

# Genome-wide sequencing and analysis reveals the complexity of *Gossypium raimondii* transcriptome

## Qin Li

E-mail: vargas@pku.edu.cn http://www.zhulab.pku.edu.cn







Read density (log2)

1. Expression profiles of important gene families involved in cotton fiber development

2. Genome-wide RNA-seq data mapping and analyses

3. non-coding RNA

# 1. Expression profiles of import gene families involved in cotton fiber development

2. Genome-wide RNA-seq data mapping and analyses

3. non-coding RNA



#### Analysis of genes involved in cotton fiber initiation and elongation

### Expression levels of SUS genes



Expression levels of 3-ketoacyl-CoA synthase (KCS) genes



Expression levels of *MYB* transcription factors



### Expression levels of ACO genes



1. Expression profiles of import gene families involved in cotton fiber development

### 2. Genome-wide RNA-seq data mapping and analyses

3. non-coding RNA

# Genome-wide and single-nucleotide resolution mapping of *G. raimondii* transcriptome



# Most RNA-seq reads are mapped to annotated exons and splicing junctions (SJs)



# SJs identified in one tissue cover 60-70% gene exon-intron annotations



### Most SJs were identified inside CDS



### Basic AS types in G. raimondii transcriptome











# IR is the most frequent AS event in *G. raimondii*



## AS events profiling across nine species



## AS events found in KNOX genes





## AS events result in homeobox domain



# Leaf lobation in different cottons







1. Expression profiles of import gene families involved in cotton fiber development

2. Genome-wide RNA-seq data mapping and analyses

3. non-coding RNA

#### The most abundant ncRNAs are 24 nt long in cotton fiber cells



#### Annotation and genome-wide mapping of non-coding RNAs

Туре	of	Raw	Distinct	Total length	% mapped in	
molecules		reads	reads	(Mb)	the genome	
miRNA		568,777	647	12.0	100	
siRNA		477,336	88,300	10.6	49.1	
rRNA		452,024	33,127 9.9		46.8	
tRNA		402,235	11,533 17		78.4	
snRNA		1,964	1,107	0.04	62.2	
snoRNA		1,548	651	0.03	61.8	



Stem-loop structures of eight pre-miRNAs with both mature miRNA (red) and miRNA\* (green) strands detected in ncRNA-seq data set Expression profiles of 463 potential miRNA target genes





Array ID	<i>P-</i> value	Annotation
CM042B0	0	acyl-CoA oxidase [Arabidopsis thaliana]
CM052B0	1.20E-06	At1g76250 [Arabidopsis thaliana]
CM109A0	2.75E-05	dynamin-related protein 4 (ADL4) [Arabidopsis
CM058H0	1.20E-04	unknown protein [Arabidopsis thaliana]
CM098F1	1.74E-04	unknown protein [Arabidopsis thaliana]
CM046C0	6.48E-04	At5g23720 [Arabidopsis thaliana]
CM045C0	8.38E-04	arabinogalactan-protein (AGP13) [Arabidopsis
CM039G0	8.89E-04	similarity to zinc metalloproteinase [Arabidopsis
CM054D0	8.89E-04	At5g50010 [Arabidopsis thaliana]
CM064H1	1.10E-03	OSJNBb0039L24.13 [Oryza sativa (japonica

0	3	5	10	15	20 DPA

Array ID	<i>P</i> -value	Annotation
CM081F0	0	?
CM039A0	7.61E-06	cytochrome P450 [Pyrus communis]
CM031D0	8.90E-06	ubiquitin family [Arabidopsis thaliana]
CM001H0	1.95E-05	alcohol dehydrogenase (ADH) [Arabidopsis thaliana]
CM071D0	2.62E-05	cytochrome P450 [Citrus sinensis]
CM113B0	5.12E-05	?
СМ002Н0	6.28E-05	BPG-independent PGAM
СМ050Н0	7.24E-05	harpin-induced protein 1 family (HIN1) [Arabidopsis
CM026F0	1.01E-04	OSJNBb0035I14.15 [Oryza sativa (japonica
СМ023Н0	1.91E-04	vacuolar H <sup>+</sup> -pyrophosphatase [Prunus persica]
CM066B0	1.92E-04	expressed protein [Arabidopsis thaliana]
CM055C0	4.93E-04	branched-chain alpha-keto acid decarboxylase E1 beta
CM024B1	6.17E-04	K <sup>+</sup> efflux antiporter [Arabidopsis thaliana]
CM023A0	1.40E-03	nodulin MtN21 family protein [Arabidopsis thaliana]

#### Summary

1.Qualitative transcript differences in key fiber development genes were found between the fibered upland cotton and the fiberless raimondii cotton.

2.As many as 35% of intron-containing genes were found alternatively spliced in *G. raimondii* trasncriptome and the most frequent AS event was intron retention. AS was potentially a new mechanism of leaf morphogenesis.

3.562 microRNA loci were identified using high-throughput non-coding RNA sequencing combining with D-genome reference sequences. Analysis of the expression profiles of 463 potential miRNA target genes indicated that miRNAs may play an important role in cotton fiber development.





В

Hu G et al. Genetics 2011

D

1 cm

### G. arboreum (AA) G. hirsutum (AADD)

### G. raimondii (DD)





# Global statistics of *G. raimondii* genome assembly and annotation

a. Assembly

Assembly			Number	N50	Longest	Size	% of the		
				(Kb)	(Mb)	(Mb)	assembly		
Contigs	All		41,307	44.9	0.3	744.4	-		
	All		4,715	2,284	12.8	775.2	100		
Cooffoldo	Anchored		281		12.8	567.2	73.2		
Scallolos	Anchore	d	228		12.8	406.3	52.4		
	and orier	nted							
b. Annotatio	b. Annotation								
Annotation Genes annotated 4		Num	ber	Size (Mb)		% of the assembly			
		40,976		115.7		14.9			
miRNA		348		0.04	04 <0.0				
rRNA		565		0.1		0.01			
tRNA		1,041		80.0		0.01			
snRNA		29		0.1		0.02			
Transposable		148,740		441.4		57.0			
elements (TEs)									