**	**-0.07**
Kurtosis	0.27	-0.19	-0.04	**0.10**	-0.14	-0.25	-0.34	**0.03**	**-0.40**
SE	0.16	0.16	0.15	**0.14**	0.20	0.17	0.19	**0.16**	**0.14**

Table 2 Uniformity index (%) of the RIL population during 2009-2011

	2011				2010				2009
	Rep1	Rep2	Rep3	**Average**	Rep1	Rep2	Rep3	**Average**	**Average**
Min	76.20	73.00	72.90	**75.23**	75.40	76.70	76.10	**76.75**	**73.30**
1st Quantile	80.75	79.50	79.70	**80.20**	80.20	80.25	79.58	**79.97**	**79.50**
Median	82.10	81.30	81.40	**81.57**	81.30	81.40	80.95	**81.00**	**81.00**
3rd Quantile	83.40	82.50	82.80	**82.62**	82.40	82.55	82.10	**82.20**	**81.90**
Max	86.70	85.50	85.60	**85.23**	84.40	85.70	84.70	**84.80**	**84.80**
Range	10.50	12.50	12.70	**10.00**	9.00	9.00	8.60	**8.05**	**11.50**
Mean	81.93	80.96	81.16	**81.36**	81.05	81.24	80.76	**80.94**	**80.63**
SD	2.07	2.14	2.15	**1.77**	1.81	1.77	1.79	**1.63**	**1.99**
Skew	-0.51	-0.51	-0.64	**-0.54**	-0.71	-0.36	-0.19	**-0.33**	**-0.57**
Kurtosis	0.01	0.26	0.31	**0.28**	0.20	-0.16	-0.45	**-0.30**	**0.51**
SE	0.15	0.16	0.15	**0.12**	0.16	0.15	0.16	**0.13**	**0.14**

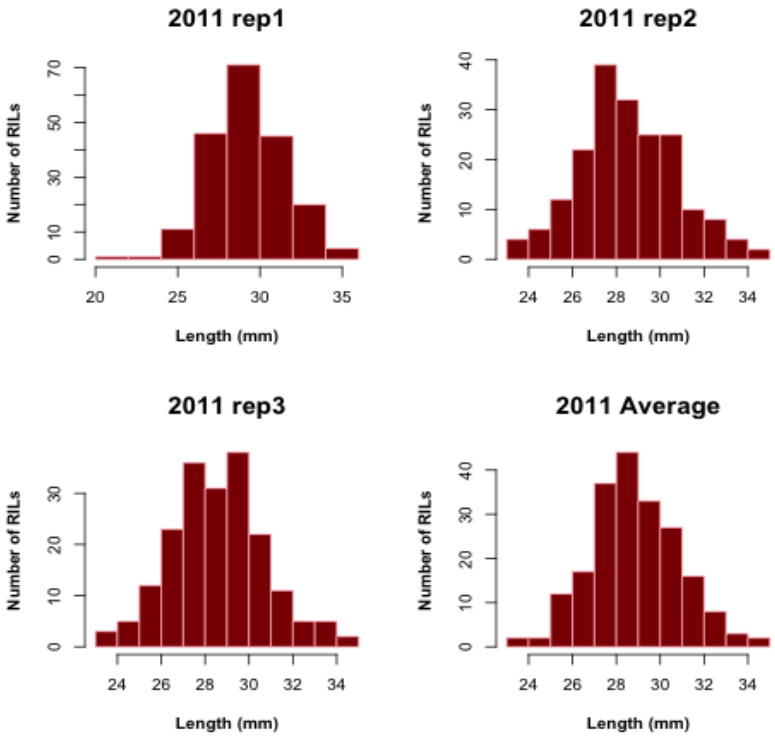

Figure 1 UHML distributions in the TAM 94L-25 x NMSI 1331 RIL population in 2011

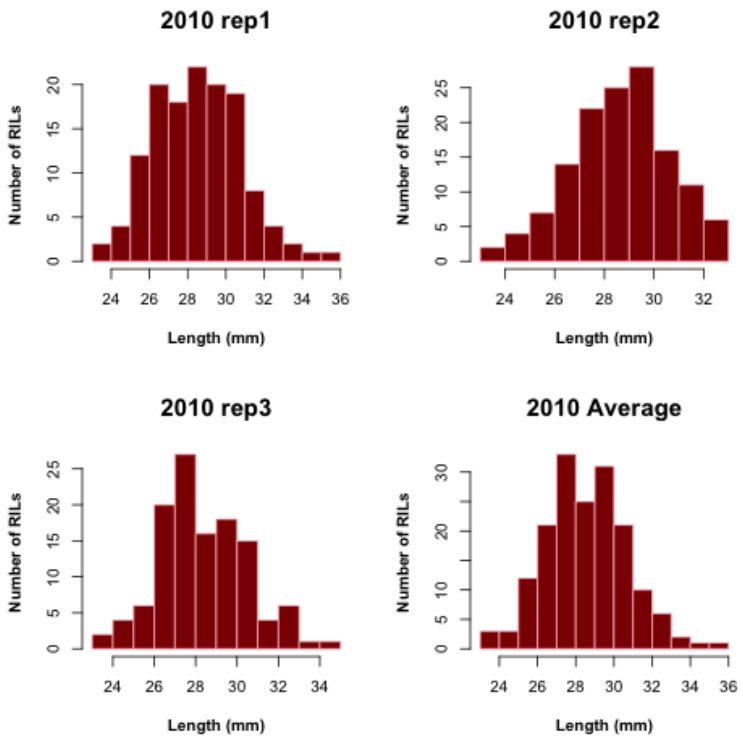

Figure 2 UHML distributions in the TAM 94L-25 x NMSI 1331 RIL population in 2010

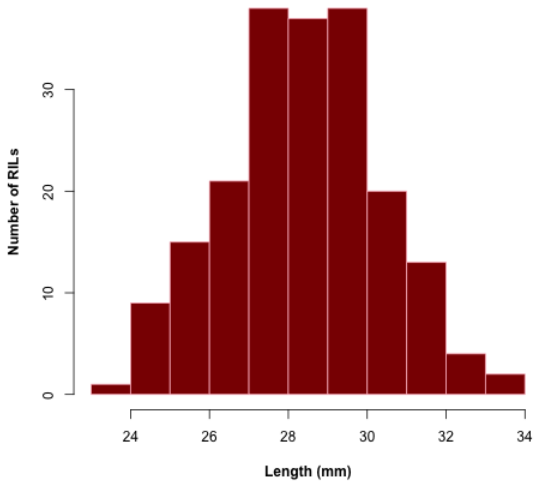

Figure 3 UHML distributions in the TAM 94L-25 x NMSI 1331 RIL population in 2009

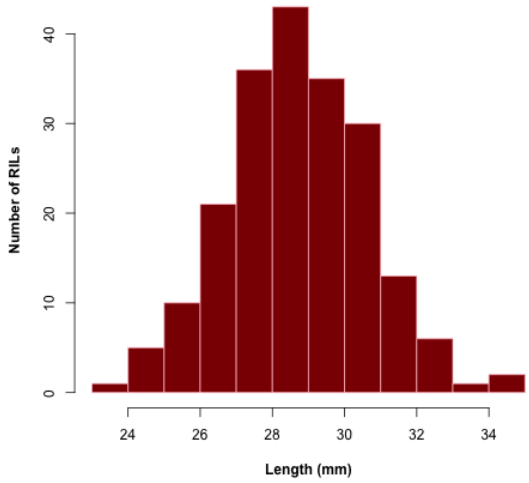

Figure 4 Average UHML distribution in the TAM 94L-25 x NMSI 1331 RIL population during 2009-11

To check the reproducibility of the fiber phenomic results, we calculated the correlations of the fiber traits, especially UHML, between replicates and between years. Significant correlations of UHMLs were observed between replicates within a year (Figs. 5 – 6) with a correlation coefficient of 0.80 – 0.85 ($P \leq 0.05$) in 2010 and of 0.76 ($P \leq 0.05$) in 2011, and between years, with a correlation coefficient of 0.67 – 0.96 ($P \leq 0.05$) (Table 3 and Fig. 7). These results indicated that the UHLM of the RIL population studied in this research was reproducible within and between years.

Table 3 Pearson's correlation of UHMLs between 2009, 2010 and 2011. "*" indicates $P \leq 0.05$.

Year	2009	2010	2011	3yr-aver
2009	1			
2010	0.72*	1		
2011	0.67*	0.91*	1	
3yr-aver	0.87*	0.96*	0.94*	1

The asterisk "* " indicates $P \leq 0.05$.

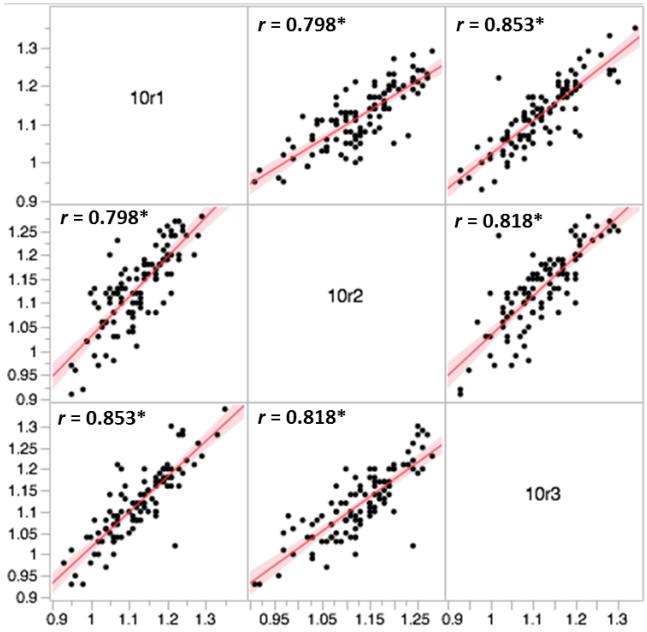

Figure 5 Correlations of UHMLs of the TAM 94L-25 x NMSI 1331 RIL population between replicates in 2010. "*" indicates $P \leq 0.05$.

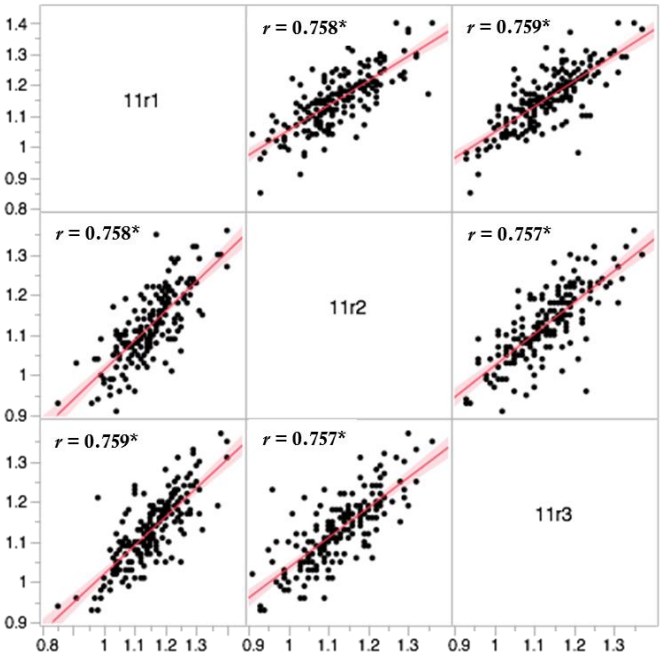

Figure 6 Correlations of UHMLs of the TAM 94L-25 x NMSI 1331 RIL population between replicates in 2011. "*" indicates $P \leq 0.05$.

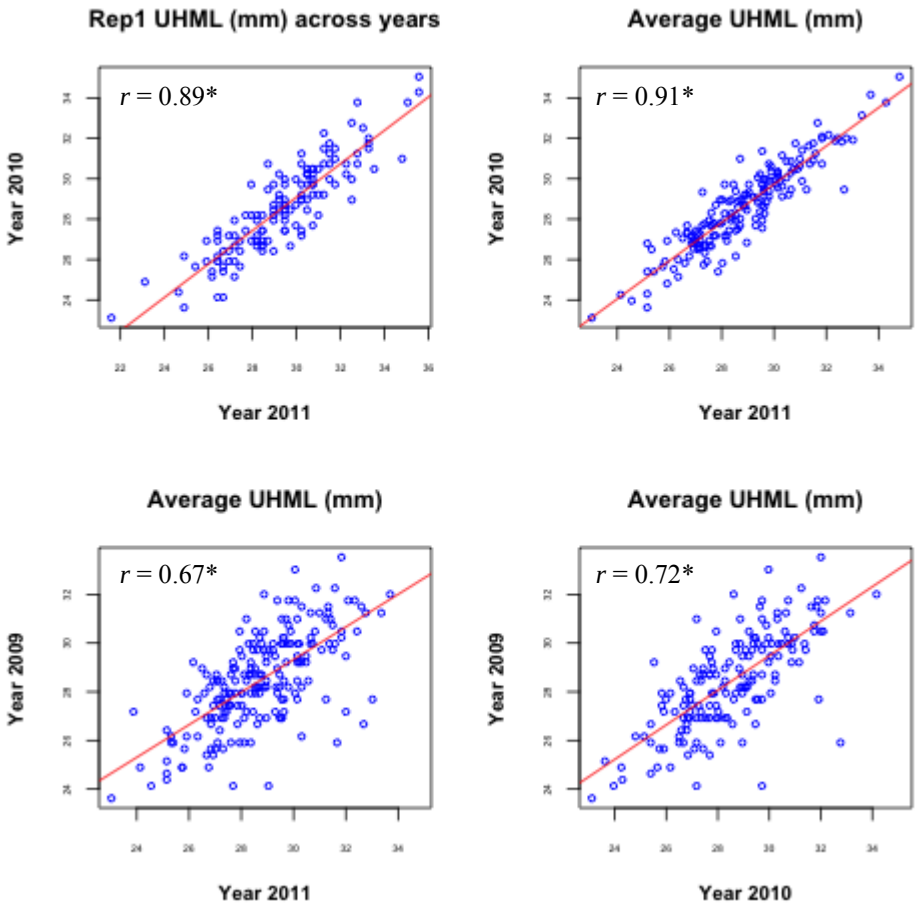

Figure 7 Correlations of UHMLs of the TAM 94L-25 x NMSI 1331 RIL population between 2010 and 2011. "*" indicates $P \leq 0.05$.

Pearson's correlation test was also used to test the association relationship between UHML and other quality and yield traits, including micronaire, UI, strength, ELONG, cotton seed yield, cotton lint yield, and lint percentage. The values used in the analyses were the average values of three replicates in 2011 among TAM 94L-25, NMSI 1331 and their RILs (Table 4). In TAM 94L-25, UHML significantly correlated with strength, UI, and ELONG, whereas UHML significantly correlated with UI only in

NMSI 1331. In the RIL population, UHML showed similar correlation relationships with fiber strength, UI, and ELONG as their upland cotton parent, TAM 94L-25, while it showed significant and negative correlations with micronaire and lint yield percentage (Fig. 8). These results agreed with the findings of Ulloa and Meredith (2000) and Karademir et al. (2010). Both of them reported strong positive correlations between UHML and strength, and negative correlations between UHML and micronaire. Hence, the results suggest that the UHML of this population could be improved simultaneously with strength and uniformity by UHML selection, while lint percentage and micronaire may be reduced.

Table 4 Pearson's correlation coefficients of HVI traits. (a) TAM 94L-25 (n = 6), (b) NMSI 1331 (n = 4), and (c) the RIL population (n = 198) in 2011. LP, Lint Percentage; MIC, Micronaire; UHML, Upper half mean length; STR, Strength; UI, Uniformity index; ELONG, Elongation. (*) indicates $P \leq 0.05$, (**) indicates $P \leq 0.01$, and (***) indicates $P \leq 0.001$ at the two-tailed test.

a. TAM 94L25

	LP	MIC	UHML	STR	UI
MIC	0.367*				
UHML	-0.042	0.110			
STR	-0.024	0.066	0.694***		
UI	0.200	0.385*	0.546**	0.560***	
ELONG	0.111	-0.061	-0.433*	-0.342	-0.097

b. NMSI 1331

	LP	MIC	UHML	STR	UI
MIC	0.495**				
UHML	0.127	-0.015			
STR	0.403*	0.923***	0.263		
UI	0.133	0.479**	0.517**	0.596***	
ELONG	-0.059	0.263	-0.266	0.286	0.133

c. 2011 RILs

	LP	MIC	UHML	STR	UI
MIC	0.627***				
UHML	-0.338***	-0.316***			
STR	-0.027	0.197**	0.554***		
UI	0.110	0.301***	0.570***	0.804***	
ELONG	0.252***	0.356***	-0.373***	0.05	0.512

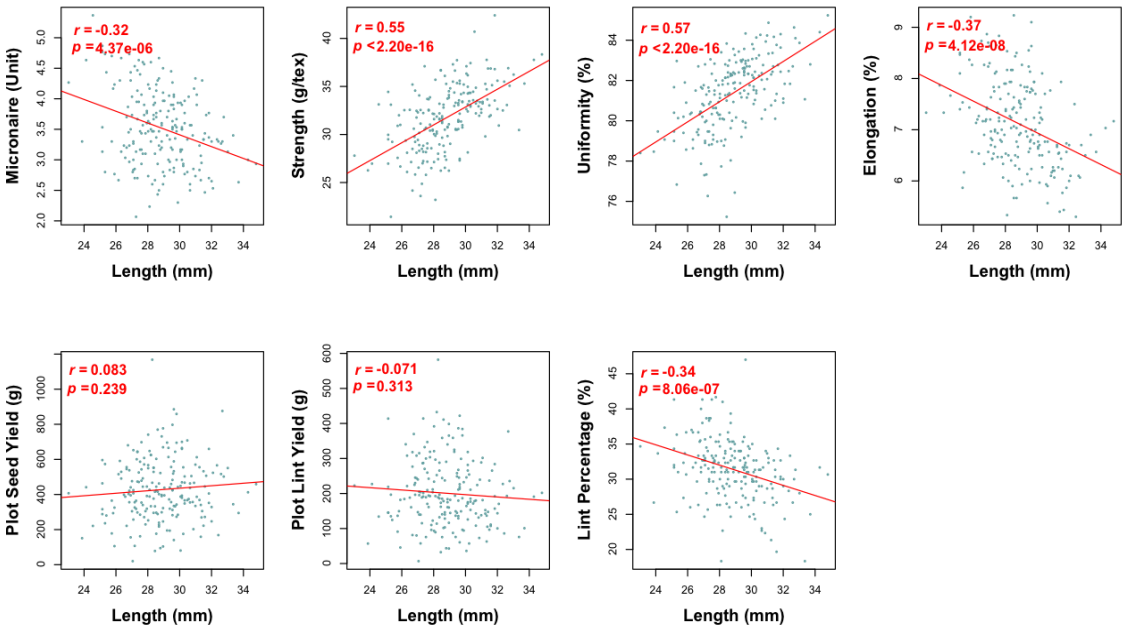

Figure 8 Correlations of UHML with other HVI traits in the TAM 94L-25 x NMSI 1331 RIL population in 2011

3.2 Transcriptome sequencing and digital expression profiling

3.2.1 Development of strategies for sequencing and expression profiling of the genes expressed in 10-dpa fibers of the RIL population

To isolate the genes controlling the UHML in the RIL population using our newly-developed genome-wide high-throughput gene and QTL cloning system, we need both the nucleotide sequences and expression profiles of all genes expressed in developing fibers of the population. We carried out the field trial for three years, with a total of seven replicates. The question was whether we should sequence the nucleotide sequences and profile the expressions of all the genes expressed in every replicate or just one of them. To answer the questions, we first sequenced and profiled the genes

expressed in 10-dpa fibers collected from eight plants of the maternal parent, TAM 94L-25 and four plants of the paternal parent, NMSI 1331 using a 100 PE (100-nucleotide paired end) sequencing module. Of the eight TAM94L-25 plants, three were collected from Rep 1 and three from Rep 2 in 2011, and two from Rep 1 in 2010. Of the four NMSI 1331 plants, two were from Rep 1 in 2011 and two from Rep 1 in 2010. A number between 11.5 million to 15.3 million 100-bp clean reads was obtained for each biological replicate of the parents. Clean reads were defined as the sequence reads whose adaptor sequences have been removed and that have less than 5% of unknown nucleotides and less than 20% of nucleotides having a sequence quality of Q10, a base calling error rate of one out of 10 times of calling.

Several computer programs specifically designed for RNA-seq assembly using short nucleotide sequence reads have been developed and available to public. To determine which of the programs is better for assembling 100-nucleotide clean reads generated from transcriptome sequencing in this study into unigene contigs, we tested two programs that have been widely used for such research, Trinity and SOAPdenovo (Table 5). The clean reads of RNA-seq from eight biological replicates of the RIL population parental lines were assembled, individually, for this experiment. Although the Trinity program can directly assemble the 100-nucleotide (bp) paired-end clean reads into unigene contigs, the SOAPdenovo program cannot do so; it can only assemble 31-, 63- or 127-nucleotide clean reads. Therefore, the 100-nucleotide paired-end clean reads of each sample were randomly clipped into 63-nucleotide short sequences and then assembled together when using SOAPdenovo. The results showed that in all assemblies,

Trinity assembled many fewer and much larger unigene contigs than SOAPdenovo. Either the mean size or N50 size of the unigene contigs assembled by Trinity was much larger than those assembled by SOAPdenovo (Figs. 9 and 10). N50 is determined as the size of the contigs when the sum of the larger contigs reaches half of the total length of the assembled contigs. Wang et al. (2012) reported that *G. raimondii*, a relative of the D subgenome of the tetraploid cottons, has an average transcript size of 2,485 bp. Therefore, the mean length and N50 size of unigene contigs assembled by SOAPdenovo appeared to be far smaller than the average transcript length of *G. raimondii*. On the other hand, because Trinity could assemble the 100-nucleotide clean reads of our RNA-seq into much larger unigene contigs and provide a well-organized downstream analysis and data visualization, it was selected and used for unigene assembly in this study.

Due to the absence of reference genome information for cultivated tetraploid cotton, unigene contigs were assembled from the clean reads of six biological replicates of the maternal parent, TAM 94L25, and four biological replicates of the paternal parent, NMSI 1331, grown in 2011. The unigenes were assembled from the clean reads that combined the clean reads from 1, 3, 4, and 6 biological replicates of the parents (Table 6). The result showed that more clean reads were combined and used, more unigene contigs were assembled, giving longer unigene contigs, larger N50 and N90 sizes.

Table 5 Preliminary examination of computer programs, Trinity and SOAPdenovo, for *de novo* assembly of 100bp clean reads generated from RNA-seq of the parental lines of the TAM 94L-25 x NMSI 1331 RIL population

Parent line	TAM 94L-25	TAM 94L-25	NMSI 1331	NMSI 1331	TAM 94L-25	TAM 94L-25	NMSI 1331	NMSI 1331
Year	2011	2011	2011	2011	2010	2010	2010	2010
Replicate	1	1	1	1	1	1	2	2
Entry ID	1080	1220	1150	1210	1060	1120	2030	2050
Clean reads (100 PE)	13772920	13390508	15291936	13264046	12157416	12482222	11473044	13731484
SOAPdenovo								
No. of contigs	120526	118520	110109	102507	118606	101141	99209	98018
Mean length (nucloetides)	229	301	297	304	303	301	302	301
Max length (nucloetides)	3156	3453	6178	4825	3623	10141	3476	4793
N50 (nucloetides)	322	325	321	325	325	324	323	326
N90 (nucloetides)	190	191	189	192	192	191	192	190
Min length (nucloetides)	100	100	100	100	100	100	100	100
Trinity								
No. of contigs	78338	76282	70897	68401	79293	67557	67463	61916
Mean length (nucloetides)	785.75	803.97	806.25	774.72	791.26	740.08	757.63	759.67
Max length (nucloetides)	9261	10898	24022	17387	10509	11847	8558	8183
N50 (nucloetides)	1223	1283	1263	1224	1249	1143	1183	1190
N90 (nucloetides)	326	325	326	312	323	307	312	313
Min length (nucloetides)	201	201	201	201	201	201	201	201

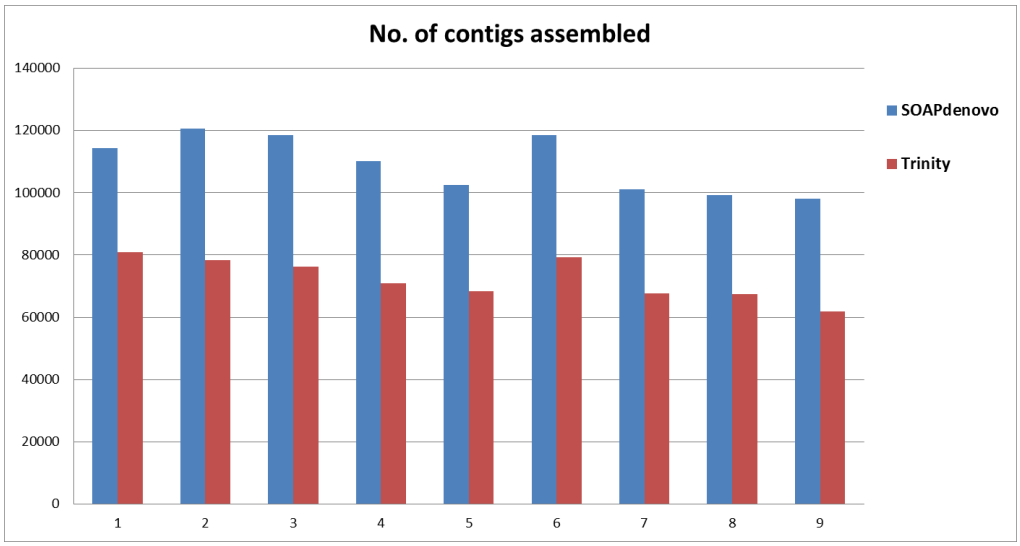

Figure 9 Unigene contigs assembled with SOAPdenovo and Trinity. The x-axial shows the biological replicates of the TAM 94L-25 x NMSI 1331 RIL population parental lines in 2011 and 2010: 1, RIL-1146; 2, 1080; 3, 1220; 4, 1150; 5, 1210; 6, 1060 (2010); 7, 1120 (2010); 8, 2030 (2010); 9, 2050 (2010).

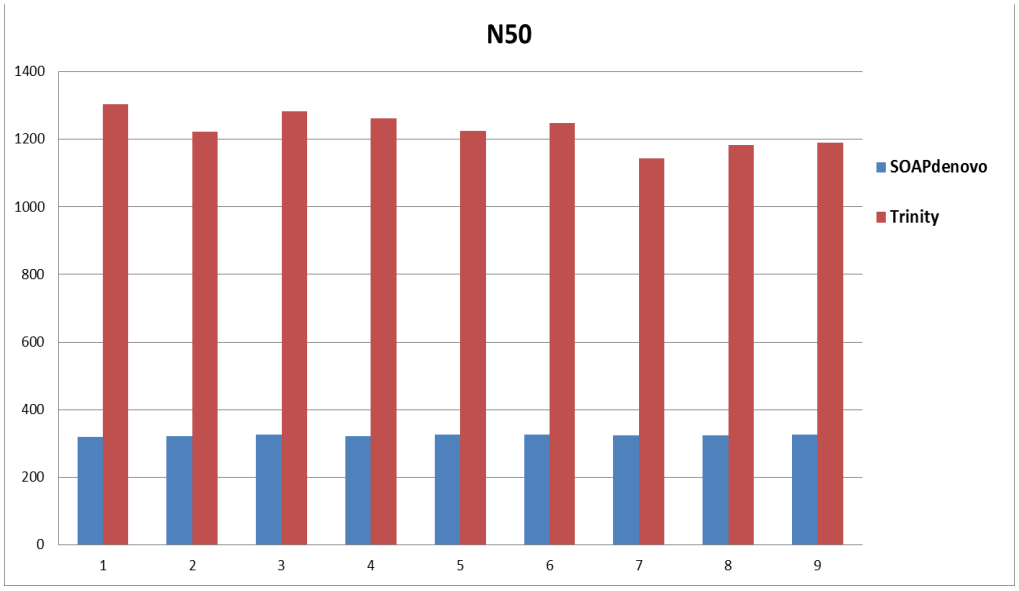

Figure 10 N50 of unigenes assembled with SOAPdenovo and Trinity. The x-axial shows the biological replicates of the TAM 94L-25 x NMSI 1331 RIL population parental lines in 2011 and 2010: 1, RIL-1146; 2, 1080; 3, 1220; 4, 1150; 5, 1210; 6, 1060 (2010); 7, 1120 (2010); 8, 2030 (2010); 9, 2050 (2010).

Therefore, the unigenes assembled from six replicates of the maternal parent, TAM 94L-25, were selected and used to be the unigene reference sequences in this study. A total of 85,755,752 100-nucleotide clean reads were generated from the six biological replicates of the maternal parent, TAM 94L-25. Assembly of these reads by Trinity resulted in a total of 159,936 unigenes, with a N50 of 1,746 bp and an average length of 1,071.6 bp, ranging from 201 bp to 19,507 bp.

Table 6 Preliminary examination of assembling unigene reference sequences from the parental lines of the RIL population using Trinity

	No. of sample assembled	No. of clean reads at each end	No. of unigenes	Average length	Max length	Min length	N50	N90
TAM 94L-25	x1	6,886,460	78,338	786	9,261	201	1,223	326
	x3	19,668,055	117,999	751	9,917	201	1,116	319
	x6	42,877,876	159,936	1,072	19,507	201	1,746	451
NMSI 1331	x1	7,618,243	77,624	787	9,261	201	1,265	316
	x4	28,882,178	132,986	1,018	29,210	201	1,685	416

To answer the question of how many biological replicates of each RIL or parent should be used to profile the expression abundances of the genes expressed in 10-dpa fibers, the clean reads of each biological replicate was aligned to the reference unigene sequences assembled from the six biological replicates of the maternal parent, TAM 94L-25 and the expression level of each gene presented in transcript per million reads (TPM) were calculated using Trinity. The expression levels of each biological replicate were compared between different plants within a replicate, between replicates within the same year, and between years (2010 and 2011) using those of the maternal parent, TAM 94L-25. The result showed that there were significant and strong association between

plants within a field trial replicate ($r = 0.9713$, $P < 0.0001$), between field trial replicates within the same year ($r = 0.9813$, $P < 0.0001$) and between years ($r = 0.9636$, $P < 0.0001$) (Figs. 11 and 12). This result indicated that it was acceptable to profile the expression of the genes expressed in 10-dpa fibers just for one biological replicate for a parent or an RIL.

Therefore, the first replicate (Rep 1) in the 2011 field trial was selected for sequencing and profiling the expression abundances of the genes expressed in 10-dpa fibers of the entire RIL population. This replicate had the widest genetic variation among all seven replicates in UHML (Table 1). It had an average UHML of 29.22 mm, with a range from 21.59 mm to 35.56 mm and a standard deviation of 2.30 mm ($CV = 7.87\%$). The average fiber uniformity of the population was 81.93% in 2011, with a range from 76.2% to 86.7%. These results indicated that it had sufficient fiber length variations for the study.

3.2.2 Transcriptome sequencing and expression profiling of the genes expressed in 10-dpa fibers of the entire RIL population

The genes expressed in 10-dpa fibers of all 198 RILs were sequenced and profiled in expression using the same 100 PE sequencing module and the same multiplex method as those used above for their parents. A total of 2.686 billion of 100-nucleotide paired-end RNA-Seq clean reads were obtained for the entire TAM 94L-25 x NMSI 1331 RIL population (Table 7). An average of 13.23 million 100-nucleotide paired-end

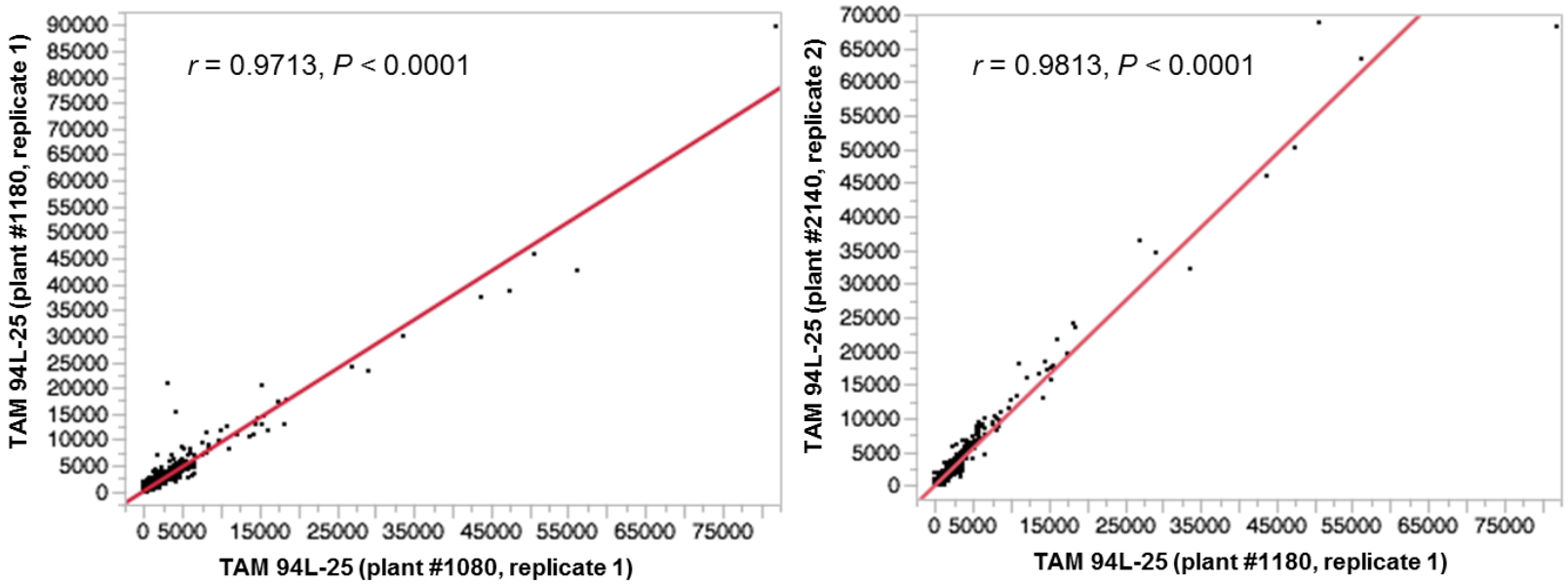

Figure 11 Correlation of expression levels (TPM) of genes in 10-dpa fibers between plants within a replicate (left) and between replicates (right) of the female parent, TAM 94L-25, in 2011

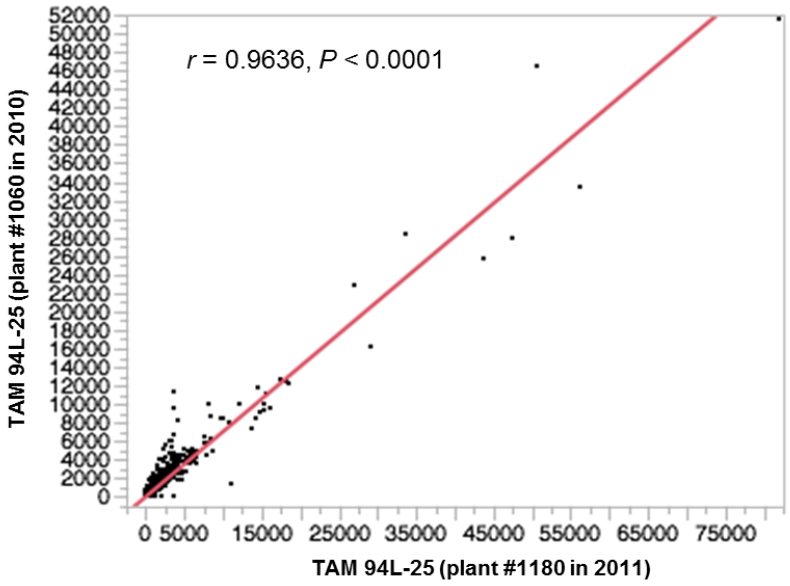

Figure 12 Correlation of expression levels (TPM) of genes in 10-dpa fibers of the female parent, TAM 94L-25, between 2010 and 2011

clean reads were sequenced for each RIL, with a range from 7 million to 19 million of clean reads (Fig. 13 and Table 7). The Q20, which represents a base calling error rate of one out of 100 times of calling, averaged 98.74%, with a range from 96.43% to 99.07% (Table 7 and Fig. 14). The clean reads of the RNA-seq had an average GC content of 44.21%, ranging from 43.13% to 50.61% among the 198 RILs and two parents (Table 7 and Fig. 15).

Table 7 Summary of RNA-Seq in 100-nucloetide paired end reads for 198 RILs and 2 parents. TAM 94L-25 had three biological replicates and NMSI 1331 had two biological replicates

	Total	Mean	Maximum	Minimum
Number of Clean Reads	2686403478	13233514.67	19044152	6608566
Q20		98.74	99.07	96.43
QC (%)		44.21	50.61	43.13

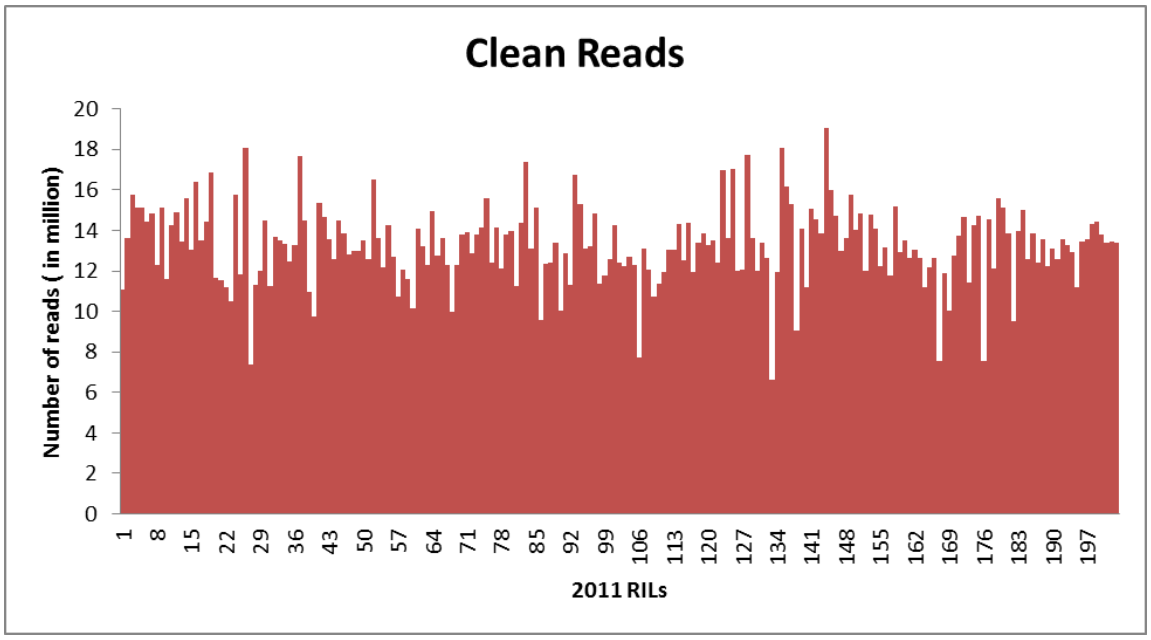

Figure 13 Distribution of numbers of 100-nuleotide (bp) clean reads derived from RNA-Seq among the 198 RILs and 2 parents

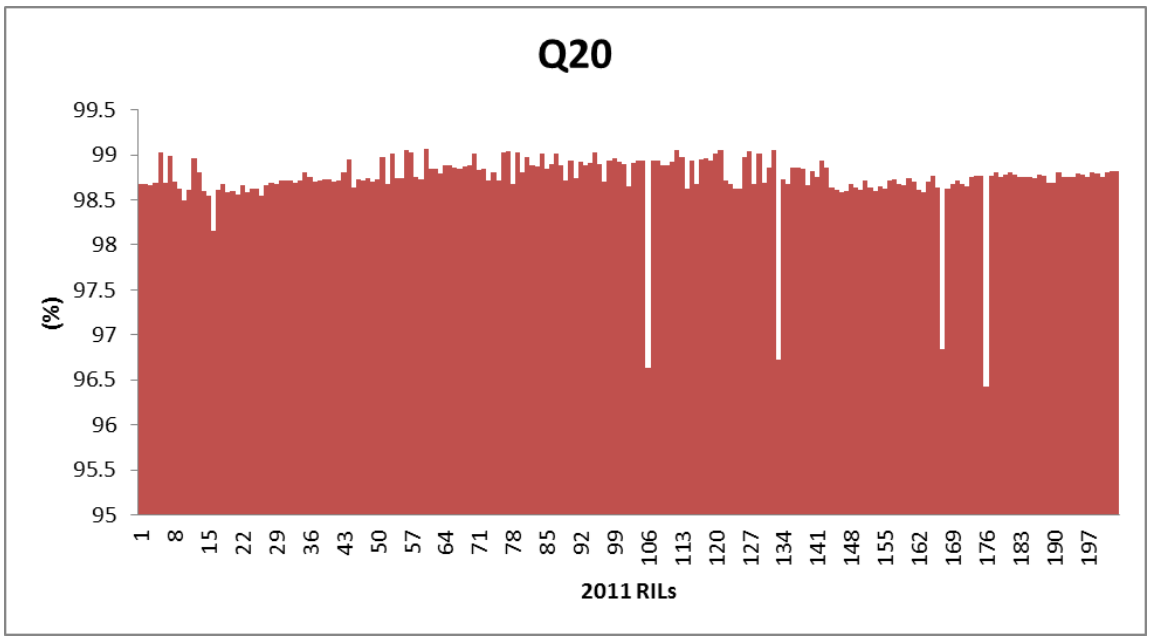

Figure 14 Distribution of Q20 of the RNA-Seq reads with a quality of Q20 or higher among the 198 RILs and 2 parents

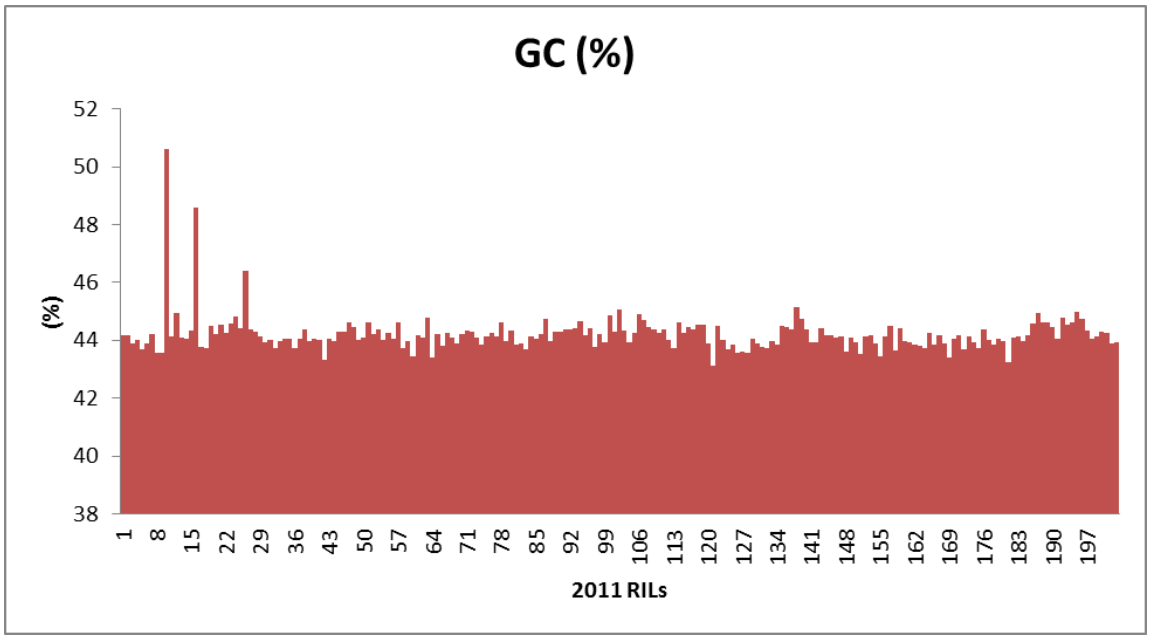

Figure 15 Distribution of GC content of the sequence reads derived from RNA-Seq among the 198 RILs and 2 parents

3.2.3 *de novo* assembly of 100-nucleotide clean reads for RILs

Unigene contigs were assembled from 100-nucleotide clean reads of the RNA-seq with the program Trinity for the 198 RILs derived from an interspecific cross between TAM 94L-25 and NMSI 1331. A total of 14.86 million unigene contigs were assembled for the RILs, with an average of 73,185 unigene contigs for each RIL. The average length of assembled unigenes for each RIL was 778.17 bp, with an average N50 size of 1,215 bp, ranging from 201 bp to 14,378 bp (Figs 16 and 17).

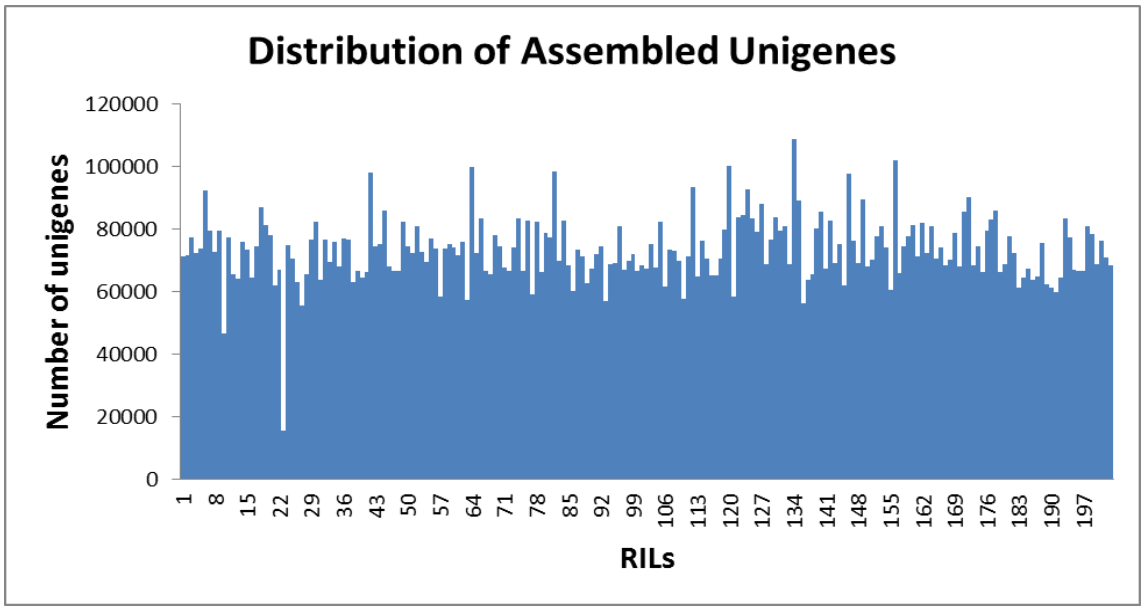

Figure 16 Unigenes assembled with Trinity for each of the 198 RILs of the TAM 94L-25 x NMSI 1331 population

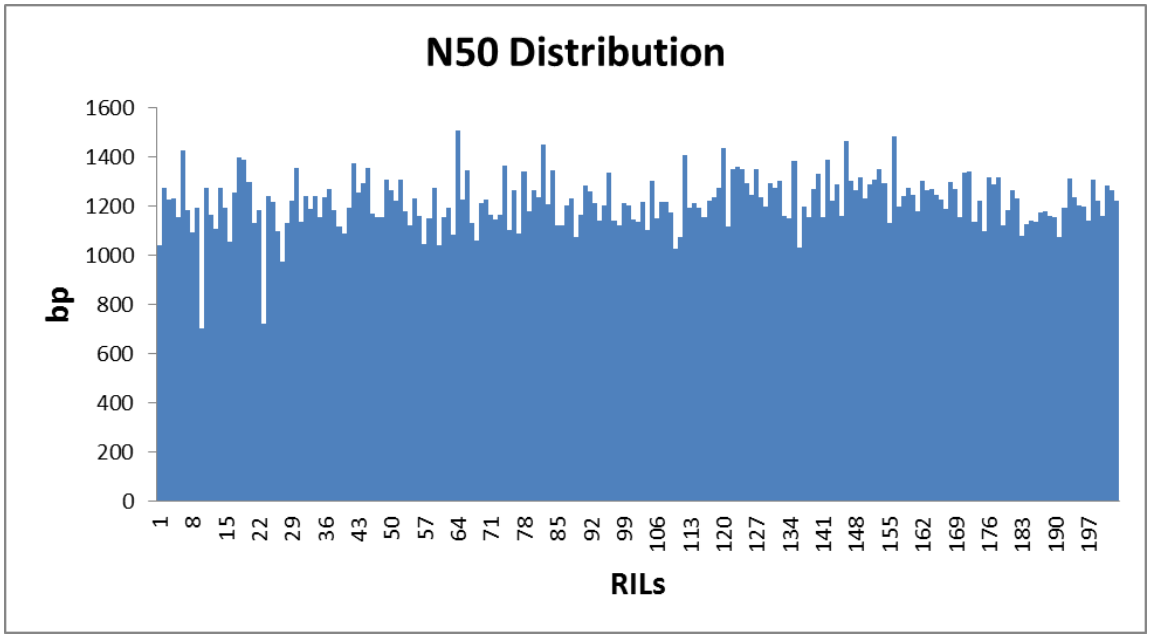

Figure 17 N50 of the unigenes assembled with Trinity for each of the 198 RILs of the TAM 94L-25 x NMSI 1331 population

3.2.4 Expression of each gene in 10-dpa fibers of the RILs and parents

To estimate the expression level of each gene expressed in 10 dpa fibers and their expression variations among the 198 RILs, the clean reads of each RIL or parent was aligned to the reference unigene sequences assembled from the six biological replicates of the maternal parent, TAM 94L-25, using Trinity. The expression profile of each gene, presented in number of transcripts per million clean reads (TPM), was obtained. The result showed that the expression of a large number of the genes expressed in 10-dpa fibers varied dramatically among different RILs. The largest expression variations were found at the unigene "comp30157_c0_seq1", with a maximum TPM of 360,800, a minimum TPM of 410, and a standard deviation of 27,936.34 across different RILs. It was also observed that 2,576 unigenes, accounting for 1.6% of the reference unigenes, had no expression levels detected across RILs. The reasons of this are probably that [1] expression levels of these unigenes might be too few, so that they were automatically excluded from TPM calculation formula in Trinity, and/or [2] most of short reads of different RILs were aligned into the gaps between these unigenes and others, thus giving no TPM values in these unigenes. Of the 159,936 reference unigenes assembled from the six TAM 94L-25 biological replicates, 18,919 expressed in the 10-dpa fibers of all the 198 RILs and 79,708 expressed in the 10-dpa fibers of more than 100 of the RILs. In order to maximize the results of this study, the unigenes expressed in the 10-dpa fibers of more than 100 RILs were used for further analysis of this study.

3.3 Isolation and analysis of genes controlling UHML

A total of 482 unigenes controlling UHML were isolated, with a confidence of 99.9994% ($P \leq$ 6E-5), using our new genome-wide high-throughput gene and QTL cloning system. After those that shared identical sequences with an overlap of more than 20 nucleotides were merged, 474 independent genes resulted (Table 8). These genes were named as the *GFL* genes with Arabic number suffixes 001 through 474 such as *GFL001* for the first gene in the list. To further verify the *GFL* genes, an extensive literature search was performed to find the genes previously cloned that have been shown to control UHML or trichomes. As a result of the search, a total of 21 genes were reported in the literature as controlling UHML or trichomes (Table 9). When these genes were searched against the list of the 474 *GFL* genes isolated in this study, 9 (42.86%) also were found in our 474 *GFL* genes. However, when the confidence of gene isolation was reduced to 99.75% ($P \leq 0.0025$), 15 (71.43%) of the 21 genes were in the list of the UHML genes isolated in this study. When the confidence of gene isolation was reduced to 95.00% ($P \leq 0.05$), all (100%) of the 21 genes were in the list of the UHML genes isolated in the process of fiber gene isolation in this study. This result has further confirmed the reliability of the *GFL* genes isolated in this study.

Table 8 *GFL* genes and their effects and action direction on UHML in cotton

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL001	0.00	29.11	3.00	26.81	3.00	-2.31	-7.92
GFL002	0.00	29.86	6.05	27.87	6.05	-1.99	-6.67
GFL003	0.00	29.18	3.58	27.24	3.58	-1.94	-6.64
GFL004	0.00	28.87	4.25	27.00	4.25	-1.87	-6.47
GFL005	0.00	29.30	4.60	27.43	4.60	-1.87	-6.37
GFL006	0.00	29.23	4.75	27.37	4.75	-1.86	-6.35
GFL007	0.04	29.73	56.57	27.87	56.53	-1.86	-6.26
GFL008	202.45	29.54	535.78	27.70	333.33	-1.84	-6.22
GFL009	3.63	30.18	45.37	28.31	41.74	-1.86	-6.17
GFL010	65.89	29.67	252.71	27.84	186.82	-1.83	-6.16
GFL011	5.80	29.79	38.38	27.97	32.58	-1.83	-6.13
GFL012	0.65	29.17	7.76	27.39	7.11	-1.79	-6.12
GFL013	2.35	29.69	11.54	27.88	9.19	-1.81	-6.10
GFL014	1.00	29.73	41.76	27.92	40.77	-1.81	-6.09
GFL015	0.00	28.80	3.91	27.06	3.91	-1.74	-6.04
GFL016	11.03	29.33	99.33	27.56	88.30	-1.76	-6.00
GFL017	0.83	29.92	8.96	28.15	8.13	-1.77	-5.93
GFL018	34.32	30.02	137.10	28.25	102.78	-1.77	-5.88
GFL019	1279.37	29.87	2365.35	28.11	1085.98	-1.76	-5.88
GFL020	42.88	29.79	101.65	28.05	58.77	-1.74	-5.83
GFL021	1.75	29.80	10.52	28.07	8.77	-1.73	-5.82
GFL022	2.46	29.43	18.09	27.73	15.64	-1.70	-5.77
GFL023	4.38	29.89	20.15	28.18	15.77	-1.71	-5.72
GFL024	0.00	29.42	68.84	27.74	68.84	-1.68	-5.70
GFL025	18.56	29.16	44.11	27.52	25.55	-1.65	-5.65
GFL026	76.73	29.09	166.91	27.46	90.18	-1.64	-5.62
GFL027	4.34	29.93	127.18	28.25	122.85	-1.68	-5.60
GFL028	0.62	29.43	7.88	27.79	7.26	-1.65	-5.60
GFL029	387.85	29.50	958.77	27.87	570.92	-1.63	-5.52
GFL030	73.21	29.59	174.55	27.96	101.34	-1.63	-5.50
GFL031	2.25	29.32	10.25	27.72	8.01	-1.60	-5.45
GFL032	16.78	29.70	108.31	28.08	91.54	-1.62	-5.45
GFL033	0.51	29.02	8.26	27.45	7.75	-1.57	-5.42
GFL034	4.95	29.33	20.22	27.74	15.27	-1.59	-5.42
GFL035	0.67	29.37	7.08	27.78	6.41	-1.59	-5.41
GFL036	0.50	29.34	17.58	27.76	17.08	-1.58	-5.39
GFL037	17.04	29.62	56.56	28.04	39.52	-1.59	-5.35
GFL038	8.82	28.97	33.52	27.42	24.70	-1.55	-5.34
GFL039	30.25	29.57	133.95	27.99	103.70	-1.58	-5.33
GFL040	16.53	29.91	59.61	28.31	43.08	-1.60	-5.33
GFL041	140.44	29.43	310.19	27.87	169.75	-1.56	-5.30
GFL042	1.47	29.43	14.57	27.87	13.11	-1.56	-5.29
GFL043	7.56	29.78	44.72	28.21	37.16	-1.57	-5.29
GFL044	0.00	29.01	3.61	27.47	3.61	-1.53	-5.29
GFL045	15.53	29.68	163.86	28.11	148.33	-1.57	-5.28
GFL046	42.97	29.67	129.71	28.11	86.74	-1.56	-5.25
GFL047	450.78	29.63	996.19	28.08	545.41	-1.55	-5.24
GFL048	24.69	29.40	78.39	27.86	53.70	-1.54	-5.23
GFL049	250.94	29.26	597.10	27.73	346.16	-1.53	-5.22
GFL050	17.51	29.30	43.33	27.78	25.83	-1.53	-5.21

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL051	59.73	29.74	165.37	28.20	105.64	-1.54	-5.18
GFL052	18.84	29.56	54.22	28.04	35.38	-1.52	-5.14
GFL053	42.44	29.27	105.85	27.77	63.41	-1.50	-5.14
GFL054	0.62	29.24	10.32	27.74	9.70	-1.50	-5.13
GFL055	7.37	29.56	20.95	28.07	13.58	-1.49	-5.04
GFL056	0.71	29.31	9.64	27.83	8.93	-1.47	-5.03
GFL057	25.91	29.90	142.14	28.40	116.23	-1.50	-5.02
GFL058	2.14	29.44	18.38	27.97	16.24	-1.48	-5.02
GFL059	0.00	29.23	3.50	27.76	3.50	-1.46	-5.01
GFL060	4.28	29.54	45.63	28.06	41.35	-1.48	-5.00
GFL061	14.42	29.56	45.71	28.09	31.29	-1.48	-4.99
GFL062	1.02	29.36	11.03	27.90	10.01	-1.47	-4.99
GFL063	9.98	29.08	65.48	27.63	55.49	-1.45	-4.99
GFL064	11.32	29.69	42.43	28.21	31.11	-1.48	-4.99
GFL065	16.31	29.50	46.11	28.03	29.79	-1.47	-4.98
GFL066	0.64	29.13	6.57	27.68	5.93	-1.45	-4.98
GFL067	15.29	29.55	85.86	28.08	70.57	-1.47	-4.96
GFL068	54.31	29.47	107.66	28.01	53.36	-1.46	-4.96
GFL069	12.88	29.02	46.95	27.59	34.07	-1.43	-4.94
GFL070	0.00	29.10	33.28	27.66	33.28	-1.44	-4.94
GFL071	0.67	29.19	6.02	27.76	5.35	-1.44	-4.92
GFL072	0.00	29.20	4.58	27.77	4.58	-1.44	-4.92
GFL073	28.83	29.45	196.14	28.01	167.31	-1.44	-4.90
GFL074	0.00	29.24	7.41	27.81	7.41	-1.43	-4.90
GFL075	25.99	29.57	63.97	28.12	37.98	-1.45	-4.90
GFL076	0.37	28.99	23.90	27.57	23.53	-1.42	-4.89
GFL077	1.31	29.67	14.42	28.22	13.11	-1.45	-4.89
GFL078	40.68	29.24	106.90	27.81	66.22	-1.43	-4.88
GFL079	234.19	29.61	485.82	28.16	251.64	-1.45	-4.88
GFL080	105.54	29.47	216.75	28.03	111.22	-1.44	-4.88
GFL081	0.37	29.64	8.59	28.19	8.22	-1.44	-4.87
GFL082	46.21	29.52	149.58	28.08	103.37	-1.44	-4.87
GFL083	1.52	29.26	8.92	27.84	7.40	-1.42	-4.86
GFL084	18.87	29.42	68.58	28.00	49.70	-1.42	-4.83
GFL085	131.36	29.92	250.14	28.47	118.78	-1.44	-4.83
GFL086	0.14	29.53	27.32	28.11	27.18	-1.42	-4.82
GFL087	4.01	29.29	22.43	27.88	18.43	-1.41	-4.81
GFL088	0.00	29.25	32.58	27.85	32.58	-1.40	-4.80
GFL089	3.47	29.56	28.98	28.15	25.50	-1.42	-4.80
GFL090	7.96	29.37	38.39	27.96	30.43	-1.41	-4.79
GFL091	22.37	29.47	118.51	28.06	96.14	-1.41	-4.78
GFL092	2.27	29.53	12.43	28.12	10.16	-1.41	-4.77
GFL093	0.00	29.19	3.50	27.80	3.50	-1.39	-4.76
GFL094	32.45	29.61	84.04	28.21	51.59	-1.41	-4.75
GFL095	1.46	29.61	9.61	28.21	8.15	-1.41	-4.75
GFL096	15.81	29.61	72.67	28.20	56.86	-1.40	-4.74
GFL097	10.33	29.63	63.53	28.23	53.20	-1.40	-4.72
GFL098	11.38	29.12	87.85	27.74	76.47	-1.37	-4.72
GFL099	48.53	29.03	133.69	27.67	85.16	-1.37	-4.71
GFL100	24.83	29.43	58.10	28.05	33.27	-1.38	-4.71

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL101	0.00	29.10	47.97	27.73	47.97	-1.37	-4.70
GFL102	42.76	29.58	536.37	28.19	493.61	-1.39	-4.70
GFL103	80.38	29.35	197.32	27.97	116.94	-1.38	-4.70
GFL104	0.00	29.02	4.34	27.66	4.34	-1.36	-4.70
GFL105	0.59	29.19	10.85	27.82	10.26	-1.37	-4.69
GFL106	47.71	29.40	253.18	28.02	205.47	-1.38	-4.69
GFL107	0.00	28.98	3.56	27.62	3.56	-1.36	-4.68
GFL108	15.82	29.56	59.18	28.18	43.37	-1.38	-4.67
GFL109	5.43	29.43	31.33	28.05	25.90	-1.37	-4.67
GFL110	33.41	29.41	80.49	28.04	47.08	-1.37	-4.66
GFL111	0.00	29.09	5.93	27.74	5.93	-1.35	-4.66
GFL112	0.00	28.99	4.83	27.65	4.83	-1.35	-4.65
GFL113	0.00	29.29	25.76	27.93	25.76	-1.36	-4.65
GFL114	3.86	29.16	18.31	27.80	14.45	-1.35	-4.64
GFL115	70.03	29.17	153.91	27.81	83.88	-1.35	-4.64
GFL116	0.00	29.03	15.07	27.69	15.07	-1.34	-4.62
GFL117	167.48	29.35	463.65	27.99	296.16	-1.36	-4.62
GFL118	0.07	29.61	21.68	28.25	21.61	-1.37	-4.62
GFL119	0.00	29.34	77.42	27.99	77.42	-1.35	-4.60
GFL120	410.29	29.43	751.09	28.08	340.80	-1.35	-4.58
GFL121	84.81	29.51	247.61	28.16	162.80	-1.35	-4.58
GFL122	0.67	29.20	6.61	27.87	5.94	-1.33	-4.57
GFL123	75.11	29.05	163.11	27.73	88.00	-1.32	-4.56
GFL124	58.80	29.53	120.20	28.19	61.40	-1.35	-4.55
GFL125	0.45	29.61	58.65	28.27	58.20	-1.35	-4.55
GFL126	51.71	29.31	132.98	27.98	81.28	-1.33	-4.55
GFL127	18.36	29.30	54.86	27.97	36.50	-1.33	-4.54
GFL128	4.73	29.55	23.52	28.21	18.79	-1.33	-4.51
GFL129	25.43	29.43	90.07	28.10	64.64	-1.33	-4.51
GFL130	252.46	29.44	550.00	28.12	297.54	-1.33	-4.50
GFL131	0.07	29.49	53.32	28.16	53.25	-1.33	-4.50
GFL132	0.00	29.23	12.25	27.92	12.25	-1.31	-4.49
GFL133	35.55	29.70	162.78	28.37	127.22	-1.33	-4.49
GFL134	0.00	29.03	4.48	27.73	4.48	-1.30	-4.48
GFL135	12.87	29.67	48.48	28.34	35.60	-1.33	-4.48
GFL136	0.00	29.33	12.03	28.02	12.03	-1.31	-4.47
GFL137	0.00	29.03	3.95	27.73	3.95	-1.30	-4.47
GFL138	18.81	29.27	46.82	27.97	28.01	-1.31	-4.46
GFL139	291.00	29.56	590.31	28.24	299.31	-1.32	-4.46
GFL140	1.58	29.55	10.74	28.23	9.16	-1.32	-4.46
GFL141	0.00	28.97	6.33	27.68	6.33	-1.29	-4.46
GFL142	1.55	29.39	9.53	28.08	7.98	-1.31	-4.45
GFL143	4.70	29.18	17.92	27.88	13.23	-1.30	-4.45
GFL144	0.29	29.24	49.35	27.94	49.06	-1.30	-4.45
GFL145	24.70	29.34	71.70	28.04	47.00	-1.30	-4.44
GFL146	0.73	29.20	6.88	27.91	6.14	-1.30	-4.44
GFL147	426.90	29.19	1019.35	27.89	592.45	-1.30	-4.44
GFL148	11.51	29.30	93.73	28.00	82.22	-1.30	-4.43
GFL149	0.00	29.06	3.40	27.77	3.40	-1.29	-4.42
GFL150	108.35	29.92	231.87	28.59	123.52	-1.32	-4.42

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL151	2.31	29.36	17.38	28.07	15.07	-1.29	-4.41
GFL152	6.99	29.73	27.24	28.43	20.25	-1.31	-4.40
GFL153	276.41	29.37	511.83	28.08	235.42	-1.29	-4.40
GFL154	0.00	29.36	10.49	28.07	10.49	-1.29	-4.40
GFL155	42.56	29.56	107.70	28.26	65.15	-1.30	-4.39
GFL156	5.22	29.84	29.32	28.53	24.10	-1.31	-4.38
GFL157	48.97	29.38	117.05	28.09	68.08	-1.29	-4.38
GFL158	77.31	29.19	173.16	27.91	95.86	-1.28	-4.38
GFL159	43.10	29.60	85.28	28.31	42.18	-1.29	-4.37
GFL160	0.00	29.38	8.79	28.09	8.79	-1.28	-4.37
GFL161	0.00	29.19	32.40	27.91	32.40	-1.27	-4.37
GFL162	0.81	29.16	34.94	27.88	34.13	-1.27	-4.37
GFL163	125.24	29.41	232.18	28.12	106.93	-1.28	-4.36
GFL164	5.75	29.32	31.53	28.05	25.77	-1.28	-4.35
GFL165	124.57	29.57	238.78	28.28	114.21	-1.29	-4.35
GFL166	37.28	29.60	78.40	28.32	41.13	-1.29	-4.34
GFL167	3.92	29.39	34.59	28.12	30.68	-1.28	-4.34
GFL168	19.07	29.63	64.77	28.34	45.70	-1.29	-4.34
GFL169	103.52	29.63	282.65	28.34	179.13	-1.29	-4.34
GFL170	88.80	29.85	175.32	28.56	86.52	-1.29	-4.32
GFL171	78.32	29.71	156.53	28.43	78.20	-1.28	-4.32
GFL172	0.00	29.11	4.95	27.85	4.95	-1.26	-4.32
GFL173	71.75	29.58	187.66	28.31	115.91	-1.28	-4.32
GFL174	7.73	29.95	27.04	28.66	19.31	-1.29	-4.30
GFL175	0.00	29.80	59.35	28.52	59.35	-1.28	-4.29
GFL176	0.19	29.34	47.64	28.09	47.46	-1.25	-4.28
GFL177	5.64	29.51	36.06	28.25	30.41	-1.26	-4.27
GFL178	10.42	29.13	58.30	27.89	47.88	-1.24	-4.27
GFL179	6.76	29.25	26.97	28.00	20.20	-1.25	-4.26
GFL180	33.83	29.19	108.60	27.95	74.77	-1.24	-4.26
GFL181	119.85	29.46	373.13	28.20	253.27	-1.25	-4.26
GFL182	155.26	29.41	375.98	28.16	220.72	-1.25	-4.26
GFL183	1.93	29.36	17.31	28.11	15.38	-1.25	-4.26
GFL184	0.00	29.31	27.10	28.06	27.10	-1.24	-4.24
GFL185	22.30	28.92	68.63	27.70	46.34	-1.22	-4.24
GFL186	0.04	29.38	28.42	28.14	28.38	-1.24	-4.23
GFL187	87.17	29.44	196.20	28.20	109.02	-1.24	-4.23
GFL188	11.39	29.65	95.53	28.39	84.14	-1.25	-4.22
GFL189	20.25	29.51	148.92	28.27	128.67	-1.24	-4.21
GFL190	3.28	29.33	13.74	28.10	10.46	-1.23	-4.21
GFL191	0.00	29.06	4.58	27.83	4.58	-1.22	-4.21
GFL192	13.65	29.22	33.61	27.99	19.97	-1.23	-4.20
GFL193	36.46	29.05	109.99	27.83	73.53	-1.22	-4.20
GFL194	60.38	28.99	262.31	27.77	201.93	-1.22	-4.20
GFL195	2.95	29.55	24.71	28.31	21.76	-1.24	-4.20
GFL196	18.04	29.96	209.02	28.70	190.98	-1.26	-4.20
GFL197	1.48	29.38	8.18	28.15	6.70	-1.23	-4.20
GFL198	6.43	28.96	23.52	27.74	17.10	-1.22	-4.20
GFL199	32.08	29.51	82.13	28.27	50.04	-1.24	-4.19
GFL200	8.69	29.84	67.68	28.59	58.99	-1.25	-4.19

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL201	126.55	29.59	340.56	28.36	214.02	-1.24	-4.19
GFL202	3.19	29.39	15.73	28.16	12.55	-1.23	-4.18
GFL203	8.12	29.20	26.93	27.98	18.82	-1.22	-4.18
GFL204	5.45	29.55	23.60	28.31	18.15	-1.23	-4.18
GFL205	44.97	29.45	133.31	28.22	88.34	-1.23	-4.17
GFL206	53.09	29.59	192.17	28.36	139.08	-1.23	-4.16
GFL207	9.11	29.37	102.61	28.15	93.51	-1.22	-4.16
GFL208	0.00	28.98	5.80	27.78	5.80	-1.21	-4.16
GFL209	6.30	29.34	30.53	28.12	24.23	-1.22	-4.16
GFL210	98.08	29.23	227.47	28.02	129.39	-1.22	-4.16
GFL211	61.03	29.34	371.95	28.12	310.93	-1.22	-4.15
GFL212	0.62	29.00	9.69	27.80	9.07	-1.20	-4.15
GFL213	0.60	29.22	8.21	28.01	7.61	-1.21	-4.14
GFL214	2.93	29.03	12.20	27.83	9.27	-1.20	-4.13
GFL215	18.48	29.63	58.75	28.41	40.27	-1.22	-4.13
GFL216	1.17	28.99	28.14	27.80	26.97	-1.20	-4.13
GFL217	179.24	29.73	401.61	28.51	222.37	-1.23	-4.12
GFL218	1.47	29.29	14.81	28.08	13.34	-1.21	-4.12
GFL219	1.94	29.78	40.26	28.55	38.31	-1.23	-4.12
GFL220	42.31	29.36	138.36	28.16	96.05	-1.21	-4.12
GFL221	9.69	29.46	35.73	28.25	26.04	-1.21	-4.11
GFL222	243.41	29.34	842.44	28.14	599.04	-1.21	-4.11
GFL223	9.28	29.38	38.79	28.17	29.51	-1.21	-4.11
GFL224	5.16	29.25	23.28	28.05	18.12	-1.20	-4.10
GFL225	171.66	29.74	536.91	28.52	365.25	-1.22	-4.10
GFL226	7.86	29.62	34.00	28.41	26.14	-1.21	-4.10
GFL227	0.71	29.40	61.51	28.20	60.79	-1.20	-4.10
GFL228	7.08	29.43	30.84	28.23	23.76	-1.21	-4.10
GFL229	17.83	29.28	75.68	28.08	57.85	-1.20	-4.10
GFL230	15.84	29.29	87.92	28.10	72.08	-1.20	-4.09
GFL231	1.61	29.40	60.32	28.20	58.71	-1.20	-4.09
GFL232	10.70	29.00	39.24	27.81	28.54	-1.18	-4.08
GFL233	242.86	29.65	518.02	28.44	275.16	-1.21	-4.08
GFL234	0.50	29.22	6.29	28.03	5.79	-1.19	-4.07
GFL235	0.00	29.43	11.58	28.23	11.58	-1.20	-4.07
GFL236	124.35	29.67	263.29	28.46	138.94	-1.21	-4.07
GFL237	123.03	29.48	223.75	28.29	100.71	-1.20	-4.06
GFL238	34.49	29.57	189.38	28.36	154.89	-1.20	-4.06
GFL239	529.39	29.18	1294.70	27.99	765.31	-1.18	-4.05
GFL240	254.93	29.16	518.15	27.98	263.23	-1.18	-4.05
GFL241	0.00	29.26	137.88	28.08	137.88	-1.18	-4.04
GFL242	115.46	28.97	291.52	27.80	176.05	-1.17	-4.03
GFL243	0.00	29.02	40.00	27.86	40.00	-1.16	-4.01
GFL244	0.00	28.87	3.30	27.71	3.30	-1.16	-4.01
GFL245	1.97	29.44	53.29	28.26	51.33	-1.18	-4.01
GFL246	7.47	29.34	32.72	28.17	25.25	-1.17	-4.00
GFL247	33.67	29.65	137.70	28.47	104.03	-1.18	-3.99
GFL248	2.93	28.76	15.99	27.62	13.06	-1.15	-3.99
GFL249	22.20	29.09	79.99	27.93	57.78	-1.16	-3.99
GFL250	24.78	29.15	70.17	27.99	45.39	-1.16	-3.98

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL251	60.64	29.74	146.73	28.56	86.09	-1.18	-3.98
GFL252	33.08	29.51	138.19	28.34	105.10	-1.17	-3.97
GFL253	259.05	29.26	660.96	28.10	401.91	-1.16	-3.97
GFL254	0.00	28.98	23.63	27.83	23.63	-1.15	-3.97
GFL255	0.00	29.24	3.50	28.08	3.50	-1.16	-3.97
GFL256	0.81	29.26	6.83	28.11	6.02	-1.16	-3.96
GFL257	1.43	29.01	9.06	27.87	7.64	-1.15	-3.95
GFL258	6.81	29.04	142.77	27.90	135.97	-1.15	-3.95
GFL259	28.56	29.19	169.26	28.04	140.69	-1.15	-3.94
GFL260	6.90	29.34	101.33	28.18	94.43	-1.16	-3.94
GFL261	8.00	29.35	37.54	28.20	29.54	-1.16	-3.94
GFL262	0.00	29.31	47.15	28.16	47.15	-1.15	-3.94
GFL263	8.41	29.84	49.42	28.67	41.01	-1.17	-3.94
GFL264	0.00	29.13	8.80	27.98	8.80	-1.14	-3.93
GFL265	7.03	29.20	49.19	28.06	42.16	-1.15	-3.93
GFL266	1.13	29.35	12.49	28.20	11.36	-1.15	-3.92
GFL267	24.65	29.39	93.17	28.24	68.52	-1.15	-3.92
GFL268	320.30	29.24	630.07	28.10	309.77	-1.15	-3.92
GFL269	0.99	29.28	47.72	28.13	46.74	-1.15	-3.92
GFL270	10.10	29.26	132.38	28.11	122.28	-1.15	-3.92
GFL271	0.00	29.17	29.94	28.03	29.94	-1.14	-3.91
GFL272	28.67	29.16	211.20	28.02	182.53	-1.14	-3.91
GFL273	20.28	29.93	75.91	28.76	55.63	-1.17	-3.90
GFL274	153.84	29.25	272.67	28.11	118.82	-1.14	-3.89
GFL275	457.12	29.36	1119.68	28.22	662.56	-1.14	-3.89
GFL276	0.00	29.06	7.30	27.93	7.30	-1.13	-3.89
GFL277	2.17	29.34	15.08	28.20	12.92	-1.14	-3.89
GFL278	51.87	29.44	124.27	28.30	72.40	-1.14	-3.89
GFL279	5.39	29.15	30.69	28.02	25.30	-1.13	-3.88
GFL280	17.47	29.05	44.24	27.92	26.77	-1.13	-3.88
GFL281	9.88	29.68	65.22	28.53	55.34	-1.15	-3.88
GFL282	738.77	28.95	1473.32	27.83	734.55	-1.12	-3.87
GFL283	9.11	29.43	65.18	28.29	56.07	-1.14	-3.87
GFL284	2.84	29.23	26.60	28.10	23.77	-1.13	-3.87
GFL285	75.81	29.46	266.62	28.33	190.81	-1.14	-3.86
GFL286	2.18	29.20	18.58	28.08	16.40	-1.12	-3.84
GFL287	22.07	29.30	91.44	28.18	69.36	-1.12	-3.83
GFL288	0.00	28.99	15.65	27.88	15.65	-1.11	-3.82
GFL289	20.00	29.46	48.36	28.34	28.36	-1.12	-3.82
GFL290	1.51	29.03	14.29	27.92	12.77	-1.11	-3.81
GFL291	21.09	29.15	111.93	28.04	90.84	-1.11	-3.81
GFL292	102.64	29.33	242.87	28.21	140.23	-1.12	-3.80
GFL293	32.35	29.60	96.62	28.48	64.27	-1.13	-3.80
GFL294	5.02	29.04	42.05	27.94	37.03	-1.10	-3.79
GFL295	174.99	29.43	426.21	28.31	251.22	-1.12	-3.79
GFL296	92.56	29.58	263.69	28.46	171.13	-1.12	-3.79
GFL297	2.10	29.33	26.90	28.22	24.80	-1.11	-3.79
GFL298	23.09	29.53	69.79	28.41	46.70	-1.12	-3.79
GFL299	0.00	29.12	18.63	28.02	18.63	-1.10	-3.77
GFL300	0.00	28.83	3.35	27.74	3.35	-1.09	-3.77

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL301	130.59	29.30	283.37	28.19	152.77	-1.10	-3.76
GFL302	4.18	29.27	38.12	28.17	33.94	-1.10	-3.75
GFL303	71.07	28.99	188.74	27.90	117.67	-1.09	-3.75
GFL304	4.61	29.33	26.80	28.23	22.19	-1.10	-3.74
GFL305	17.02	29.23	58.04	28.13	41.02	-1.09	-3.74
GFL306	2.64	29.42	53.41	28.33	50.77	-1.10	-3.74
GFL307	2.32	29.46	11.00	28.36	8.68	-1.10	-3.73
GFL308	85.23	29.46	243.05	28.36	157.82	-1.10	-3.73
GFL309	15.41	29.32	59.63	28.23	44.22	-1.09	-3.71
GFL310	0.79	29.16	6.63	28.08	5.84	-1.08	-3.70
GFL311	4.47	29.32	76.02	28.23	71.55	-1.08	-3.69
GFL312	0.64	28.90	10.09	27.83	9.45	-1.07	-3.69
GFL313	5.93	29.24	19.33	28.16	13.40	-1.08	-3.69
GFL314	17.86	29.02	113.99	27.95	96.12	-1.07	-3.68
GFL315	1.37	29.31	11.60	28.24	10.23	-1.08	-3.68
GFL316	6.92	29.23	18.85	28.15	11.92	-1.07	-3.67
GFL317	305.35	29.40	543.75	28.32	238.40	-1.08	-3.67
GFL318	147.79	29.29	485.06	28.21	337.27	-1.07	-3.66
GFL319	2.27	29.44	16.74	28.37	14.47	-1.08	-3.66
GFL320	0.00	29.09	42.12	28.03	42.12	-1.06	-3.65
GFL321	0.00	28.87	3.75	27.82	3.75	-1.05	-3.64
GFL322	15.04	29.20	466.97	28.14	451.93	-1.06	-3.64
GFL323	1.52	29.22	15.14	28.16	13.62	-1.06	-3.64
GFL324	118.96	29.75	235.09	28.67	116.13	-1.08	-3.64
GFL325	94.37	29.11	190.93	28.05	96.56	-1.06	-3.64
GFL326	19.77	29.22	538.85	28.16	519.07	-1.06	-3.64
GFL327	43.32	29.25	91.78	28.19	48.47	-1.06	-3.63
GFL328	70.37	29.54	182.16	28.47	111.80	-1.07	-3.63
GFL329	64.86	29.47	137.13	28.39	72.27	-1.07	-3.63
GFL330	103.64	28.92	245.59	27.87	141.94	-1.05	-3.63
GFL331	93.90	29.16	289.27	28.11	195.37	-1.06	-3.63
GFL332	33.20	29.20	95.08	28.15	61.88	-1.06	-3.62
GFL333	0.00	28.96	32.78	27.92	32.78	-1.05	-3.62
GFL334	0.00	29.05	5.71	28.00	5.71	-1.05	-3.61
GFL335	226.44	28.94	491.33	27.90	264.89	-1.04	-3.61
GFL336	112.41	29.30	330.14	28.24	217.73	-1.06	-3.60
GFL337	13.00	29.20	300.59	28.15	287.59	-1.05	-3.60
GFL338	50.41	29.24	122.32	28.19	71.91	-1.05	-3.60
GFL339	141.91	29.40	378.40	28.34	236.49	-1.06	-3.60
GFL340	0.00	28.87	11.83	27.83	11.83	-1.04	-3.59
GFL341	26.04	29.51	81.76	28.45	55.72	-1.06	-3.59
GFL342	2.43	29.18	24.22	28.14	21.78	-1.05	-3.59
GFL343	77.76	29.47	259.63	28.41	181.87	-1.06	-3.59
GFL344	0.00	28.64	3.57	27.61	3.57	-1.03	-3.58
GFL345	0.57	29.07	8.47	28.03	7.90	-1.04	-3.58
GFL346	0.00	28.90	35.92	27.86	35.92	-1.03	-3.58
GFL347	1.21	29.21	12.13	28.17	10.92	-1.04	-3.58
GFL348	0.00	28.86	3.58	27.83	3.58	-1.03	-3.57
GFL349	0.02	28.70	30.48	27.68	30.46	-1.02	-3.56
GFL350	464.82	29.47	939.01	28.42	474.20	-1.05	-3.56

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL351	3.99	29.40	20.27	28.35	16.28	-1.05	-3.56
GFL352	23.84	28.98	106.31	27.95	82.47	-1.03	-3.56
GFL353	0.85	29.50	32.69	28.45	31.84	-1.05	-3.56
GFL354	713.39	29.50	1722.58	28.46	1009.18	-1.05	-3.55
GFL355	142.13	29.35	273.83	28.31	131.70	-1.04	-3.55
GFL356	5.74	29.47	46.66	28.43	40.91	-1.05	-3.55
GFL357	0.00	29.26	49.80	28.22	49.80	-1.04	-3.55
GFL358	0.00	29.25	33.94	28.21	33.94	-1.04	-3.54
GFL359	0.00	28.99	32.56	27.96	32.56	-1.02	-3.54
GFL360	0.00	29.17	13.84	28.13	13.84	-1.03	-3.53
GFL361	0.00	29.16	13.42	28.13	13.42	-1.03	-3.52
GFL362	0.47	29.18	55.29	28.15	54.81	-1.03	-3.51
GFL363	0.00	29.09	19.77	28.07	19.77	-1.02	-3.51
GFL364	0.00	29.01	16.63	28.00	16.63	-1.01	-3.50
GFL365	0.97	29.36	52.42	28.34	51.45	-1.02	-3.49
GFL366	47.76	29.46	101.35	28.44	53.59	-1.02	-3.47
GFL367	350.09	29.37	633.23	28.35	283.13	-1.02	-3.47
GFL368	0.00	28.97	21.83	27.96	21.83	-1.00	-3.46
GFL369	11.43	29.11	35.96	28.11	24.53	-1.01	-3.46
GFL370	42.35	29.43	104.25	28.41	61.90	-1.02	-3.46
GFL371	8.71	29.45	72.28	28.43	63.56	-1.02	-3.45
GFL372	0.03	29.22	27.76	28.21	27.73	-1.01	-3.44
GFL373	0.00	29.04	75.21	28.04	75.21	-1.00	-3.44
GFL374	0.00	29.09	12.49	28.09	12.49	-1.00	-3.43
GFL375	79.07	29.27	159.60	28.26	80.53	-1.01	-3.43
GFL376	0.00	29.11	24.82	28.12	24.82	-1.00	-3.42
GFL377	24.60	29.54	212.37	28.53	187.77	-1.01	-3.41
GFL378	2.93	28.73	61.51	27.75	58.58	-0.98	-3.41
GFL379	0.00	28.80	55.03	27.82	55.03	-0.98	-3.40
GFL380	1.05	29.22	94.33	28.23	93.28	-0.99	-3.40
GFL381	101.14	29.21	204.97	28.23	103.83	-0.98	-3.37
GFL382	0.36	29.29	19.93	28.31	19.56	-0.99	-3.37
GFL383	0.00	28.99	14.51	28.02	14.51	-0.98	-3.37
GFL384	0.00	28.78	3.38	27.81	3.38	-0.97	-3.36
GFL385	198.79	29.30	499.05	28.33	300.25	-0.98	-3.34
GFL386	0.00	28.84	25.23	27.88	25.23	-0.96	-3.33
GFL387	19.24	29.19	55.46	28.22	36.22	-0.97	-3.32
GFL388	46.29	29.53	125.29	28.55	79.00	-0.98	-3.32
GFL389	5.16	29.37	27.61	28.39	22.45	-0.98	-3.32
GFL390	0.05	29.08	14.60	28.12	14.54	-0.96	-3.31
GFL391	0.00	28.76	3.30	27.81	3.30	-0.95	-3.31
GFL392	7.24	29.02	44.77	28.07	37.53	-0.95	-3.29
GFL393	1.17	28.89	27.65	27.95	26.47	-0.95	-3.28
GFL394	14.09	28.99	65.35	28.04	51.26	-0.95	-3.27
GFL395	1.22	28.98	9.52	28.04	8.30	-0.94	-3.25
GFL396	29.79	29.45	72.78	28.49	42.99	-0.95	-3.24
GFL397	0.00	28.95	44.46	28.02	44.46	-0.94	-3.23
GFL398	13.49	29.08	34.27	28.14	20.78	-0.94	-3.23
GFL399	8.49	29.45	38.29	28.51	29.80	-0.94	-3.21
GFL400	0.00	28.92	5.11	27.99	5.11	-0.93	-3.20

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL401	28.27	29.09	76.37	28.16	48.10	-0.93	-3.19
GFL402	0.00	28.95	13.51	28.03	13.51	-0.92	-3.17
GFL403	0.00	29.03	10.36	28.12	10.36	-0.91	-3.15
GFL404	0.00	28.93	12.26	28.02	12.26	-0.91	-3.14
GFL405	0.49	29.05	10.45	28.14	9.96	-0.91	-3.13
GFL406	0.00	28.88	18.89	27.98	18.89	-0.90	-3.12
GFL407	9.25	29.22	34.04	28.31	24.79	-0.91	-3.12
GFL408	0.00	29.02	33.65	28.12	33.65	-0.90	-3.10
GFL409	290.07	29.52	668.39	28.61	378.32	-0.91	-3.08
GFL410	14.22	29.48	66.44	28.58	52.22	-0.91	-3.08
GFL411	0.00	29.01	4.34	28.12	4.34	-0.89	-3.06
GFL412	0.00	28.88	24.43	28.00	24.43	-0.88	-3.05
GFL413	0.00	29.09	38.53	28.21	38.53	-0.88	-3.02
GFL414	0.00	28.75	4.79	27.89	4.79	-0.86	-2.98
GFL415	0.00	29.04	9.30	28.18	9.30	-0.86	-2.98
GFL416	0.00	28.89	15.95	28.04	15.95	-0.85	-2.96
GFL417	7.47	29.18	43.18	28.32	35.71	-0.86	-2.93
GFL418	0.00	29.09	10.20	28.27	10.20	-0.81	-2.79
GFL419	0.00	29.11	4.82	28.31	4.82	-0.80	-2.74
GFL420	0.69	28.86	7.21	28.10	6.52	-0.76	-2.64
GFL421	0.00	28.55	15.30	29.43	15.30	0.88	3.08
GFL422	0.00	28.81	8.85	29.70	8.85	0.89	3.09
GFL423	0.00	28.64	17.44	29.57	17.44	0.92	3.23
GFL424	2.15	28.39	10.32	29.33	8.17	0.94	3.33
GFL425	0.00	28.56	38.43	29.51	38.43	0.95	3.33
GFL426	0.00	28.65	26.66	29.61	26.66	0.97	3.38
GFL427	0.49	28.45	16.21	29.42	15.72	0.97	3.40
GFL428	0.00	28.63	10.93	29.62	10.93	0.98	3.43
GFL429	0.00	28.61	23.12	29.61	23.12	0.99	3.47
GFL430	34.06	28.30	57008.56	29.31	56974.50	1.01	3.57
GFL431	68.98	28.20	272.05	29.22	203.07	1.02	3.63
GFL432	95.71	28.33	228.11	29.38	132.40	1.04	3.68
GFL433	1.29	28.56	11.43	29.63	10.14	1.06	3.72
GFL434	0.00	28.64	21.62	29.72	21.62	1.08	3.75
GFL435	0.00	28.94	7.83	30.03	7.83	1.09	3.76
GFL436	0.00	28.71	12.28	29.79	12.28	1.08	3.77
GFL437	0.00	28.55	12.20	29.63	12.20	1.08	3.78
GFL438	3.25	28.29	12.51	29.36	9.26	1.07	3.79
GFL439	8.58	28.15	70.99	29.23	62.41	1.08	3.82
GFL440	0.00	28.52	14.24	29.61	14.24	1.09	3.82
GFL441	0.00	28.75	70.79	29.86	70.79	1.10	3.83
GFL442	0.83	28.13	11.56	29.21	10.72	1.08	3.85
GFL443	0.59	28.45	7.26	29.56	6.66	1.11	3.89
GFL444	0.00	28.71	3.45	29.84	3.45	1.13	3.94
GFL445	0.00	28.20	25.56	29.32	25.56	1.12	3.97
GFL446	5.06	28.40	80.68	29.55	75.62	1.15	4.04
GFL447	0.00	28.66	10.72	29.81	10.72	1.16	4.04
GFL448	0.00	28.46	31.34	29.64	31.34	1.18	4.15
GFL449	0.28	28.51	32.40	29.70	32.12	1.19	4.17
GFL450	1.06	28.22	33.82	29.43	32.76	1.21	4.29

Table 8 Continued

Gene ID	TPM[a]	Length (mm)[a]	TPM[b]	Length (mm)[b]	Δ TPM	Length (mm)	Δ ratio (%)
GFL451	2.00	28.29	16.64	29.51	14.64	1.22	4.31
GFL452	36.98	28.33	127.91	29.57	90.93	1.24	4.38
GFL453	1.34	28.08	28.52	29.32	27.18	1.24	4.42
GFL454	0.60	27.95	13.76	29.19	13.16	1.24	4.44
GFL455	1.36	28.12	18.66	29.38	17.29	1.26	4.48
GFL456	1.13	28.08	30.91	29.36	29.78	1.28	4.54
GFL457	0.89	28.39	46.57	29.68	45.68	1.30	4.58
GFL458	0.13	28.30	11.75	29.60	11.63	1.30	4.61
GFL459	73.85	28.01	372.19	29.30	298.34	1.29	4.61
GFL460	0.00	27.93	40.01	29.21	40.01	1.29	4.61
GFL461	0.00	28.44	4.83	29.76	4.83	1.32	4.66
GFL462	0.76	28.39	50.70	29.73	49.94	1.34	4.70
GFL463	0.00	28.54	9.31	29.89	9.31	1.35	4.72
GFL464	0.00	28.41	4.58	29.76	4.58	1.35	4.77
GFL465	0.00	28.49	10.34	29.85	10.34	1.36	4.77
GFL466	0.02	28.33	57.91	29.78	57.89	1.44	5.10
GFL467	39.42	28.71	208.15	30.20	168.73	1.49	5.19
GFL468	1.23	28.44	10.87	29.93	9.64	1.48	5.21
GFL469	0.63	28.35	14.55	29.84	13.91	1.49	5.26
GFL470	0.00	28.49	8.44	30.00	8.44	1.50	5.27
GFL471	1.49	27.95	47.92	29.44	46.43	1.50	5.36
GFL472	0.00	28.46	14.57	30.02	14.57	1.56	5.48
GFL473	3.23	28.45	151.10	30.22	147.87	1.77	6.23
GFL474	268.78	28.33	1136.29	30.25	867.52	1.93	6.80

a: The mean of 30 RILs that showed the least active expression for the *GFL* gene.

b: The mean of 30 RILs that showed the most active expression for the *GFL* gene.

To determine the action directions and effects of the *GFL* genes on UHML, the expression data of the genes in the 198 RILs of the TAM 94L-25 x NMSI 1331 population were extracted from the gene expression profiles of 10-dpa fibers of the population generated by RNA-seq in this study and subjected to statistical analysis against the genetic variation of the UHML of the population. The results showed that 429 (88.6%) of the *GFL* genes were shown to contribute to UHML negatively, whereas 54 (11.4%) contributed to UHML positively (Table 8). The *GFL* genes that negatively contributed UHML, including *GFL001* through *GFL420*, were found to have shorter UHML when their expression levels increased or expression modes were turned on. For *GFL* genes that positively contributed UHML, their higher expression levels were found to result in longer UHML. Furthermore, to determine the effect of each *GFL* gene, 30 RILs that showed the most active expression and 30 RILs that showed the least active expression for each gene among the 198 RILs of the population were compiled into two differently expressed groups of RILs for each *GFL* gene. The UHML of the two RIL groups with different expression levels for the *GFL* gene were subjected to statistical analysis (t-test). The result showed that the *GFL* genes had a positive or negative effect of 2.64 – 7.92% on UHML. In other words, when the gene is turned on or more actively expressed in 10-dpa fibers, the UHML of the RIL will be increased or decreased by 2.64 – 7.92%. Of the *GFL* genes that had negative effects on UHML, *GFL001* showed the largest negative effect on UHML (-7.92%) when it was actively expressed in 10-dpa fibers. Of the *GFL* genes with positive effects on UHML, *GFL474* increased UHML by 6.80% when its expression level increased from 268.78 TPM to 1,136.29 TPM.

Table 9 List of published genes controlling fiber length or trichome development

Gene	Gene description	GenBank no.	Reference	*P value* ≤ 6.25e-05	≤ 0.0025	≤ 0.05
XET/XTH	Xyloglucan endotransglycosylase/hydrolase	HM749062.1	Lee *et al.* 2010	✔	✔	✔
PHYA	Phytochrome	HM143740.1	Abdurakhmonov *et al.* 2014	✔	✔	✔
GhACT1	Actin	AY305723.1	Li *et al.* 2005		✔	✔
GhADF1	Actin-deploymerizing factor	AI731080	Wang *et al.* 2008	✔	✔	✔
GhPEL	Pectate lyase	DQ073046.1	Wang *et al.* 2010		✔	✔
F3H	Flavanone 3-hydroxylase	GU434116.1	Tan *et al.* 2013	✔	✔	✔
GhPFN2	Profilin	GU237487.1	Wang *et al.* 2010			✔
GhFLA1	Fasciclin-like arabinogalactan protein	EF672627.1	Huang *et a l.* 2013	✔	✔	✔
GhAGP4	Arabinoglactan	EF470295.1	Li *et al.* 2010		✔	✔
GbTCP	TCP transcription factor	DQ359121.1	Hao *et al.* 2012			✔
GhMYB109	R2R3 MYB transcription factor	AJ549758.1	Pu *et al.* 2008		✔	✔
GhMYB25	R2R3 MYB transcription factor	AF336283.1	Machado *et al.* 2009		✔	✔
GhRDL1	*AtRD22-Like 1* gene	AY072821.1	Xu *et al.* 2013			✔
WLIM1a	LIN-11, Isl1 and MEC-3 domain protein	JX648310.1	Han *et al.* 2013	✔	✔	✔
GhVIN1	Vacuolar Invertase	FJ915120.1	Wang *et al.* 2010		✔	✔
GhSusA1	Sucrose synthase	HQ702185	Jiang *et al.* 2012	✔	✔	✔
GhGA20ox1-3	Gibberellin 20-oxidase	AY603789.1	Xiao *et al.* 2010			✔
GhDET2	Steroid 5a-reductase	DQ116446.1	Luo *et al.* 2007		✔	✔
GhTTG1	TRANSPARENT TESTA GLABRA1	AF336281.1	Humphries *et al.* 2005	✔	✔	✔
GhTTG3	TRANSPARENT TESTA GLABRA1	AF530911.1	Humphries *et al* . 2005	✔		✔

✔ : published fiber genes that were isolated by the new high-throughput gene/QTL cloning systems at different significant levels.

To determine what the *GFL* genes are and what they do in the 10-dpa developing fibers, the biochemical functions of the *GFL* genes were inferred by BLASTX to the NCBI's non-redundant protein database. Three hundred ninety-nine of the 474 *GFL* genes (84.18%) were annotated (Table 10). Of the 399 annotated *GFL* genes, 353 (88.47%) genes were annotated with gene descriptions under an E-value cutoff of 1.0E-06 and 116 were assigned to at least one enzyme code, with a total of 164 assigned enzyme codes. These results totally agreed with the 21 genes previously cloned by different scientists that encoded diverged proteins or enzymes (Table 9). The *GFL* genes isolated in this study encode a diversity of proteins or enzymes, such as cytochromes, alcohol dehydrogenase transcription factor Myb/SANT-like family protein, 2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein, AMP-dependent synthetase and ligase family protein, cellulose synthase, Galactosyltransferase family protein, etc. (Table 10).

GO analysis assigned the *GFL* genes to 3,117 GO items, with an average of 3.775 GO terms per gene and a standard deviation of 1.498. Two hundred ninety *GFL* genes were assigned to the category of biological process, 298 *GFL* genes were assigned to the category of molecular function and 316 *GFL* genes were assigned to the category of cellular component (Figs. 18 – 20). Of the 290 *GFL* genes assigned to the category of biological process, 32% participate in cellular and metabolic processes (Fig. 18). Of the 298 *GFL* genes assigned to the category of molecular function, 85% are involved in binding and catalytic activities (Fig. 19). Of the 316 *GFL* genes assigned to the category of cellular component, 75% take part in cell and organelle components (Fig. 20).

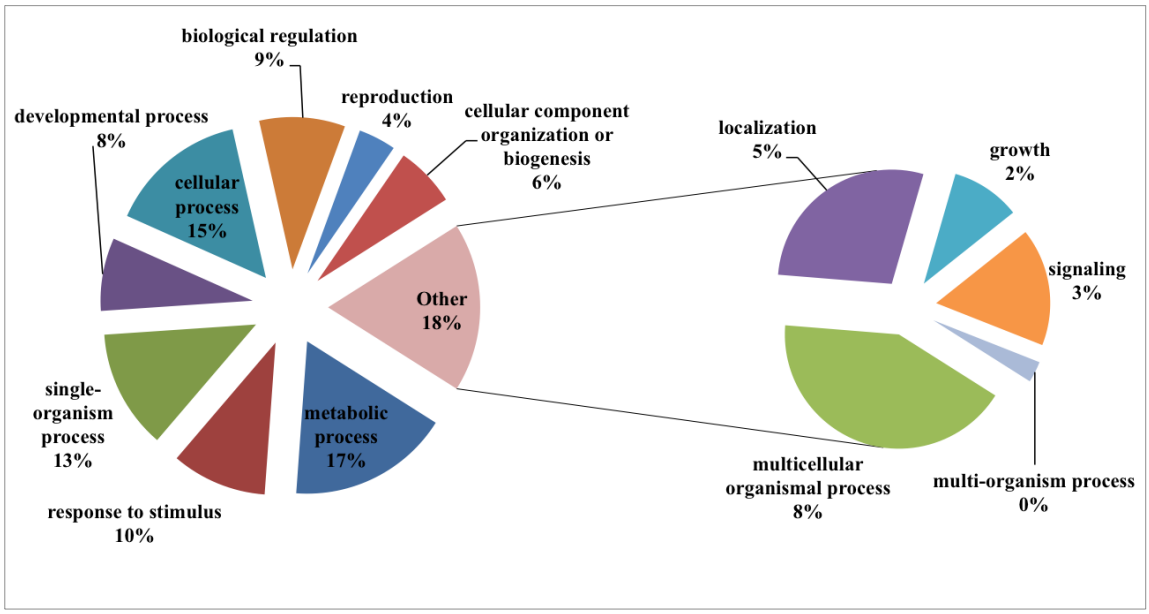

Figure 18 GO items of the *GFL* genes assigned to the functional category of biological process (level 2)

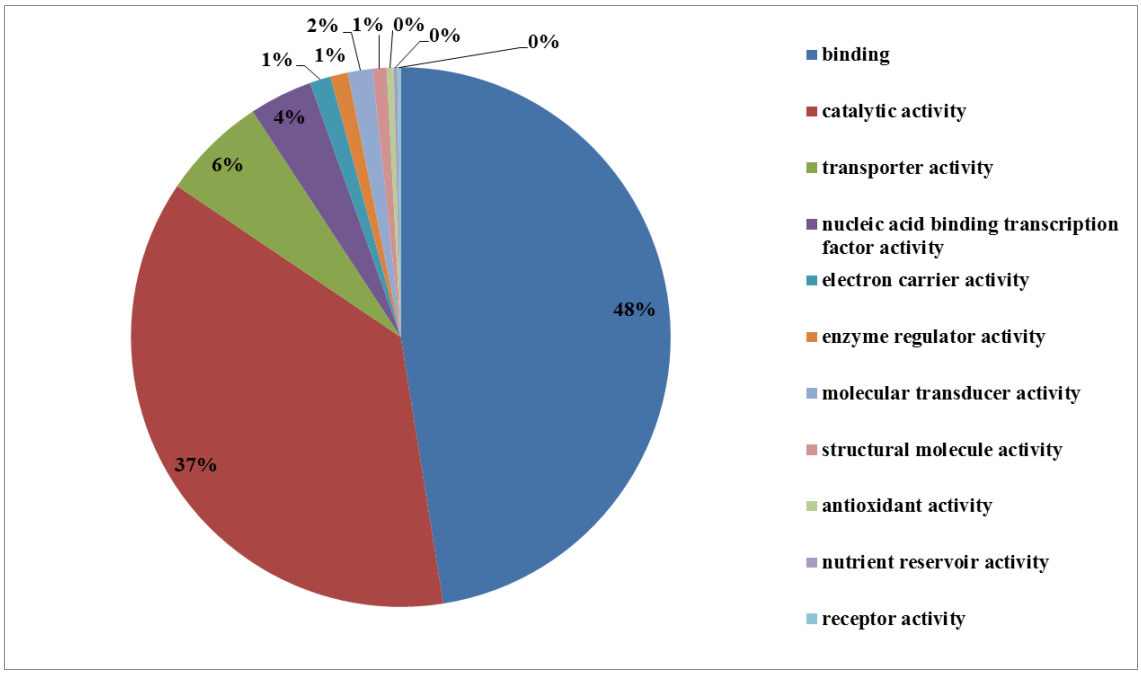

Figure 19 GO items of the *GFL* genes assigned to the functional category of molecular function (level 2)

Table 10 Annotation of *GFL* genes controlling UHML in 10-dpa tetraploid cotton species

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL015	-	448	0	
GFL001	-	309	1	-
GFL420	-	334	0	
GFL111	-	300	0	
GFL234	-	595	0	
GFL451	-	932	0	
GFL122	-	202	0	
GFL056	-	225	0	
GFL464	-	627	0	
GFL212	-	533	0	
GFL348	-	232	0	
GFL151	-	381	0	
GFL172	-	332	0	
GFL140	-	481	0	
GFL384	-	300	0	
GFL149	-	232	0	
GFL323	-	283	0	
GFL012	-	464	0	
GFL059	-	414	0	
GFL442	-	636	0	
GFL042	-	960	0	
GFL002	-	413	0	
GFL107	-	372	0	
GFL132	-	789	0	
GFL345	-	376	0	
GFL112	-	358	0	
GFL034	-	457	0	
GFL141	-	449	0	
GFL005	-	727	0	
GFL443	-	615	0	
GFL197	-	344	0	
GFL047	-	475	0	
GFL344	-	431	0	
GFL224	-	343	0	
GFL297	-	387	0	
GFL028	-	789	0	
GFL022	-	452	0	
GFL315	-	242	0	
GFL254	-	540	0	

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL351	-	381	0	
GFL411	-	403	0	
GFL036	-	428	0	
GFL074	-	474	0	
GFL093	-	288	0	
GFL386	-	1249	0	
GFL221	-	367	0	
GFL369	-	613	0	
GFL003	-	249	0	
GFL073	-	1199	0	
GFL334	-	434	0	
GFL087	-	431	0	
GFL300	-	410	0	
GFL305	-	377	0	
GFL261	-	396	0	
GFL441	-	542	0	
GFL382	-	240	0	
GFL359	-	2300	0	
GFL067	-	507	0	
GFL368	-	715	0	
GFL370	-	699	0	
GFL391	-	350	0	
GFL248	-	256	0	
GFL414	-	348	0	
GFL167	-	1148	0	
GFL428	-	812	0	
GFL309	-	685	0	
GFL255	-	235	0	
GFL072	-	229	0	
GFL389	-	360	0	
GFL179	-	312	0	
GFL033	-	349	4	EC:2.7.1.23
GFL244	-	230	0	
GFL444	-	323	0	
GFL066	-	296	0	
GFL312	-	806	0	
GFL263	14-3-3f protein	766	21	-
GFL459	14-3-3h protein	1335	1	-
GFL102	2-oxoglutarate and Fe(II)-dependent oxygenase superfamily protein, putative	1277	1	EC:1.14.11.0

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL380	20S proteasome subunit alpha-1	1274	15	EC:3.4.25.0
GFL426	3-ketoacyl-acyl carrier protein synthase III, III isoform 1	3682	16	EC:2.3.1.41
GFL242	51 kDa subunit of complex I	1962	17	EC:1.6.5.3
GFL456	6-phosphogluconolactonase 1 isoform 1	769	7	EC:3.1.1.31
GFL241	AAA-type ATPase family protein	1769	16	EC:3.6.4.3
GFL214	AarF domain-containing kinase isoform 2	310	4	-
GFL032	ABI-1-like 1 isoform 1	1205	8	-
GFL397	AC002423_8T23E23.16	4555	4	EC:2.7.7.49
GFL004	AC007323_16T25K16.5	357	10	EC:3.6.1.1
GFL205	Acetyl-CoA synthetase isoform 1	1805	12	EC:6.2.1.1
GFL336	Actin depolymerizing factor 5	797	5	-
GFL121	actin-97	301	3	-
GFL454	Acyl-CoA binding protein 4 isoform 3	337	15	-
GFL376	Adenosine-5'-phosphosulfate (APS) kinase 3 isoform 1	1124	11	EC:2.7.1.25; EC:2.7.2.4; EC:1.3.1.74
GFL181	ADP-glucose pyrophosphorylase family protein isoform 1	2365	6	EC:2.7.7.13
GFL430	AF218378_1protein kinase	1212	4	-
GFL190	Alba DNA/RNA-binding protein	262	6	-
GFL398	Alcohol dehydrogenase transcription factor Myb/SANT-like family protein	801	1	
GFL039	Alkaline-phosphatase-like family protein isoform 1	1983	7	EC:3.6.1.9; EC:3.1.4.1
GFL372	Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase	2446	3	-
GFL409	Alpha/beta-Hydrolases superfamily protein isoform 1	2606	5	-
GFL302	Alpha/beta-Hydrolases superfamily protein isoform 2	1284	13	-
GFL153	Alpha/beta-Hydrolases superfamily protein, putative isoform 1	1048	1	-
GFL400	AMP-dependent synthetase and ligase family protein	559	10	EC:2.3.1.86; EC:1.13.12.7; EC:6.2.1.12
GFL466	AMP-dependent synthetase and ligase family protein isoform 1	3083	9	EC:6.2.1.3
GFL355	angustifolia	2839	26	EC:1.1.1.29
GFL168	Ankyrin repeat domain-containing protein 50 isoform 3	864	1	-
GFL063	Ankyrin repeat family protein	1665	17	-
GFL223	Arabinanase/levansucrase/invertase, putative	1894	3	
GFL326	Arginine decarboxylase	2697	17	EC:4.1.1.19
GFL159	Arginine methyltransferase 11	635	9	EC:2.1.1.125
GFL403	Asparagine synthase family protein, putative	1607	4	-
GFL058	ASYMMETRIC LEAVES 2-like 1	1263	6	-
GFL269	Auxin efflux carrier family protein isoform 1	1140	2	-
GFL128	Basic helix-loop-helix DNA-binding superfamily protein	1341	12	-
GFL225	Basic leucine-zipper 44	1357	12	-
GFL250	Basic-leucine zipper transcription factor family protein isoform 3	1419	12	-
GFL282	Beige-related and WD-40 repeat-containing protein isoform 2	390	2	-
GFL152	Beige/BEACH domain,WD domain, G-beta repeat protein	9339	10	-
GFL185	beta-mannosidase	1891	4	EC:3.2.1.152

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL272	Bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1, putative	1592	8	EC:4.2.1.109; EC:3.1.3.18; EC:3.1.3.77
GFL327	Binding to TOMV RNA 1L (long form) isoform 1	1072	10	EC:3.1.2.15
GFL360	Biotin carboxyl carrier protein subunit of of Het-ACCase (BCCP1), putative isoform 3	478	20	EC:6.4.1.2
GFL105	BR-signaling kinase 2 isoform 1	1705	19	EC:2.7.11.0; EC:2.7.10.0; EC:2.7.1.40; EC:3.2.1.15
GFL069	Bromodomain 4, putative	800	2	-
GFL299	BSD domain (BTF2-like transcription factors, Synapse-associated proteins and DOS2-like proteins) isoform 4, partial	1161	5	-
GFL280	BSD domain-containing protein, putative	834	3	
GFL048	BTB-POZ and MATH domain 2	3930	8	-
GFL158	BTB/POZ domain with WD40/YVTN repeat-like protein	1347	7	-
GFL288	BTB/POZ domain-containing protein	2314	1	-
GFL045	BTB/POZ domain-containing protein isoform 1	876	1	-
GFL247	C2H2 zinc-finger protein SERRATE isoform 2, partial	2660	17	-
GFL377	Calcium-binding EF-hand family protein	588	10	-
GFL096	Calcium-binding EF-hand family protein, putative isoform 1	1027	4	-
GFL050	Calcium-dependent lipid-binding (CaLB domain) family protein isoform 1	442	11	-
GFL278	Calcium-dependent lipid-binding family protein isoform 2, partial	731	6	-
GFL182	Calcium-dependent protein kinase 13 isoform 2	2758	6	EC:2.7.11.17
GFL440	Calcium-dependent protein kinase 6 isoform 2	2317	17	EC:2.7.11.17
GFL014	CDK inhibitor P21 binding protein, putative	1635	2	-
GFL086	Cellulose-synthase-like C12	2781	7	EC:2.4.1.12
GFL251	Chloride channel F isoform 2	1256	7	-
GFL071	Chloride channel protein CLC-d isoform 6	247	14	-
GFL407	Chloroplast isoform 2, partial	879	4	-
GFL343	Chloroplast-targeted copper chaperone-like protein	749	0	
GFL267	Chloroplast-targeted copper chaperone-like protein	642	9	EC:3.1.3.0
GFL268	Chromatin remodeling complex subunit isoform 1	3709	17	-
GFL230	Cleavage stimulation factor 64 kDa subunit, putative	2348	7	-
GFL354	CLIP-associated protein isoform 1	4930	21	-
GFL201	Cobalt ion binding	1482	5	-
GFL029	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family isoform 2	2075	27	EC:1.10.2.2
GFL245	Concanavalin A-like lectin protein kinase family protein	907	2	-
GFL266	conserved hypothetical protein	833	4	EC:1.6.5.3
GFL054	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	333	7	EC:2.4.2.26
GFL052	CRT-like transporter 3 isoform 1	1801	7	-
GFL298	Cullin 1	431	27	-
GFL264	Cyclin b2,4 isoform 2	1343	16	-
GFL329	Cyclin family protein isoform 2	1887	10	-
GFL169	Cyclin p1,1 isoform 1	1049	5	-

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL450	Cyclin p4,1	893	10	EC:2.7.11.22
GFL319	Cysteine-rich RLK 42	1143	7	EC:2.7.10.0
GFL196	Cytochrome P450	2101	3	-
GFL455	Cytochrome P450, putative	914	9	EC:1.14.13.0
GFL038	D6 protein kinase like 2	323	13	EC:2.7.11.0
GFL025	DEA(D/H)-box RNA helicase family protein isoform 3	596	0	
GFL240	Developmental regulator, ULTRAPETALA	1340	7	-
GFL145	DHHC-type zinc finger family protein	1036	2	-
GFL171	DNA binding/zinc ion binding protein	734	3	-
GFL082	DNA glycosylase superfamily protein	1627	10	EC:3.2.2.21; EC:3.2.2.20
GFL412	DNA glycosylase superfamily protein isoform 2	1376	9	-
GFL422	DNAse I-like superfamily protein isoform 1	1288	6	-
GFL421	DNAse I-like superfamily protein, putative isoform 2	1439	5	-
GFL425	Domain of Uncharacterized protein function isoform 1	1545	4	-
GFL139	Drought-responsive family protein	1178	6	-
GFL166	Dynamin-related protein 3A isoform 2	702	14	-
GFL306	E3 ubiquitin-protein ligase UPL5	4323	12	EC:2.7.7.49; EC:6.3.2.19
GFL293	EIN2-like protein, nramp transporter isoform 1	747	25	-
GFL203	Electron transport complex protein rnfC, putative isoform 3	783	0	
GFL194	Emb:CAB89363.1	2931	1	-
GFL387	Embryo defective 1703, putative isoform 2	2326	5	-
GFL262	Embryo defective 2423, putative	3267	5	-
GFL410	Endonuclease or glycosyl hydrolase	2036	2	-
GFL419	Endonuclease/exonuclease/phosphatase family protein	374	2	-
GFL084	Enhanced disease resistance 2 isoform 1	772	14	-
GFL103	ENTH/VHS	2417	2	-
GFL416	Ethylene-insensitive 3f isoform 1	1287	3	-
GFL364	Exostosin family protein	1813	3	-
GFL173	Extra-large G-protein 1	3645	14	-
GFL142	F-box/RNI-like superfamily protein isoform 1	1090	16	EC:6.3.2.19
GFL010	Far1-related sequence 3 isoform 1	373	1	-
GFL447	Fatty acyl-ACP thioesterases B isoform 2	1415	9	EC:3.1.2.14
GFL274	FKBP-type peptidyl-prolyl cis-trans isomerase family protein	1249	10	EC:5.2.1.8
GFL026	Flavin-binding monooxygenase family protein isoform 1	1815	13	EC:4.6.1.2; EC:1.14.13.8
GFL011	Floral homeotic protein DEFICIENS isoform 1	970	14	-
GFL362	FMN-linked oxidoreductases superfamily protein isoform 1	1638	9	-
GFL328	Formin-like protein 5 isoform 1	3499	10	-
GFL431	Fructose-2,6-bisphosphatase isoform 1	4773	9	EC:2.7.1.105; EC:3.1.3.46
GFL257	Galactose oxidase/kelch repeat superfamily protein	223	21	EC:6.3.2.19

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL385	Galactosyltransferase family protein	2022	6	EC:2.4.1.134
GFL191	GATA transcription factor 2, putative	693	9	-
GFL217	GCIP-interacting family protein isoform 1	1352	1	-
GFL134	GDP-fucose protein O-fucosyltransferase 1 isoform 1	437	1	EC:2.4.1.221
GFL342	Geminivirus rep interacting kinase 2	2260	9	EC:2.7.10.0; EC:2.7.11.17
GFL258	Global transcription factor group E4, putative isoform 1	2300	8	-
GFL178	GLU-ADT subunit B isoform 1	2137	11	EC:6.3.5.7
GFL404	glucosamine-fructose-6-phosphate aminotransferase, putative	631	9	EC:2.6.1.16
GFL070	Glucose-inhibited division family A protein isoform 1	2399	9	EC:2.1.1.74
GFL357	Glutamate-ammonia ligases,catalytics,glutamate-ammonia ligases isoform 3	2240	17	EC:6.3.1.2
GFL363	Glutamyl-tRNA(Gln) amidotransferase subunit A isoform 1	2115	14	EC:6.3.5.0; EC:3.5.1.4
GFL130	Glutathione-disulfide reductase isoform 1	1879	26	EC:1.8.1.7; EC:1.1.1.44; EC:1.1.1.30
GFL055	Glycosyl transferase family 1 protein isoform 1	443	7	-
GFL108	Glycosyltransferase isoform 2	773	7	EC:2.4.1.43
GFL160	GRAM domain family protein, putative isoform 1	418	0	
GFL206	GRIP and coiled-coil domain-containing protein 2, putative isoform 1	2632	14	-
GFL155	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein isoform 1	1396	8	EC:3.1.3.0
GFL291	HAT dimerization domain-containing protein isoform 2	3182	3	-
GFL365	HD domain class transcription factor isoform 2	3321	16	-
GFL040	Heat shock factor protein HSF8, putative	2609	9	-
GFL249	Heat shock protein DnaJ, N-terminal with domain of Uncharacterized protein function (DUF1977) isoform 1	2021	4	-
GFL136	Helicase/SANT-associated, putative isoform 5	450	4	
GFL120	Hexokinase 1 isoform 2	2081	18	EC:2.7.1.2; EC:2.7.1.4
GFL043	Histidine acid phosphatase family protein isoform 1	2327	7	EC:3.1.3.2; EC:3.1.3.62
GFL424	histone deacetylase	254	15	EC:3.5.1.98
GFL175	Histone deacetylase 6 isoform 1	1415	21	EC:3.5.1.98
GFL390	Homeodomain-like superfamily protein	563	9	-
GFL448	Hydroxyproline-rich glycoprotein family protein, putative	2322	3	
GFL237	Hydroxyproline-rich glycoprotein family protein, putative isoform 1	1122	10	-
GFL110	Hydroxyproline-rich glycoprotein family protein, putative isoform 2	1178	2	-
GFL177	hypothetical protein	481	16	EC:2.7.11.22
GFL284	hypothetical protein M569_00896, partial	994	2	
GFL406	hypothetical protein PRUPE_ppa008947mg	1010	8	-
GFL228	hypothetical protein PRUPE_ppa008972mg	574	17	-
GFL075	hypothetical protein PRUPE_ppa014254mg	294	5	-
GFL281	hypothetical protein PRUPE_ppa022115mg	2165	7	
GFL154	hypothetical protein TAANSRALLha_1144N5.t00004	723	10	
GFL077	hypothetical protein VITISV_002159	1014	5	-

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL035	hypothetical protein VITISV_032906	555	3	-
GFL174	Inhibitor-3	450	7	-
GFL304	Inositol monophosphatase family protein isoform 2	739	11	EC:3.1.3.25; EC:3.1.3.7
GFL060	Insulinase (Peptidase family M16) family protein isoform 1	1051	9	EC:3.4.24.0
GFL433	Integral membrane single C2 domain protein, putative isoform 1	715	5	-
GFL378	Integrase-type DNA-binding superfamily protein	1195	13	-
GFL394	Integrase-type DNA-binding superfamily protein isoform 1	1464	7	-
GFL318	IQ calmodulin-binding motif family protein	1834	2	-
GFL413	IQ-domain 17 isoform 2	1466	2	-
GFL127	Iron-sulfur cluster biosynthesis family protein isoform 1	886	5	-
GFL163	Isopropyl malate isomerase large subunit 1	1809	14	EC:4.2.1.33; EC:5.4.4.0
GFL308	K+ uptake permease 7 isoform 1	3403	8	-
GFL210	Kinase domain-containing protein isoform 1	4374	7	EC:2.7.12.1; EC:2.7.11.0
GFL215	Kinase superfamily protein with octicosapeptide/Phox/Bem1p domain, putative isoform 1	660	7	EC:2.7.10.2; EC:2.7.12.1; EC:2.7.11.0
GFL310	Leucine-rich repeat containing-like protein isoform 1	991	2	-
GFL276	Leucine-rich repeat containing-like protein isoform 1	1016	7	
GFL461	Leucine-rich repeat-containing protein, putative	576	2	-
GFL437	Lipases,hydrolases, acting on ester bonds isoform 1	1818	5	-
GFL156	lipoxygenase, putative	1773	11	EC:1.13.11.12
GFL202	LRR and NB-ARC domains-containing disease resistance protein, putative	345	4	
GFL465	LRR receptor-like serine/threonine-protein kinase, putative	541	15	-
GFL289	Lupus la ribonucleoprotein, putative isoform 3	496	6	-
GFL239	Major facilitator superfamily protein isoform 1	2404	4	-
GFL375	Mannose-P-dolichol utilization defect 1 protein isoform 1	1312	1	-
GFL271	MATE efflux family protein isoform 2	2482	5	-
GFL150	Metacaspase 1 isoform 1	1379	11	EC:3.4.22.0
GFL295	Metal-dependent phosphohydrolase isoform 1	1295	4	EC:3.1.4.0
GFL439	Methyltransferase family protein, putative	1073	6	-
GFL246	Microtubule-associated protein 6, putative	1030	2	-
GFL366	MIF4G domain-containing protein / MA3 domain-containing protein	742	13	-
GFL023	Mitochondrial editing factor 22	1919	3	-
GFL064	Mitochondrial substrate carrier family protein	1679	5	-
GFL358	Mitochondrial transcription termination factor family protein	1758	6	-
GFL332	Mitochondrial-processing peptidase subunit beta, mitochondrial, putative	3545	6	EC:3.4.24.0
GFL401	Monogalactosyl diacylglycerol synthase 1	1914	13	EC:2.4.1.46
GFL399	MRNA splicing factor, thioredoxin-like U5 snRNP	842	5	-
GFL232	MSCS-like 3 isoform 1	1187	10	-
GFL216	Myb-like transcription factor family protein, putative	958	6	-
GFL235	Na+/H+ antiporter 6 isoform 3	536	12	-

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL432	NAD kinase 1 isoform 1	2282	8	EC:2.7.1.86; EC:2.7.1.23
GFL458	NAD(P)-binding Rossmann-fold superfamily protein	1127	3	-
GFL018	NADH-ubiquinone oxidoreductase 21 kDa subunit	778	8	-
GFL463	NB-ARC domain-containing disease resistance protein, putative	915	7	
GFL097	NB-ARC domain-containing disease resistance-like protein isoform 2	3483	2	-
GFL226	Neutral invertase isoform 1	2791	10	EC:3.2.1.26; EC:3.2.1.48; EC:3.2.1.97
GFL285	Nicotinamidase 2	911	5	EC:3.5.1.19; EC:3.3.2.1
GFL231	No lysine kinase 3 isoform 1	2426	11	EC:2.7.11.25
GFL330	Non-intrinsic ABC protein 6 isoform 1	1531	17	-
GFL192	NPL4-like protein 1	631	9	EC:1.11.1.7
GFL030	Nuclear transcription factor Y subunit B-10 isoform 2	704	9	-
GFL109	nuclease, putative	1554	2	
GFL275	Nucleotide-sugar transporter family protein	1674	10	-
GFL273	Nucleotide-sugar transporter family protein	1350	5	-
GFL200	Nucleotidyltransferase family protein isoform 5	2249	2	EC:2.7.7.0
GFL335	O-acetylserine lyase B isoform 1	1665	29	EC:2.5.1.47
GFL147	O-fucosyltransferase family protein	810	3	-
GFL256	O-fucosyltransferase family protein isoform 1	2190	5	-
GFL021	Oberon 2 isoform 1	2176	14	-
GFL322	OEP37_PEARecName: Full=Outer envelope pore protein 37, chloroplastic; AltName: Full=Chloroplastic outer envelope pore protein of 37 kDa; Short=PsOEP37; Flags: Precursor	446	10	-
GFL393	P-loop containing nucleoside triphosphate hydrolases superfamily protein	601	11	-
GFL188	P-loop containing nucleoside triphosphate hydrolases superfamily protein	1754	7	EC:3.6.1.3
GFL016	P-loop containing nucleoside triphosphate hydrolases superfamily protein isoform 1	2153	15	-
GFL083	P4H isoform 1	1626	10	EC:1.14.11.2
GFL089	PAP10	1654	14	EC:3.1.3.2
GFL164	PATATIN-like protein 9, IIIB isoform 1	1559	7	-
GFL341	Pentatricopeptide repeat (PPR) superfamily protein, putative	1236	2	-
GFL311	Pentatricopeptide repeat (PPR) superfamily protein, putative isoform 1	1242	2	-
GFL395	Pentatricopeptide repeat-containing protein, putative	2231	2	-
GFL170	Peptidase M20/M25/M40 family protein isoform 5	999	13	-
GFL438	Peptide n-glycanase, putative isoform 2	1910	10	EC:3.5.1.52
GFL388	PfkB-like carbohydrate kinase family protein	230	9	EC:2.7.7.49; EC:2.7.1.4
GFL402	Phenylalanyl-tRNA synthetase / phenylalanine--tRNA ligase, putative isoform 1	881	10	EC:6.1.1.20
GFL017	Phloem protein 2-A15	426	3	-
GFL468	Phloem protein 2-like a10, putative	1647	1	-
GFL471	Phosphatase 2A, regulatory subunit PR55, BETA isoform 1	1271	7	-
GFL113	Phosphatidylinositol-4-phosphate 5-kinase family protein, putative isoform 3, partial	441	11	EC:2.7.1.150; EC:2.7.1.68
GFL162	Phosphoprotein phosphatase	294	1	-

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL453	Photosystem II subunit Q-2	1063	18	-
GFL472	Phototropic-responsive NPH3 family protein isoform 1	1439	11	-
GFL374	Plant calmodulin-binding protein-related, putative isoform 1	809	2	-
GFL148	Pleiotropic drug resistance 12 isoform 2	1351	8	EC:3.6.3.41; EC:3.6.3.28; EC:3.6.3.31
GFL313	Poly(A) binding protein 8 isoform 2	213	5	-
GFL092	Polyketide cyclase / dehydrase and lipid transport protein	438	9	-
GFL144	Polymerase gamma 2 isoform 4	3667	14	EC:2.7.7.7
GFL009	Polynucleotide adenylyltransferase family protein isoform 1	2493	7	EC:2.7.7.25; EC:2.7.7.56
GFL208	predicted protein	217	13	EC:5.1.3.5; EC:5.1.3.2
GFL321	predicted protein	250	4	-
GFL317	predicted protein	2265	7	-
GFL220	predicted protein	2116	7	EC:2.7.11.0; EC:2.7.12.1
GFL469	predicted protein	594	2	
GFL041	PREDICTED: DEAD-box ATP-dependent RNA helicase 35 isoform 1	2281	10	-
GFL044	PREDICTED: GMP synthase	295	16	EC:3.4.21.0; EC:3.2.1.0; EC:6.3.5.2; EC:6.3.5.4
GFL462	PREDICTED: GTP-binding protein SAR1A-like	1320	26	-
GFL435	PREDICTED: monosaccharide-sensing protein 2-like	509	7	-
GFL307	PREDICTED: probable sugar phosphate/phosphate translocator At3g11320-like	398	14	-
GFL347	PREDICTED: putative ribonuclease H protein At1g65750-like	982	14	
GFL183	PREDICTED: ruBisCO large subunit-binding protein subunit beta, chloroplastic	308	20	-
GFL124	PREDICTED: serine/threonine-protein phosphatase PP2A catalytic subunit-like	1437	11	-
GFL353	PREDICTED: TATA-binding protein-associated factor 2N-like isoform X1	1290	2	-
GFL467	PREDICTED: uncharacterized amino-acid permease C15C4.04c-like	1814	19	-
GFL333	PREDICTED: uncharacterized protein LOC100815379	715	6	-
GFL460	PREDICTED: uncharacterized protein LOC101205308, partial	2432	2	-
GFL213	proline-rich protein	302	2	
GFL445	Protochlorophyllide oxidoreductase C, C,PORC	1546	22	EC:1.6.99.1; EC:1.3.1.33
GFL024	PRP38 family protein isoform 1	1690	5	-
GFL331	putative p-coumarate 3-hydroxylase	1746	15	EC:1.14.13.21
GFL279	putative PDF1-interacting protein 2, partial	384	0	
GFL436	RAC-like 9 isoform 1	1808	28	EC:3.1.4.12
GFL131	Radiation sensitive 17, putative	2688	8	-
GFL270	RAN GTPase activating protein 1 isoform 1	2547	20	-
GFL405	Receptor protein kinase, putative	630	4	-
GFL294	Receptor-like protein kinase 1, putative	699	7	EC:2.7.11.0
GFL418	REF4-related 1	4599	8	-
GFL316	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain isoform 1	334	19	-
GFL367	Regulatory particle AAA-ATPase 2A	1688	31	EC:3.6.4.3
GFL195	retrotransposon protein, putative, unclassified	514	5	

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL079	Reversibly glycosylated polypeptide 3	703	10	EC:5.4.99.30; EC:2.4.1.186
GFL106	Ribonucleoside-diphosphate reductase small chain A isoform 1	1313	14	EC:1.17.4.1
GFL119	Ribosomal protein L30/L7 family protein isoform 1	848	4	-
GFL198	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	321	10	-
GFL252	Ribosomal protein S24e family protein isoform 1	683	12	-
GFL027	RING/FYVE/PHD zinc finger superfamily protein, putative isoform 1	3936	7	-
GFL099	RING/FYVE/PHD-type zinc finger family protein isoform 1	2085	4	-
GFL187	RING/U-box superfamily protein	1667	2	-
GFL076	RING/U-box superfamily protein isoform 1	1459	15	EC:6.3.2.19
GFL068	RING/U-box superfamily protein isoform 1	1051	4	EC:6.3.2.19
GFL260	RING/U-box superfamily protein isoform 1	2126	4	-
GFL157	RING/U-box superfamily protein, putative	1858	2	-
GFL114	RNA binding protein, putative	1523	1	-
GFL470	RNA helicase-like protein	551	11	-
GFL046	RNA-binding KH domain-containing protein	1590	3	-
GFL051	RNA-binding KH domain-containing protein isoform 1	2264	1	-
GFL233	Root hair defective 3 GTP-binding protein (RHD3) isoform 2	2360	27	-
GFL116	Root hair defective 6-like 2, putative	848	0	
GFL379	Root phototropism protein, putative isoform 2	1757	4	-
GFL013	RP non-ATPase subunit 8A	363	20	-
GFL429	Rubisco methyltransferase family protein	2287	1	-
GFL053	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	278	3	-
GFL081	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	2448	5	EC:2.1.1.0
GFL184	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein isoform 1	1666	4	EC:2.1.1.0
GFL085	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein isoform 1	2722	8	EC:2.1.1.0
GFL125	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein isoform 1	2973	8	EC:2.1.1.0
GFL473	SAD1/UNC-84 domain protein 2, putative	1831	5	-
GFL090	Sec14p-like phosphatidylinositol transfer family protein	673	8	-
GFL361	Serine/threonine-protein kinase 24 isoform 1	479	12	EC:2.7.11.25
GFL356	Serine/threonine-protein kinase-like protein, putative	436	0	
GFL339	Seryl-tRNA synthetase isoform 1	3268	20	EC:6.1.1.11
GFL324	SH3 domain-containing protein	1596	7	-
GFL265	SIN3-like 2, putative isoform 1	910	16	-
GFL057	Splicing factor PWI domain-containing protein / RNA recognition motif-containing protein isoform	4443	8	-
GFL352	Squamosa promoter-binding protein, putative isoform 1	2753	4	-
GFL019	Steroid binding protein, putative	1356	17	EC:1.3.1.74
GFL381	Stress responsive alpha-beta barrel domain protein, putative isoform 2	1297	12	-
GFL301	Sugar transporter, putative isoform 2, partial	1715	11	EC:1.3.1.74
GFL189	Sulfate transporter 4.1 isoform 1	2382	5	-

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL292	TATA-box binding protein, putative	1677	7	EC:1.3.1.74
GFL061	Teosinte branched 1, putative isoform 4, partial	418	0	
GFL031	Terpenoid synthases superfamily protein	594	9	-
GFL138	Tetratricopeptide repeat (TPR)-like superfamily protein	1591	4	-
GFL126	Tetratricopeptide repeat (TPR)-like superfamily protein	1752	1	-
GFL176	Tetratricopeptide repeat (TPR)-like superfamily protein isoform 3	1315	6	-
GFL277	Tetratricopeptide repeat (TPR)-like superfamily protein, putative	1007	2	-
GFL290	Tetratricopeptide repeat-like superfamily protein	720	2	-
GFL165	Tetratricopeptide repeat-like superfamily protein	1807	11	-
GFL062	Tetratricopeptide repeat-like superfamily protein isoform 2	1117	3	-
GFL474	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein	2362	17	EC:4.1.1.1
GFL006	Tir-nbs resistance protein	1012	6	EC:3.6.1.15
GFL238	Topoisomerase II-associated protein PAT1, putative isoform 2	1585	0	
GFL283	Toxicos en levadura 4, putative	1280	4	EC:6.3.2.19
GFL180	Transcription factor	1250	14	EC:2.4.2.30
GFL078	Transcription factor jumonji family protein / zinc finger family protein	2150	14	-
GFL091	Transcription regulator NOT2/NOT3/NOT5 family protein	2129	3	-
GFL427	Transducin/WD40 repeat-like superfamily protein	1158	2	-
GFL243	Transducin/WD40 repeat-like superfamily protein isoform 1	2437	20	EC:1.3.1.74; EC:6.3.2.19
GFL204	Transducin/WD40 repeat-like superfamily protein isoform 2	724	0	
GFL320	Transducin/WD40 repeat-like superfamily protein, putative isoform 1	2336	1	-
GFL117	translation factor sui1, putative	477	11	-
GFL007	Translation initiation factor 3B1 isoform 2	1897	11	-
GFL452	Translation initiation factor IF2/IF5 isoform 1	698	7	-
GFL100	Translocase of outer membrane 22-I	658	14	-
GFL186	Transmembrane amino acid transporter family protein isoform 1	1715	4	-
GFL417	Transmembrane amino acid transporter family protein isoform 4	994	6	-
GFL236	Trichome birefringence-like 13	2563	3	-
GFL383	Trigger factor, putative	2126	13	EC:5.2.1.8
GFL207	Tryptophan aminotransferase related 2, putative	2016	16	EC:2.6.1.28; EC:2.6.1.27; EC:4.4.1.4
GFL350	TUDOR-SN protein 1 isoform 1	3617	23	-
GFL101	Type one serine/threonine protein phosphatase 2	1498	8	-
GFL115	Type one serine/threonine protein phosphatase 4 isoform 1	1035	6	-
GFL218	Type-b response regulator, putative	832	1	-
GFL227	Ubiquitin carboxyl-terminal hydrolase 19, putative isoform 1	2462	8	EC:3.1.2.15; EC:3.4.22.0
GFL373	Ubiquitin ligase protein cop1, putative isoform 6	2413	18	EC:1.3.1.74
GFL209	Ubiquitin-specific protease family C19-related protein	340	1	-
GFL211	UDP-glucuronic acid decarboxylase 3	1287	8	EC:4.1.1.35
GFL080	UDP-sugar pyrophospharylase	2056	8	EC:2.7.7.9; EC:2.7.7.11; EC:2.7.7.10; EC:2.7.7.44

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL133	Uncharacterized protein isoform 1	1727	3	
GFL396	Uncharacterized protein isoform 1	1579	6	-
GFL098	Uncharacterized protein isoform 1	1516	2	
GFL037	Uncharacterized protein isoform 1	1161	1	
GFL008	Uncharacterized protein isoform 1	1993	1	-
GFL219	Uncharacterized protein isoform 1	1889	2	-
GFL296	Uncharacterized protein isoform 1	2936	1	-
GFL259	Uncharacterized protein isoform 1	3595	4	
GFL340	Uncharacterized protein isoform 1	1269	2	-
GFL199	Uncharacterized protein isoform 1	1096	2	-
GFL088	Uncharacterized protein isoform 2	1787	4	-
GFL457	Uncharacterized protein isoform 2	2383	6	
GFL286	Uncharacterized protein isoform 2, partial	702	6	-
GFL146	Uncharacterized protein isoform 3	516	13	-
GFL104	Uncharacterized protein isoform 4	496	0	
GFL338	Uncharacterized protein isoform 6	926	1	-
GFL408	Uncharacterized protein TCM_000519	1754	1	-
GFL065	Uncharacterized protein TCM_001288	754	3	
GFL434	Uncharacterized protein TCM_002116	494	7	-
GFL135	Uncharacterized protein TCM_004747	839	2	-
GFL423	Uncharacterized protein TCM_006792	1272	7	-
GFL287	Uncharacterized protein TCM_007249	936	3	
GFL129	Uncharacterized protein TCM_014639	1245	2	-
GFL137	Uncharacterized protein TCM_015620	551	2	
GFL049	Uncharacterized protein TCM_016081	2142	2	-
GFL325	Uncharacterized protein TCM_017339	1301	1	-
GFL123	Uncharacterized protein TCM_025519	995	1	-
GFL095	Uncharacterized protein TCM_025926	316	3	
GFL446	Uncharacterized protein TCM_025992	779	0	
GFL449	Uncharacterized protein TCM_027411	1940	1	-
GFL349	Uncharacterized protein TCM_029945	816	3	
GFL118	Uncharacterized protein TCM_029952	1549	3	-
GFL253	Uncharacterized protein TCM_031178	1316	3	
GFL193	Uncharacterized protein TCM_037753	701	2	-
GFL337	Uncharacterized protein TCM_041352	626	0	
GFL346	unnamed protein product	1425	7	-
GFL314	Uridylyltransferase-related	2483	11	-
GFL161	Vacuolar sorting protein 9 domain isoform 1	1502	10	-
GFL303	Vascular plant one zinc finger protein isoform 1	2112	10	-

Table 10 Continued

Gene ID	Seq. Description	Seq. Length	#GOs	Enzyme Codes
GFL371	VIRE2-interacting protein 1, putative	700	10	-
GFL020	Vps51/Vps67 family (components of vesicular transport) protein isoform 1	843	6	-
GFL143	Winged-helix DNA-binding transcription factor family protein, putative	443	1	-
GFL229	WW domain-containing protein, putative isoform 1	1945	4	-
GFL222	WWE protein-protein interaction domain protein family, putative isoform 1	2786	17	EC:2.4.2.30
GFL415	XB3 in isoform 1	1904	10	EC:6.3.2.19; EC:3.4.21.0
GFL094	Zinc finger (C2H2 type, AN1-like) family protein	1175	4	-
GFL392	zinc finger protein	1384	21	-

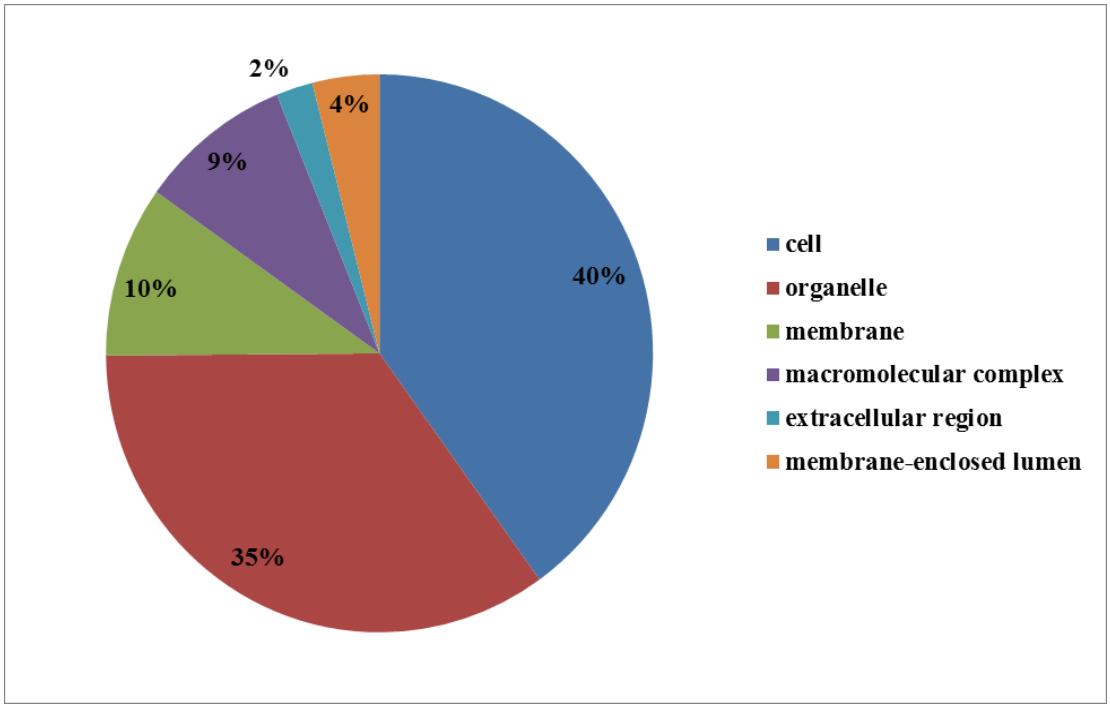

Figure 20 GO items of the *GFL* genes assigned to genes the functional category of cellular component (level 2)

Furthermore, a search was made using the KEGG database to determine the biological pathways that are associated with the *GFL* genes. The 474 *GFL* genes were mapped to 60 KEGG pathways that are categorized into 13 different biological processes, including carbohydrate metabolism, nucleotide metabolism and lipid metabolism (Table 11). The two biological pathways that included the largest numbers of the *GFL* genes were found to be the "Starch and Sucrose Metabolism" (KEGG map00500) of the carbohydrate metabolism category and the "Purine Metabolism" (KEGG map00230) of the nucleotide metabolism category (Kanehisa and Goto 2000;

Kanehisa et al. 2014). Nine of the *GFL* genes were found to encode key enzymes in starch and sucrose metabolism, including *GFL108*, *GFL226*, *GFL080*, *GFL039*, *GFL086*, *GFL395*, *GFL120*, *GFL211* and *GFL069*. Nine *GFL* genes encode enzymes in purine metabolism, including *GFL106*, *GFL006*, *GFL144*, *GFL026*, *GFL044*, *GFL069*, *GFL039*, *GFL374* and *GFL376*.

Of the 60 pathways to which the *GFL* genes were mapped, 95% were metabolism-related, including carbohydrate metabolism, nucleotide metabolism, lipid metabolism, cofactors and vitamins metabolism, energy metabolism, etc. (Table 11). Among these metabolism-related pathways, the carbohydrate and nucleotide metabolisms represented the two major biological processes in which the largest numbers of pathways are included (Fig. 21). Only six of the *GFL* genes were mapped to three KEGG non-metabolism pathways that belong to genetic information processing, environmental information processing, and organismal systems.

Table 11 KEGG pathways in which the *GFL* genes are involved

Function	No. of pathways	KEGG category	No. of pathways	KEGG pathway	Seq. in the pathway
Metabolism	57	Carbohydrate metabolism	12	Amino sugar and nucleotide sugar metabolism	8
				Ascorbate and aldarate metabolism	1
				Butanoate metabolism	1
				Fructose and mannose metabolism	4
				Galactose metabolism	4
				Glycolysis / Gluconeogenesis	4
				Glyoxylate and dicarboxylate metabolism	3
				Inositol phosphate metabolism	3
				Pentose and glucuronate interconversions	2
				Pentose phosphate pathway	2
				Pyruvate metabolism	3
				Starch and sucrose metabolism	9
		Nucleotide metabolism	10	Alanine, aspartate and glutamate metabolism	3
				Arginine and proline metabolism	4
				Cysteine and methionine metabolism	3
				Glycine, serine and threonine metabolism	2
				Lysine biosynthesis	1
				Phenylalanine metabolism	3
				Purine metabolism	9
				Pyrimidine metabolism	2
				Tryptophan metabolism	2
				Valine, leucine and isoleucine biosynthesis	1
		Lipid metabolism	7	alpha-Linolenic acid metabolism	1
				Fatty acid biosynthesis	4
				Fatty acid degradation	1
				Glycerolipid metabolism	1
				Linoleic acid metabolism	1
				Sphingolipid metabolism	1
				Synthesis and degradation of ketone bodies	1
		Metabolism of cofactors and vitamins	7	Biotin metabolism	1
				Nicotinate and nicotinamide metabolism	4
				Pantothenate and CoA biosynthesis	1
				Porphyrin and chlorophyll metabolism	1
				Riboflavin metabolism	3
				Thiamine metabolism	1
				Ubiquinone and other terpenoid-quinone biosynthesis	1

Table 11 Continued

Function	No. of pathways	KEGG category	No. of pathways	KEGG pathway	Seq. in the pathway
		Biosynthesis of other secondary metabolites	6	Aflatoxin biosynthesis	1
				Butirosin and neomycin biosynthesis	1
				Flavone and flavonol biosynthesis	1
				Flavonoid biosynthesis	1
				Phenylpropanoid biosynthesis	2
				Streptomycin biosynthesis	2
		Energy metabolism	5	Carbon fixation pathways in prokaryotes	2
				Methane metabolism	3
				Nitrogen metabolism	1
				Oxidative phosphorylation	4
				Sulfur metabolism	3
		Xenobiotics biodegradation and metabolism	4	Aminobenzoate degradation	3
				Drug metabolism - cytochrome P450	1
				Drug metabolism - other enzymes	1
				Styrene degradation	1
		Glycan biosynthesis and metabolism	3	Glycosaminoglycan biosynthesis - chondroitin sulfate /	2
				Glycosaminoglycan biosynthesis - heparan sulfate / hep	2
				Other types of O-glycan biosynthesis	1
		Metabolism of terpenoids and polyketides	2	Biosynthesis of siderophore group nonribosomal peptid	1
				Tetracycline biosynthesis	1
		Metabolism of other amino acids	1	Glutathione metabolism	2
Genetic Information Processing	1	Translation	1	Aminoacyl-tRNA biosynthesis	3
Environmental Information Processing	1	Signal transduction	1	Phosphatidylinositol signaling system	2
Organismal Systems	1	Immune system	1	T cell receptor signaling pathway	1

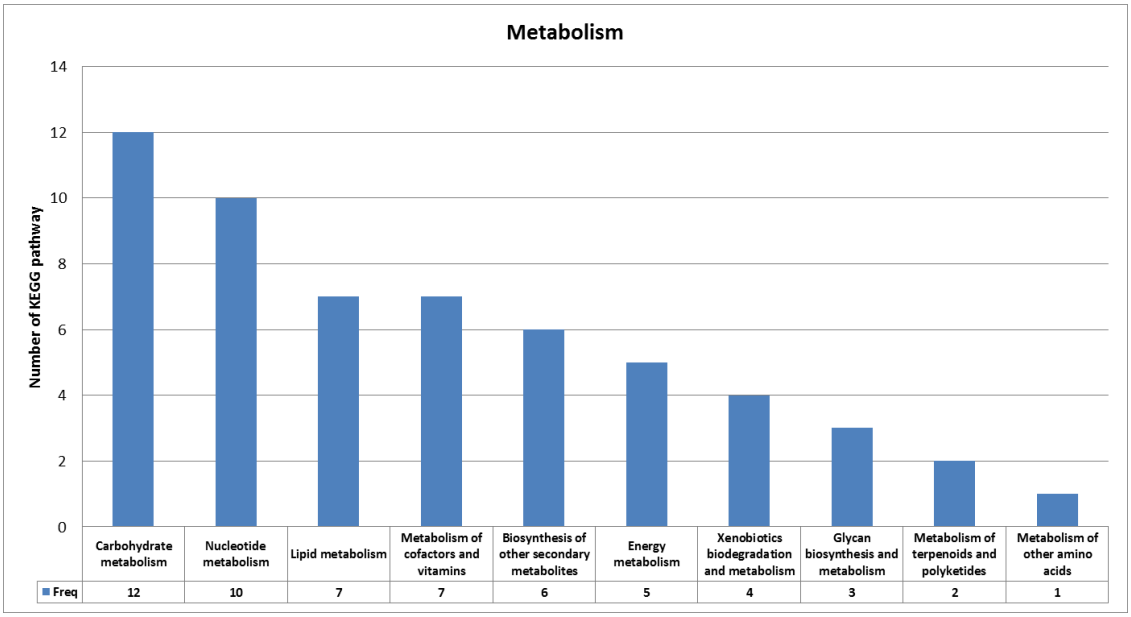

Figure 21 Metabolism-related KEGG pathways in which the *GFL* genes are involved

3.4 Molecular basis of UHML development

3.4.1 Interactions of the 474 *GFL* genes

Is there any biological relationship among the 474 *GFL* genes? In other words, do they work independent or as a team to contribute to the UFML? We hypothesized that they should be independent of each other in expression if they contribute to the UFML independently or that they should be correlated in expression or action if they work as a team to contribute to the UHML. To this hypothesis, the *GFL* genes were subjected to co-regulation network analysis using the expression data of *GFL* genes in 10-dpa fibers of the 198 RILs of the TAM 94L-25 x MNSI 1331 population. Consequently, all of the 474 *GFL* genes were found in a single regulation network ($P \leq 0.05$) (Fig. 22). The formation of a single interaction network from all 474 *GFL* genes further confirmed the

GFL genes. The *GFL* gene network consisted of 474 genes (nodes) and 51,993 interactions (edges), accounting for 46.38% of all 112,101 possible interactions of the 474 *GFL* genes. In the gene network, each of the 474 *GFL* genes interacted with an average of 220 other *GFL* genes, varying from 20 (*GFL412*) to 383 (*GFL165*) (Table 12). The number of edges connected from a node had significant correlations (r =0.1037, $P \leq 0.05$) with the absolute gene effects on UHML. More gene-gene interactions derived from a *GFL* gene and other *GFL* genes contributed more gene effects on UHML. A gene involved in a large number of gene-gene interactions contributed more gene effects on UHML than a gene involved in fewer gene-gene interactions.

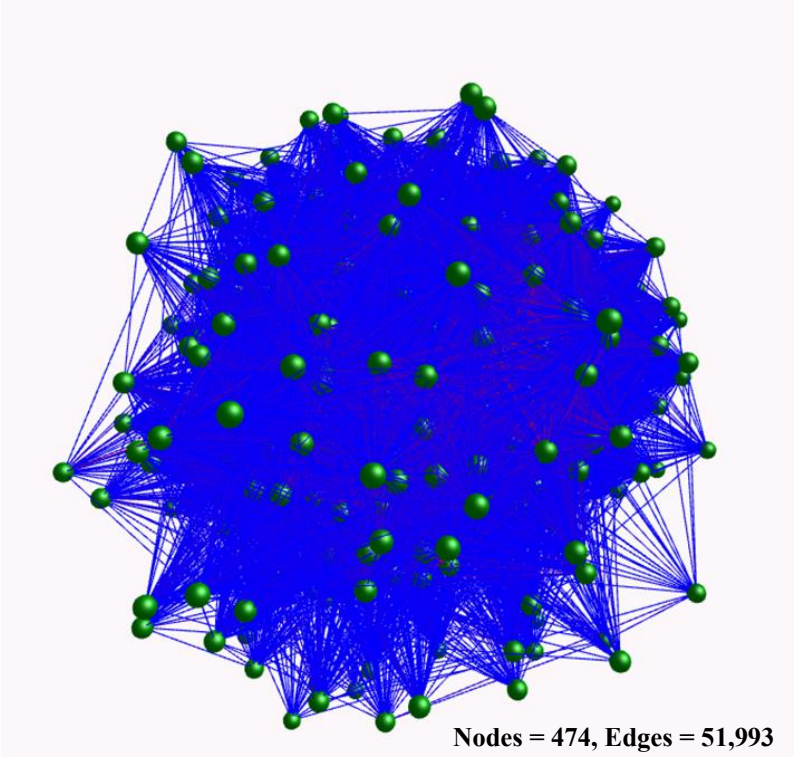

Figure 22 The co-regulation network of the 474 *GFL* genes

Table 12 Number of edges that a single *GFL* gene has in the 474 *GFL* network of the entire RIL population

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL412	20	-3.05	3.05	7.27
GFL364	23	-3.50	3.50	3.90
GFL425	23	3.33	3.33	7.72
GFL429	24	3.47	3.47	6.08
GFL434	24	3.75	3.75	6.70
GFL104	25	-4.70	4.70	1.20
GFL437	27	3.78	3.78	2.77
GFL447	28	4.04	4.04	3.34
GFL470	28	5.27	5.27	2.15
GFL421	29	3.08	3.08	4.26
GFL423	33	3.23	3.23	4.83
GFL472	36	5.48	5.48	4.24
GFL093	40	-4.76	4.76	0.86
GFL004	42	-6.47	6.47	1.15
GFL426	42	3.38	3.38	7.69
GFL340	43	-3.59	3.59	3.07
GFL299	44	-3.77	3.77	5.46
GFL461	44	4.66	4.66	1.13
GFL428	46	3.43	3.43	3.77
GFL264	47	-3.93	3.93	2.33
GFL457	49	4.58	4.58	18.91
GFL460	49	4.61	4.61	15.24
GFL344	50	-3.58	3.58	0.90
GFL448	50	4.15	4.15	11.18
GFL471	51	5.36	5.36	21.04
GFL415	59	-2.98	2.98	2.57
GFL191	61	-4.21	4.21	1.16
GFL436	61	3.77	3.77	3.55
GFL465	61	4.77	4.77	2.78
GFL320	62	-3.65	3.65	11.19
GFL397	63	-3.23	3.23	10.80
GFL440	64	3.82	3.82	3.89
GFL464	69	4.77	4.77	1.89
GFL276	70	-3.89	3.89	2.14
GFL439	71	3.82	3.82	31.40
GFL402	72	-3.17	3.17	4.45
GFL184	73	-4.24	4.24	7.51
GFL384	74	-3.36	3.36	0.88
GFL416	75	-2.96	2.96	5.29
GFL449	75	4.17	4.17	15.76
GFL072	78	-4.92	4.92	1.19
GFL458	78	4.61	4.61	4.72
GFL374	80	-3.43	3.43	4.30
GFL422	81	3.09	3.09	2.06
GFL441	82	3.83	3.83	16.43
GFL442	82	3.85	3.85	5.73
GFL116	84	-4.62	4.62	4.01
GFL243	84	-4.01	4.01	12.49

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL300	84	-3.77	3.77	0.79
GFL361	84	-3.52	3.52	5.29
GFL444	87	3.94	3.94	1.03
GFL254	90	-3.97	3.97	9.05
GFL386	90	-3.33	3.33	7.34
GFL459	91	4.61	4.61	221.72
GFL349	93	-3.56	3.56	12.44
GFL433	93	3.72	3.72	5.35
GFL469	93	5.26	5.26	5.21
GFL466	94	5.10	5.10	16.41
GFL101	98	-4.70	4.70	23.33
GFL446	99	4.04	4.04	29.73
GFL122	100	-4.57	4.57	2.63
GFL373	103	-3.44	3.44	24.84
GFL413	103	-3.02	3.02	14.18
GFL001	104	-7.92	7.92	0.75
GFL007	104	-6.26	6.26	23.88
GFL021	104	-5.82	5.82	5.25
GFL358	104	-3.54	3.54	13.80
GFL379	104	-3.40	3.40	21.39
GFL112	105	-4.65	4.65	1.54
GFL244	107	-4.01	4.01	1.03
GFL266	108	-3.92	3.92	5.39
GFL337	109	-3.60	3.60	110.58
GFL453	109	4.42	4.42	10.78
GFL113	111	-4.65	4.65	8.26
GFL263	113	-3.94	3.94	27.04
GFL411	114	-3.06	3.06	1.51
GFL435	114	3.76	3.76	2.97
GFL363	115	-3.51	3.51	4.99
GFL090	118	-4.79	4.79	23.16
GFL149	118	-4.42	4.42	1.05
GFL451	118	4.31	4.31	7.93
GFL279	121	-3.88	3.88	16.94
GFL146	122	-4.44	4.44	2.97
GFL058	123	-5.02	5.02	8.18
GFL152	123	-4.40	4.40	15.87
GFL408	124	-3.10	3.10	9.83
GFL070	127	-4.94	4.94	10.54
GFL467	127	5.19	5.19	100.01
GFL102	128	-4.70	4.70	303.56
GFL290	128	-3.81	3.81	6.33
GFL125	129	-4.55	4.55	22.74
GFL430	130	3.57	3.57	23676.20
GFL368	131	-3.46	3.46	7.03
GFL406	131	-3.12	3.12	6.94
GFL420	131	-2.64	2.64	3.03
GFL455	131	4.48	4.48	7.32

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL005	132	-6.37	6.37	1.29
GFL006	132	-6.35	6.35	1.38
GFL351	133	-3.56	3.56	10.44
GFL132	134	-4.49	4.49	4.21
GFL131	137	-4.50	4.50	22.00
GFL208	137	-4.16	4.16	2.28
GFL427	137	3.40	3.40	5.69
GFL462	137	4.70	4.70	23.65
GFL443	140	3.89	3.89	2.73
GFL456	140	4.54	4.54	14.82
GFL017	142	-5.93	5.93	4.07
GFL118	142	-4.62	4.62	8.53
GFL241	142	-4.04	4.04	63.19
GFL310	142	-3.70	3.70	3.26
GFL360	143	-3.53	3.53	4.43
GFL378	143	-3.41	3.41	24.06
GFL463	144	4.72	4.72	2.15
GFL074	147	-4.90	4.90	2.29
GFL390	147	-3.31	3.31	6.64
GFL273	148	-3.90	3.90	46.92
GFL391	148	-3.31	3.31	0.85
GFL044	150	-5.29	5.29	0.90
GFL172	150	-4.32	4.32	1.55
GFL200	151	-4.19	4.19	36.90
GFL468	152	5.21	5.21	4.63
GFL033	153	-5.42	5.42	3.73
GFL315	153	-3.68	3.68	5.83
GFL346	153	-3.58	3.58	10.60
GFL015	154	-6.04	6.04	0.81
GFL234	155	-4.07	4.07	2.70
GFL380	156	-3.40	3.40	39.29
GFL424	156	3.33	3.33	5.60
GFL059	157	-5.01	5.01	0.97
GFL419	157	-2.74	2.74	1.94
GFL474	157	6.80	6.80	635.38
GFL088	158	-4.80	4.80	13.97
GFL297	158	-3.79	3.79	9.22
GFL345	159	-3.58	3.58	3.91
GFL096	160	-4.74	4.74	35.58
GFL322	160	-3.64	3.64	283.10
GFL359	160	-3.54	3.54	11.45
GFL211	162	-4.15	4.15	170.65
GFL333	162	-3.62	3.62	10.12
GFL394	162	-3.27	3.27	36.64
GFL286	163	-3.84	3.84	8.95
GFL031	164	-5.45	5.45	5.30
GFL209	164	-4.16	4.16	13.80
GFL395	164	-3.25	3.25	4.82
GFL438	164	3.79	3.79	7.09
GFL119	165	-4.60	4.60	26.88

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL262	166	-3.94	3.94	16.65
GFL137	169	-4.47	4.47	0.94
GFL259	169	-3.94	3.94	93.27
GFL221	170	-4.11	4.11	22.28
GFL260	170	-3.94	3.94	49.06
GFL107	171	-4.68	4.68	1.18
GFL111	171	-4.66	4.66	1.51
GFL308	171	-3.73	3.73	164.44
GFL353	172	-3.56	3.56	13.76
GFL133	173	-4.49	4.49	90.41
GFL403	174	-3.15	3.15	3.19
GFL075	176	-4.90	4.90	42.41
GFL272	176	-3.91	3.91	103.67
GFL032	177	-5.45	5.45	56.10
GFL089	177	-4.80	4.80	13.39
GFL372	179	-3.44	3.44	9.05
GFL454	179	4.44	4.44	6.60
GFL376	180	-3.42	3.42	8.50
GFL404	181	-3.14	3.14	4.74
GFL473	181	6.23	6.23	56.23
GFL086	183	-4.82	4.82	11.09
GFL024	184	-5.70	5.70	22.23
GFL348	184	-3.57	3.57	1.01
GFL418	184	-2.79	2.79	2.74
GFL235	185	-4.07	4.07	5.16
GFL336	187	-3.60	3.60	211.49
GFL383	188	-3.37	3.37	4.39
GFL003	189	-6.64	6.64	0.79
GFL431	189	3.63	3.63	155.80
GFL010	191	-6.16	6.16	151.67
GFL342	191	-3.59	3.59	12.06
GFL181	193	-4.26	4.26	228.73
GFL307	193	-3.73	3.73	5.50
GFL287	194	-3.83	3.83	54.36
GFL305	194	-3.74	3.74	34.30
GFL094	195	-4.75	4.75	57.64
GFL255	195	-3.97	3.97	1.31
GFL405	197	-3.13	3.13	4.21
GFL347	198	-3.58	3.58	5.94
GFL140	199	-4.46	4.46	4.95
GFL352	199	-3.56	3.56	66.02
GFL445	199	3.97	3.97	9.27
GFL452	199	4.38	4.38	78.55
GFL154	200	-4.40	4.40	4.19
GFL288	200	-3.82	3.82	5.78
GFL042	201	-5.29	5.29	6.48
GFL174	201	-4.30	4.30	16.26
GFL314	202	-3.68	3.68	66.98
GFL393	202	-3.28	3.28	13.66

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL054	205	-5.13	5.13	4.17
GFL304	205	-3.74	3.74	14.39
GFL321	205	-3.64	3.64	1.19
GFL023	206	-5.72	5.72	10.70
GFL106	207	-4.69	4.69	145.10
GFL450	208	4.29	4.29	15.40
GFL334	211	-3.61	3.61	2.08
GFL407	211	-3.12	3.12	20.81
GFL048	212	-5.23	5.23	46.22
GFL067	212	-4.96	4.96	47.88
GFL377	212	-3.41	3.41	86.70
GFL036	213	-5.39	5.39	8.76
GFL156	213	-4.38	4.38	14.35
GFL256	213	-3.96	3.96	3.24
GFL014	214	-6.09	6.09	18.98
GFL362	214	-3.51	3.51	22.24
GFL141	215	-4.46	4.46	2.29
GFL302	215	-3.75	3.75	21.53
GFL369	216	-3.46	3.46	22.65
GFL091	217	-4.78	4.78	73.38
GFL177	218	-4.27	4.27	17.40
GFL388	219	-3.32	3.32	82.55
GFL207	220	-4.16	4.16	38.98
GFL219	221	-4.12	4.12	16.89
GFL231	221	-4.09	4.09	28.36
GFL354	221	-3.55	3.55	1134.56
GFL039	222	-5.33	5.33	83.28
GFL382	224	-3.37	3.37	6.94
GFL081	225	-4.87	4.87	3.69
GFL239	225	-4.05	4.05	909.74
GFL270	225	-3.92	3.92	56.50
GFL365	226	-3.49	3.49	19.28
GFL357	228	-3.55	3.55	22.11
GFL134	230	-4.48	4.48	1.58
GFL245	230	-4.01	4.01	19.38
GFL410	230	-3.08	3.08	38.95
GFL121	231	-4.58	4.58	160.67
GFL198	231	-4.20	4.20	12.95
GFL013	232	-6.10	6.10	5.92
GFL012	233	-6.12	6.12	3.29
GFL283	233	-3.87	3.87	29.33
GFL311	234	-3.69	3.69	39.29
GFL341	234	-3.59	3.59	53.53
GFL381	234	-3.37	3.37	153.14
GFL022	235	-5.77	5.77	9.63
GFL043	236	-5.29	5.29	25.09
GFL180	236	-4.26	4.26	68.70
GFL225	236	-4.10	4.10	318.42
GFL316	236	-3.67	3.67	11.65

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL186	237	-4.23	4.23	11.37
GFL389	237	-3.32	3.32	14.83
GFL144	238	-4.45	4.45	20.29
GFL009	240	-6.17	6.17	21.79
GFL066	240	-4.98	4.98	2.60
GFL228	241	-4.10	4.10	18.14
GFL250	241	-3.98	3.98	46.95
GFL020	243	-5.83	5.83	69.10
GFL178	243	-4.27	4.27	32.98
GFL248	246	-3.99	3.99	8.49
GFL319	247	-3.66	3.66	7.58
GFL018	248	-5.88	5.88	87.13
GFL176	249	-4.28	4.28	20.47
GFL002	250	-6.67	6.67	2.56
GFL056	250	-5.03	5.03	4.13
GFL076	251	-4.89	4.89	9.71
GFL196	251	-4.20	4.20	132.80
GFL335	251	-3.61	3.61	362.13
GFL385	251	-3.34	3.34	309.12
GFL392	251	-3.29	3.29	24.38
GFL055	252	-5.04	5.04	13.87
GFL109	252	-4.67	4.67	15.50
GFL202	253	-4.18	4.18	8.67
GFL226	253	-4.10	4.10	19.76
GFL162	254	-4.37	4.37	13.30
GFL160	255	-4.37	4.37	3.55
GFL301	255	-3.76	3.76	203.29
GFL356	255	-3.55	3.55	23.86
GFL169	256	-4.34	4.34	178.52
GFL204	256	-4.18	4.18	13.02
GFL212	256	-4.15	4.15	3.91
GFL040	257	-5.33	5.33	35.93
GFL095	257	-4.75	4.75	4.78
GFL213	257	-4.14	4.14	2.67
GFL281	257	-3.88	3.88	37.56
GFL285	257	-3.86	3.86	185.44
GFL214	258	-4.13	4.13	6.88
GFL224	258	-4.10	4.10	12.52
GFL011	259	-6.13	6.13	17.14
GFL193	259	-4.20	4.20	69.99
GFL246	259	-4.00	4.00	16.88
GFL414	259	-2.98	2.98	1.52
GFL432	259	3.68	3.68	163.63
GFL077	260	-4.89	4.89	6.40
GFL136	260	-4.47	4.47	5.06
GFL175	260	-4.29	4.29	26.98
GFL271	260	-3.91	3.91	12.50
GFL167	261	-4.34	4.34	15.35
GFL179	261	-4.26	4.26	15.05

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL323	261	-3.64	3.64	7.55
GFL142	262	-4.45	4.45	4.41
GFL400	262	-3.20	3.20	1.70
GFL035	263	-5.41	5.41	3.16
GFL063	263	-4.99	4.99	35.51
GFL161	263	-4.37	4.37	12.73
GFL326	263	-3.64	3.64	165.88
GFL227	264	-4.10	4.10	30.14
GFL105	265	-4.69	4.69	3.95
GFL150	265	-4.42	4.42	162.00
GFL232	265	-4.08	4.08	24.41
GFL257	265	-3.95	3.95	4.50
GFL399	265	-3.21	3.21	22.46
GFL197	266	-4.20	4.20	3.91
GFL324	266	-3.64	3.64	174.98
GFL371	267	-3.45	3.45	31.16
GFL019	268	-5.88	5.88	1801.61
GFL188	268	-4.22	4.22	40.51
GFL183	269	-4.26	4.26	8.24
GFL318	269	-3.66	3.66	300.14
GFL062	270	-4.99	4.99	5.59
GFL218	270	-4.12	4.12	5.79
GFL296	270	-3.79	3.79	171.60
GFL108	271	-4.67	4.67	30.97
GFL083	272	-4.86	4.86	4.02
GFL045	274	-5.28	5.28	73.60
GFL073	274	-4.90	4.90	96.62
GFL028	275	-5.60	5.60	3.25
GFL057	275	-5.02	5.02	71.15
GFL182	275	-4.26	4.26	259.33
GFL230	275	-4.09	4.09	46.89
GFL195	276	-4.20	4.20	12.61
GFL253	276	-3.97	3.97	470.65
GFL110	277	-4.66	4.66	55.77
GFL275	277	-3.89	3.89	737.16
GFL367	277	-3.47	3.47	487.68
GFL016	278	-6.00	6.00	43.27
GFL060	279	-5.00	5.00	21.39
GFL092	279	-4.77	4.77	6.55
GFL274	279	-3.89	3.89	209.27
GFL252	280	-3.97	3.97	82.34
GFL306	280	-3.74	3.74	23.32
GFL229	281	-4.10	4.10	43.50
GFL223	282	-4.11	4.11	21.02
GFL417	282	-2.93	2.93	23.60
GFL409	283	-3.08	3.08	480.68
GFL205	284	-4.17	4.17	84.51
GFL265	285	-3.93	3.93	26.70
GFL355	285	-3.55	3.55	208.29

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL128	286	-4.51	4.51	12.08
GFL164	287	-4.35	4.35	17.28
GFL216	287	-4.13	4.13	11.57
GFL071	288	-4.92	4.92	2.49
GFL098	289	-4.72	4.72	47.02
GFL069	290	-4.94	4.94	26.95
GFL151	290	-4.41	4.41	8.37
GFL190	291	-4.21	4.21	7.69
GFL249	292	-3.99	3.99	47.49
GFL258	292	-3.95	3.95	66.18
GFL050	293	-5.21	5.21	28.99
GFL087	293	-4.81	4.81	11.75
GFL240	293	-4.05	4.05	383.81
GFL251	293	-3.98	3.98	97.39
GFL284	293	-3.87	3.87	12.73
GFL267	294	-3.92	3.92	53.50
GFL312	294	-3.69	3.69	3.82
GFL157	295	-4.38	4.38	79.72
GFL215	295	-4.13	4.13	36.53
GFL034	297	-5.42	5.42	10.54
GFL206	297	-4.16	4.16	114.14
GFL280	297	-3.88	3.88	31.04
GFL261	298	-3.94	3.94	21.25
GFL084	299	-4.83	4.83	41.39
GFL277	300	-3.89	3.89	7.28
GFL027	302	-5.60	5.60	59.64
GFL120	302	-4.58	4.58	571.22
GFL143	302	-4.45	4.45	10.80
GFL398	302	-3.23	3.23	22.10
GFL171	303	-4.32	4.32	116.06
GFL238	304	-4.06	4.06	106.30
GFL126	305	-4.55	4.55	87.27
GFL331	305	-3.63	3.63	163.49
GFL387	305	-3.32	3.32	34.85
GFL135	307	-4.48	4.48	27.42
GFL138	307	-4.46	4.46	31.46
GFL201	307	-4.19	4.19	233.72
GFL291	307	-3.81	3.81	64.36
GFL046	308	-5.25	5.25	77.96
GFL189	308	-4.21	4.21	68.62
GFL192	309	-4.20	4.20	21.92
GFL194	309	-4.20	4.20	149.05
GFL269	309	-3.92	3.92	13.72
GFL203	310	-4.18	4.18	16.20
GFL293	310	-3.80	3.80	57.99
GFL375	310	-3.43	3.43	117.48
GFL401	310	-3.19	3.19	51.76
GFL061	311	-4.99	4.99	27.04
GFL097	312	-4.72	4.72	31.79

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL139	312	-4.46	4.46	419.40
GFL170	312	-4.32	4.32	129.10
GFL173	312	-4.32	4.32	123.70
GFL233	312	-4.08	4.08	362.30
GFL153	313	-4.40	4.40	387.83
GFL294	313	-3.79	3.79	17.95
GFL100	314	-4.71	4.71	39.73
GFL309	314	-3.71	3.71	33.79
GFL038	316	-5.34	5.34	19.69
GFL047	316	-5.24	5.24	699.51
GFL313	316	-3.69	3.69	11.50
GFL082	317	-4.87	4.87	96.49
GFL114	317	-4.64	4.64	10.23
GFL242	317	-4.03	4.03	199.04
GFL068	318	-4.96	4.96	79.35
GFL343	318	-3.59	3.59	149.66
GFL117	319	-4.62	4.62	283.58
GFL026	320	-5.62	5.62	121.52
GFL168	320	-4.34	4.34	38.97
GFL185	321	-4.24	4.24	41.89
GFL289	321	-3.82	3.82	32.05
GFL295	323	-3.79	3.79	313.96
GFL317	323	-3.67	3.67	422.43
GFL052	324	-5.14	5.14	33.68
GFL220	327	-4.12	4.12	78.81
GFL303	328	-3.75	3.75	129.99
GFL129	329	-4.51	4.51	51.12
GFL222	331	-4.11	4.11	516.41
GFL085	332	-4.83	4.83	187.35
GFL158	334	-4.38	4.38	122.63
GFL236	334	-4.07	4.07	193.65
GFL366	335	-3.47	3.47	74.05
GFL247	336	-3.99	3.99	76.76
GFL217	337	-4.12	4.12	273.81
GFL298	337	-3.79	3.79	43.21
GFL148	338	-4.43	4.43	50.30
GFL350	338	-3.56	3.56	689.50
GFL037	340	-5.35	5.35	33.28
GFL064	341	-4.99	4.99	25.65
GFL338	342	-3.60	3.60	82.23
GFL155	343	-4.39	4.39	71.14
GFL328	343	-3.63	3.63	121.77
GFL339	343	-3.60	3.60	224.23
GFL065	344	-4.98	4.98	29.06
GFL079	344	-4.88	4.88	341.33
GFL187	345	-4.23	4.23	134.42
GFL199	345	-4.19	4.19	52.34
GFL396	345	-3.24	3.24	49.35
GFL051	346	-5.18	5.18	104.06

Table 12 Continued

Gene ID	Number of edges	LEN ratio (%)	Abs(LEN ratio)	Aver TPM
GFL147	347	-4.44	4.44	670.11
GFL278	347	-3.89	3.89	85.34
GFL053	348	-5.14	5.14	69.38
GFL145	348	-4.44	4.44	43.26
GFL030	349	-5.50	5.50	115.26
GFL127	351	-4.54	4.54	36.11
GFL159	351	-4.37	4.37	63.05
GFL237	351	-4.06	4.06	172.01
GFL078	352	-4.88	4.88	69.40
GFL025	353	-5.65	5.65	29.98
GFL282	353	-3.87	3.87	1106.15
GFL370	354	-3.46	3.46	67.70
GFL124	355	-4.55	4.55	87.88
GFL029	357	-5.52	5.52	644.92
GFL080	357	-4.88	4.88	159.40
GFL123	357	-4.56	4.56	119.43
GFL330	357	-3.63	3.63	172.65
GFL049	358	-5.22	5.22	417.68
GFL099	358	-4.71	4.71	88.91
GFL103	358	-4.70	4.70	130.65
GFL329	358	-3.63	3.63	97.91
GFL210	359	-4.16	4.16	159.98
GFL292	360	-3.80	3.80	167.16
GFL332	361	-3.62	3.62	59.87
GFL008	362	-6.22	6.22	345.32
GFL166	363	-4.34	4.34	57.59
GFL041	365	-5.30	5.30	211.70
GFL130	365	-4.50	4.50	397.25
GFL327	365	-3.63	3.63	65.35
GFL163	369	-4.36	4.36	177.25
GFL268	370	-3.92	3.92	464.80
GFL115	374	-4.64	4.64	109.54
GFL325	382	-3.64	3.64	138.46
GFL165	383	-4.35	4.35	177.36

3.4.2 Variations of *GFL* gene networks and UHMLs

3.4.2.1 Variation of the 474 *GFL* gene network among RILs with different UHMLs

The next question was whether or not the *GFL* gene network is stable across RILs with different UHML phenotypes and if not, what would happen to the UHML. To answer this question, the 198 RILs and the parents of the TAM 94L-25 x NMSI 1331 population were grouped in a quantile of 20% into five groups based on their UHMLs, with each group having 40 lines. The group having the shortest fibers was defined as G1 while the group having the longest fibers was defined as G5. The co-regulation networks of the 474 *GFL* genes were constructed for each group (Fig. 23) and examined in terms of number of nodes, number of edges and the node components of the network. Under a significance level of $P \leq 0.05$, 465 (98.10%) of the 474 *GFL* genes were included in the gene network for G1, G2 and G3, but the numbers of edges varied from 19,443 for G1, 19,327 for G2 and 27,580 for G3. From G3 to G4, the number of nodes in the network was increased by one, whereas the number of nodes in the network was reduced by four from G4 to G5. The numbers of edges varied dramatically - 27,580, 16,835 and 16,195 from G3, G4, and G5 (Table 13). Consequently, UHML was increased by 2.64 mm for G3, when 8,137 edges were added to the G1 network. The loss of three nodes from the G3 network led to the loss of 11,385 edges for the G5 network, which resulted in a UHML increase of 2.92 mm for G5. These results indicated that UHML was determined not only by the variation of number of nodes, but also by the variation of numbers of

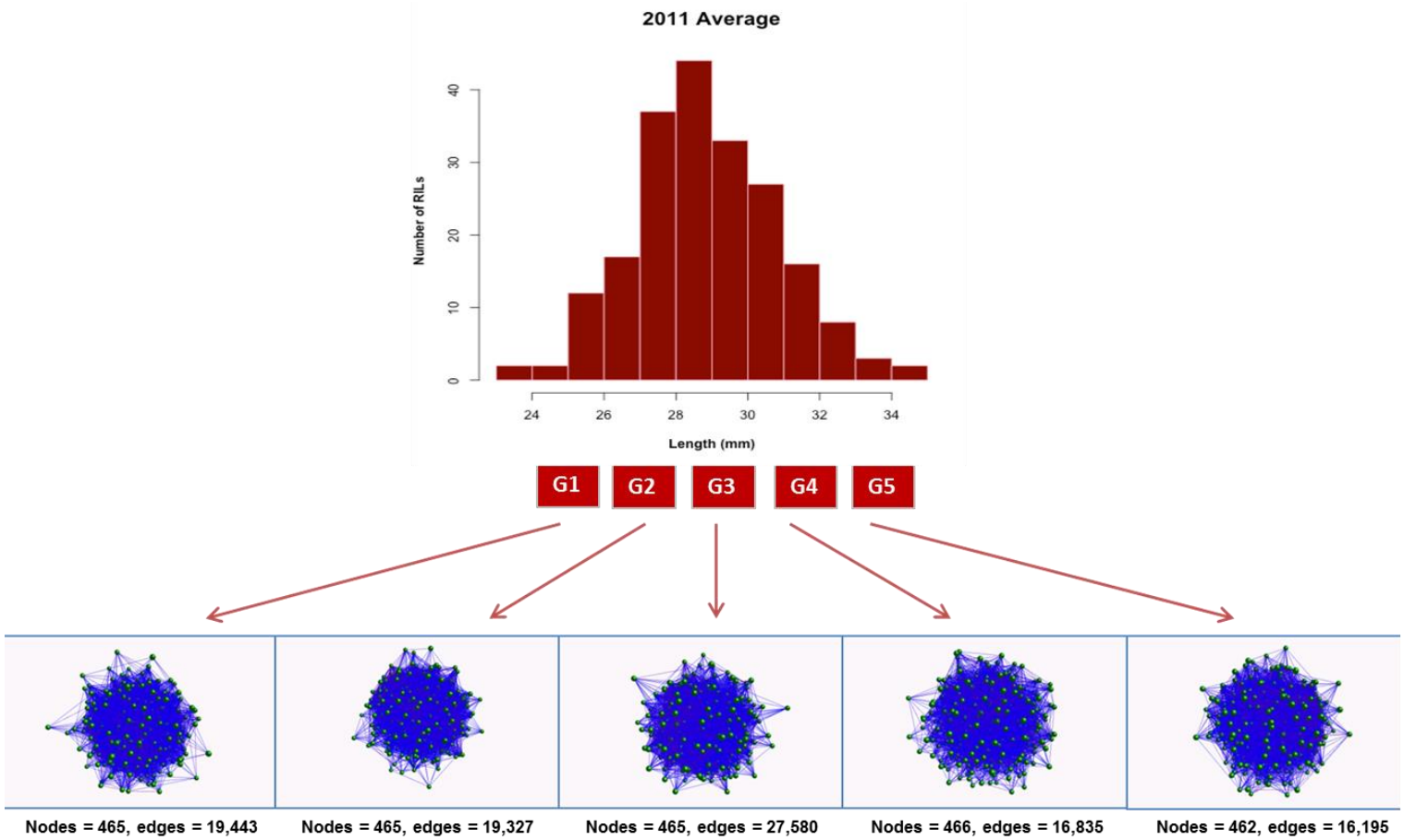

Figure 23 Association of the variations of *GFL* gene networks with the UHML of the RILs of the TAM 94L-25 x NMSI 1331 population

Table 13 Number of edges and nodes in the gene x gene networks among groups with different UHMLs

	G1	G2	G3	G4	G5
Average length within the group (mm)	26.16	27.76	28.80	29.85	31.72
Total node	474	474	474	474	474
No. of significant edges	19443	19327	27580	16835	16195
No. of node in edges*	465	465	465	466	462
No. of unique edges*	6043	5564	9841	4763	5033
Ratio of unique edges (%)	31.08	28.79	35.68	28.29	31.08
Markov clusters	6	7	5	10	8

(*) significant at the 0.05probability level

edges. Therefore, the dynamic of GFL genes and their interactions led to UHML variations between groups.

Moreover, the composition of nodes in the *GFL* gene networks may also vary with network variation, thus influencing UHML. It was noted that although the numbers of nodes of the G1, G2 and G3 networks were the same, the node compositions of the networks varied among the three groups of RILs (Table 14). For example, both the gene networks of G1 and G2 consisted of 465 *GFL* genes (nodes), but three of the nodes, *GFL115*, *GFL351* and *GFL457*, were found to be specific for the G1 network and three of them, *GFL093*, *GFL088* and *GFL434*, were specific for the G2 network. Therefore, there were a total of six nodes different between the G1 and G2 networks, accounting for 0.65% [6/(465 + 465)] of the nodes constituting the networks. Similar changes in the node compositions of the *GFL* gene networks were observed between other pairs of UHML RIL groups. Among all 10 possible pairs of the RIL groups with different UHMLs, the node compositions of the networks changed by three nodes between G3 and G4, and by 11 nodes between G3 and G5. These findings suggested that the node

Table 14 Pairwise comparison of the nodes of the *GFL* gene networks between the RIL groups with different UHMLs

Subject	G1	G2	G3	G4	G5
Mean of UHLMd of RIL group (in)	1.03	1.093	1.134	1.175	1.249
No. of edges in the *GFL* gene network	19443	19327	27580	16835	16195
No. of nodes in the *GFL* gene network	465	465	465	466	462
Nodes specific for RIL group network pair	*GFL115*	*GFL093*			
	GFL351	*GFL088*			
	GFL457	*GFL434*			
	GFL234		*GFL088*		
	GFL428		*GFL470*		
	GFL383			*GFL088*	
	GFL428			*GFL434*	
				GFL470	
	GFL286				*GFL450*
	GFL036				*GFL434*
	GFL272				
	GFL115				
	GFL457				
		GFL234	*GFL457*		
		GFL093	*GFL351*		
		GFL428	*GFL115*		
		GFL434	*GFL470*		
		GFL093		*GFL457*	
		GFL383		*GFL470*	
		GFL428		*GFL351*	
				GFL115	
		GFL272			*GFL351*
		GFL093			*GFL450*
		GFL088			
		GFL036			
		GFL286			
			GFL383	*GFL234*	
				GFL434	
			GFL286		*GFL234*
			GFL088		*GFL450*
			GFL036		*GFL434*
			GFL272		*GFL428*
			GFL457		
			GFL115		
			GFL470		
				GFL286	*GFL450*
				GFL036	*GFL383*
				GFL088	*GFL428*
				GFL272	
				GFL457	
				GFL470	
				GFL115	

variation caused the variations of number of edges between RIL groups with the same number of nodes and that the variations of gene networks could be attributed to the node component changes of the networks. Such changes of node composition of the networks may also play a role in UHML.

Although the numbers of edges varied dramatically among the *GFL* gene networks of different fiber-length groups of RILs (G1 through G5), the largest number of edges was observed for the group with a median UHML (G3). For the two RIL groups with the shortest and longest fibers (G1 and G5), the numbers of edges of their networks were fewer than those of the G3 network by 8,137 and 11,385 edges, respectively (Table 13). This difference may be attributed to both the node constitution of edges, as observed above, and the number of edges connecting each gene in the networks of different fiber-length groups. To determine the variations in number of edges for each gene (node) constituting the *GFL* gene networks across the RIL groups with different UHMLs (G1 through G5), the network of each group was scrutinized (Table 15). The results showed that the number of edges of each *GFL* gene varied across the gene networks of the RIL groups with different UHMLs. For example, gene *GFL325* had the largest numbers of edges in the gene networks of G1, G2, and G3 that had shorter fibers, whereas it had fewer edges in the gene networks of G4 and G5 that had longer fibers. Similarly, variations in number of edges for individual genes were observed across the gene networks of different fiber-length RIL groups. Some of the genes did not have any edges in some fiber-length RIL groups, e.g., *GFL115* did not have any edges in G2 and G5, whereas it had 2, 1, and 1 edges in G1, G3, and G4, respectively.

Table 15 Edge number variation of each gene constituting the *GFL* gene networks among the five fiber-length RIL groups, and its Pearson's correlation coefficients with that of UHML

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL001	10	13	15	2	11	-0.206
GFL002	22	9	81	14	39	0.211
GFL003	3	3	1	1	1	-0.824
GFL004	22	7	20	7	11	-0.494
GFL005	59	11	12	53	10	-0.437
GFL006	16	26	50	34	21	0.172
GFL007	0	0	0	0	0	NA
GFL008	50	27	66	39	33	-0.272
GFL009	15	25	57	27	43	0.565
GFL010	36	69	22	37	54	0.112
GFL011	59	23	44	48	53	0.107
GFL012	95	36	109	14	16	-0.641
GFL013	74	45	38	42	24	-0.916*
GFL014	28	21	36	9	43	0.284
GFL015	68	32	46	29	19	-0.866
GFL016	76	86	52	63	67	-0.444
GFL017	29	72	22	47	34	-0.089
GFL018	29	41	45	19	17	-0.550
GFL019	83	95	115	73	77	-0.295
GFL020	44	79	71	70	46	-0.064
GFL021	2	4	3	2	1	-0.534
GFL022	10	38	48	12	11	-0.186
GFL023	4	10	18	10	2	-0.136
GFL024	28	23	46	7	19	-0.350
GFL025	83	73	125	66	74	-0.171
GFL026	83	118	122	106	68	-0.312
GFL027	114	106	93	105	68	-0.866
GFL028	16	22	27	16	19	0.036
GFL029	122	135	158	111	95	-0.519
GFL030	87	94	151	77	104	0.114
GFL031	38	24	17	25	36	-0.027
GFL032	5	9	14	8	5	-0.062
GFL033	13	19	17	1	28	0.294
GFL034	0	0	0	0	0	NA
GFL035	4	12	33	16	13	0.303
GFL036	3	2	3	2	0	NA
GFL037	123	102	144	97	101	-0.400
GFL038	32	30	48	27	11	-0.575
GFL039	33	103	131	17	38	-0.201
GFL040	19	21	13	26	40	0.762

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL041	125	137	209	134	97	-0.256
GFL042	0	0	0	0	0	NA
GFL043	29	33	85	25	18	-0.190
GFL044	25	30	7	27	17	-0.326
GFL045	14	15	36	21	12	-0.014
GFL046	64	75	141	105	109	0.598
GFL047	26	59	103	38	32	-0.044
GFL048	59	70	86	92	55	0.055
GFL049	142	126	237	149	119	-0.117
GFL050	79	83	67	73	90	0.293
GFL051	126	122	162	65	96	-0.472
GFL052	9	11	27	10	15	0.241
GFL053	90	79	111	63	42	-0.695
GFL054	23	22	23	30	13	-0.406
GFL055	25	15	22	17	6	-0.832
GFL056	25	11	38	9	18	-0.214
GFL057	43	45	68	52	25	-0.354
GFL058	36	51	37	27	16	-0.758
GFL059	27	11	15	11	39	0.359
GFL060	86	81	122	77	79	-0.161
GFL061	140	160	140	68	79	-0.775
GFL062	78	58	39	60	53	-0.553
GFL063	3	2	1	5	1	-0.181
GFL064	15	6	16	5	12	-0.203
GFL065	28	22	35	19	21	-0.415
GFL066	35	27	46	14	24	-0.436
GFL067	20	26	65	44	22	0.127
GFL068	44	39	58	34	34	-0.404
GFL069	129	119	100	24	40	-0.866
GFL070	16	19	22	16	20	0.357
GFL071	106	79	82	31	37	-0.896*
GFL072	15	6	41	9	7	-0.166
GFL073	40	45	91	49	66	0.430
GFL074	16	6	9	3	14	-0.158
GFL075	29	111	48	71	30	-0.162
GFL076	62	72	106	62	41	-0.368
GFL077	3	5	2	3	2	-0.486
GFL078	111	125	186	101	76	-0.375
GFL079	124	150	172	141	118	-0.175
GFL080	104	98	147	71	52	-0.584

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL081	23	20	16	22	28	0.462
GFL082	60	53	77	46	38	-0.561
GFL083	67	59	108	14	36	-0.453
GFL084	3	3	2	4	2	-0.259
GFL085	140	118	171	112	119	-0.321
GFL086	27	27	45	63	23	0.162
GFL087	68	66	165	78	104	0.316
GFL088	0	3	8	2	0	NA
GFL089	62	91	47	68	38	-0.546
GFL090	11	15	5	6	13	-0.089
GFL091	40	11	36	32	34	0.073
GFL092	46	70	85	77	33	-0.191
GFL093	0	3	0	0	0	NA
GFL094	33	31	86	62	51	0.416
GFL095	11	52	22	24	22	-0.010
GFL096	39	30	40	51	25	-0.219
GFL097	44	37	50	27	45	-0.087
GFL098	51	34	45	31	9	-0.891*
GFL099	143	136	222	132	142	-0.033
GFL100	36	62	72	34	22	-0.413
GFL101	4	5	3	2	6	0.204
GFL102	20	32	8	12	9	-0.617
GFL103	132	134	203	111	145	0.036
GFL104	16	18	18	14	17	-0.100
GFL105	19	18	28	9	14	-0.400
GFL106	54	63	52	36	46	-0.599
GFL107	60	21	69	23	39	-0.300
GFL108	18	25	38	20	35	0.558
GFL109	27	82	63	36	54	0.130
GFL110	10	16	12	9	7	-0.566
GFL111	22	64	23	11	52	0.153
GFL112	17	9	30	4	10	-0.293
GFL113	3	7	14	6	11	0.574
GFL114	101	57	134	72	25	-0.570
GFL115	2	0	1	1	0	NA
GFL116	30	50	39	12	29	-0.365
GFL117	97	73	153	70	73	-0.244
GFL118	8	10	4	4	6	-0.531
GFL119	31	13	58	52	12	-0.086
GFL120	14	16	14	13	10	-0.795

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL121	57	54	98	65	37	-0.256
GFL122	18	32	26	14	16	-0.401
GFL123	68	64	80	52	24	-0.772
GFL124	22	19	24	19	19	-0.433
GFL125	23	33	23	22	8	-0.732
GFL126	21	34	29	23	18	-0.394
GFL127	85	74	143	96	47	-0.301
GFL128	3	8	6	3	5	0.004
GFL129	76	74	99	67	87	0.223
GFL130	100	94	125	68	51	-0.689
GFL131	9	1	1	4	2	-0.558
GFL132	6	4	2	2	1	-0.943*
GFL133	20	29	41	21	26	0.100
GFL134	12	5	18	11	3	-0.387
GFL135	93	81	121	79	81	-0.246
GFL136	115	32	165	64	36	-0.394
GFL137	64	20	41	12	19	-0.730
GFL138	36	35	32	37	15	-0.753
GFL139	35	54	57	62	33	-0.011
GFL140	19	62	11	81	51	0.429
GFL141	57	33	124	52	101	0.454
GFL142	23	37	74	18	30	-0.017
GFL143	96	54	168	122	35	-0.239
GFL144	132	22	100	45	50	-0.525
GFL145	141	171	214	115	140	-0.205
GFL146	45	24	26	34	7	-0.813
GFL147	53	51	75	37	37	-0.463
GFL148	30	33	55	21	28	-0.178
GFL149	26	14	25	5	3	-0.809
GFL150	25	51	54	56	45	0.547
GFL151	20	24	32	12	22	-0.119
GFL152	16	12	21	10	15	-0.132
GFL153	155	134	197	121	106	-0.516
GFL154	40	14	34	17	10	-0.719
GFL155	25	20	29	19	17	-0.572
GFL156	63	66	40	43	58	-0.363
GFL157	13	17	15	15	12	-0.329
GFL158	145	175	203	148	96	-0.520
GFL159	56	93	111	80	90	0.457
GFL160	37	31	71	9	33	-0.183

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL161	58	52	105	38	63	-0.007
GFL162	69	51	89	30	24	-0.654
GFL163	168	152	168	123	75	-0.887*
GFL164	35	16	33	2	21	-0.457
GFL165	85	85	123	85	82	-0.073
GFL166	162	100	219	151	141	-0.017
GFL167	37	41	58	18	14	-0.598
GFL168	67	68	137	45	80	0.036
GFL169	20	29	53	28	29	0.210
GFL170	95	92	137	90	73	-0.336
GFL171	64	95	117	80	66	-0.083
GFL172	67	12	33	38	34	-0.368
GFL173	89	104	152	93	90	-0.059
GFL174	9	32	51	27	21	0.189
GFL175	3	4	13	3	2	-0.119
GFL176	67	67	33	25	19	-0.909*
GFL177	82	25	37	19	30	-0.691
GFL178	14	20	9	1	14	-0.321
GFL179	10	18	27	17	25	0.702
GFL180	3	1	3	2	2	-0.235
GFL181	0	0	0	0	0	NA
GFL182	41	47	71	31	52	0.107
GFL183	76	53	34	68	73	0.076
GFL184	7	30	17	9	12	-0.123
GFL185	8	1	11	6	4	-0.185
GFL186	14	11	27	23	4	-0.221
GFL187	130	142	173	177	94	-0.253
GFL188	104	35	82	109	117	0.429
GFL189	16	13	27	17	11	-0.205
GFL190	95	24	133	62	66	-0.117
GFL191	9	16	14	23	13	0.400
GFL192	53	42	44	37	25	-0.965**
GFL193	5	13	10	12	4	-0.146
GFL194	13	27	45	20	22	0.153
GFL195	41	60	97	40	64	0.210
GFL196	51	60	73	37	50	-0.248
GFL197	68	95	89	33	26	-0.699
GFL198	27	24	24	5	12	-0.770
GFL199	93	136	190	125	134	0.328
GFL200	7	9	17	7	5	-0.215

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL201	20	54	76	56	20	-0.024
GFL202	6	3	11	3	12	0.477
GFL203	36	62	81	39	37	-0.151
GFL204	13	38	56	22	29	0.180
GFL205	39	62	102	89	69	0.519
GFL206	5	2	6	2	5	0.018
GFL207	6	18	7	13	14	0.393
GFL208	75	51	22	26	19	-0.891*
GFL209	11	13	16	12	12	0.085
GFL210	78	84	135	68	80	-0.063
GFL211	15	8	9	14	11	-0.173
GFL212	80	45	141	40	56	-0.211
GFL213	13	4	7	10	5	-0.504
GFL214	11	13	30	4	11	-0.125
GFL215	63	81	151	47	81	0.033
GFL216	36	51	70	29	41	-0.082
GFL217	23	29	53	38	25	0.119
GFL218	39	16	55	1	16	-0.439
GFL219	48	38	109	65	58	0.230
GFL220	81	62	135	49	71	-0.152
GFL221	27	60	47	43	56	0.568
GFL222	81	75	114	56	77	-0.178
GFL223	14	19	21	23	13	0.003
GFL224	68	118	165	43	60	-0.257
GFL225	42	40	87	31	49	0.049
GFL226	16	33	61	69	87	0.979**
GFL227	13	12	23	4	3	-0.544
GFL228	34	62	21	84	27	-0.003
GFL229	88	88	81	58	37	-0.932*
GFL230	112	31	131	73	56	-0.324
GFL231	16	13	38	18	7	-0.221
GFL232	49	82	130	38	50	-0.155
GFL233	107	140	122	138	135	0.632
GFL234	1	4	0	2	2	NA
GFL235	11	9	15	16	16	0.787
GFL236	6	3	2	3	2	-0.789
GFL237	70	71	72	54	56	-0.770
GFL238	61	55	113	16	31	-0.406
GFL239	88	95	111	104	45	-0.535
GFL240	14	19	18	9	12	-0.468

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL241	32	24	34	12	12	-0.771
GFL242	167	97	165	125	55	-0.716
GFL243	8	11	23	8	5	-0.215
GFL244	16	20	32	22	84	0.827
GFL245	5	11	19	8	9	0.159
GFL246	107	101	51	79	36	-0.853
GFL247	137	114	172	89	108	-0.400
GFL248	9	10	21	12	12	0.244
GFL249	23	33	69	60	24	0.144
GFL250	7	5	3	2	1	-0.973**
GFL251	15	34	18	25	20	0.051
GFL252	76	73	114	129	87	0.410
GFL253	36	31	71	27	26	-0.216
GFL254	7	7	23	1	7	-0.098
GFL255	59	46	45	25	12	-0.975**
GFL256	27	20	18	53	25	0.237
GFL257	4	15	20	17	8	0.204
GFL258	111	73	147	61	76	-0.372
GFL259	28	46	8	19	12	-0.577
GFL260	9	24	9	5	17	0.039
GFL261	46	55	77	50	61	0.347
GFL262	30	7	41	18	37	0.279
GFL263	32	20	15	17	11	-0.908*
GFL264	16	5	1	4	6	-0.577
GFL265	70	58	117	16	62	-0.217
GFL266	15	4	10	12	16	0.305
GFL267	74	73	170	53	43	-0.267
GFL268	146	151	223	143	116	-0.294
GFL269	128	108	193	73	116	-0.191
GFL270	36	25	27	11	17	-0.824
GFL271	38	49	48	40	18	-0.648
GFL272	3	4	3	2	0	NA
GFL273	15	30	90	20	33	0.141
GFL274	86	130	99	84	56	-0.607
GFL275	5	11	5	6	6	-0.130
GFL276	3	1	3	4	7	0.795
GFL277	72	81	86	91	72	0.114
GFL278	134	127	176	113	96	-0.494
GFL279	15	11	12	12	6	-0.880*
GFL280	1	5	7	2	5	0.375

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL281	5	6	9	3	2	-0.516
GFL282	93	94	167	75	63	-0.320
GFL283	22	44	30	39	39	0.555
GFL284	38	31	65	52	44	0.337
GFL285	113	104	170	107	39	-0.541
GFL286	11	2	12	3	0	NA
GFL287	73	103	122	56	94	0.036
GFL288	8	7	26	20	25	0.788
GFL289	45	58	120	56	62	0.165
GFL290	33	26	11	17	27	-0.328
GFL291	48	12	58	26	25	-0.309
GFL292	58	39	74	35	50	-0.195
GFL293	32	56	123	66	67	0.364
GFL294	141	111	150	99	139	-0.080
GFL295	38	31	32	26	19	-0.973**
GFL296	66	63	43	74	42	-0.466
GFL297	74	76	67	62	59	-0.914*
GFL298	112	123	209	89	130	0.030
GFL299	34	27	22	20	16	-0.977**
GFL300	33	27	39	19	19	-0.647
GFL301	17	48	28	18	19	-0.254
GFL302	23	28	44	13	26	-0.084
GFL303	23	13	17	4	13	-0.619
GFL304	84	15	72	24	53	-0.286
GFL305	23	36	73	42	38	0.287
GFL306	1	1	2	2	1	0.202
GFL307	22	14	17	18	17	-0.380
GFL308	56	76	59	51	24	-0.750
GFL309	32	23	39	19	23	-0.426
GFL310	3	4	1	2	3	-0.206
GFL311	31	45	87	39	26	-0.122
GFL312	61	33	75	30	37	-0.417
GFL313	19	13	30	12	10	-0.392
GFL314	74	54	62	53	29	-0.916*
GFL315	26	18	83	41	51	0.431
GFL316	40	75	78	29	28	-0.414
GFL317	9	8	8	9	3	-0.760
GFL318	60	63	59	37	31	-0.874
GFL319	60	85	64	103	74	0.376
GFL320	6	2	15	3	5	-0.045

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL321	6	7	12	6	4	-0.284
GFL322	72	75	65	25	37	-0.793
GFL323	59	45	169	34	48	-0.101
GFL324	55	60	61	70	52	0.006
GFL325	179	183	241	152	135	-0.473
GFL326	13	26	32	12	19	0.012
GFL327	44	40	50	33	32	-0.648
GFL328	17	31	36	22	30	0.403
GFL329	114	131	195	127	95	-0.204
GFL330	23	28	28	20	17	-0.633
GFL331	12	9	19	12	13	0.187
GFL332	138	106	190	95	121	-0.188
GFL333	21	5	30	6	3	-0.493
GFL334	64	26	85	52	34	-0.288
GFL335	85	104	90	129	66	-0.163
GFL336	51	59	56	73	59	0.517
GFL337	20	35	29	14	37	0.313
GFL338	91	77	147	90	75	-0.140
GFL339	167	172	226	168	151	-0.223
GFL340	35	34	5	22	27	-0.322
GFL341	32	17	46	17	24	-0.215
GFL342	2	10	23	8	3	-0.017
GFL343	38	21	55	19	21	-0.380
GFL344	18	36	21	15	27	0.042
GFL345	11	33	52	13	26	0.132
GFL346	3	7	30	22	8	0.282
GFL347	30	29	50	41	51	0.803
GFL348	43	22	21	72	12	-0.178
GFL349	8	4	8	6	7	-0.032
GFL350	91	96	123	79	68	-0.487
GFL351	1	0	2	1	2	NA
GFL352	43	21	37	32	17	-0.667
GFL353	31	33	9	13	16	-0.673
GFL354	45	60	64	67	77	0.977**
GFL355	50	53	87	61	46	-0.043
GFL356	19	41	82	12	28	-0.038
GFL357	90	23	76	28	33	-0.581
GFL358	27	17	111	24	21	-0.041
GFL359	5	2	2	2	5	0.036
GFL360	17	3	43	6	6	-0.205

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL361	31	12	38	15	7	-0.576
GFL362	9	10	41	11	10	0.017
GFL363	10	13	56	17	38	0.492
GFL364	5	34	13	7	9	-0.190
GFL365	10	1	7	2	3	-0.563
GFL366	113	128	212	151	147	0.356
GFL367	55	77	82	56	65	0.039
GFL368	29	20	8	11	11	-0.802
GFL369	17	20	48	17	11	-0.181
GFL370	34	28	38	28	29	-0.371
GFL371	48	71	97	55	96	0.611
GFL372	23	27	50	22	8	-0.390
GFL373	30	24	87	8	17	-0.212
GFL374	10	19	15	17	19	0.716
GFL375	68	107	110	77	90	0.174
GFL376	19	18	29	8	19	-0.165
GFL377	15	28	19	26	22	0.372
GFL378	21	37	9	12	9	-0.601
GFL379	15	42	29	45	23	0.208
GFL380	68	79	48	31	31	-0.842
GFL381	13	17	26	19	13	0.018
GFL382	29	31	33	40	35	0.722
GFL383	4	2	1	0	1	NA
GFL384	9	4	61	23	11	0.114
GFL385	13	22	31	18	17	0.096
GFL386	1	3	3	2	2	0.220
GFL387	12	10	14	9	6	-0.709
GFL388	71	85	91	60	67	-0.359
GFL389	27	22	60	18	17	-0.225
GFL390	1	5	10	6	5	0.426
GFL391	13	24	18	10	26	0.380
GFL392	66	103	124	52	61	-0.278
GFL393	22	21	34	47	12	-0.031
GFL394	14	18	8	12	8	-0.651
GFL395	5	10	11	1	2	-0.480
GFL396	35	48	53	17	21	-0.547
GFL397	17	10	17	15	12	-0.329
GFL398	50	25	66	28	7	-0.612
GFL399	14	33	53	55	18	0.180
GFL400	16	10	15	4	6	-0.758

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL401	73	62	120	53	37	-0.430
GFL402	21	60	23	6	10	-0.497
GFL403	10	4	7	6	6	-0.470
GFL404	40	23	32	20	21	-0.763
GFL405	27	23	59	11	6	-0.421
GFL406	22	45	58	15	59	0.420
GFL407	25	50	48	28	29	-0.142
GFL408	17	28	8	25	13	-0.217
GFL409	95	135	126	82	90	-0.370
GFL410	25	16	70	26	20	-0.032
GFL411	33	25	11	14	11	-0.870
GFL412	6	5	14	14	7	0.306
GFL413	3	3	4	5	4	0.680
GFL414	48	66	75	52	47	-0.195
GFL415	11	12	10	33	9	0.175
GFL416	10	11	63	19	26	0.277
GFL417	36	48	95	51	50	0.202
GFL418	33	67	36	8	66	0.155
GFL419	39	16	66	19	23	-0.241
GFL420	16	3	11	4	3	-0.701
GFL421	1	4	5	8	12	0.992***
GFL422	5	16	24	6	24	0.543
GFL423	4	11	6	11	22	0.864
GFL424	5	7	7	6	2	-0.569
GFL425	9	17	50	8	19	0.115
GFL426	24	12	22	11	23	-0.045
GFL427	6	4	8	3	6	-0.053
GFL428	1	1	0	0	1	NA
GFL429	12	17	12	13	24	0.693
GFL430	5	10	4	7	29	0.753
GFL431	13	28	35	27	14	-0.013
GFL432	38	46	90	33	31	-0.179
GFL433	30	12	26	32	23	0.029
GFL434	0	1	0	2	2	NA
GFL435	24	24	36	19	48	0.639
GFL436	43	69	30	14	36	-0.451
GFL437	6	3	7	4	3	-0.480
GFL438	14	4	13	8	9	-0.271
GFL439	39	18	23	12	11	-0.867
GFL440	18	10	14	8	27	0.394

Table 15 Continued

	G1	G2	G3	G4	G5	
Number of edges	19443	19327	27580	16835	16195	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
	Number of edges					*r*
GFL441	8	2	33	6	10	0.087
GFL442	48	14	34	13	11	-0.746
GFL443	13	31	81	19	12	-0.085
GFL444	12	7	12	5	11	-0.159
GFL445	66	77	136	82	86	0.251
GFL446	34	10	21	6	26	-0.243
GFL447	29	29	14	24	17	-0.665
GFL448	27	15	57	19	22	-0.076
GFL449	13	6	41	6	10	-0.076
GFL450	0	0	0	0	2	NA
GFL451	8	13	11	5	7	-0.427
GFL452	83	110	128	94	38	-0.526
GFL453	20	20	33	22	21	0.088
GFL454	37	14	21	19	54	0.414
GFL455	5	7	3	7	3	-0.348
GFL456	3	5	46	48	27	0.600
GFL457	1	0	3	2	0	NA
GFL458	22	37	25	16	30	0.006
GFL459	29	8	16	14	20	-0.256
GFL460	2	4	2	2	4	0.384
GFL461	8	15	52	23	8	0.030
GFL462	31	12	10	15	31	0.072
GFL463	29	17	24	16	62	0.600
GFL464	20	36	16	38	51	0.745
GFL465	15	17	78	12	20	0.024
GFL466	47	14	79	16	26	-0.246
GFL467	60	50	104	26	30	-0.422
GFL468	39	7	19	18	26	-0.219
GFL469	98	22	68	33	21	-0.706
GFL470	0	0	1	2	0	NA
GFL471	3	23	5	11	8	0.002
GFL472	11	6	11	8	9	-0.179
GFL473	7	14	7	7	17	0.522
GFL474	46	46	31	21	33	-0.679

(*) significant at $P \leq 0.05$; (**) significant at $P \leq 0.01$; (***) significant at $P \leq 0.001$

To determine whether number of edges from a node is associated with UHML, Pearson's correlation analysis was used to test the relationship between number of edges from a single node across different fiber-length groups and its corresponding average UHML in each group. In consideration of the lack of edges from a single gene in some fiber-length groups and the sufficient number of samples in statistical analyses, *GFL* genes that did not have edges in any one of the fiber-length groups were excluded from the test. Significant correlations between number of edges from 20 *GFL* genes and UHML were observed. These *GFL* genes included *GFL013*, *GFL071*, *GFL098*, *GFL132*, *GFL163*, *GFL176*, *GFL192*, *GFL208*, *GFL226*, *GFL229*, *GFL250*, *GFL255*, *GFL263*, *GFL279*, *GFL295*, *GFL297*, *GFL299*, *GFL314*, *GFL354*, and *GFL421*. The gain or loss of edges for these genes has significant influence on UHML, thus the interaction of genes may be an important determinant of UHML. For example, *GFL354* was observed as having an increasing number of edges from shorter to longer fiber groups, whereas *GFL255* had reduced number of edges from shorter to longer fiber groups. Of the *GFL* genes that did not have significant correlations between the number of edges and UHML, 72 tended to have an increasing number of edges from the short fiber groups (G1 and G2) to the middle length fiber group (G3) and then have an decreasing number of edges from the middle length fiber group (G3) to the longer fiber groups (G4 and G5). For instance, UHML was increased by 1.60 mm when *GFL325* gained four additional edges relative to the network of G1. When *GFL325* gained 58 additional edges in the G2 network, UHML was increased by 1.04 mm. However, when 106 edges were eliminated from *GFL325* in the G3 network, UHML was increased by

2.92 mm. Such tendencies were not only found for *GFL325*, but also observed for many of other gene nodes in the networks. Therefore, UHML could be improved by either gain or loss of nodes or edges from individual nodes in different controlling patterns, thus suggesting another aspect of the mechanism underlying UHML development.

3.4.2.2 Shared nodes and edges among the *GFL* gene networks of fiber-length groups

To further scrutinize the impacts of the edge variation of the gene networks on UHML, we further dissected and compared the *GFL* gene networks of the five UHML RIL groups. From G1 to G5, a total of 2,242 edges (Fig. 24) were found to be shared among the *GFL* gene networks of the UHML groups. Further examination showed that these edges were connected by 211 nodes in the networks. The nodes and edges shared among the networks are listed in Table A-1. Of the shared nodes, *GFL339* interacted with the largest number of *GFL* genes (53 nodes) in the networks of all five fiber-length groups, from the shortest UHML G1 through the longest UHML G5, followed by *GFL340*, *GF0341*, *GFL342*, etc. This finding suggested that *GFL339* might play a leading role in the network of each fiber-length group. Of the 2,242 shared edges, 38 were found in the networks of all five fiber-length RIL groups.

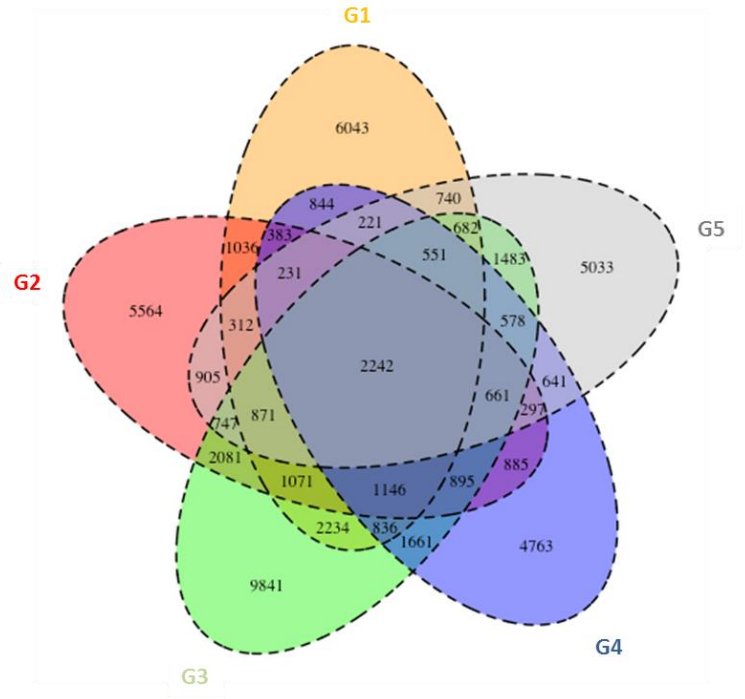

Figure 24 Numbers of unique and shared edges among the *GFL* gene networks of different UHML RIL groups

3.4.2.3 Nodes and edges unique to the *GFL* network of each UHML RIL group

The above systems analysis suggested that the variation of number of edges in the *GFL* gene networks played an important role in UHML. Therefore, to investigate the dynamics of the networks in UHML development, the edges that were only observed in or unique to the network of each fiber-length group were studied (Table 16). There were 6,043 (31.08% of the total edges of the network) edges unique for G1, 5,564 (28.79%) for G2, 9,841 (35.68%) for G3, 4,763 (28.29%) for G4, and 5033 (31.08%) for G5 (Fig. 24). Some of the unique edges might be contributed by the genes that are playing critical roles in the networks, i.e., that have a large number of edges according to our results of

Table 16 Summary of the edges of each gene unique to the *GFL* gene networks of each fiber-length group derived from 474 *GFL* genes, and their Pearson correlation coefficients with of UHML and corresponding number of unique edges of *GFL* genes

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL001	9	11	12	1	10	-0.200
GFL002	11	3	39	4	8	-0.066
GFL003	2	1	0	1	1	NA
GFL004	13	4	15	3	7	-0.383
GFL005	47	7	8	40	8	-0.436
GFL006	12	21	41	25	19	0.235
GFL007	0	0	0	0	0	NA
GFL008	7	3	14	3	2	-0.342
GFL009	4	12	22	8	14	0.400
GFL010	19	35	11	17	28	0.084
GFL011	25	11	15	19	21	-0.032
GFL012	56	18	61	10	8	-0.652
GFL013	40	24	21	19	11	-0.951*
GFL014	18	14	21	6	32	0.405
GFL015	51	20	33	17	13	-0.826
GFL016	25	37	11	14	25	-0.267
GFL017	11	35	16	22	25	0.316
GFL018	10	7	13	2	4	-0.588
GFL019	13	13	27	11	18	0.213
GFL020	7	20	21	16	4	-0.228
GFL021	0	0	1	0	1	NA
GFL022	2	18	21	3	5	-0.120
GFL023	3	1	4	1	2	-0.246
GFL024	21	16	37	4	14	-0.317
GFL025	14	12	24	9	16	0.051
GFL026	13	15	30	20	9	-0.113
GFL027	37	23	17	35	20	-0.450
GFL028	6	11	14	8	11	0.400
GFL029	18	16	26	14	9	-0.534
GFL030	14	11	27	14	13	-0.006

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL031	18	19	8	13	17	-0.219
GFL032	0	4	6	2	2	NA
GFL033	10	16	15	0	24	NA
GFL034	0	0	0	0	0	NA
GFL035	4	5	24	6	4	-0.003
GFL036	1	0	1	2	0	NA
GFL037	18	7	17	12	11	-0.364
GFL038	6	4	10	5	2	-0.420
GFL039	15	43	75	7	15	-0.177
GFL040	7	8	4	7	17	0.669
GFL041	12	7	40	11	4	-0.165
GFL042	0	0	0	0	0	NA
GFL043	11	9	43	9	10	-0.035
GFL044	21	24	6	19	17	-0.275
GFL045	4	4	15	5	4	0.009
GFL046	4	9	29	14	20	0.597
GFL047	7	7	43	6	10	0.042
GFL048	16	11	19	20	15	0.214
GFL049	16	18	51	14	20	0.041
GFL050	19	21	14	21	21	0.227
GFL051	13	8	32	7	10	-0.116
GFL052	1	2	9	2	2	0.084
GFL053	15	15	36	11	5	-0.343
GFL054	10	8	7	14	3	-0.405
GFL055	7	2	6	6	2	-0.483
GFL056	6	4	15	2	6	-0.054
GFL057	8	12	16	14	5	-0.200
GFL058	16	23	16	12	14	-0.498
GFL059	25	8	9	8	31	0.216
GFL060	21	14	36	16	19	-0.052

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL061	35	48	34	11	14	-0.749
GFL062	33	24	16	21	19	-0.752
GFL063	3	1	0	3	0	NA
GFL064	4	2	1	1	0	NA
GFL065	3	3	6	0	6	NA
GFL066	13	7	23	7	9	-0.201
GFL067	12	7	30	13	9	-0.041
GFL068	11	9	13	4	6	-0.624
GFL069	49	37	35	8	12	-0.900*
GFL070	10	18	21	11	13	0.005
GFL071	54	42	31	11	16	-0.912*
GFL072	12	3	36	7	6	-0.121
GFL073	3	7	27	6	15	0.388
GFL074	10	5	6	2	10	-0.075
GFL075	13	62	21	30	18	-0.141
GFL076	22	27	55	21	17	-0.176
GFL077	2	2	0	0	0	NA
GFL078	10	11	48	11	9	-0.036
GFL079	13	22	31	16	10	-0.236
GFL080	19	20	41	9	6	-0.428
GFL081	10	10	9	16	13	0.601
GFL082	16	13	16	7	13	-0.459
GFL083	25	25	50	7	10	-0.434
GFL084	1	1	1	2	0	NA
GFL085	28	12	30	15	14	-0.497
GFL086	19	22	29	47	18	0.207
GFL087	16	9	56	23	18	0.117
GFL088	0	3	7	1	0	NA
GFL089	16	27	6	19	11	-0.341
GFL090	6	8	4	5	10	0.415

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL091	28	5	21	21	16	-0.218
GFL092	14	20	33	28	11	-0.031
GFL093	0	3	0	0	0	NA
GFL094	12	16	42	23	16	0.161
GFL095	5	25	7	8	12	0.014
GFL096	9	3	11	16	7	0.189
GFL097	11	5	10	4	12	0.085
GFL098	11	8	9	3	4	-0.861
GFL099	22	13	50	17	27	0.145
GFL100	9	17	20	6	5	-0.421
GFL101	3	5	1	0	2	NA
GFL102	8	24	4	8	8	-0.264
GFL103	11	4	36	5	20	0.235
GFL104	14	16	14	11	14	-0.343
GFL105	4	6	11	1	1	-0.404
GFL106	17	15	20	11	13	-0.523
GFL107	38	13	44	14	24	-0.312
GFL108	3	6	13	2	11	0.444
GFL109	10	34	38	13	10	-0.215
GFL110	1	7	2	1	2	-0.179
GFL111	12	31	7	7	21	0.003
GFL112	11	7	25	3	9	-0.142
GFL113	3	5	13	5	9	0.488
GFL114	25	16	53	20	6	-0.344
GFL115	1	0	0	0	0	NA
GFL116	22	31	24	9	26	-0.164
GFL117	10	4	42	11	18	0.230
GFL118	6	7	4	2	3	-0.776
GFL119	14	6	36	39	7	0.091
GFL120	2	4	1	0	2	NA

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL121	11	8	37	9	7	-0.108
GFL122	11	23	18	11	11	-0.294
GFL123	13	11	24	12	6	-0.350
GFL124	4	4	2	1	1	-0.897*
GFL125	19	26	17	20	6	-0.714
GFL126	8	12	13	4	1	-0.665
GFL127	18	9	38	23	5	-0.216
GFL128	2	2	0	2	0	NA
GFL129	10	8	14	4	14	0.211
GFL130	11	12	23	9	11	-0.080
GFL131	8	1	0	4	2	NA
GFL132	3	2	2	2	0	NA
GFL133	9	10	13	6	8	-0.328
GFL134	4	2	12	3	0	NA
GFL135	23	21	38	31	16	-0.148
GFL136	28	4	64	12	9	-0.224
GFL137	38	15	21	6	9	-0.830
GFL138	10	11	7	15	1	-0.497
GFL139	3	12	13	14	5	0.145
GFL140	8	35	7	51	27	0.429
GFL141	22	16	63	21	35	0.254
GFL142	3	16	31	5	7	-0.025
GFL143	23	12	58	23	9	-0.183
GFL144	51	13	33	13	17	-0.664
GFL145	24	15	35	7	15	-0.370
GFL146	36	16	18	29	6	-0.710
GFL147	7	2	16	2	3	-0.228
GFL148	2	3	15	0	9	NA
GFL149	22	7	19	3	3	-0.742
GFL150	5	9	9	7	7	0.220

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL151	5	4	6	2	6	0.068
GFL152	9	9	12	6	11	0.135
GFL153	26	16	44	13	15	-0.318
GFL154	26	5	21	9	6	-0.635
GFL155	5	4	7	4	6	0.261
GFL156	23	24	15	15	23	-0.222
GFL157	1	5	4	3	2	0.021
GFL158	19	34	42	14	13	-0.367
GFL159	7	14	19	15	16	0.679
GFL160	12	5	29	3	6	-0.217
GFL161	23	25	49	13	25	-0.070
GFL162	19	11	35	9	9	-0.329
GFL163	38	23	27	11	7	-0.936*
GFL164	9	2	9	2	2	-0.598
GFL165	13	10	16	6	12	-0.213
GFL166	33	17	36	15	12	-0.651
GFL167	10	11	19	6	2	-0.533
GFL168	9	12	44	12	11	0.028
GFL169	4	8	20	6	10	0.265
GFL170	16	15	36	16	14	-0.073
GFL171	9	12	27	6	11	-0.022
GFL172	44	9	19	19	24	-0.387
GFL173	14	17	36	17	18	0.133
GFL174	5	15	40	16	16	0.271
GFL175	1	0	7	0	1	NA
GFL176	26	23	14	8	7	-0.941*
GFL177	48	13	11	8	17	-0.640
GFL178	2	7	3	0	4	NA
GFL179	3	8	16	7	10	0.444
GFL180	2	0	1	0	0	NA

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL181	0	0	0	0	0	NA
GFL182	11	10	23	5	22	0.391
GFL183	28	14	9	29	21	-0.036
GFL184	4	18	11	6	7	-0.111
GFL185	3	1	5	3	1	-0.259
GFL186	5	4	12	7	2	-0.185
GFL187	14	15	24	26	18	0.480
GFL188	21	7	7	27	28	0.482
GFL189	2	4	7	4	3	0.150
GFL190	32	9	54	16	28	-0.022
GFL191	6	12	9	18	11	0.520
GFL192	18	13	14	6	6	-0.914*
GFL193	1	2	3	1	1	-0.161
GFL194	1	3	12	9	5	0.440
GFL195	8	20	33	14	18	0.257
GFL196	12	21	27	12	15	-0.039
GFL197	27	41	43	13	12	-0.588
GFL198	8	4	8	4	1	-0.778
GFL199	13	8	34	14	13	0.062
GFL200	4	4	9	3	4	-0.063
GFL201	4	11	21	13	7	0.162
GFL202	3	1	6	2	5	0.388
GFL203	8	9	17	3	6	-0.279
GFL204	3	11	28	6	8	0.086
GFL205	7	8	21	19	20	0.815
GFL206	3	0	3	1	3	NA
GFL207	1	8	3	2	5	0.200
GFL208	56	39	16	17	12	-0.905*
GFL209	2	4	5	3	5	0.659
GFL210	11	9	30	4	6	-0.229

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL211	5	1	1	1	3	-0.336
GFL212	20	15	67	11	14	-0.118
GFL213	6	1	1	2	4	-0.214
GFL214	6	5	23	3	5	-0.082
GFL215	12	22	60	8	17	-0.019
GFL216	8	6	18	3	13	0.221
GFL217	2	2	14	3	3	0.073
GFL218	13	5	29	0	6	NA
GFL219	17	13	50	27	19	0.146
GFL220	10	9	43	8	15	0.093
GFL221	13	28	19	20	34	0.727
GFL222	9	4	16	4	13	0.258
GFL223	3	3	5	8	6	0.752
GFL224	13	28	59	10	17	-0.063
GFL225	9	12	35	8	6	-0.143
GFL226	9	12	28	34	52	0.976**
GFL227	2	6	11	2	2	-0.146
GFL228	11	28	10	45	13	0.163
GFL229	21	26	20	18	9	-0.818
GFL230	32	6	45	12	13	-0.338
GFL231	6	6	24	13	3	-0.038
GFL232	14	21	46	12	15	-0.073
GFL233	13	17	15	14	33	0.771
GFL234	1	2	0	1	1	NA
GFL235	5	7	12	14	13	0.877
GFL236	2	0	0	0	1	NA
GFL237	14	11	15	7	4	-0.825
GFL238	10	7	45	2	3	-0.176
GFL239	21	25	29	30	10	-0.404
GFL240	2	6	2	0	0	NA

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL241	19	14	20	9	5	-0.824
GFL242	52	14	36	17	14	-0.709
GFL243	6	8	17	6	2	-0.304
GFL244	10	12	16	13	65	0.801
GFL245	4	8	12	4	5	-0.068
GFL246	41	38	13	18	10	-0.876
GFL247	19	16	36	5	11	-0.350
GFL248	3	6	10	7	5	0.275
GFL249	6	11	30	14	9	0.122
GFL250	3	1	0	1	0	NA
GFL251	4	14	5	5	5	-0.205
GFL252	14	17	29	30	16	0.270
GFL253	11	7	36	8	5	-0.159
GFL254	3	3	18	1	7	0.150
GFL255	41	23	28	17	8	-0.942*
GFL256	15	15	11	38	15	0.250
GFL257	2	3	7	5	2	0.087
GFL258	25	11	40	11	13	-0.318
GFL259	15	31	4	8	8	-0.488
GFL260	5	15	5	4	10	0.060
GFL261	8	12	23	9	18	0.456
GFL262	19	4	29	10	26	0.305
GFL263	19	12	8	9	9	-0.798
GFL264	13	2	1	3	4	-0.564
GFL265	20	9	36	6	13	-0.226
GFL266	13	4	8	8	15	0.297
GFL267	15	20	81	15	10	-0.094
GFL268	12	26	41	21	12	-0.080
GFL269	25	8	61	7	15	-0.159
GFL270	15	9	13	4	8	-0.666

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL271	11	14	22	11	5	-0.406
GFL272	2	3	2	1	0	NA
GFL273	12	15	68	13	20	0.087
GFL274	15	36	15	14	8	-0.491
GFL275	0	6	1	3	2	NA
GFL276	2	1	3	4	6	0.895*
GFL277	21	26	19	25	18	-0.335
GFL278	22	18	30	15	8	-0.627
GFL279	4	1	1	6	5	0.429
GFL280	0	1	2	1	0	NA
GFL281	4	1	3	1	1	-0.690
GFL282	14	9	46	16	10	-0.045
GFL283	9	12	10	11	12	0.659
GFL284	5	6	15	14	8	0.408
GFL285	26	28	50	17	8	-0.485
GFL286	7	2	9	2	0	NA
GFL287	24	34	45	15	26	-0.161
GFL288	4	3	13	9	14	0.808
GFL289	6	6	36	4	5	-0.056
GFL290	16	11	8	11	16	0.034
GFL291	13	6	16	7	4	-0.568
GFL292	8	3	9	3	10	0.221
GFL293	2	11	45	17	13	0.249
GFL294	29	6	24	6	26	-0.061
GFL295	14	9	7	10	3	-0.872
GFL296	19	14	14	17	14	-0.544
GFL297	28	21	23	18	23	-0.529
GFL298	17	8	56	9	19	0.033
GFL299	26	24	15	14	11	-0.938*
GFL300	22	19	26	15	11	-0.718

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL301	5	19	11	6	3	-0.379
GFL302	9	7	24	4	6	-0.180
GFL303	8	2	4	2	1	-0.816
GFL304	41	5	40	13	32	-0.101
GFL305	8	14	37	13	11	0.054
GFL306	0	0	0	0	1	NA
GFL307	14	8	12	9	9	-0.595
GFL308	26	29	23	14	15	-0.831
GFL309	5	2	5	2	1	-0.707
GFL310	1	3	0	0	3	NA
GFL311	14	22	54	17	10	-0.131
GFL312	10	6	19	6	10	-0.001
GFL313	5	4	18	4	2	-0.169
GFL314	21	19	16	13	11	-0.983**
GFL315	17	11	51	14	29	0.263
GFL316	13	36	37	11	11	-0.299
GFL317	2	1	2	2	0	NA
GFL318	24	24	19	8	4	-0.933*
GFL319	11	16	17	40	27	0.708
GFL320	3	0	13	3	4	NA
GFL321	3	2	8	3	3	0.041
GFL322	30	29	30	13	16	-0.791
GFL323	17	16	92	11	16	-0.041
GFL324	6	6	13	17	10	0.552
GFL325	24	15	38	12	12	-0.398
GFL326	5	11	16	3	9	0.053
GFL327	3	5	6	2	3	-0.242
GFL328	3	5	4	3	6	0.579
GFL329	12	13	34	9	11	-0.094
GFL330	3	5	4	3	0	NA

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL331	0	3	3	3	1	NA
GFL332	17	6	43	8	13	-0.077
GFL333	10	3	17	3	1	-0.454
GFL334	41	13	54	33	12	-0.400
GFL335	9	18	16	36	14	0.355
GFL336	15	16	15	27	11	-0.014
GFL337	11	27	15	11	26	0.378
GFL338	12	7	40	15	13	0.089
GFL339	22	8	45	18	18	-0.011
GFL340	27	29	5	16	22	-0.321
GFL341	9	8	16	3	8	-0.206
GFL342	2	4	17	7	1	-0.018
GFL343	11	6	14	4	4	-0.573
GFL344	15	33	16	13	23	0.017
GFL345	3	12	29	8	11	0.200
GFL346	3	4	20	12	7	0.312
GFL347	18	12	23	21	31	0.794
GFL348	30	17	16	58	8	-0.123
GFL349	7	3	8	3	6	-0.128
GFL350	10	7	21	9	13	0.223
GFL351	1	0	2	1	2	NA
GFL352	20	5	18	18	13	-0.107
GFL353	26	25	7	9	11	-0.748
GFL354	14	14	13	17	28	0.825
GFL355	8	9	22	12	12	0.290
GFL356	8	12	39	5	9	-0.053
GFL357	57	12	42	19	14	-0.664
GFL358	22	9	91	16	11	-0.093
GFL359	4	2	0	0	2	NA
GFL360	7	2	34	6	4	-0.050

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL361	22	11	30	13	6	-0.536
GFL362	7	4	29	2	7	-0.031
GFL363	3	9	46	14	27	0.491
GFL364	5	31	11	6	5	-0.298
GFL365	6	0	2	0	2	NA
GFL366	12	14	38	20	22	0.381
GFL367	3	13	16	10	13	0.573
GFL368	20	13	5	6	4	-0.894*
GFL369	7	10	32	10	4	-0.110
GFL370	7	6	4	3	3	-0.926*
GFL371	10	7	21	6	15	0.250
GFL372	14	20	32	11	5	-0.421
GFL373	19	9	66	5	14	-0.095
GFL374	5	14	13	13	14	0.724
GFL375	18	22	24	18	19	-0.082
GFL376	11	9	21	4	16	0.165
GFL377	1	12	5	9	9	0.516
GFL378	7	17	6	3	2	-0.583
GFL379	8	34	17	35	16	0.204
GFL380	31	42	18	20	21	-0.601
GFL381	2	1	6	4	0	NA
GFL382	8	10	12	17	16	0.908*
GFL383	4	2	1	0	1	NA
GFL384	5	2	52	15	9	0.128
GFL385	2	6	9	6	5	0.368
GFL386	1	3	3	1	1	-0.250
GFL387	4	2	3	2	1	-0.868
GFL388	21	21	26	14	25	0.108
GFL389	9	9	38	7	8	-0.058
GFL390	0	4	7	5	4	NA

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL391	8	11	8	6	13	0.389
GFL392	14	27	37	7	11	-0.296
GFL393	13	11	24	35	12	0.234
GFL394	7	9	2	6	5	-0.397
GFL395	1	4	7	0	1	NA
GFL396	4	7	10	0	3	NA
GFL397	15	8	13	13	9	-0.456
GFL398	18	8	23	5	3	-0.618
GFL399	6	16	27	29	10	0.257
GFL400	10	3	8	1	1	-0.768
GFL401	16	13	37	12	10	-0.205
GFL402	14	50	20	5	7	-0.449
GFL403	5	2	5	3	3	-0.391
GFL404	26	14	17	15	11	-0.840
GFL405	15	12	39	6	3	-0.345
GFL406	16	24	38	11	40	0.495
GFL407	11	19	16	7	7	-0.540
GFL408	13	21	6	20	9	-0.233
GFL409	13	31	33	14	24	0.149
GFL410	14	3	39	10	7	-0.111
GFL411	23	17	8	11	9	-0.839
GFL412	5	3	12	13	5	0.253
GFL413	2	2	4	5	4	0.760
GFL414	15	15	21	8	14	-0.260
GFL415	9	9	9	29	7	0.181
GFL416	7	8	57	18	19	0.241
GFL417	9	6	31	11	7	-0.019
GFL418	16	35	16	4	36	0.213
GFL419	29	11	58	12	19	-0.167
GFL420	10	2	5	2	2	-0.741

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL421	1	4	5	7	11	0.994***
GFL422	4	10	13	4	15	0.573
GFL423	4	11	6	9	20	0.824
GFL424	2	2	3	1	1	-0.555
GFL425	7	15	46	7	14	0.067
GFL426	20	8	18	6	18	-0.115
GFL427	2	0	6	0	3	NA
GFL428	0	0	0	0	0	NA
GFL429	9	15	10	8	19	0.537
GFL430	4	6	2	6	25	0.767
GFL431	3	6	15	11	4	0.159
GFL432	12	9	35	7	7	-0.172
GFL433	17	7	16	22	20	0.506
GFL434	0	1	0	2	2	NA
GFL435	11	11	22	11	31	0.744
GFL436	33	50	20	7	20	-0.591
GFL437	6	3	6	4	2	-0.674
GFL438	6	2	4	2	6	0.033
GFL439	31	12	11	8	3	-0.904*
GFL440	13	9	11	7	19	0.408
GFL441	5	2	21	3	4	-0.037
GFL442	41	11	26	9	8	-0.768
GFL443	6	16	54	7	6	-0.073
GFL444	8	4	8	3	8	-0.026
GFL445	26	20	35	13	20	-0.346
GFL446	22	4	15	5	19	-0.076
GFL447	24	25	11	21	15	-0.576
GFL448	17	11	42	11	13	-0.110
GFL449	7	5	28	4	6	-0.056
GFL450	0	0	0	0	2	NA

Table 16 Continued

	G1	G2	G3	G4	G5	
UHM (mm)	26.16	27.76	28.80	29.85	31.72	
No. of unique edges*	6043	5564	9841	4763	5033	
% of unique edges (%)	31.08	28.79	35.68	28.29	31.08	
Gene ID	Number of unique edges					*r*
GFL451	3	8	7	4	5	0.058
GFL452	12	18	45	18	25	0.322
GFL453	9	9	17	5	13	0.184
GFL454	26	11	13	13	34	0.321
GFL455	1	3	1	3	2	0.314
GFL456	1	2	18	23	9	0.521
GFL457	1	0	3	2	0	NA
GFL458	17	23	16	13	16	-0.432
GFL459	14	6	4	9	9	-0.312
GFL460	0	2	2	2	2	NA
GFL461	3	9	36	9	7	0.081
GFL462	22	4	6	9	19	-0.011
GFL463	13	7	8	5	34	0.592
GFL464	15	22	10	27	38	0.769
GFL465	8	13	60	9	12	0.022
GFL466	30	9	62	11	15	-0.218
GFL467	17	25	52	7	17	-0.143
GFL468	24	3	7	5	13	-0.369
GFL469	55	13	44	14	10	-0.702
GFL470	0	0	1	2	0	NA
GFL471	2	19	2	10	7	0.060
GFL472	8	4	8	6	7	-0.032
GFL473	2	5	3	4	10	0.823
GFL474	17	11	11	11	15	-0.199

(*) significant at $P \leq 0.05$; (**) significant at $P \leq 0.01$; (***) significant at $P \leq 0.001$

above analysis. For instance, 16 unique edges for the G3 network were contributed by *GFL165* (Table 16) that had 383 edges in the *GFL* network of the entire RIL population (Table 12).

Two network-specific nodes were found, one (*GFL450*) specific for the network of G5 and the other (*GFL093*) for that of G2 (Table 16). Unique edges connected from *GFL450* were only observed in G5, and those connected from *GFL093* were only in G2. As shown in Table 8, *GFL093* had a negative effect on UHML whereas *GFL450* had positive effect on UHML when they had increased expressions. The uniqueness of the *GFL* genes for the networks may play a critical role in the UHMLs of the RIL groups, though additional studies will be needed to test the hypothesis.

3.5 *GFL* genes of diploid cottons and molecular mechanisms of fiber length evolution in the process of cotton polyploidization

Cultivated tetraploid cottons originated through a process of hybridization and polyploidization between an A-genome diploid species and a D-genome diploid species (Zhang et al. 2014). It has been documented that fiber (seed trichomes) lengths vary dramatically from diploids to tetraploid cottons. Therefore, it was hypothesized that the *GFL* genes isolated from the tetraploid cottons must have different networks from those of diploid cottons and these differences may contribute to the observed UHML variation between diploid and tetraploid cottons. To test this hypothesis, 10-dpa fibers (for A-genome diploid species) or seed coats (for the D-genome diploid species) were collected from two A-genome diploid species and five D-genome diploid species. The genes

expressed in the tissues were sequenced and profiled in expression. The orthologous of all 474 *GFL* genes of the tetraploid cottons were identified from the genes expressed in 10-dpa fibers or seed coats of the diploids, suggesting the conservation of the *GFL* genes during the process of polyploidization from diploids to tetraploids.

Gene networks were constructed from the *GFL* orthologous genes of the A- and D-genome diploids, respectively. As a result, 428 of the 474 A-genome diploid *GFL* orthologous genes were constructed into a single network, with a total of 5,775 edges (Fig. 25) and 436 of the 474 D-genome diploid *GFL* orthologous genes were constructed into a single network, with a total of 5,395 edges (Fig. 26), under a significance level of $P \leq 0.05$ (Table 17). The number of nodes for the A-genome diploid network was 46 fewer than the number of node s of the tetraploids, and the number of nodes for the D-genome diploid network was 38 less than that of tetraploids. The number of nodes of the *GFL* gene network increased by 10.75% and 8.72% during the process of polyploidization from diploid cottons, A-genome and D-genome, to tetraploid cottons. Furthermore, the number of edges varied dramatically from diploids to tetraploids. Of all possible interactions (edges), 46.38% were found in the network of tetraploid cottons, whereas only 5.69 and 6.32% were observed in the networks of A- and D-genome diploids, respectively. The process of polyploidization from diploids to tetraploids thus led to an increase of edges by nearly 10-fold (Fig. 27). In comparison among the D-, A- and AD-genome species that have a fiber length from short to longer fibers, the numbers of nodes varied by 1.87% between D- and A-genome species and 10.75% or 8.72% between A- or D-genome species and AD-genome species whereas the numbers of edges

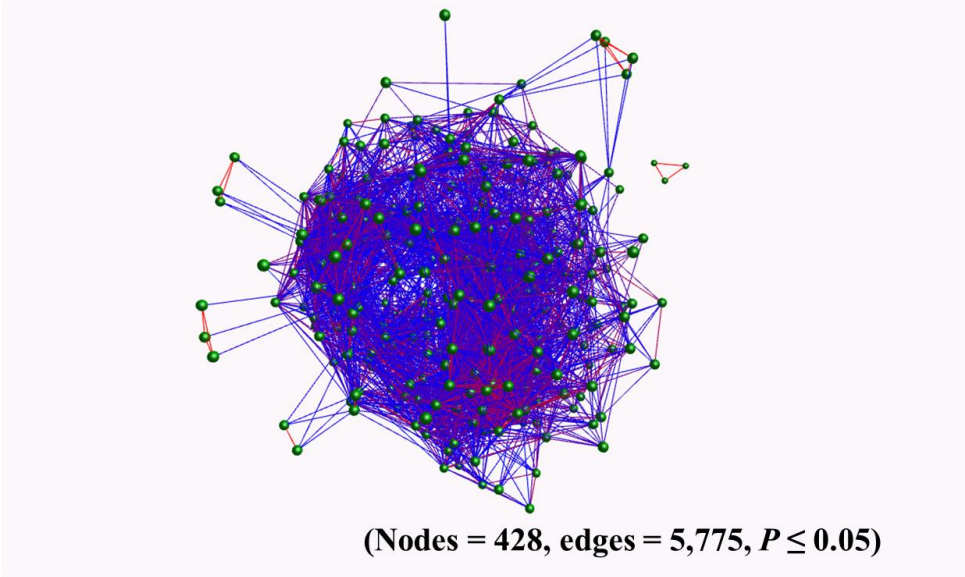

Figure 25 The network of the *GFL* orthologous genes in the A-genome diploids

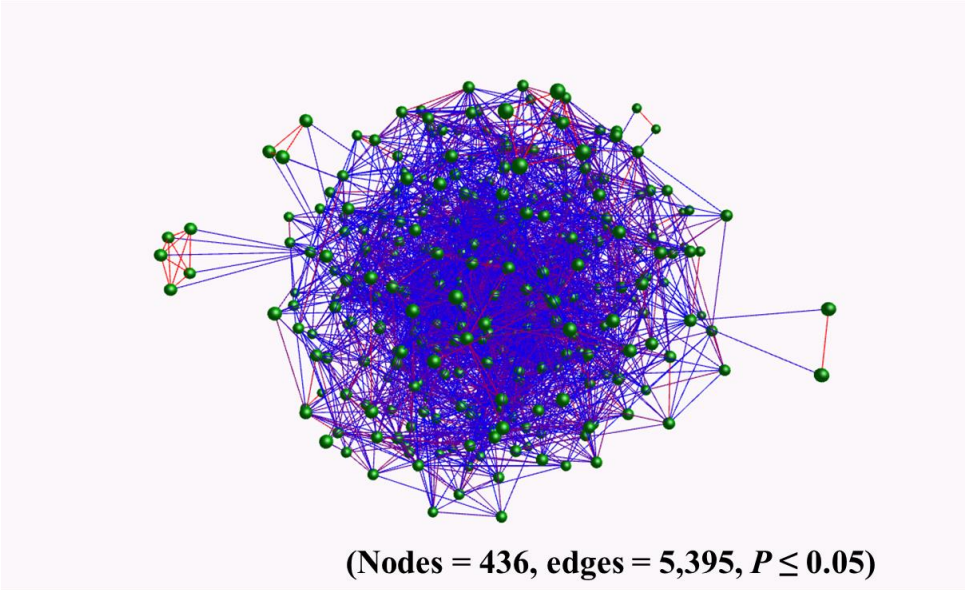

Figure 26 The network of the *GFL* orthologous genes in the D-genome diploids

Table 17 Comparison of the networks of the *GFL* genes among A- and D-genome diploid species and AD-genome tetraploid cottons

subject	tetraploid (AADD)	diploid (AA)	diploid (DD)
No. of RILs, accessions or species	200	8	9
# of *GFL* genes identified	474	474	474
# of edges in network	51,993	5,775	5,395
# of nodes in network	474	428	436
# of all possible edges	112101	91378	94830
% of edges in network	46.38	6.32	5.69
Markov clusters	3	29	31

varied by 7.04% between D- and A-genome species and by over 8-fold between A- or D-genome species and AD-genome species. This result further suggested the importance of the edge number variation of the networks in UHML performance.

Careful examination of the network edges revealed that 238 edges were shared among the networks of the A-genome diploids, D-genome diploids, and AD-genome tetraploid cottons (Fig. 28; Tables 18 and 19). The shared edges were contributed by 176 *GFL* genes, with an average of 2.7 edges per gene and a range from 1 to 12 edges (Table 19). There were 384 (238 + 146) edges shared between the networks of A- and D-genome species, whereas there were 3,262 edges common between the networks of A- and AD-genome species, and 3,006 edges common between the networks of D- and AD-genome species (Fig. 28). The network of AD-genome cotton shared more edges with that of A-genome species than D-genome species by 8.52%. In comparison, the network of AD-genome species have also many more network-specific edges (45,963) than that of either A (2,368)- or D (2,514)-genome species. These changes may have contributed to the difference of UHML among the A-, D- and AD-genome species.

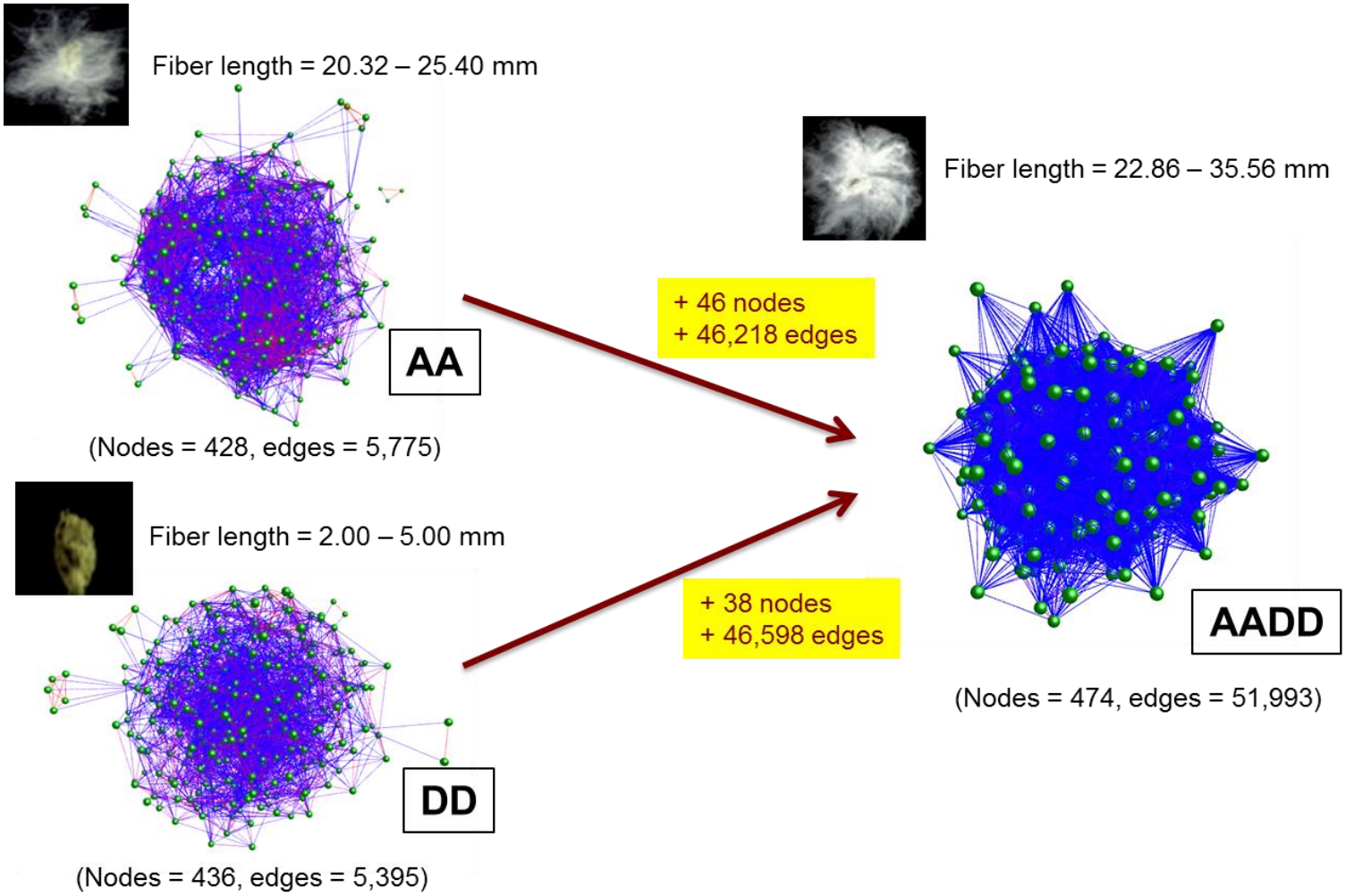

Figure 27 Node and edge number variations of the *GFL* gene network from diploids (AA and DD) to tetraploid cottons (AADD). Fiber pictures were from Paterson et al. (2012).

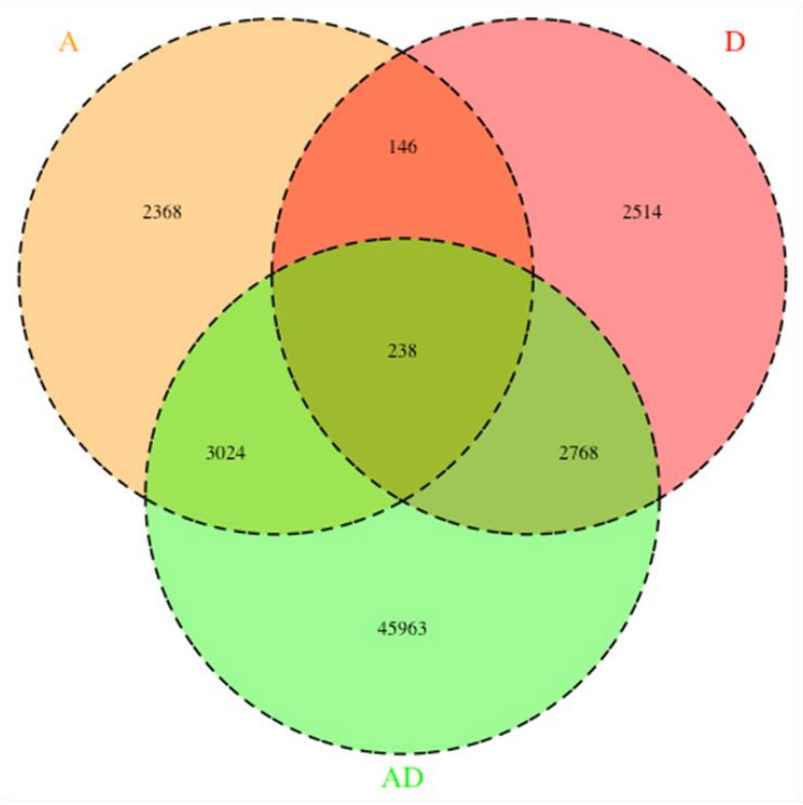

Figure 28 Variation of numbers of edges among the *GFL* networks of A-, D- and AD-genome species

Table 18 Correlation coefficients ($P \leq 0.05$) of edges shared among the *GFL* networks of tetrapploid (AADD), A-genome diploid (AA) and D-genome diploid (DD) species

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL409	*GFL123*	0.762	0.700	0.321
GFL242	*GFL370*	0.810	0.762	0.307
GFL322	*GFL274*	0.786	0.783	0.336
GFL322	*GFL207*	0.738	0.850	0.212
GFL339	*GFL256*	0.791	0.879	0.289
GFL269	*GFL309*	0.810	0.728	0.548
GFL061	*GFL220*	0.755	0.703	0.324
GFL285	*GFL232*	0.833	0.733	0.304
GFL285	*GFL301*	0.905	0.767	0.287
GFL158	*GFL157*	0.826	0.700	0.439
GFL087	*GFL163*	0.755	0.895	0.202
GFL087	*GFL026*	0.874	0.762	0.150
GFL143	*GFL319*	0.850	0.854	0.256
GFL143	*GFL230*	0.810	0.932	0.206
GFL143	*GFL371*	0.738	0.895	0.159
GFL143	*GFL220*	0.738	0.733	0.225
GFL143	*GFL095*	0.857	0.714	0.235
GFL143	*GFL098*	0.762	0.850	0.258
GFL143	*GFL331*	0.762	0.733	0.306
GFL143	*GFL128*	0.790	0.711	0.222
GFL143	*GFL115*	0.905	0.800	0.412
GFL452	*GFL055*	0.762	0.700	0.155
GFL049	*GFL190*	0.934	0.776	0.310
GFL049	*GFL180*	0.755	0.733	0.427
GFL049	*GFL034*	0.896	0.731	0.279
GFL366	*GFL173*	0.762	0.750	0.514
GFL366	*GFL190*	0.743	0.819	0.275
GFL366	*GFL127*	0.786	0.800	0.348
GFL366	*GFL237*	0.857	0.700	0.452
GFL366	*GFL394*	0.762	0.733	0.270
GFL308	*GFL295*	0.762	0.767	0.318
GFL069	*GFL246*	0.850	0.748	0.386
GFL069	*GFL127*	0.786	0.929	0.190

Table 18 Continued

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL069	*GFL159*	0.762	0.874	0.196
GFL069	*GFL120*	0.738	0.753	0.301
GFL323	*GFL129*	0.755	0.717	0.264
GFL012	*GFL202*	0.811	0.717	0.238
GFL012	*GFL084*	0.779	0.883	0.179
GFL334	*GFL244*	0.811	0.790	0.176
GFL334	*GFL060*	0.756	0.717	0.165
GFL334	*GFL223*	0.756	0.800	0.179
GFL268	*GFL065*	0.766	0.837	0.458
GFL233	*GFL325*	0.881	0.800	0.322
GFL233	*GFL244*	0.792	0.849	0.142
GFL233	*GFL223*	0.738	0.767	0.607
GFL135	*GFL079*	0.786	0.703	0.492
GFL135	*GFL190*	0.826	0.826	0.246
GFL135	*GFL026*	0.905	0.753	0.191
GFL135	*GFL030*	0.762	0.787	0.373
GFL135	*GFL130*	0.905	0.728	0.277
GFL135	*GFL225*	0.857	0.837	0.316
GFL135	*GFL343*	0.810	0.845	0.333
GFL135	*GFL207*	0.810	0.703	0.372
GFL135	*GFL198*	0.976	0.812	0.244
GFL135	*GFL236*	0.762	0.778	0.325
GFL135	*GFL007*	0.905	0.728	0.168
GFL013	*GFL081*	0.756	0.874	0.204
GFL325	*GFL244*	0.856	0.832	0.164
GFL325	*GFL163*	0.833	0.750	0.571
GFL325	*GFL060*	0.810	0.817	0.340
GFL325	*GFL046*	0.738	0.833	0.383
GFL197	*GFL080*	0.786	0.727	0.317
GFL232	*GFL329*	0.738	0.767	0.273
GFL319	*GFL103*	0.934	0.770	0.348
GFL079	*GFL293*	0.905	0.700	0.386
GFL079	*GFL052*	0.833	0.717	0.288
GFL048	*GFL349*	0.764	0.785	0.275

Table 18 Continued

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL219	*GFL329*	0.786	0.700	0.339
GFL219	*GFL230*	0.738	0.763	0.226
GFL244	*GFL060*	0.792	0.916	0.188
GFL246	*GFL120*	0.778	0.904	0.273
GFL304	*GFL448*	0.939	0.902	0.153
GFL304	*GFL016*	0.810	0.854	0.208
GFL163	*GFL389*	0.905	0.800	0.253
GFL163	*GFL038*	0.786	0.865	0.321
GFL163	*GFL330*	0.952	0.729	0.511
GFL173	*GFL190*	0.778	0.937	0.288
GFL085	*GFL162*	0.791	0.700	0.186
GFL392	*GFL321*	0.854	0.817	0.221
GFL267	*GFL190*	0.862	0.836	0.309
GFL267	*GFL026*	0.810	0.750	0.312
GFL267	*GFL225*	0.857	0.733	0.335
GFL267	*GFL343*	0.786	0.850	0.809
GFL267	*GFL198*	0.905	0.733	0.248
GFL267	*GFL180*	0.946	0.733	0.266
GFL373	*GFL400*	0.851	0.797	0.256
GFL336	*GFL377*	0.862	0.800	0.332
GFL170	*GFL188*	0.762	0.833	0.159
GFL329	*GFL230*	0.738	0.831	0.310
GFL329	*GFL121*	0.810	0.733	0.212
GFL329	*GFL168*	0.857	0.700	0.318
GFL329	*GFL220*	0.762	0.800	0.515
GFL329	*GFL095*	0.857	0.798	0.305
GFL329	*GFL081*	0.756	0.717	0.279
GFL329	*GFL152*	0.881	0.733	0.143
GFL329	*GFL098*	0.738	0.750	0.406
GFL329	*GFL309*	0.810	0.800	0.388
GFL329	*GFL115*	0.786	0.900	0.614
GFL190	*GFL026*	0.850	0.836	0.146
GFL190	*GFL030*	0.743	0.852	0.280
GFL190	*GFL225*	0.970	0.903	0.180
GFL190	*GFL237*	0.778	0.751	0.262

Table 18 Continued

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL190	*GFL343*	0.743	0.836	0.267
GFL190	*GFL198*	0.755	0.785	0.260
GFL190	*GFL180*	0.873	0.743	0.255
GFL060	*GFL230*	0.786	0.848	0.409
GFL060	*GFL363*	0.756	0.785	0.142
GFL060	*GFL220*	0.810	0.700	0.386
GFL060	*GFL073*	0.786	0.800	0.369
GFL050	*GFL124*	0.790	0.870	0.211
GFL464	*GFL147*	0.894	0.811	0.171
GFL092	*GFL203*	0.862	0.718	0.282
GFL029	*GFL424*	0.764	0.756	0.141
GFL103	*GFL051*	0.881	0.783	0.720
GFL103	*GFL456*	0.833	0.767	0.285
GFL188	*GFL044*	0.773	0.725	0.195
GFL188	*GFL377*	0.738	0.883	0.502
GFL247	*GFL398*	0.810	0.800	0.296
GFL247	*GFL331*	0.738	0.733	0.383
GFL230	*GFL220*	0.833	0.831	0.439
GFL230	*GFL095*	0.905	0.735	0.322
GFL230	*GFL081*	0.781	0.763	0.186
GFL230	*GFL115*	0.881	0.729	0.518
GFL026	*GFL394*	0.810	0.833	0.334
GFL026	*GFL343*	0.762	0.733	0.283
GFL026	*GFL124*	0.738	0.733	0.486
GFL026	*GFL185*	0.762	0.717	0.386
GFL026	*GFL236*	0.762	0.717	0.538
GFL404	*GFL171*	0.833	0.767	0.171
GFL404	*GFL052*	0.881	0.800	0.141
GFL114	*GFL363*	0.805	0.712	0.166
GFL114	*GFL318*	0.762	0.867	0.375
GFL078	*GFL043*	0.857	0.800	0.178
GFL078	*GFL220*	0.738	0.700	0.455
GFL121	*GFL327*	0.786	0.711	0.241
GFL121	*GFL227*	0.761	0.783	0.219
GFL121	*GFL211*	0.833	0.800	0.195

Table 18 Continued

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL121	*GFL193*	0.810	0.883	0.424
GFL311	*GFL175*	0.810	0.739	0.331
GFL016	*GFL275*	0.738	0.800	0.435
GFL171	*GFL020*	0.847	0.745	0.498
GFL171	*GFL253*	0.762	0.733	0.278
GFL171	*GFL377*	0.810	0.717	0.381
GFL171	*GFL045*	0.762	0.800	0.295
GFL171	*GFL052*	0.857	0.750	0.298
GFL127	*GFL020*	0.798	0.762	0.170
GFL127	*GFL280*	0.819	0.717	0.271
GFL020	*GFL401*	0.822	0.778	0.325
GFL020	*GFL130*	0.749	0.728	0.304
GFL020	*GFL192*	0.798	0.738	0.453
GFL020	*GFL198*	0.872	0.745	0.251
GFL020	*GFL178*	0.847	0.843	0.246
GFL020	*GFL185*	0.872	0.762	0.208
GFL338	*GFL073*	0.786	0.750	0.218
GFL046	*GFL256*	0.741	0.854	0.236
GFL030	*GFL343*	0.905	0.733	0.344
GFL354	*GFL371*	0.786	0.778	0.332
GFL354	*GFL211*	0.762	0.800	0.453
GFL354	*GFL236*	0.833	0.767	0.183
GFL205	*GFL105*	0.738	0.703	0.374
GFL159	*GFL185*	0.952	0.778	0.301
GFL159	*GFL063*	0.905	0.787	0.158
GFL371	*GFL306*	0.738	0.762	0.157
GFL014	*GFL385*	0.762	0.717	0.139
GFL053	*GFL214*	0.778	0.728	0.337
GFL168	*GFL327*	0.905	0.820	0.412
GFL168	*GFL115*	0.905	0.750	0.546
GFL130	*GFL198*	0.810	0.800	0.378
GFL130	*GFL240*	0.810	0.733	0.415
GFL130	*GFL120*	0.905	0.733	0.383
GFL182	*GFL389*	0.738	0.733	0.145

Table 18 Continued

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL182	*GFL275*	0.738	0.833	0.321
GFL225	*GFL343*	0.738	0.850	0.371
GFL367	*GFL120*	0.952	0.817	0.656
GFL417	*GFL398*	0.790	0.728	0.243
GFL220	*GFL073*	0.738	0.767	0.414
GFL220	*GFL095*	0.952	0.849	0.357
GFL220	*GFL309*	0.810	0.717	0.463
GFL220	*GFL115*	0.929	0.750	0.561
GFL106	*GFL180*	0.771	0.717	0.438
GFL407	*GFL152*	0.738	0.767	0.170
GFL165	*GFL038*	0.905	0.898	0.393
GFL453	*GFL066*	0.810	0.723	0.140
GFL324	*GFL198*	0.762	0.817	0.305
GFL095	*GFL098*	0.738	0.723	0.279
GFL095	*GFL327*	0.976	0.705	0.253
GFL095	*GFL309*	0.857	0.748	0.139
GFL095	*GFL115*	0.976	0.874	0.372
GFL123	*GFL134*	0.790	0.709	0.222
GFL123	*GFL124*	0.833	0.867	0.500
GFL123	*GFL387*	0.778	0.800	0.395
GFL237	*GFL394*	0.857	0.867	0.325
GFL237	*GFL343*	0.810	0.750	0.330
GFL410	*GFL194*	0.857	0.833	0.277
GFL389	*GFL084*	0.810	0.817	0.160
GFL150	*GFL377*	0.738	0.845	0.384
GFL150	*GFL052*	0.881	0.812	0.346
GFL459	*GFL471*	0.833	0.729	0.423
GFL459	*GFL427*	0.755	0.809	0.221
GFL398	*GFL098*	0.976	0.700	0.195
GFL398	*GFL065*	0.826	0.762	0.314
GFL022	*GFL055*	0.766	0.833	0.176
GFL169	*GFL424*	0.909	0.781	0.158
GFL301	*GFL211*	0.810	0.867	0.195
GFL301	*GFL034*	0.822	0.782	0.270

Table 18 Continued

Gene ID[a]	Gene ID[a]	AA	DD	AADD
GFL108	*GFL055*	0.738	0.983	0.204
GFL108	*GFL042*	0.784	0.762	0.159
GFL377	*GFL045*	0.952	0.767	0.322
GFL377	*GFL052*	0.762	0.867	0.228
GFL394	*GFL198*	0.810	0.800	0.210
GFL138	*GFL034*	0.970	0.714	0.225
GFL471	*GFL331*	0.762	0.700	0.171
GFL251	*GFL331*	0.781	0.733	0.384
GFL343	*GFL198*	0.857	0.833	0.310
GFL343	*GFL185*	0.881	0.750	0.220
GFL038	*GFL330*	0.905	0.759	0.367
GFL038	*GFL387*	0.898	0.729	0.263
GFL327	*GFL309*	0.833	0.736	0.543
GFL327	*GFL115*	0.929	0.745	0.548
GFL148	*GFL164*	0.786	0.700	0.285
GFL045	*GFL052*	0.762	0.817	0.218
GFL045	*GFL331*	0.762	0.883	0.277
GFL164	*GFL387*	0.826	0.733	0.361
GFL164	*GFL383*	0.738	0.822	0.156
GFL055	*GFL032*	0.881	0.733	0.177
GFL211	*GFL331*	0.857	0.900	0.390
GFL211	*GFL120*	0.738	0.750	0.491
GFL309	*GFL128*	0.743	0.929	0.409
GFL309	*GFL115*	0.833	0.767	0.454
GFL105	*GFL202*	0.756	0.843	0.242
GFL105	*GFL084*	0.810	0.782	0.254
GFL198	*GFL185*	0.833	0.833	0.250
GFL198	*GFL180*	0.755	0.717	0.345
GFL052	*GFL331*	0.738	0.783	0.394
GFL330	*GFL250*	0.881	0.746	0.442
GFL202	*GFL084*	0.805	0.700	0.325
GFL185	*GFL034*	0.786	0.748	0.226
GFL180	*GFL034*	0.827	0.782	0.161

a: the gene nodes forming the shared edges.

Table 19 List of nodes (*GFL* genes) that have edges common among the networks of the A-, D- and AD-genome species

GeneID	# of edges	% of shared edges
GFL256	2	0.84
GFL257	2	0.84
GFL258	1	0.42
GFL259	1	0.42
GFL260	1	0.42
GFL261	5	2.10
GFL262	1	0.42
GFL263	1	0.42
GFL264	1	0.42
GFL265	2	0.84
GFL266	9	3.78
GFL267	6	2.52
GFL268	1	0.42
GFL269	1	0.42
GFL270	3	1.26
GFL271	12	5.04
GFL272	2	0.84
GFL273	8	3.36
GFL274	5	2.10
GFL275	6	2.52
GFL276	1	0.42
GFL277	2	0.84
GFL278	1	0.42
GFL279	3	1.26
GFL280	4	1.68
GFL281	1	0.42
GFL282	1	0.42
GFL283	10	4.20
GFL284	1	0.42
GFL285	8	3.36
GFL286	2	0.84
GFL287	1	0.42
GFL288	4	1.68

Table 19 Continued

GeneID	# of edges	% of shared edges
GFL289	3	1.26
GFL290	2	0.84
GFL291	2	0.84
GFL292	1	0.42
GFL293	5	2.10
GFL294	5	2.10
GFL295	1	0.42
GFL296	1	0.42
GFL297	1	0.42
GFL298	1	0.42
GFL299	1	0.42
GFL300	1	0.42
GFL301	4	1.68
GFL302	4	1.68
GFL303	1	0.42
GFL304	6	2.52
GFL305	4	1.68
GFL306	3	1.26
GFL307	2	0.84
GFL308	2	0.84
GFL309	1	0.42
GFL310	6	2.52
GFL311	8	3.36
GFL312	1	0.42
GFL313	2	0.84
GFL314	3	1.26
GFL315	1	0.42
GFL316	2	0.84
GFL317	2	0.84
GFL318	3	1.26
GFL319	1	0.42
GFL320	11	4.62
GFL321	2	0.84
GFL322	1	0.42

Table 19 Continued

GeneID	# of edges	% of shared edges
GFL323	1	0.42
GFL324	5	2.10
GFL325	3	1.26
GFL326	3	1.26
GFL327	4	1.68
GFL328	1	0.42
GFL329	2	0.84
GFL330	1	0.42
GFL331	1	0.42
GFL332	2	0.84
GFL333	2	0.84
GFL334	2	0.84
GFL335	3	1.26
GFL336	4	1.68
GFL337	7	2.94
GFL338	1	0.42
GFL339	2	0.84
GFL340	1	0.42
GFL341	1	0.42
GFL342	1	0.42
GFL343	2	0.84
GFL344	1	0.42
GFL345	3	1.26
GFL346	1	0.42
GFL347	2	0.84
GFL348	2	0.84
GFL349	4	1.68
GFL350	9	3.78
GFL351	1	0.42
GFL352	3	1.26
GFL353	7	2.94
GFL354	1	0.42
GFL355	1	0.42
GFL356	1	0.42

Table 19 Continued

GeneID	# of edges	% of shared edges
GFL357	1	0.42
GFL358	1	0.42
GFL359	3	1.26
GFL360	3	1.26
GFL361	1	0.42
GFL362	1	0.42
GFL363	1	0.42
GFL364	1	0.42
GFL365	1	0.42
GFL366	1	0.42
GFL367	1	0.42
GFL368	1	0.42
GFL369	2	0.84
GFL370	3	1.26
GFL371	1	0.42
GFL372	5	2.10
GFL373	5	2.10
GFL374	10	4.20
GFL375	4	1.68
GFL376	1	0.42
GFL377	1	0.42
GFL378	1	0.42
GFL379	3	1.26
GFL380	9	3.78
GFL381	8	3.36
GFL382	5	2.10
GFL383	1	0.42
GFL384	1	0.42
GFL385	1	0.42
GFL386	2	0.84
GFL387	2	0.84
GFL388	1	0.42
GFL389	1	0.42
GFL390	3	1.26

Table 19 Continued

GeneID	# of edges	% of shared edges
GFL391	1	0.42
GFL392	4	1.68
GFL393	1	0.42
GFL394	1	0.42
GFL395	1	0.42
GFL396	1	0.42
GFL397	3	1.26
GFL398	1	0.42
GFL399	1	0.42
GFL400	5	2.10
GFL401	2	0.84
GFL402	3	1.26
GFL403	1	0.42
GFL404	1	0.42
GFL405	1	0.42
GFL406	7	2.94
GFL407	1	0.42
GFL408	12	5.04
GFL409	1	0.42
GFL410	3	1.26
GFL411	1	0.42
GFL412	1	0.42
GFL413	1	0.42
GFL414	7	2.94
GFL415	3	1.26
GFL416	2	0.84
GFL417	6	2.52
GFL418	2	0.84
GFL419	4	1.68
GFL420	1	0.42
GFL421	1	0.42
GFL422	2	0.84
GFL423	1	0.42
GFL424	3	1.26

Table 19 Continued

GeneID	# of edges	% of shared edges
GFL425	2	0.84
GFL426	4	1.68
GFL427	3	1.26
GFL428	1	0.42
GFL429	4	1.68
GFL430	1	0.42
GFL431	1	0.42

4. DISCUSSION AND CONCLUSION

UHML has long been assumed to be one of the most important fiber traits for cotton commerce. However, it is unknown how many genes control UHML in the cotton genome. Although a number of genes differentially expressed in developing fibers have been identified and studied extensively (Ji et al. 2003; Arpat et al. 2004; Chaudhary et al. 2008; Hovav et al. 2008; Yang et al. 2008; Rapp et al. 2010; Lacape et al. 2012; Li et al. 2013; Yoo and Wendel 2014), most of them focused on comparative studies of genes differentially expressed between different fiber developmental stages, different tissues, and/or between cultivated and wild type species. So far, a total of 21 genes that control UHML or trichome development have been cloned in cotton and other plant species. However, the underlying molecular mechanisms of UHML development in cultivated cotton species are still to be answered.

We, in this study, used a new high-throughput gene/QTL cloning system recently developed in Dr. Hongbin Zhang's Laboratory to isolate the genes controlling UHML from cultivated tetraploid cottons. A total of 474 *GFL* genes controlling UHML were isolated and characterized. This number of genes is 20-fold more than the total number of fiber-length genes that have been reported worldwide in the past 20 years. Each of the 474 *GFL* genes has relatively small effects on UHML, ranging from 2.64% to 7.92%. Of the 474 *GFL* genes, 88.6% have negative effects on UHML, and 11.4% have positive effects.

Over 70% of the 474 *GFL* genes were annotated using the available gene description with an E-value ≤ 1.0E-06. The *GFL* genes encode a diversity of proteins or enzymes. The GO analysis has revealed that 61.18% of the *GFL* genes are assigned to the Biological Process category, 62.87% to the Molecular Function category and 66.67% to the Cellular Component category. Among the GO terms assigned, the cellular process and metabolism of the Biological Process category, and the binding and catalytic of the Molecular Function category are the most important biological processes in which the *GFL* genes are involved. The enzymes or proteins encoded by the 474 *GFL* genes are mapped to 60 KEGG metabolic pathways (Table 11). "Starch and sucrose metabolism" and "Purine metabolism" are the two pathways in which most of the *GFL* genes are involved. Therefore, fiber cell extension is likely determined by a number of biological processes and metabolic pathways in which the 474 *GFL* genes are involved.

Study of the molecular mechanisms underlying fiber UHML reveals that the 474 *GFL* genes interact, forming a single interaction network, in which each *GFL* gene interacts with at least one other *GFL* gene. Within the network, the number of interaction edges of a *GFL* gene (node) varies from 20 to 383. The number of its edges or interactions with other genes in the network is significantly correlated with the effect of a *GFL* gene on fiber UHML phenotype, suggesting the roles of the gene x gene interactions in fiber UHML development. Therefore, the molecular mechanisms of controlling fiber UHML phenotype involve not only *GFL* genes *per se*, but also the degree, number and directions of their interactions.

Furthermore, this study shows that the *GFL* gene network varies among cotton RILs with different fiber UHML. Although the variations in the number of nodes (genes) are limited between cotton RILs with different UHMLs, the variations in the number of edges (interactions) are dramatic. For example, the numbers of nodes in the *GFL* gene networks remained constant across the fiber-length RIL groups G1, G2 and G3, even though their node compositions were slightly different. But, the numbers of edges across these groups differed by 42% between G1 and G3 and by 43% between G2 and G3. Variation in the number of nodes was found between the *GFL* gene networks of G4 and G5; however, the maximum of their node variations was only 11 genes, only accounting for 2% of the 474 *GFL* genes, which is much smaller than the variation of their interaction edges between the networks. Therefore, we assume that gene x gene interaction, or epistasis within the *GFL* gene network play important roles in the cotton fiber UHML development.

For most of the 474 *GFL* genes, the gain or loss of edges from the gene-gene interaction networks among different fiber-length RIL groups was not significantly correlated with the variation of UHML; however, 20 (4.22%) of them did significantly correlate, negatively or positively, with the variation of UHML ($P \leq 0.05$). These results suggest that the variation of interactions among the *GFL* genes contribute to the fiber UHML, at least for some of the genes. Therefore, we hypothesize that the unique gene-gene interactions of the *GFL* gene network may be major players in the regulation of UHML. The discovery of the large number of unique gene-gene interactions for each fiber-length RIL group, especially for the longest and shortest fiber length groups, might

be an indication of their roles in UHML development. However, further studies are needed to address how these unique interaction edges regulate fiber UHML and what molecular mechanisms trigger the variation of the unique gene x gene interactions.

One of the major economical traits that have evolved dramatically from diploid to tetraploid cotton is fiber length. This study shows that the fiber length change from diploid cottons to tetraploid cottons, as those of fiber UHMLs among different fiber-length groups, has also resulted from the variation of the numbers of *GFL* genes and their interactions. In this study, a total of 428 *GFL* orthologous genes were identified in the 10-dpa developing fibers of the A-genome diploid species, forming a total of 5,775 interaction edges, and a total of 436 *GFL* orthologous genes were identified in the 10-dpa developing fibers of the D-genome diploid species, forming a total of 5,395 interaction edges. While 8 – 10% of the number of nodes in the *GFL* gene networks has increased from A- or D- genome diploid species to tetraploid cottons, an increase of 8-fold of gene-gene interactions were observed from the diploid cottons to tetraploid cottons. This result implies that process of polyploidization might facilitate the interactions of the *GFL* genes, even though the numbers of genes controlling the trait are relatively consistent. This change might have contributed to the increased fiber UHML in the modern cultivated tetraploid cottons.

Cotton fiber UHML is a typical quantitative trait, as indicated by its continuous phenotypic variation (Fig. 4). The results of this study have provided a first line of evidence for the hypothesis that quantitative traits are controlled by numerous genes each of which has small effects. Moreover, this study has added several new findings to

the molecular basis of quantitative genetics. First, the genes controlling a quantitative trait contribute to the trait not only positively, but also negatively; the trait performance is the result of balance between the genes having positive and negative effects. Second, the genes controlling a quantitative trait encode a variety of proteins or enzymes that are involved in multiple biological processes and metabolic pathways. Third, the genes controlling a quantitative trait interact, forming an interaction network; therefore, the performance of a quantitative trait is the consequence of interactions among the genes controlling the traits. Within the gene networks, although the number of genes and functions of genes may vary among different genotypes, the gene x gene interactions play important roles in the performance of a quantitative trait. Finally, the genes involved in the gene network controlling a quantitative trait are different in number of gene x gene interactions, thus playing different roles in the trait development.

The 474 cloned *GFL* genes and findings resulted from this study have, for the first time worldwide, provided new concepts, new knowledge and new tools for development of new methods enabling enhanced cotton breeding, such as gene-based breeding. It is based on not only the genotypes or alleles of the target genes to be selected for, like marker- or genomics-assisted breeding, but also the action, action direction (positive or negative), gene x gene interaction, gene x nongene element interaction (e.g. miRNA) and G x E interactions of the genes controlling the target trait. In addition, since the markers used for gene-based breeding are derived from the cloned genes controlling the trait, the risk of genetic recombination between the markers and target genes is minimized. Therefore, gene-based breeding is far more powerful and

efficient than the currently-used marker- or genomics-assisted breeding for enhanced plant breeding.

REFERENCES

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

Argiriou A, Kalivas A, Michailidis G and Tsaftaris A. 2012. Characterization of PROFILIN genes from allotetraploid (*Gossypium hirsutum*) cotton and its diploid progenitors and expression analysis in cotton genotypes differing in fiber characteristics. *Molecular Biology Reports* **39**(4): 3523-3532.

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA and Wilkins TA. 2004. Functional genomics of cell elongation in developing cotton fibers. *Plant Molecular Biology* **54**(6): 911-929.

Avci U, Pattathil S, Singh B, Brown VL, Hahn MG and Haigler CH. 2013. Cotton fiber cell walls of *Gossypium hirsutum* and *Gossypium barbadense* have differences related to loosely-bound xyloglucan. *PloS One* **8**(2): e56315.

Basra AS and Malik CP. 1984. Development of the cotton fiber. *International Review of Cytology* **89**: 66-113.

Braden CA and Smith CW. 2004. Fiber length development in near-long staple upland cotton. *Crop Science* **44**(5): 1553 -1559

Chaudhary B, Hovav R, Rapp R, Verma N, Udall JA and Wendel JF. 2008. Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*. *Evolution & Development* **10**(5): 567-582.

Chen X, Guo W, Liu B, Zhang Y, Song X, Cheng Y, Zhang L and Zhang T. 2012. Molecular mechanisms of fiber differential development between *G. barbadense* and *G. hirsutum* revealed by genetical genomics. *PloS One* **7**(1): e30056.

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

Conesa A and Gotz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* **2008**: 1-13.

Dhindsa RS, Beasley CA and Ting IP. 1975. Osmoregulation in cotton fiber: accumulation of potassium and malate during growth. *Plant Physiology* **56**(3): 394-398.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J and Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**(10): 3420-3435.

Gotz S, Arnold R, Sebastian-Leon P, Martin-Rodriguez S, Tischler P, Jehl MA, Dopazo J, Rattei T and Conesa A. 2011. B2G-FAR, a species-centered GO annotation repository. *Bioinformatics* **27**(7): 919-924.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**(7): 644-652.

Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ and Chen XY. 2008. The HD-Zip IV gene GaHOX1 from cotton is a functional homologue of the *Arabidopsis* GLABRA2. *Physiologia Plantarum* **134**(1): 174-182.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N and Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**(8): 1494-1512.

Hao J, Tu L, Hu H, Tan J, Deng F, Tang W, Nie Y and Zhang X. 2012. GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *Journal of Experimental Botany* **63**(17): 6267-6281.

Hovav R, Udall JA, Hovav E, Rapp R, Flagel L and Wendel JF. 2008. A majority of cotton genes are expressed in single-celled fiber. *Planta* **227**(2): 319-329.

Huang GQ, Gong SY, Xu WL, Li W, Li P, Zhang CJ, Li DD, Zheng Y, Li FG and Li XB. 2013. A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. *Plant Physiology* **161**(3): 1278-1290.

Humphries JA, Walker AR, Timmis JN and Orford SJ. 2005. Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* TRANSPARENT TESTA GLABRA1 (TTG1) gene. *Plant Molecular Biology* **57**(1): 67-81.

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC and Zhu YX. 2003. Isolation and analyses of genes preferentially expressed during early cotton fiber

development by subtractive PCR and cDNA array. *Nucleic Acids Research* **31**(10): 2534-2543.

Jiang C, Wright RJ, El-Zik KM and Paterson AH. 1998. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proceedings of the National Academy of Sciences of the United States of America* **95**(8): 4419-4424.

Kanehisa M and Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**:27-30.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M and Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**: D199-D205.

Karademir E, Karademir C, Ekininci R and Gencer O. 2010. Relationship between yield, fiber length and other fiber-related traits in advanced cotton strains. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **38**(3): 111-116.

Kim HJ, Triplett BA, Zhang HB, Lee MK, Hinchliffe DJ, Li P and Fang DD. 2012. Cloning and characterization of homeologous cellulose synthase catalytic subunit 2 genes from allotetraploid cotton (*Gossypium hirsutum* L.). *Gene* **494**(2): 181-189.

Lacape JM, Claverie M, Vidal RO, Carazzolle MF, Guimaraes Pereira GA, Ruiz M, Pre M, Llewellyn D, Al-Ghazi Y, Jacobs J, Dereeper A, Huguet S, Giband M and Lanaud C. 2012. Deep sequencing reveals differences in the transcriptional landscapes of fibers from two cultivated species of cotton. *PloS One* **7**(11): e48855.

Lee J, Burns TH, Light G, Sun Y, Fokar M, Kasukabe Y, Fujisawa K, Maekawa Y and Allen RD. 2010. Xyloglucan endotransglycosylase/hydrolase genes in cotton and their role in fiber elongation. *Planta* **232**(5): 1191-1205.

Li B and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.

Li X, Yuan D, Zhang J, Lin Z and Zhang X. 2013. Genetic mapping and characteristics of genes specifically or preferentially expressed during fiber development in cotton. *PloS One* **8**(1): e54444.

Li XB, Fan XP, Wang XL, Cai L and Yang WC. 2005. The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. *The Plant Cell* **17**(3): 859-875.

Liu Q, Talbot M and Llewellyn DJ. 2013. Pectin methylesterase and pectin remodelling differ in the fibre walls of two *Gossypium* species with very different fibre properties. *PloS One* **8**(6): e65131.

Luo M, Xiao Y, Li X, Lu X, Deng W, Li D, Hou L, Hu M, Li Y and Pei Y. 2007. GhDET2, a steroid 5alpha-reductase, plays an important role in cotton fiber cell initiation and elongation. *The Plant Journal* **51**(3): 419-430.

Machado A, Wu Y, Yang Y, Llewellyn DJ and Dennis ES. 2009. The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *The Plant Journal* **59**(1): 52-62.

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

Meredith WR and Bridge RR. 1972. Heterosis and gene action in cotton, *Gossypium hirsutum* L.. *Crop Science* **12**(3): 304-310.

Michailidis G, Argiriou A, Darzentas N and Tsaftaris A. 2009. Analysis of xyloglucan endotransglycosylase/hydrolase (XTH) genes from allotetraploid (*Gossypium hirsutum*) cotton and its diploid progenitors expressed during fiber elongation. *Journal of Plant Physiology* **166**(4): 403-416.

NCC. 2012. National Cotton Council of America: United States cotton production. Available at http://www.cotton.org/econ/world/detail.cfm (verified 22 Feb. 2014).

Padmalatha KV, Patil DP, Kumar K, Dhandapani G, Kanakachari M, Phanindra ML, Kumar S, Mohan TC, Jain N, Prakash AH, Vamadevaiah H, Katageri IS, Leelavathi S, Reddy MK, Kumar PA and Reddy VS. 2012. Functional genomics of fuzzless-lintless mutant of *Gossypium hirsutum* L. cv. MCU5 reveal key genes and pathways involved in cotton fibre initiation and elongation. *BMC Genomics* **13**: 624.

Pang CY, Wang H, Pang Y, Xu C, Jiao Y, Qin YM, Western TL, Yu SX and Zhu YX. 2010. Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and *Arabidopsis* root hair elongation. *Molecular & Cellular Proteomics* **9**(9): 2019-2033.

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, Yoo MJ, Byers R, Chen W, Doron-Faigenboim A, Duke MV, Gong L, Grimwood J, Grover C, Grupp K, Hu G, Lee TH, Li J, Lin L, Liu T, Marler BS, Page JT, Roberts AW, Romanel E, Sanders WS,

Szadkowski E, Tan X, Tang H, Xu C, Wang J, Wang Z, Zhang D, Zhang L, Ashrafi H, Bedon F, Bowers JE, Brubaker CL, Chee PW, Das S, Gingle AR, Haigler CH, Harker D, Hoffmann LV, Hovav R, Jones DC, Lemke C, Mansoor S, ur Rahman M, Rainville LN, Rambani A, Reddy UK, Rong JK, Saranga Y, Scheffler BE, Scheffler JA, Stelly DM, Triplett BA, Van Deynze A, Vaslin MF, Waghmare VN, Walford SA, Wright RJ, Zaki EA, Zhang T, Dennis ES, Mayer KF, Peterson DG, Rokhsar DS, Wang X and Schmutz J. 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**(7429): 423-427.

Pu L, Li Q, Fan X, Yang W and Xue Y. 2008. The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development. *Genetics* **180**(2): 811-820.

Qu J, Ye J, Geng YF, Sun YW, Gao SQ, Zhang BP, Chen W and Chua NH. 2012. Dissecting functions of KATANIN and WRINKLED1 in cotton fiber development by virus-induced gene silencing. *Plant Physiology* **160**(2): 738-748.

Rapp RA, Haigler CH, Flagel L, Hovav RH, Udall JA and Wendel JF. 2010. Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *BMC Biology* **8**: 139.

Reiter WD. 2002. Biosynthesis and properties of the plant cell wall. *Current Opinion in Plant Biology* **5**(6): 536-542.

Ruan YL, Llewellyn DJ and Furbank RT. 2001. The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K^+ transporters and expansin. *The Plant Cell* **13**(1): 47-60.

Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, Cheng J, Wei LP, Wang ZY and Zhu YX. 2006. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *The Plant Cell* **18**(3): 651-664.

Smart LB, Vojdani F, Maeshima M and Wilkins TA. 1998. Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. *Plant Physiology* **116**(4): 1539-1549.

Smith CW and Cothren JT. 1999. Cotton : origin, history, technology, and production (ed. CW Smith, JT Cothren), pp. xiii, 850 p., [858] p. of plates : ill. (some col.) ; 826 cm. Wiley, New York.

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

Tan J, Tu L, Deng F, Hu H, Nie Y and Zhang X. 2013. A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant Physiology* **162**(1): 86-95.

Theocharidis A, van Dongen S, Enright AJ and Freeman TC. 2009. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nature Protocols* **4**(10): 1535-1550.

Ulloa M and Meredith WR, Jr. 2000. Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. *The Journal of Cotton Science* **4**(3): 161-170.

USDA. 2014. Cotton and wool outlook No. CWS-14C. Washington, D.C.

Wang H, Guo Y, Lv F, Zhu H, Wu S, Jiang Y, Li F, Zhou B, Guo W and Zhang T. 2010a. The essential role of GhPEL gene, encoding a pectate lyase, in cell wall loosening by depolymerization of the de-esterified pectin during fiber elongation in cotton. *Plant Molecular Biology* **72**(4-5): 397-406.

Wang HY, Wang J, Gao P, Jiao GL, Zhao PM, Li Y, Wang GL and Xia GX. 2009. Down-regulation of GhADF1 gene expression affects cotton fibre properties. *Plant Biotechnology Journal* **7**(1): 13-23.

Wang J, Wang HY, Zhao PM, Han LB, Jiao GL, Zheng YY, Huang SJ and Xia GX. 2010b. Overexpression of a profilin (GhPFN2) promotes the progression of developmental phases in cotton fibers. *Plant & Cell Physiology* **51**(8): 1276-1290.

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu YX, Wang J and Yu S. 2012. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics* **44**(10): 1098-1103.

Wang L, Li XR, Lian H, Ni DA, He YK, Chen XY and Ruan YL. 2010c. Evidence that high activity of vacuolar invertase is required for cotton fiber and Arabidopsis root elongation through osmotic dependent and independent pathways, respectively. *Plant Physiology* **154**(2): 744-756.

Wang L and Ruan YL. 2010. Unraveling mechanisms of cell expansion linking solute transport, metabolism, plasmodesmtal gating and cell wall dynamics. *Plant Signaling & Behavior* **5**(12): 1561-1564.

Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ and Chen XY. 2004. Control of plant trichome development by a cotton fiber MYB gene. *The Plant Cell* **16**(9): 2323-2334.

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

Wendel JF, Olson PD and Stewart JM. 1989. Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. *American Journal of Botany* **76**(12): 1795-1806.

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

Xiao YH, Li DM, Yin MH, Li XB, Zhang M, Wang YJ, Dong J, Zhao J, Luo M, Luo XY, Hou L, Hu L and Pei Y. 2010. Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. *Journal of Plant Physiology* **167**(10): 829-837.

Xu SM, Brill E, Llewellyn DJ, Furbank RT and Ruan YL. 2012. Overexpression of a potato sucrose synthase gene in cotton accelerates leaf expansion, reduces seed abortion, and enhances fiber production. *Molecular Plant* **5**(2): 430-441.

Yang YW, Bian SM, Yao Y and Liu JY. 2008. Comparative proteomic analysis provides new insights into the fiber elongating process in cotton. *Journal of Proteome Research* **7**(11): 4623-4637.

Yoo MJ and Wendel JF. 2014. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genetics* **10**(1): e1004073.

APPENDIX A

Table A-1 Nodes and numbers of edges shared among the *GFL* gene networks of different fiber-length groups

geneID	No. of edges	geneID	No. of edges	geneID	No. of edges	geneID	No. of edges	geneID	No. of edges
GFL339	53	*GFL049*	18	*GFL417*	8	*GFL201*	4	*GFL057*	2
GFL340	47	*GFL159*	18	*GFL097*	8	*GFL120*	4	*GFL193*	1
GFL341	47	*GFL060*	18	*GFL355*	8	*GFL401*	4	*GFL323*	1
GFL342	44	*GFL350*	18	*GFL277*	7	*GFL082*	4	*GFL190*	1
GFL343	43	*GFL210*	18	*GFL106*	7	*GFL240*	4	*GFL316*	1
GFL344	41	*GFL222*	17	*GFL381*	7	*GFL065*	4	*GFL018*	1
GFL345	39	*GFL171*	16	*GFL188*	7	*GFL056*	3	*GFL076*	1
GFL346	37	*GFL170*	16	*GFL068*	7	*GFL151*	3	*GFL100*	1
GFL347	37	*GFL237*	15	*GFL284*	7	*GFL267*	3	*GFL092*	1
GFL348	37	*GFL048*	14	*GFL370*	7	*GFL195*	3	*GFL430*	1
GFL349	36	*GFL147*	14	*GFL196*	7	*GFL224*	3	*GFL221*	1
GFL350	36	*GFL053*	14	*GFL242*	7	*GFL105*	3	*GFL459*	1
GFL351	35	*GFL089*	13	*GFL205*	7	*GFL253*	3	*GFL177*	1
GFL352	35	*GFL139*	13	*GFL392*	7	*GFL133*	3	*GFL180*	1
GFL353	34	*GFL292*	13	*GFL217*	6	*GFL138*	3	*GFL062*	1
GFL354	33	*GFL130*	13	*GFL124*	6	*GFL396*	3	*GFL098*	1
GFL355	33	*GFL354*	13	*GFL225*	6	*GFL211*	3	*GFL136*	1
GFL356	33	*GFL258*	13	*GFL016*	6	*GFL275*	3	*GFL249*	1
GFL357	32	*GFL375*	12	*GFL108*	6	*GFL331*	3	*GFL052*	1
GFL358	31	*GFL297*	11	*GFL050*	6	*GFL008*	3	*GFL250*	1
GFL359	31	*GFL020*	11	*GFL148*	6	*GFL317*	3	*GFL238*	1
GFL360	30	*GFL285*	11	*GFL414*	6	*GFL183*	3	*GFL341*	1
GFL361	30	*GFL293*	11	*GFL309*	6	*GFL110*	3	*GFL302*	1
GFL362	29	*GFL371*	11	*GFL215*	6	*GFL328*	3	*GFL126*	1
GFL363	28	*GFL080*	11	*GFL123*	6	*GFL312*	3	*GFL303*	1
GFL129	27	*GFL319*	10	*GFL047*	5	*GFL212*	2	*GFL204*	1
GFL019	26	*GFL388*	10	*GFL114*	5	*GFL472*	2	*GFL162*	1
GFL078	26	*GFL289*	10	*GFL087*	5	*GFL141*	2	*GFL156*	1
GFL338	24	*GFL061*	10	*GFL330*	5	*GFL038*	2	*GFL219*	1
GFL269	24	*GFL336*	10	*GFL127*	5	*GFL207*	2	*GFL167*	1
GFL165	24	*GFL073*	10	*GFL283*	5	*GFL407*	2	*GFL432*	1
GFL367	24	*GFL261*	10	*GFL229*	5	*GFL467*	2	*GFL301*	1
GFL274	23	*GFL239*	10	*GFL230*	5	*GFL232*	2	*GFL144*	1
GFL324	22	*GFL327*	10	*GFL203*	5	*GFL295*	2	*GFL389*	1
GFL026	22	*GFL220*	10	*GFL169*	5	*GFL155*	2	*GFL473*	1
GFL046	21	*GFL287*	9	*GFL377*	4	*GFL382*	2	*GFL431*	1
GFL335	21	*GFL135*	9	*GFL096*	4	*GFL385*	2	*GFL463*	1
GFL030	20	*GFL252*	9	*GFL192*	4	*GFL032*	2	*GFL265*	1
GFL117	20	*GFL168*	9	*GFL157*	4	*GFL322*	2	*GFL209*	1
GFL409	20	*GFL025*	9	*GFL246*	4	*GFL318*	2		
GFL329	20	*GFL027*	9	*GFL216*	4	*GFL182*	2		
GFL173	20	*GFL121*	8	*GFL445*	4	*GFL189*	2		
GFL282	19	*GFL150*	8	*GFL314*	4	*GFL194*	2		