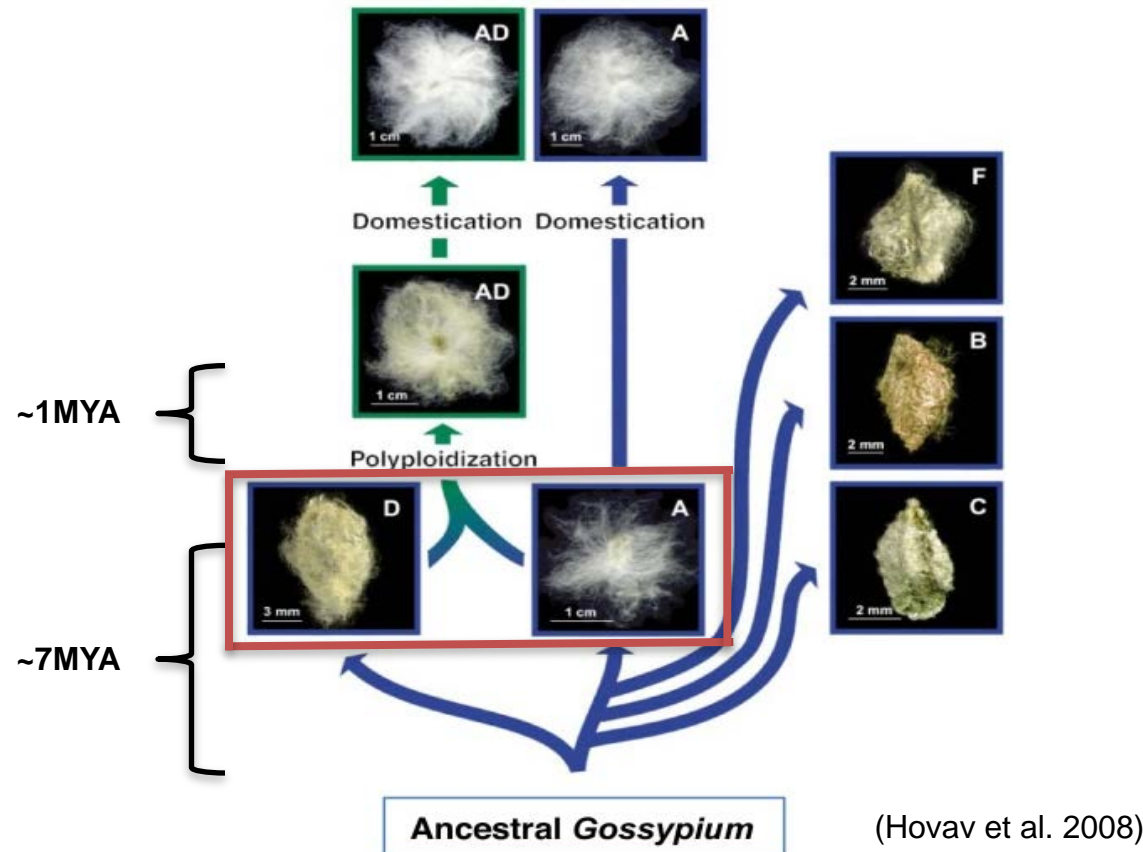


# Transcriptome Analysis of Extant Cotton Progenitors and Identification of Genome-Specific-Single Nucleotide Polymorphism (GNP)

Gyoungju Nah (Dr. Z. Jeffrey Chen Lab)  
University of Texas at Austin

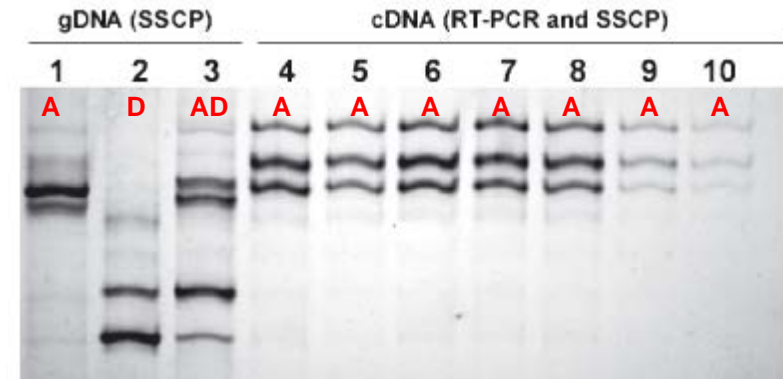
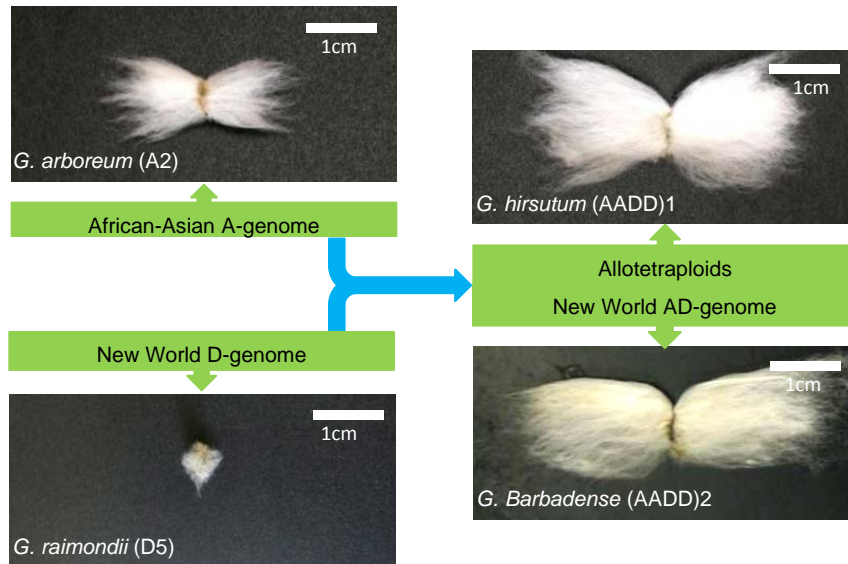


# A and D genomes as Extant Parents of AADD allotetraploid



1. What is the difference in transcriptome between *G. arboreum* (AA) and *G. raimondii* (DD)?  
 → Needs AA and DD EST information

# Allelic Expression During Cotton Fiber Development



**TC67135 (cyclin D)**  
**A –allele specific expression**  
**in fiber-bearing ovule**

(Yang et al. 2006)

**2. How does A- and D- allelic expression contribute to fiber development in allopolyploids?**  
**→ Needs GNP information**

# Work Flow

Young leaves, roots, bolls, ovules, and fibers

454/Roche Titanium sequencing:  
1,699,776 reads from *G. arboreum* (AA)  
1,464,815 reads from *G. raimondii* (DD)

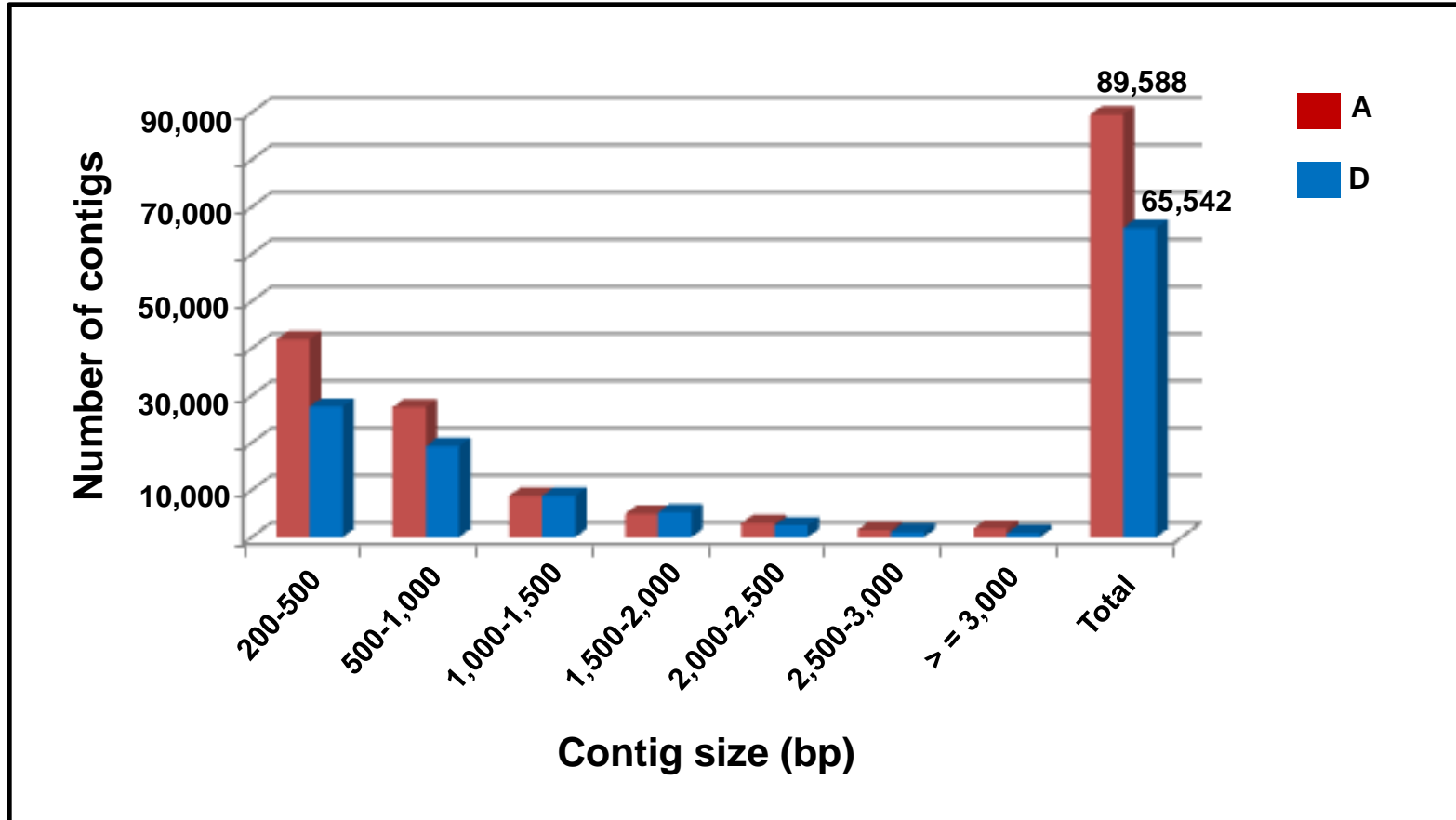
Assembly of 454 reads (Chen lab)  
AA: **62,609** contigs (avr. 1,032 bp)  
DD: **34,908** contigs (avr. 1,107 bp)

Assembly of 454 reads (Udall lab)  
AA: 89,185 contigs (avr. **629** bp)  
DD: 68,984 contigs (avr. **676** bp)

After merge of Chen and Udall ESTs:  
A: 89,588 contigs (avr. 806 bp)  
D: 65,542 contigs (avr. 840 bp)

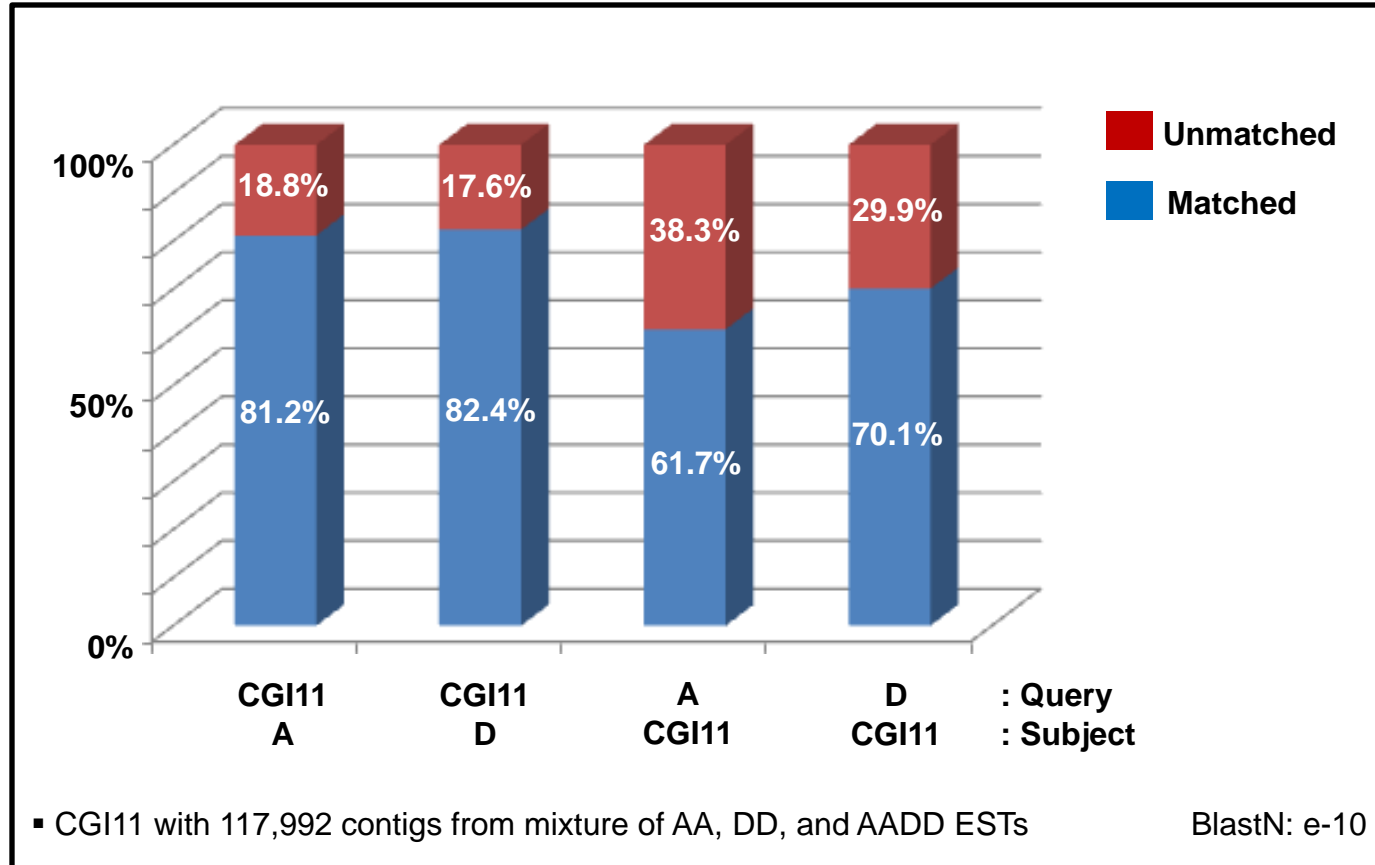
- Merge from two labs increased both number and size of AA and DD ESTs

# Contig Size Distribution of AA and DD EST libraries



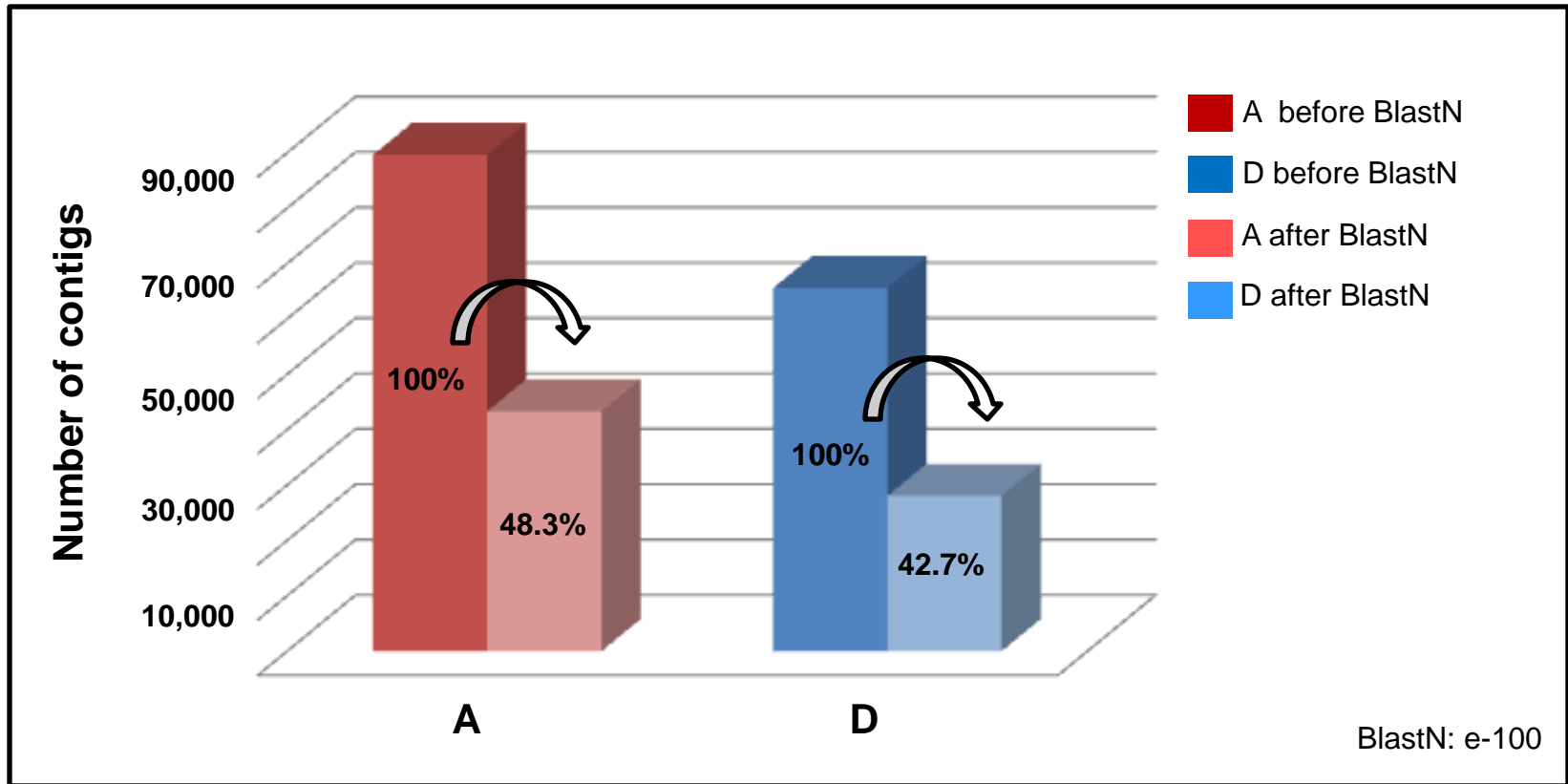
- Transcriptome size of A-subgenome is ~27% larger than that of D-subgenome
- The majority of A (88%) and D (84%) contigs range from 200-1,500bp

# AA and DD EST Coverage



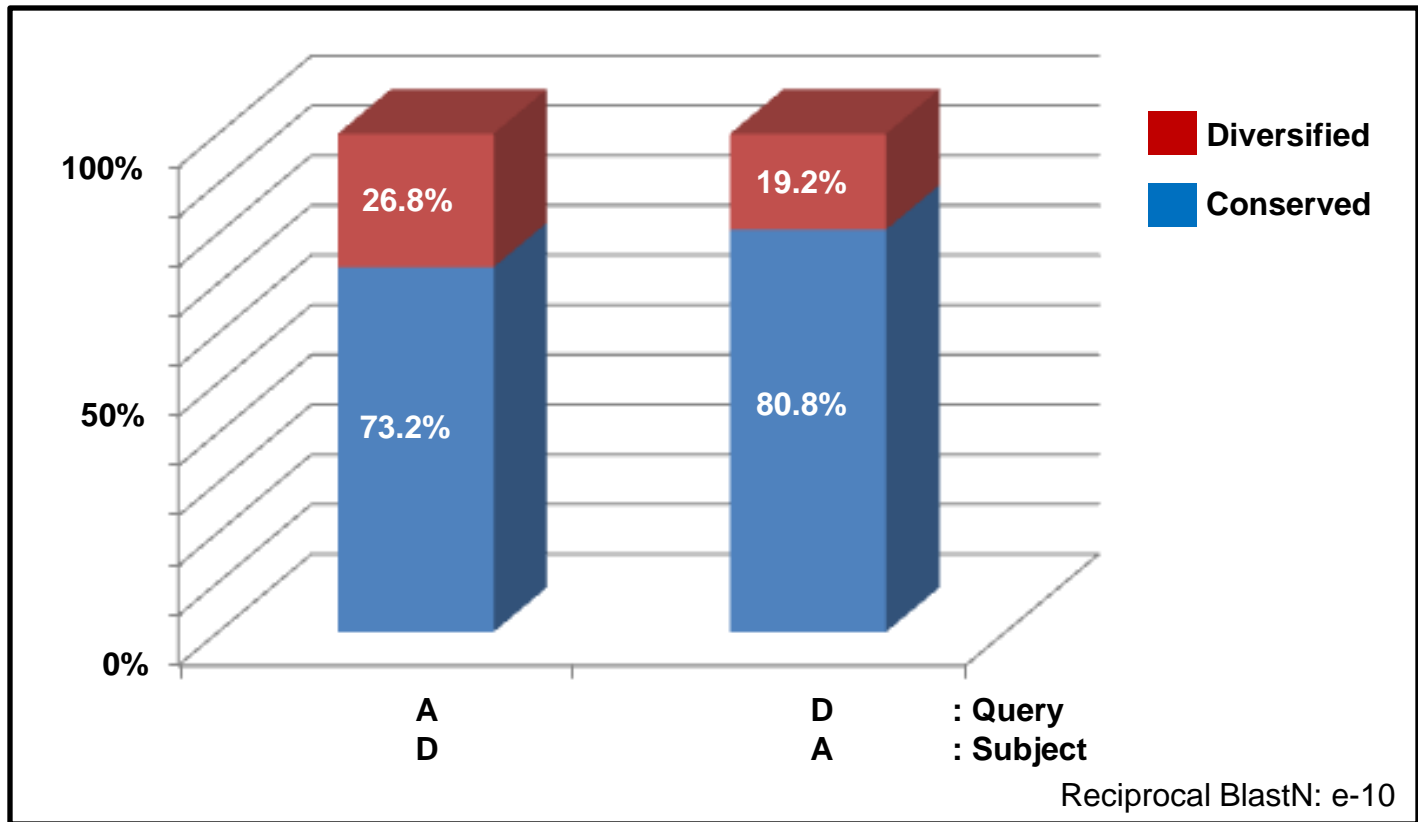
- New A and D ESTs include ~80% of entries in CGI11
- New A and D ESTs provide additional ~30-38% ESTs that are not present in CGI11

# Redundancy in AA and DD ESTs



- A and D transcriptomes are highly redundant, indicating the presence of ~50% of isoforms and paralogs in the cotton genome

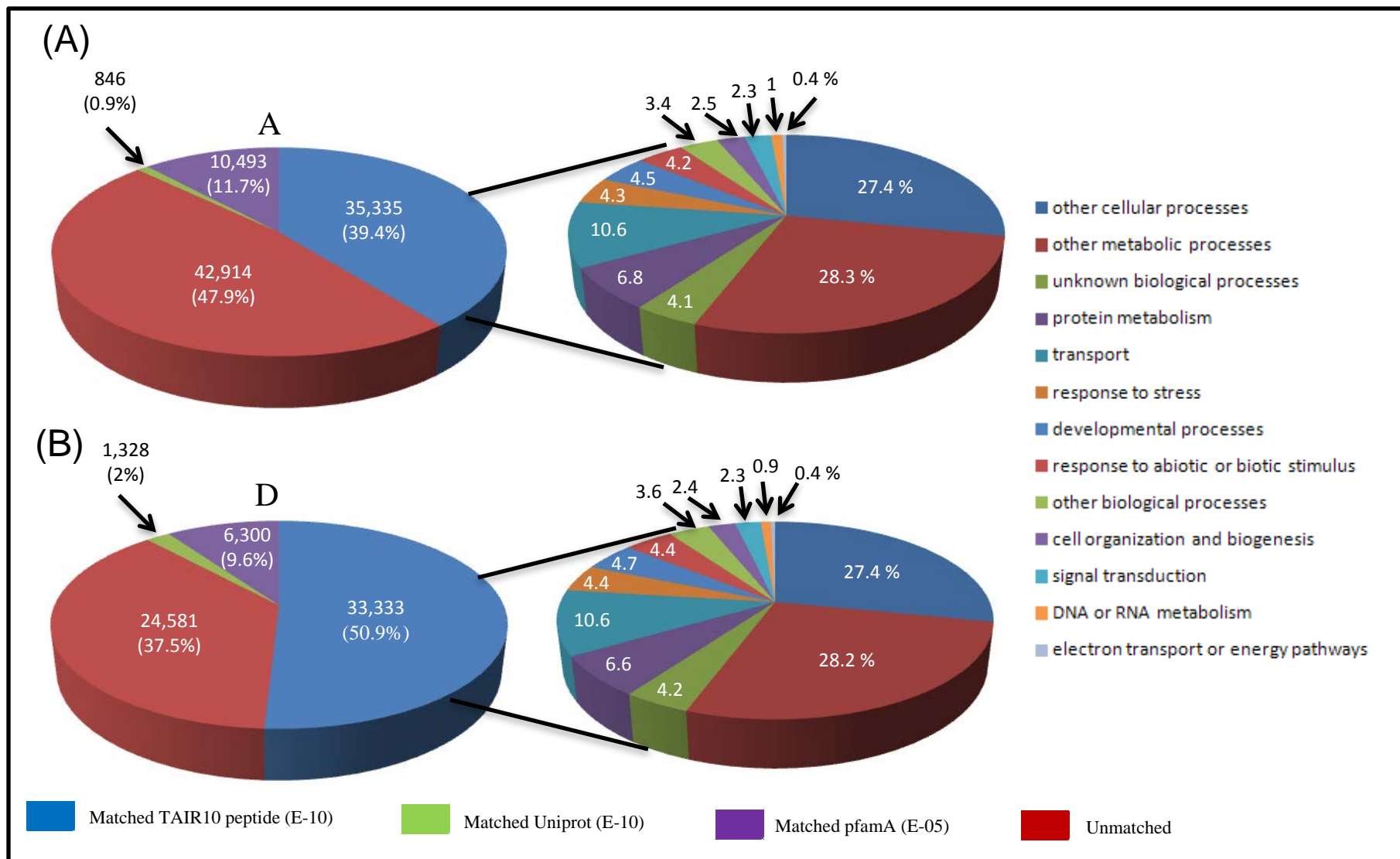
# Estimation of Diversification of AA and DD ESTs



- Either one of the libraries does not cover the entire transcriptome
- This diversification was estimated as 27% in A and 19% in D

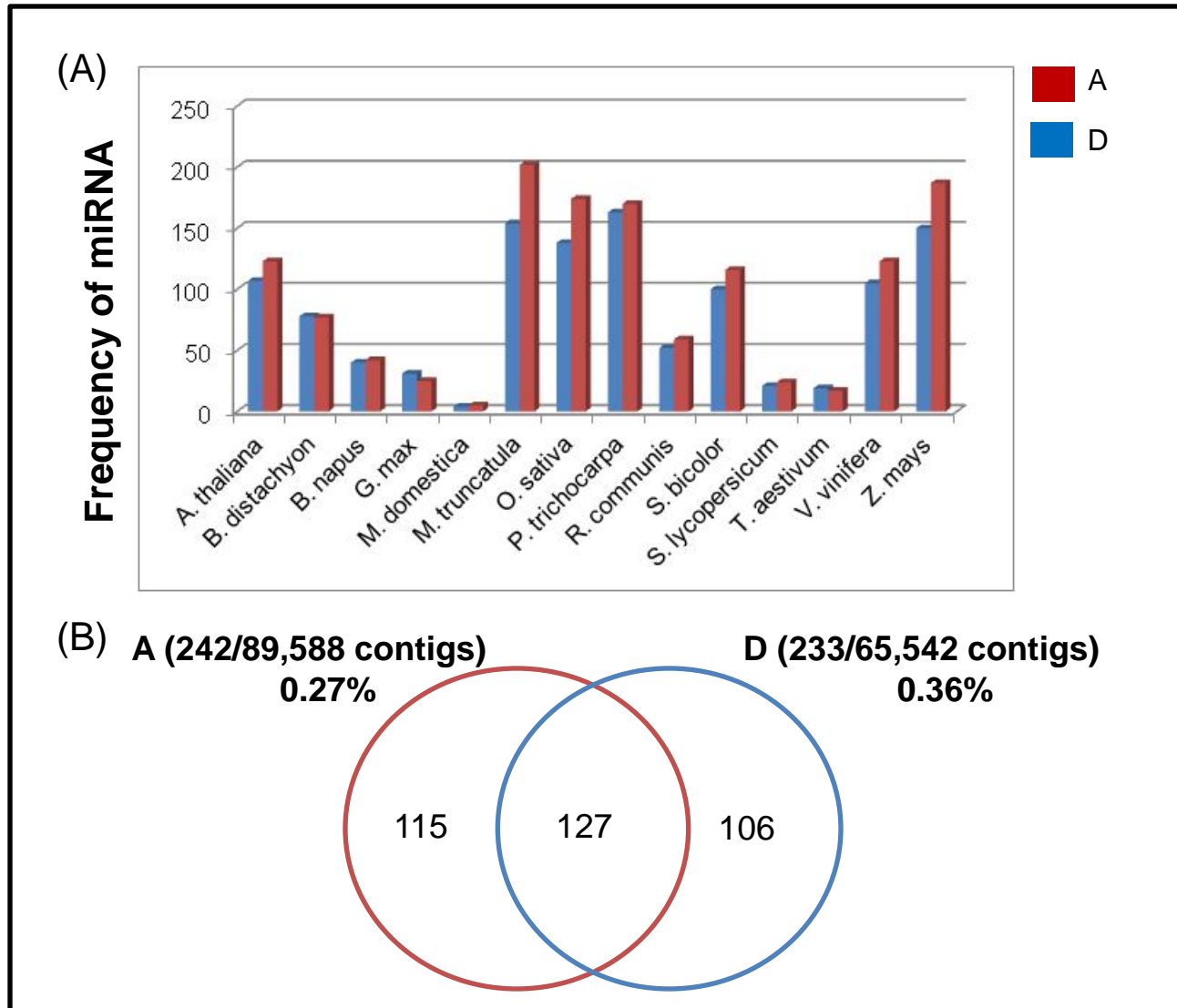


# AA and DD ESTs Against Known Protein Databases



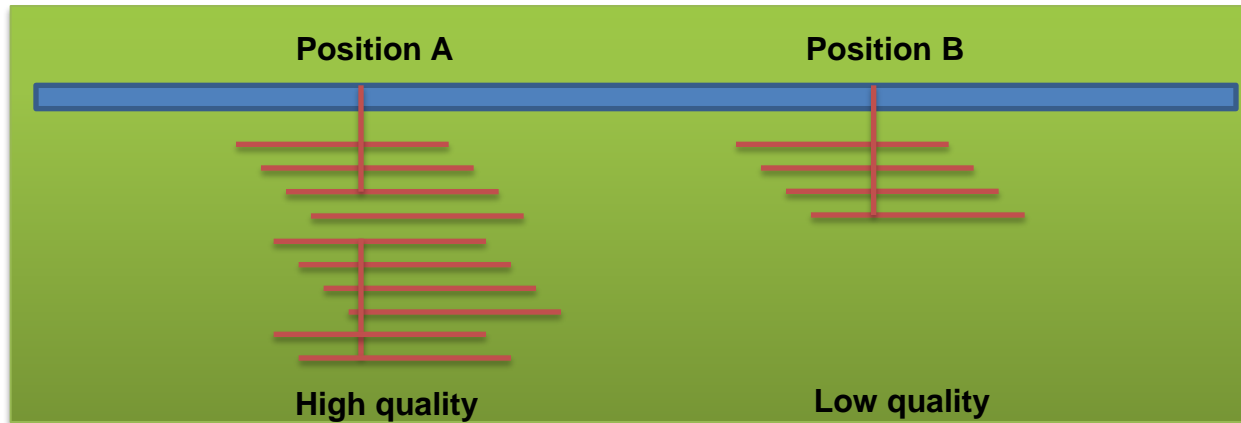
- D contains a higher portion of *Ath* protein homologs than A
- Both A and D ESTs are enriched with cellular process and metabolic process

# miRNA Targets in AA and DD ESTs



- miRNA regulation in DD genome is higher than AA, suggesting that in allotetraploid, miRNA might play important role for D-allele regulation

# Selection of High Quality GNPs



**GNP Selection (Criteria:  $\geq 8X$  coverage,  $\geq 90\%$  consensus,  $Q \geq 25$ )**

- **Number of GNP-containing contigs is 11,000**

# Allele-Separable Genes-I

## Epigenetic-associated genes

Cotton EST ID	TAIR ID	Gene Name
UDcontig30230	AT1G48410.1	AGO1
cuDContig2017	AT1G48410.1	AGO1
cuDContig12969	AT1G48410.1	AGO1
cuDContig19558	AT1G31280.1	AGO2
cuDContig4863	AT2G27040.2	OCP11
cuDContig7218	AT2G27040.2	OCP11
cuDContig17255	AT2G27880.1	AGO5
UDcontig10468	AT1G01040.2	SUS1
cuDContig12083	AT3G03300.3	DCL2
UDcontig8529	AT3G03300.3	DCL2
cuDContig3778	AT1G14790.1	RDR1
cuDContig17499	AT5G14620.1	DRM2
cuDContig11889	AT4G19020.1	CMT2
cuDContig7567	AT1G69770.1	CMT3
cuDContig13379	AT1G77300.1	SDG8
cuDContig637	AT1G73100.1	SUVH3
UDcontig12385	AT2G22740.1	SUVH6
cuDContig13480	AT3G12680.1	HUA1
cuDContig19226	AT1G05460.1	SDE3
cuDContig7543	AT1G01920.2	SET-domain
cuDContig6339	AT1G05120.1	SNF2-domain

## Clock-related genes

Cotton EST ID	TAIR ID	Gene Name
UDcontig31044	AT2G46830.1	CCA1
CDcontig25809	AT1G01060.4	LHY1
UDcontig11814	AT1G01060.4	LHY1
UDcontig14538	AT1G01060.5	LHY1

# Allele-Separable Genes-II

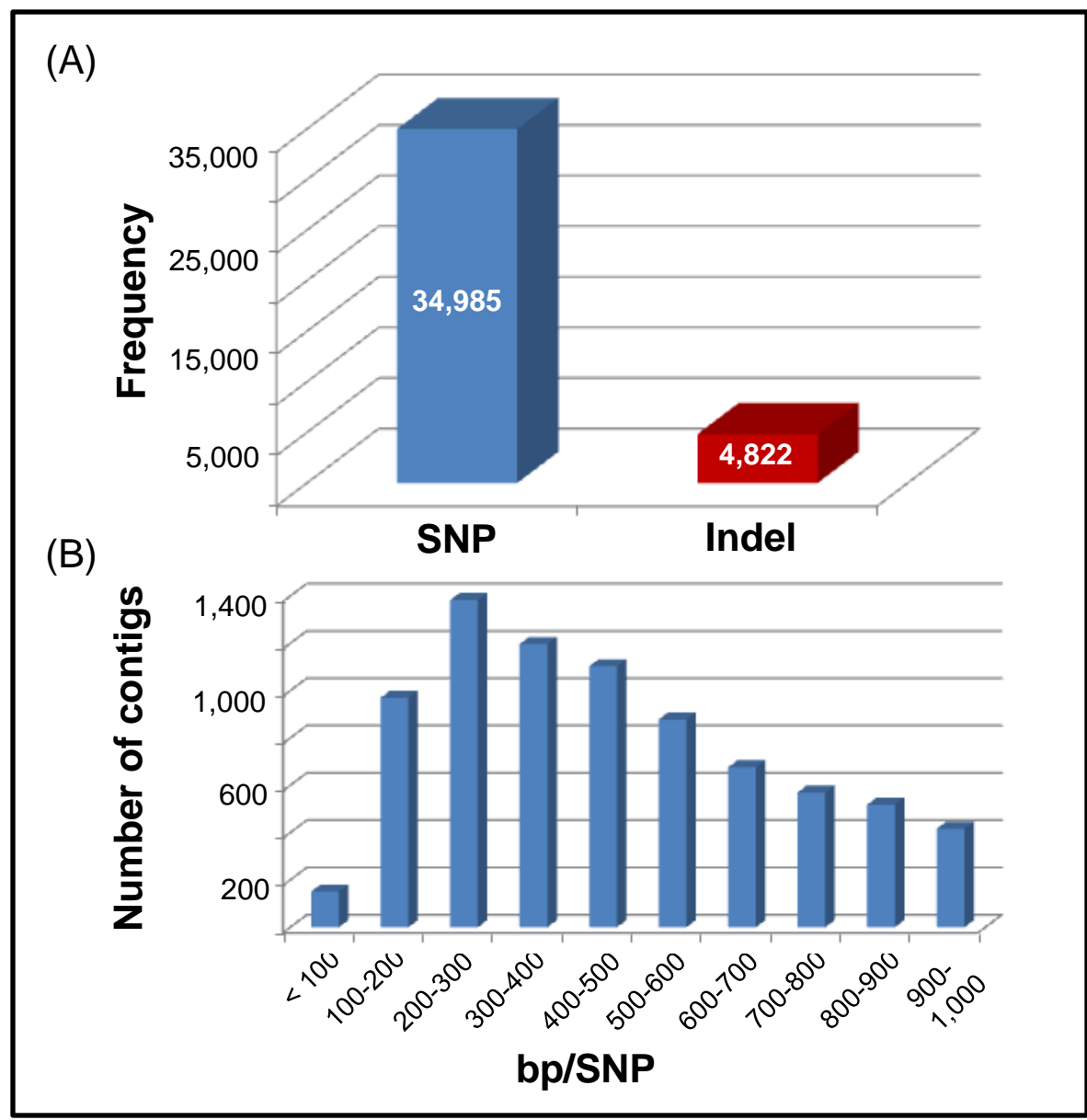
## Myb-related genes

Cotton EST ID	TAIR ID	Gene Name
cuDContig15019	AT1G22640.1	MYB3
cuDContig7688	AT1G68670.1	MYB-domain
cuDContig4930	AT1G74840.1	MYB-domain
cuDContig5155	AT2G01060.1	MYB-domain
UDcontig50103	AT2G03500.1	MYB-domain
cuDContig13445	AT2G23290.1	AtMYB70
cuDContig3509	AT2G38090.1	MYB-domain
cuDContig517	AT2G38090.1	MYB-domain
cuDContig13450	AT2G38090.1	MYB-domain
cuDContig4163	AT2G47190.1	MYB2
cuDContig13595	AT3G09600.1	MYB-domain
cuDContig1160	AT3G10760.1	MYB-domain
cuDContig3516	AT3G13040.2	MYB-domain
cuDContig15142	AT3G18100.1	MYB4R1
cuDContig5233	AT4G09460.1	AtMYB6
UDcontig45109	AT4G32730.2	PC-MYB1
cuDContig16746	AT4G32730.2	PC-MYB1
cuDContig3991	AT4G37260.1	MYB73
cuDContig770	AT4G38620.1	MYB4
cuDContig16699	AT5G04760.1	MYB-domain
cuDContig16484	AT5G15310.2	ATMYB16
UDcontig12659	AT5G45420.1	MYB-domain
cuDContig19287	AT5G52660.1	MYB-domain
cuDContig5755	AT5G52660.2	MYB-domain
cuDContig16730	AT5G67300.1	MYBR1

## Fiber-related genes

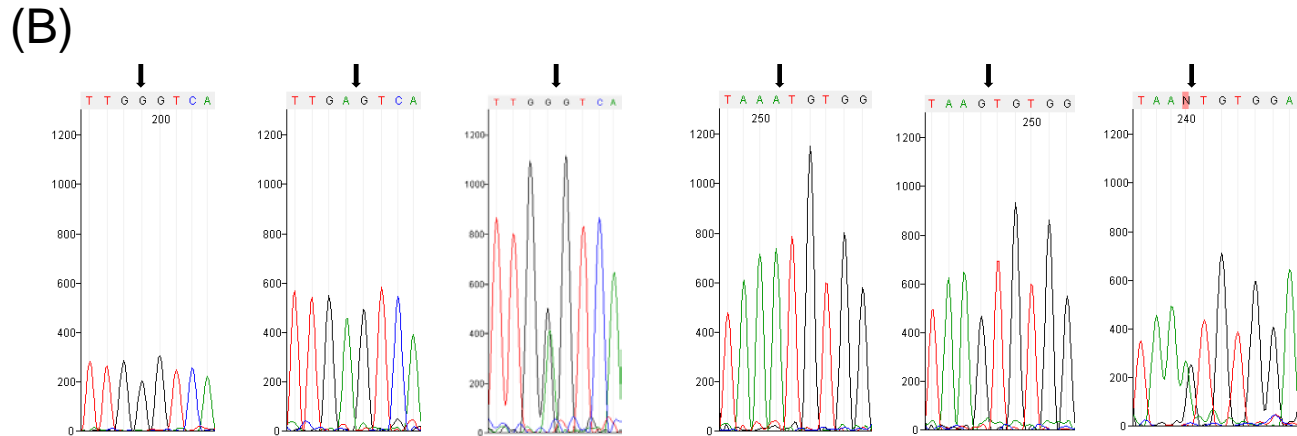
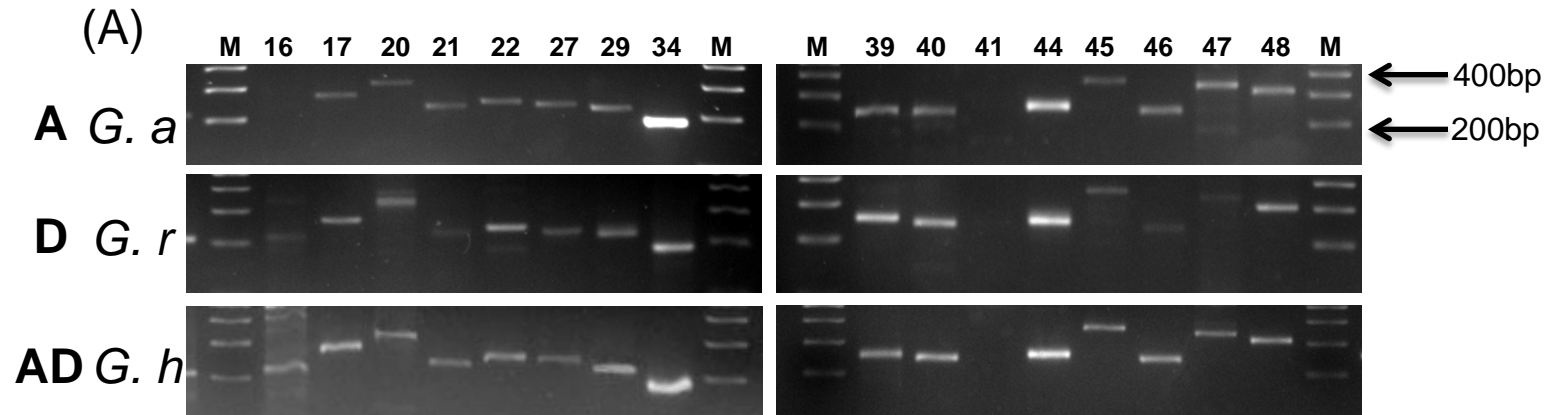
Cotton EST ID	TAIR ID	Gene Name
cuDContig1242	AT1G05010.1	EFE
cuDContig2761	AT1G05010.1	EFE
cuDContig4895	AT1G07890.8	MEE6
cuDContig7679	AT1G12910.1	ATAN11
UDcontig42026	AT1G62660.1	BFRUCT3
cuDContig4618	AT2G01570.1	RGA1
cuDContig14416	AT2G01570.1	RGA1
cuDContig1899	AT2G28950.1	ATHEXP
cuDContig3387	AT2G40610.1	EXP8
cuDContig7169	AT3G43190.1	SUS4
cuDContig12426	AT4G03010.1	Leucine-rich
cuDContig5067	AT4G22880.2	TT18
cuDContig3717	AT5G13710.2	SMT1
cuDContig389	AT5G24520.2	TTG1
cuDContig5774	AT5G25610.1	RD22

# GNP Identification and Characterization



GNP=SNP+Indel

# GNP Experimental Validation



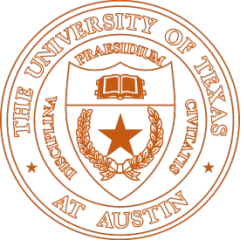
Exp#  
22 cuDContig11665 cuAContig4409  
*G. arboreum* *G. romandii* *G. hirsutum*

Exp#  
48 CDcontig15250 cuAContig1068  
*G. arboreum* *G. romandii* *G. hirsutum*

# Conclusions

- **We generated AA and DD EST libraries from extant progenitors of allotetraploid AADD cotton, which provides an important genomic resource for cotton fiber research and crop improvement**
- **Comparative analysis of AA and DD ESTs provided some new insights into transcriptome divergence between *G. arboreum* (AA) and *G. raimondii* (DD) genomes**
- **Analysis of miRNA targets in AA and DD ESTs suggests that miRNA-mediated gene regulation plays a role in expression of target genes from A and D subgenomes in allopolyploids**
- **We developed a pipeline of GNPs that can discriminate between a large number of AA and DD ESTs (~11,000), including many involved in the fiber development and epigenetic pathways.**





# Acknowledgement

## University of Texas at Austin

Dr. Jeffrey Z. Chen

Dr. Yuki Guan

## Brigham Young University

Dr. Joshua Udall

## Texas A&M University

Dr. David Stelly

## UT GSAF

Dr. Scott Hunicke-Smith

## Texas Advanced Computing Center

