Cotton genomics is providing rational ways for gene finding and genetic improvement

Xianlong Zhang

2018-5-29

- **The two cultivated tetraploid cotton species:**

  Gossypium hirsutum

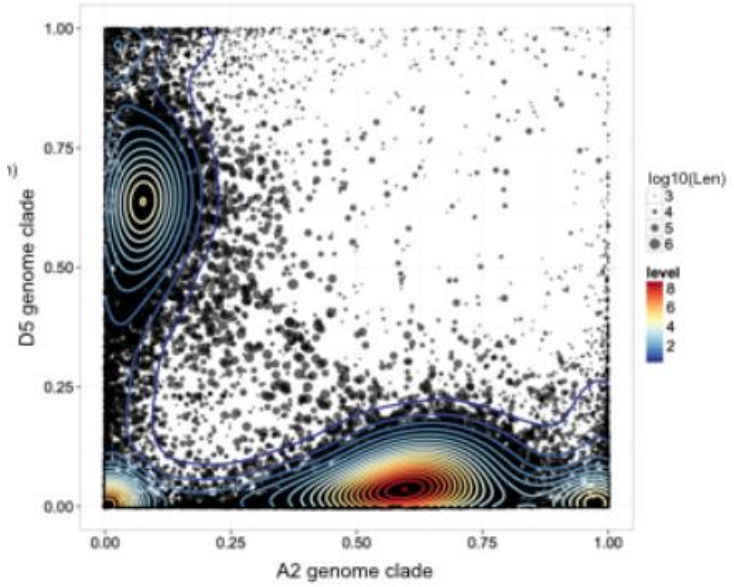
  G. barbadense

- **Using Sea Island cotton (much better fiber quality) to improve Upland cotton (relatively low quality with high yield, covers 95% areas) is our strategy**

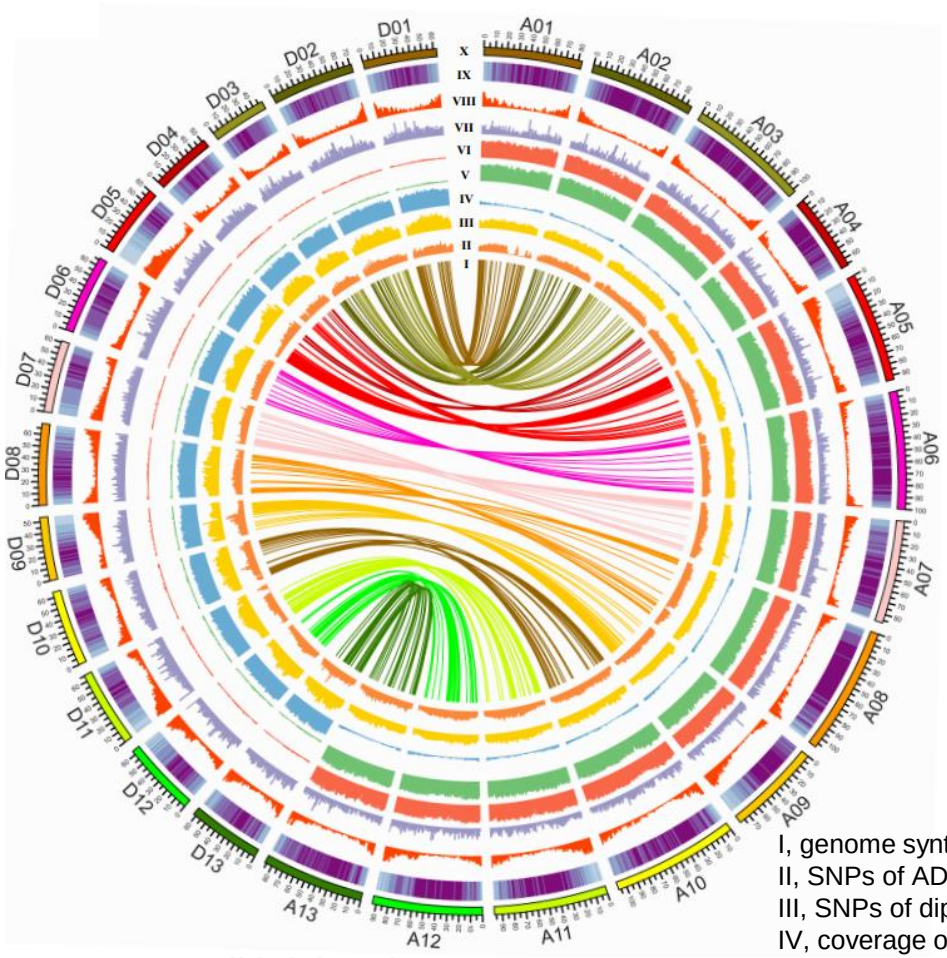
Therefore we sequenced the genome of *G. barbadense*

The genome sequence of *G. barbadense*

Subgenome assignment using diploid data

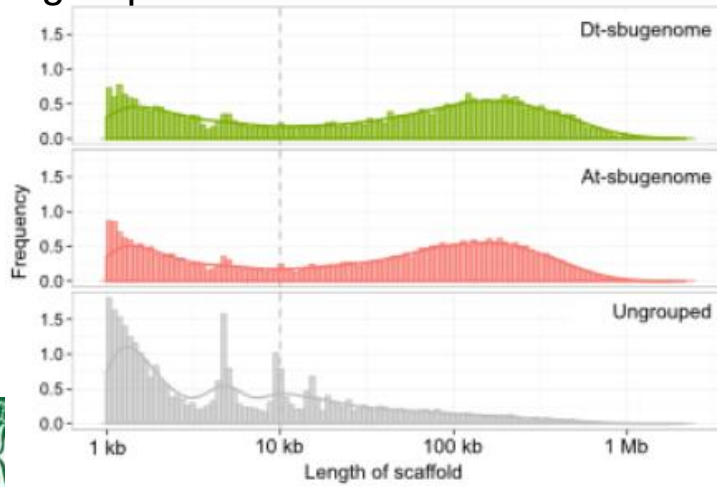


The anchored draft genome sequence



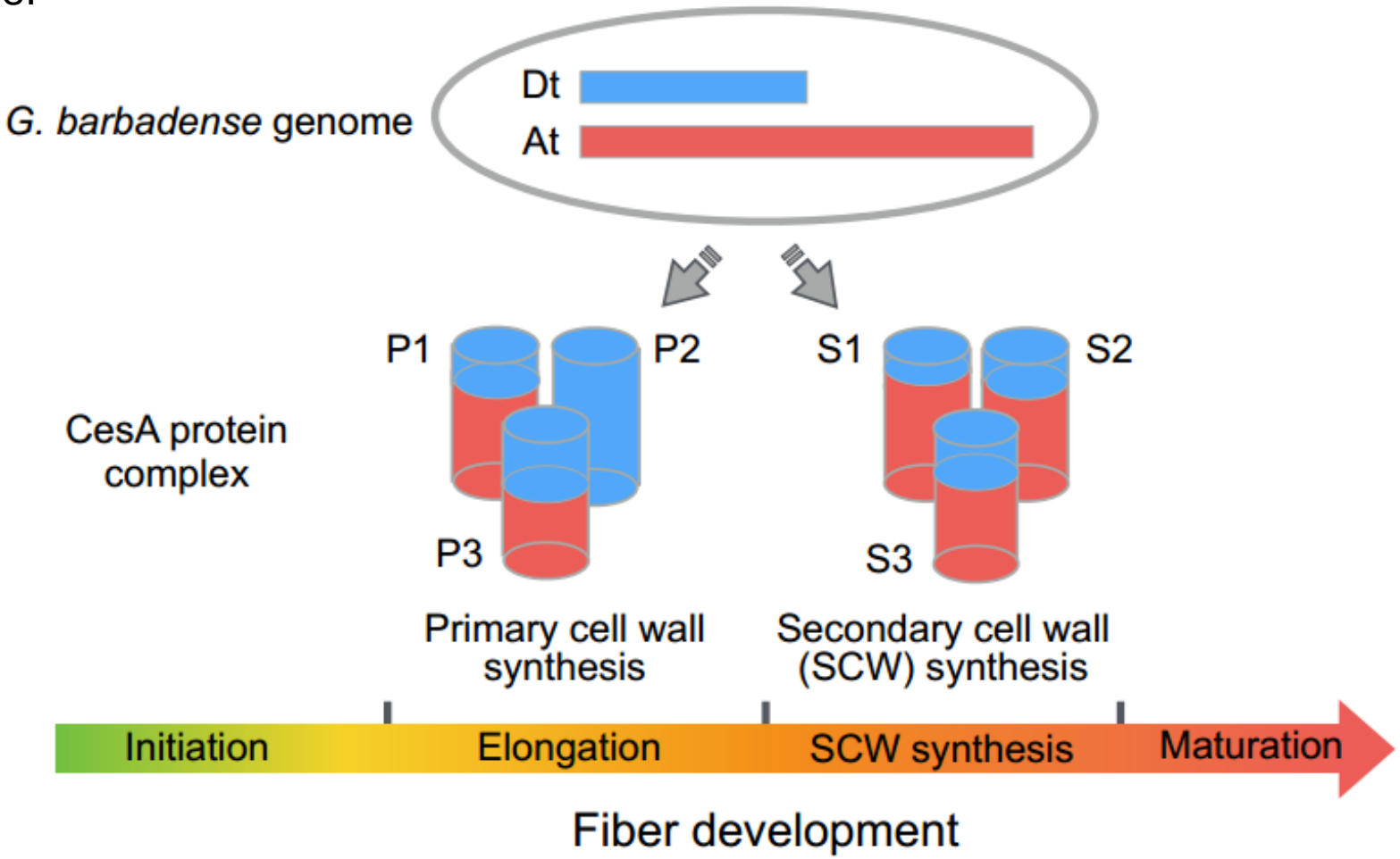
I, genome synteny
II, SNPs of AD1
III, SNPs of diploids
IV, coverage of D5
V, coverage of A1
VI, coverage of A2
VII, coverage of AD1
VIII, gene density
IX, TE density
X, pseudochromosome

Ungrouped scaffolds exhibit small sizes.



“Relay race” model of *CesAs* in allotetraploid cotton fiber development

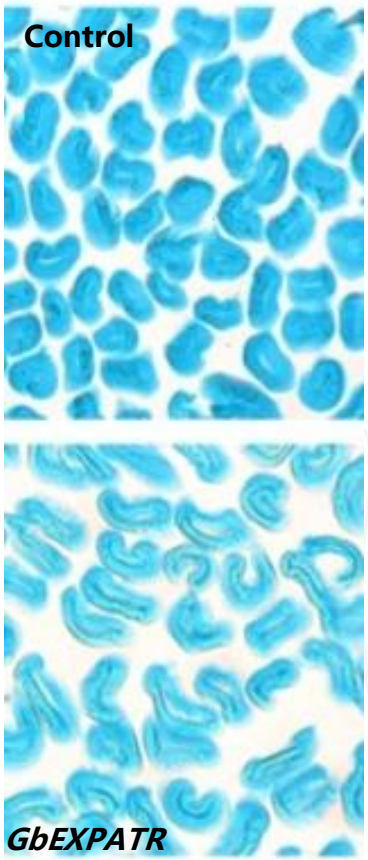
CesA genes in Dt-subgenome contribute largely in the elongation stage, whereas At-subgenomic CesAs play a predominant role in secondary cell wall synthesis stage.

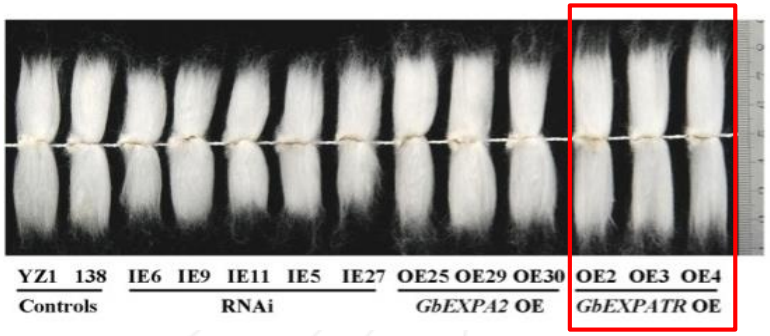


An example: gene from Gb is really able to improve fiber quality

GbEXPATR

The transgenic cotton has longer, finer and stronger fiber. And these traits can be perfectly repeated. *EXPRATR* changes cell wall properties and improves fiber quality by reducing the cellulose, increasing the callose.

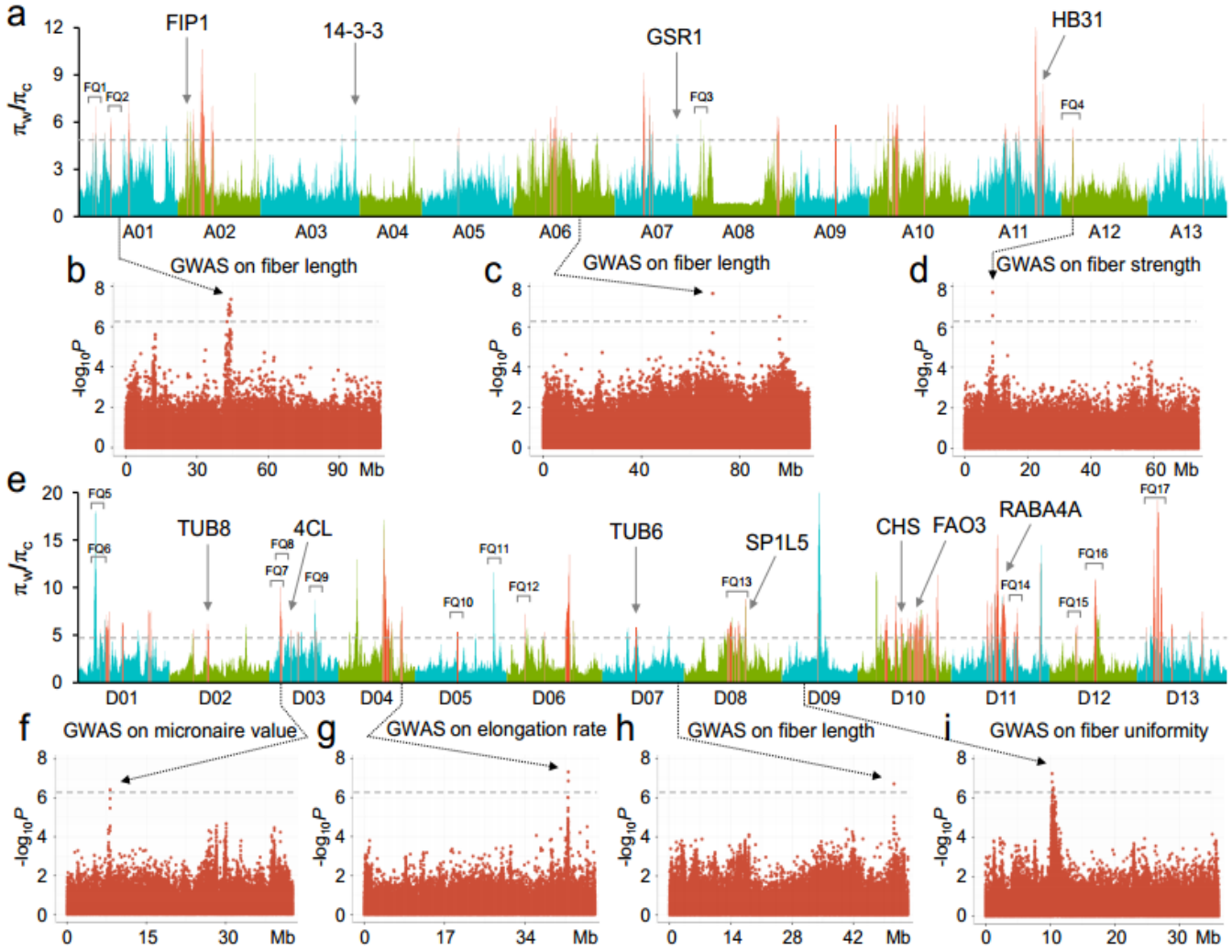




OE Lines	Fiber lenghth	Micronaire	Fiber Strenght
YZ1	27.96±0.22	5.23±0.08	25.56±0.37
138	27.50±0.76	5.23±0.07	25.52±0.55
OE2	**29.60±0.57**	**5.14±0.16**	**26.54±0.66**
OE3	**29.79±0.71**	**4.78±0.25**	**27.03±0.90**
OE4	**30.11±0.86**	**4.61±0.17**	**27.01±0.62**

Li et al., Plant Biotechnol J, 2016

We also try to find good fiber genes in *G. hirsutum*: Selection signals and GWAS on fiber quality-related traits



93 domestication sweeps, **1777** genes (549 in the At and 1228 in the Dt)

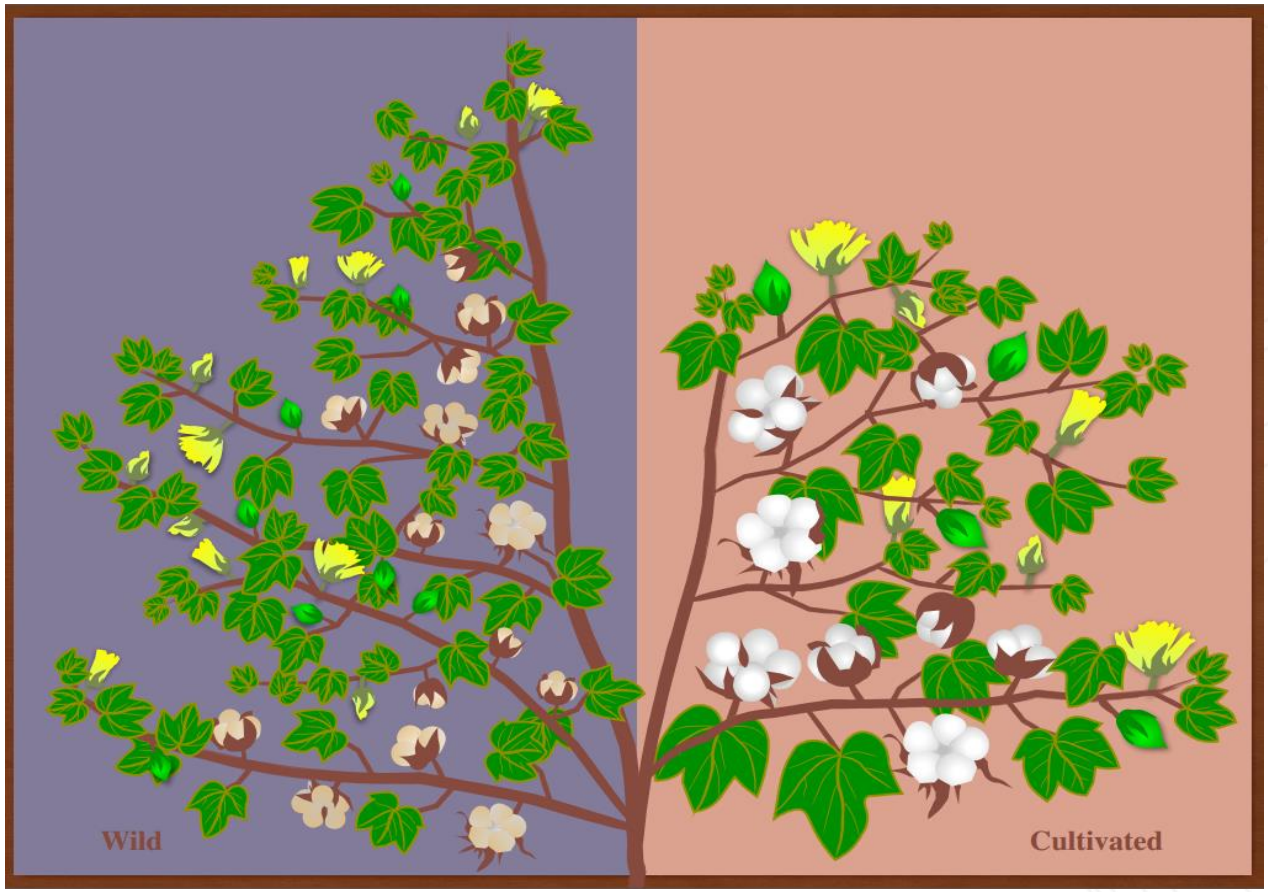
25 QTL hotspots associated with major agronomic traits, of which **19** in the Dt.

19 significant GWAS signals associated with fiber quality-related traits, of which **11** in the Dt subgenome.

We resequenced 285 accessions.

Asymmetric subgenome domestication for long white fiber

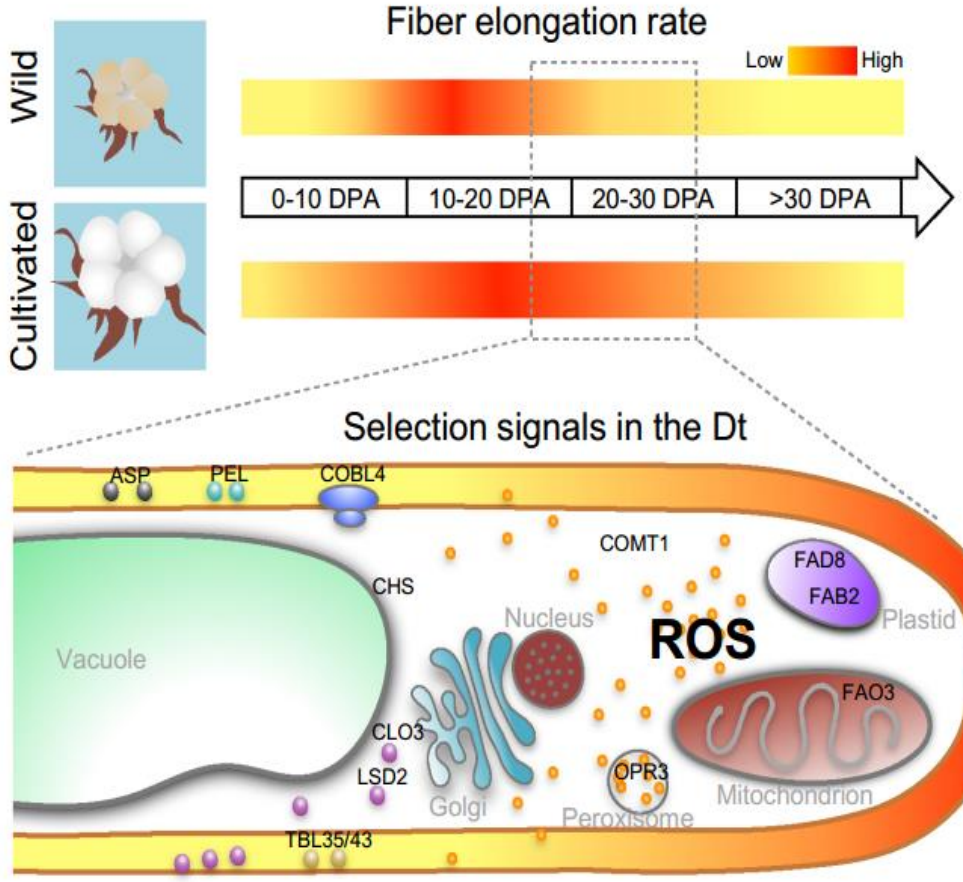
A comparison of fiber quality between wild and cultivated cotton



The effects of domestication on fiber length and color

What is the genetic basis underlying these changes?

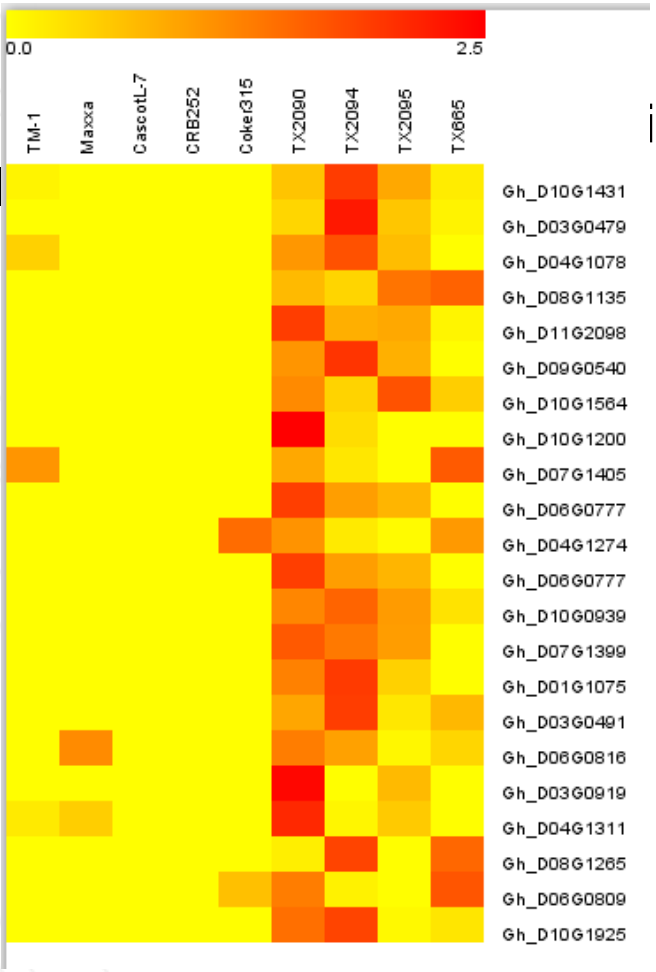
Asymmetric subgenome domestication for long fiber trait



Down-regulation of ROS-related genes

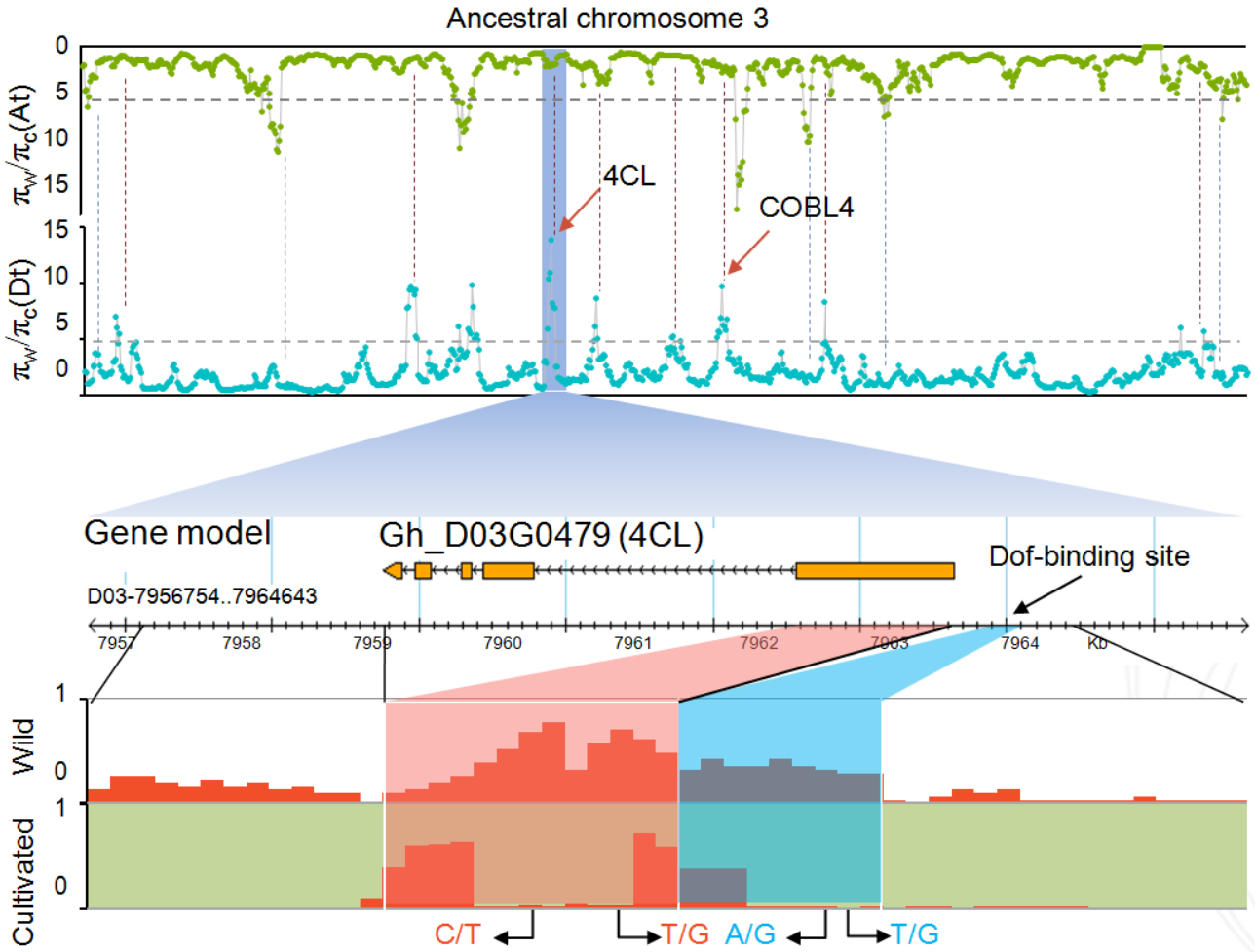
17 **ith**

sel

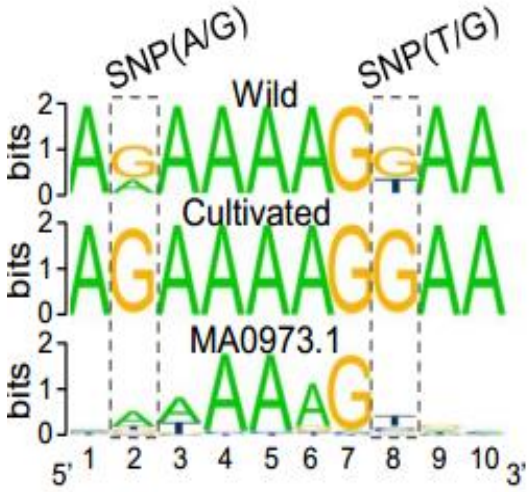


High concentration of ROS in wild cotton fiber development will terminate fiber elongation and cause developmental transition to secondary cell wall synthesis.

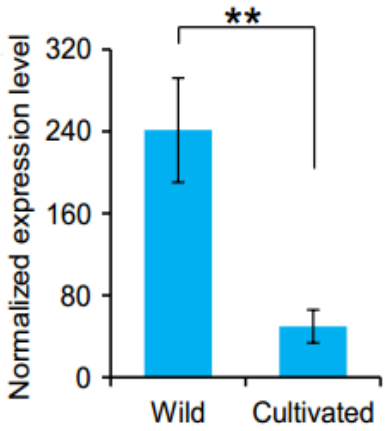
Asymmetric subgenome domestication for white fiber trait



Selection signals at the Dof-binding site



Down-regulation of 4CL

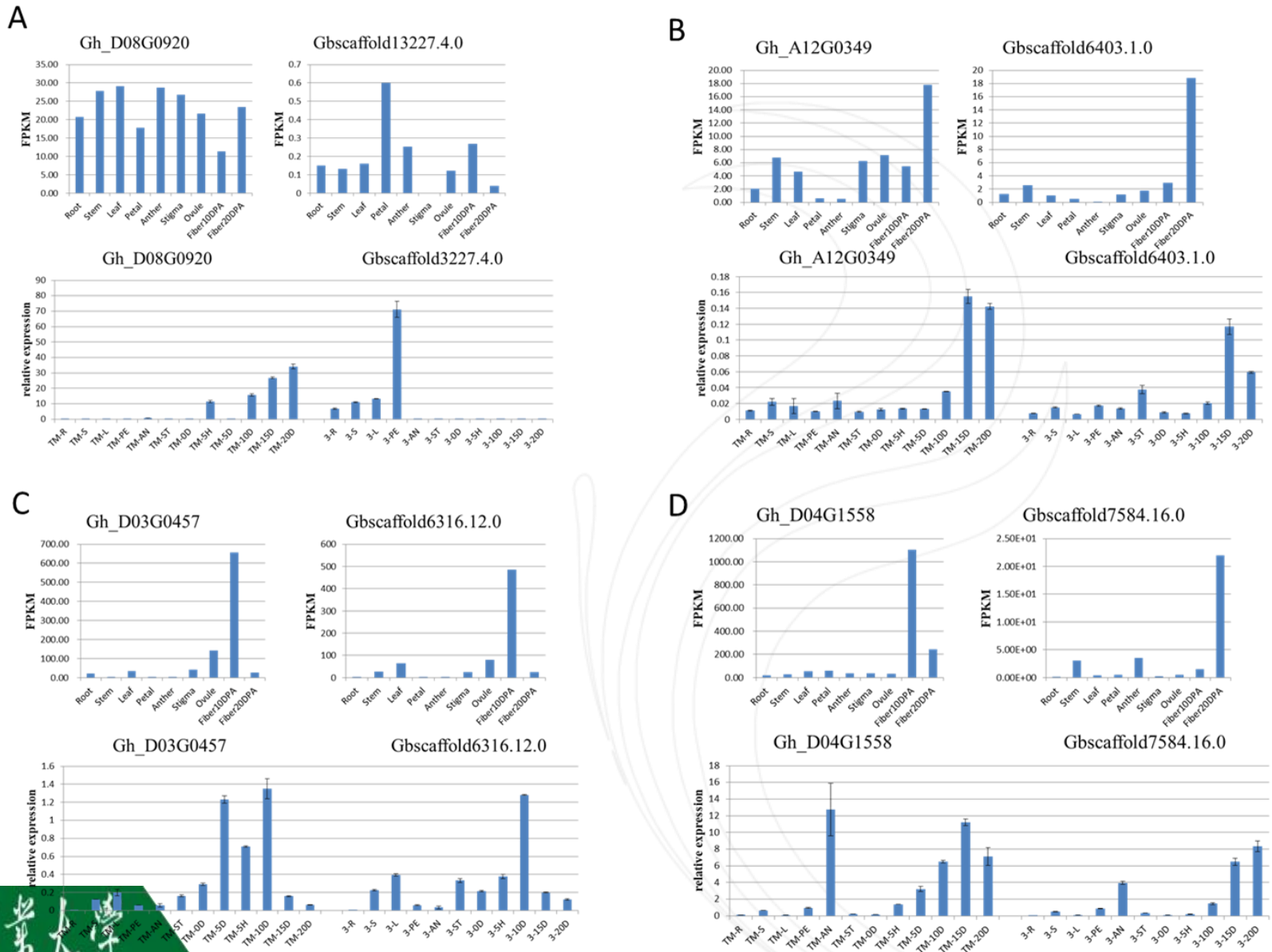


Selection signal at the *4CL* loci, the flavonoid metabolic pathway, may have driven the white fiber trait characteristic.

Candidate genes identified by GWAS

Trait	Chr			Candidate gene	Ortholog	Annotation
Fiber length	A06	GENE	1	Gh_A06G1270	AT1G11300	protein serine/threonine kinases;protein kinases;ATP binding;sugar binding;kinases;carbohydrate binding
	D07	GENE	2	Gh_D07G2042	AT2G32720	cytochrome B5 isoform B
	D07	GENE	3	Gh_D07G2049	AT4G29040	regulatory particle AAA-ATPase 2A
Fiber uniformity	D09	GENE	4	Gh_D09G0300	AT3G01640	glucuronokinase G
	D09	GENE	5	Gh_D09G0301	AT5G14420	RING domain ligase2
	D09	GENE	6	Gh_D09G0302	AT5G14410	
Micronaire value	D03	GENE	7	Gh_D03G0457	AT4G39330	cinnamyl alcohol dehydrogenase 9
	D10	GENE	8	Gh_D10G0649	AT3G16170	AMP-dependent synthetase and ligase family protein
	D10	GENE	9	Gh_D10G0652	AT1G79820	Major facilitator superfamily protein
	D10	GENE	10	Gh_D10G1264	AT3G07490	ARF-GAP domain 11
Fiber elongation rate	A07	GENE	11	Gh_A07G1543	AT2G32970	unknown protein
	D04	GENE	12	Gh_D04G1558	AT1G50010	tubulin alpha-2 chain
	D04	GENE	13	Gh_D04G1562	AT1G10200	GATA type zinc finger transcription factor family protein
	D04	GENE	14	Gh_D04G1574	AT2G14900	Gibberellin-regulated family protein
Fiber strength	A12	GENE	15	Gh_A12G0349	AT4G38620	myb domain protein 4
	A12	GENE	16	Gh_A12G0351	AT2G16700	actin depolymerizing factor 5
Short fiber rate	A01	GENE	17	Gh_A01G0543	AT4G27270	Quinone reductase family protein
	A06	GENE	18	Gh_A06G0802	AT4G33010	glycine decarboxylase P-protein 1
	A11	GENE	19	Gh_A11G1722	AT4G36750	Quinone reductase family protein
	A12	GENE	20	Gh_A12G0249	AT3G18820	RAB GTPase homolog G3F
	A12	GENE	21	Gh_A12G0257	AT2G21660	cold, circadian rhythm, and rna binding 2
	D08	GENE	22	Gh_D08G0920	AT4G25770	alpha/beta-Hydrolases superfamily protein
	D11	GENE	23	Gh_D11G3117	AT5G15750	Alpha-L RNA-binding motif/Ribosomal protein S4 family protein

The expression profiles of genes selected to CRISPR/Cas9



Flavonoid metabolism

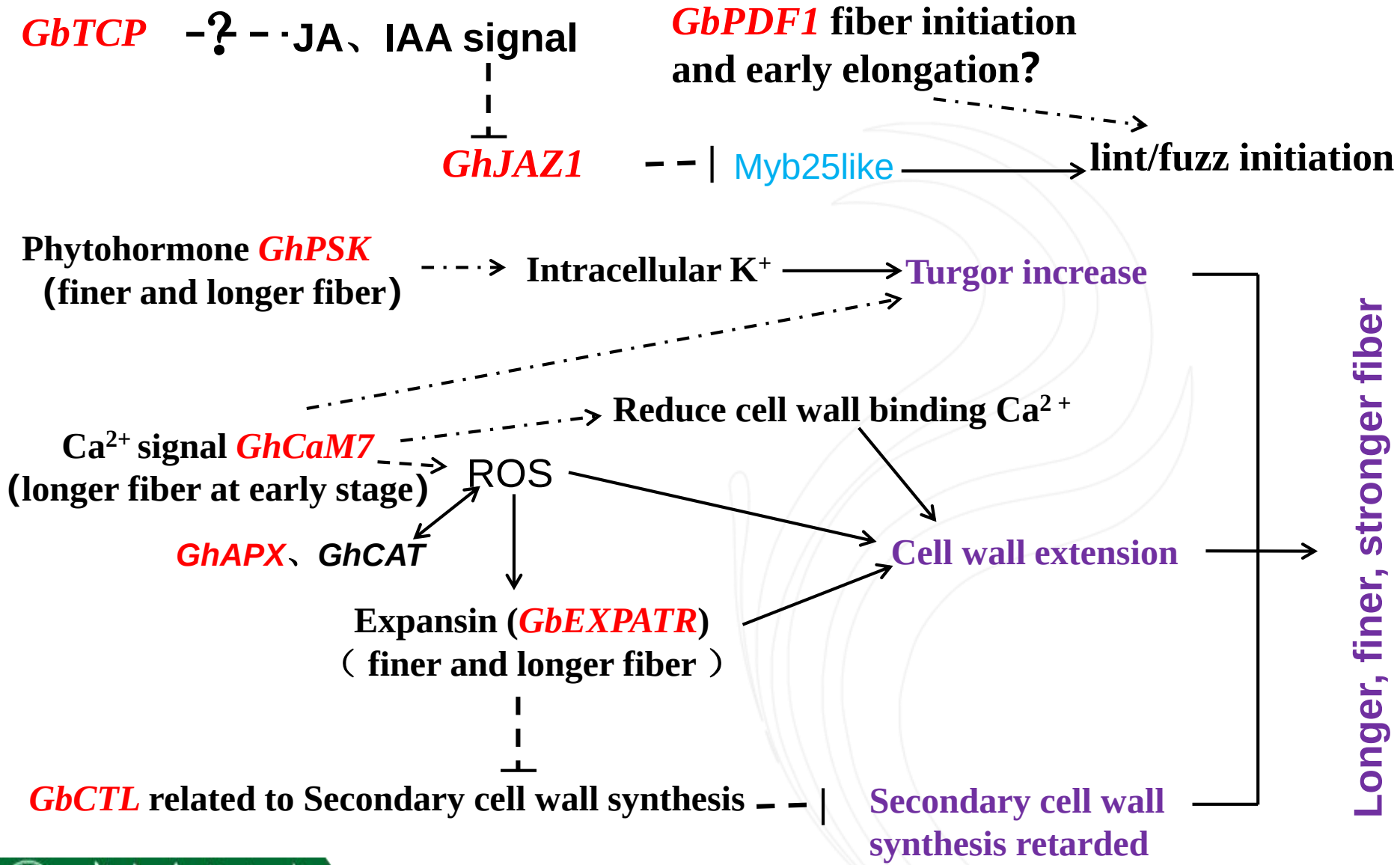
Cytoskeleton

Transcription factors

Cell wall related genes

Total: >100 genes for CRISPR/Cas9

The regulation network will be more detailed



GhF3H , flavonoid biosynthetic , cell wall synthesis?

Comparison of the sequenced cotton genomes

Category	*G. raimondii* (D5) (BGI)	*G. raimondii* (D5) (JGI)	*G. arboreum* (A2) (BGI)	*G. hirsutum* (AD1) (BGI)	*G. hirsutum* (AD1) (NBI)	*G. barbadense* (AD2) (HAU)	*G. barbadense* (AD2) (CAS)
Assembly strategy	WGS	WGS+BAC	WGS+BAC	WGS+BAC	WGS+BAC	WGS	WGS+BAC
Total assembly size, Mb	775.2	761.4	1,694	2,173	2546	2573	2,171
Total scaffold number	4,715	1,084	7,914	8,591	40,407	29,751	6,772
scaffold N50, Mb	2.3	18.8	0.66	0.76	1.6	0.26	0.5
Anchored and oriented scaffolds, Mb	406.3	748.7	1,532	1,923	1,934	1,997	1,950
Percentage of repeat sequences	57.00%	61.00%	68.50%	67.20%	64.80%	69.10%	63.20%
Number of protein-coding genes	40,976	37,505	40,134	76,943	70,478	80,876	77,526

Low continuity of tetraploid cotton genome assembly

Deficiencies of next-generation sequencing (NGS)

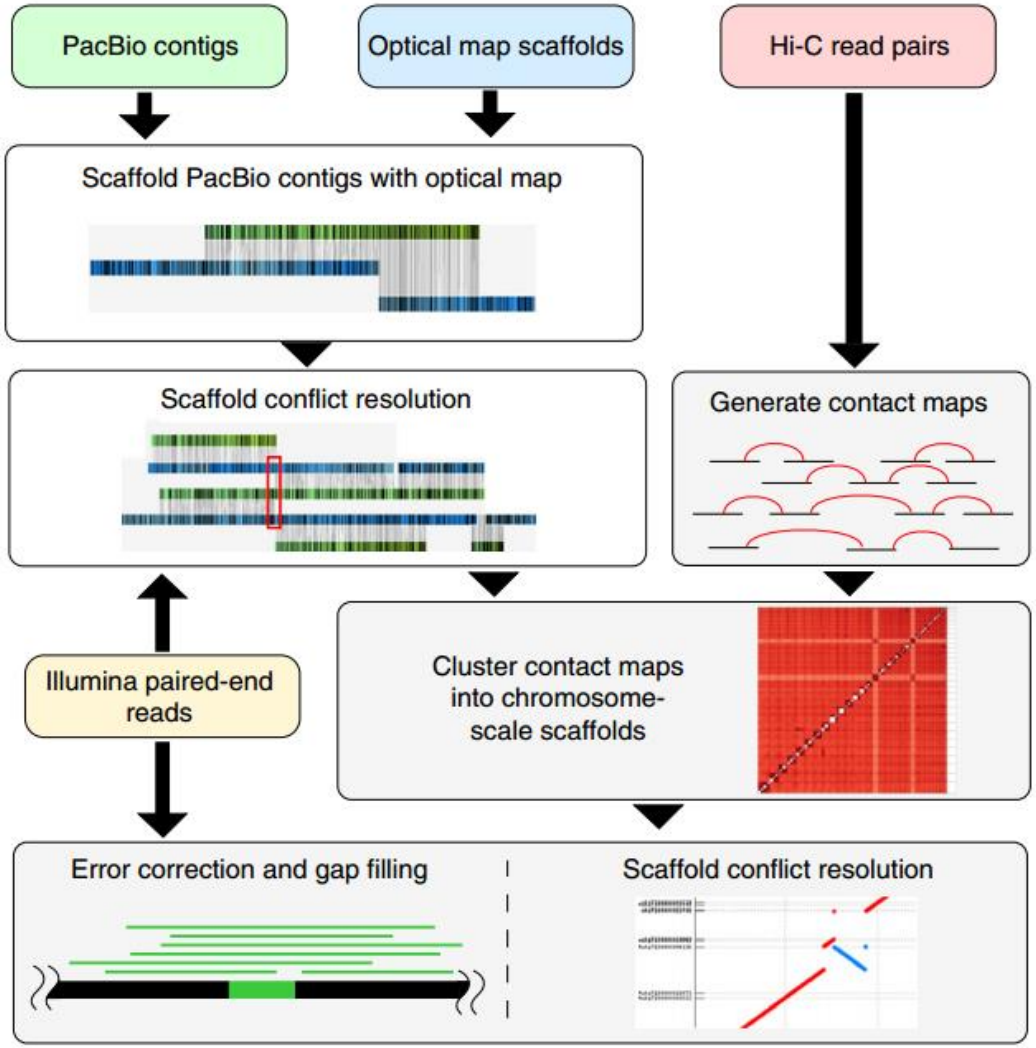
1. The NGS-based approach leaves assembly discontinuity and many assembly gaps in scaffolds;

2. The NGS-based approach fails to assemble pericentromeric regions because of sequencing bias;

3. The mixed assembly between subgenomes;

4. Genome is fragmentary with scaffold anchoring errors.

Call for reference-grade tetraploid genome sequence



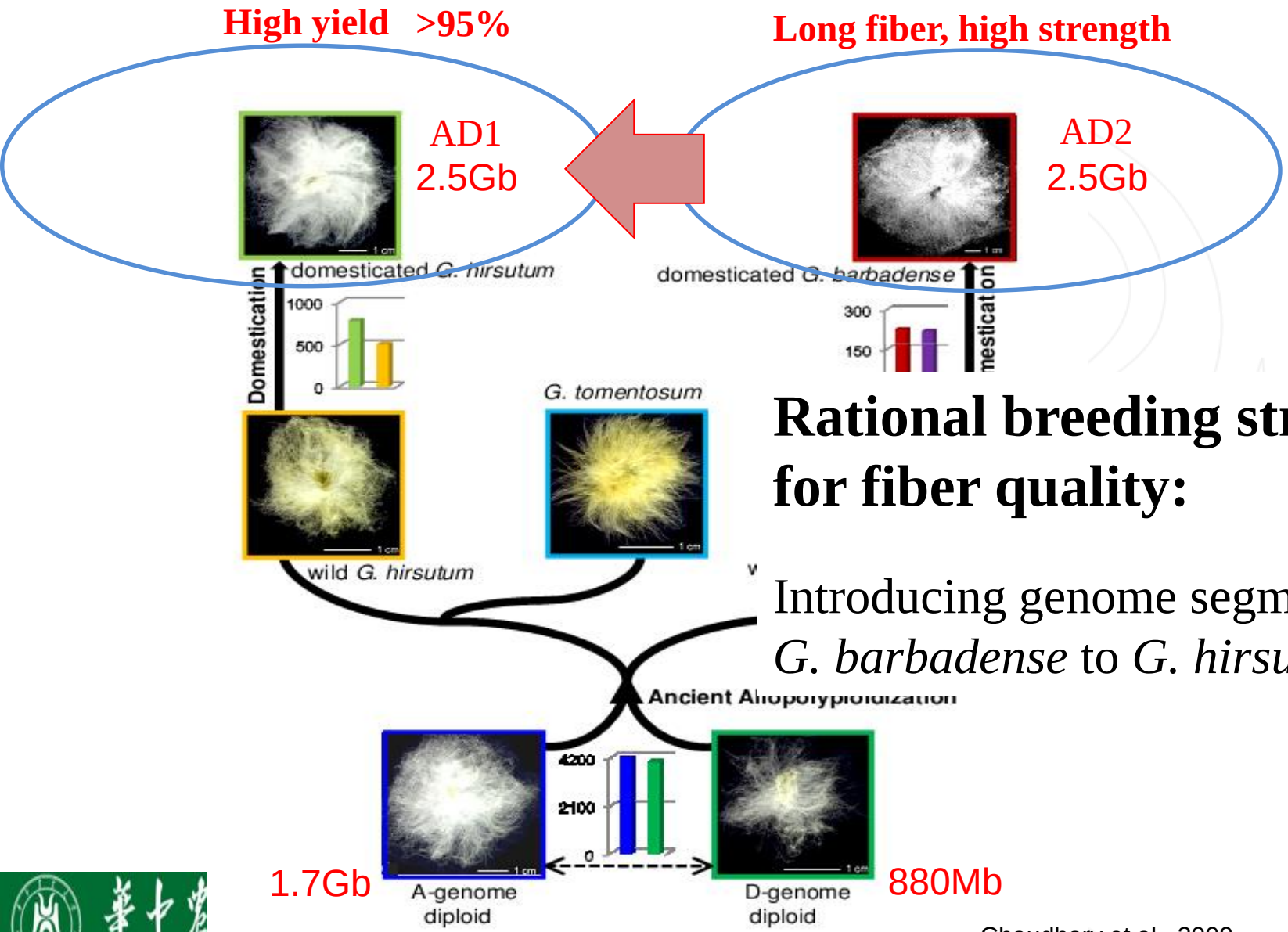
PacBio/Nanopore sequencing
Long reads and unbiased sequencing

BioNano optical maps
Construct scaffolds by using contigs,
Correct assembly errors
Estimate gaps

Hi-C technique
Construct super-scaffolds and chromosome-scale assembly

Bickhart et al. Nat Genet, 2017

Genome-enabled gene transfer from *G. barbadense* to *G. hirsutum*



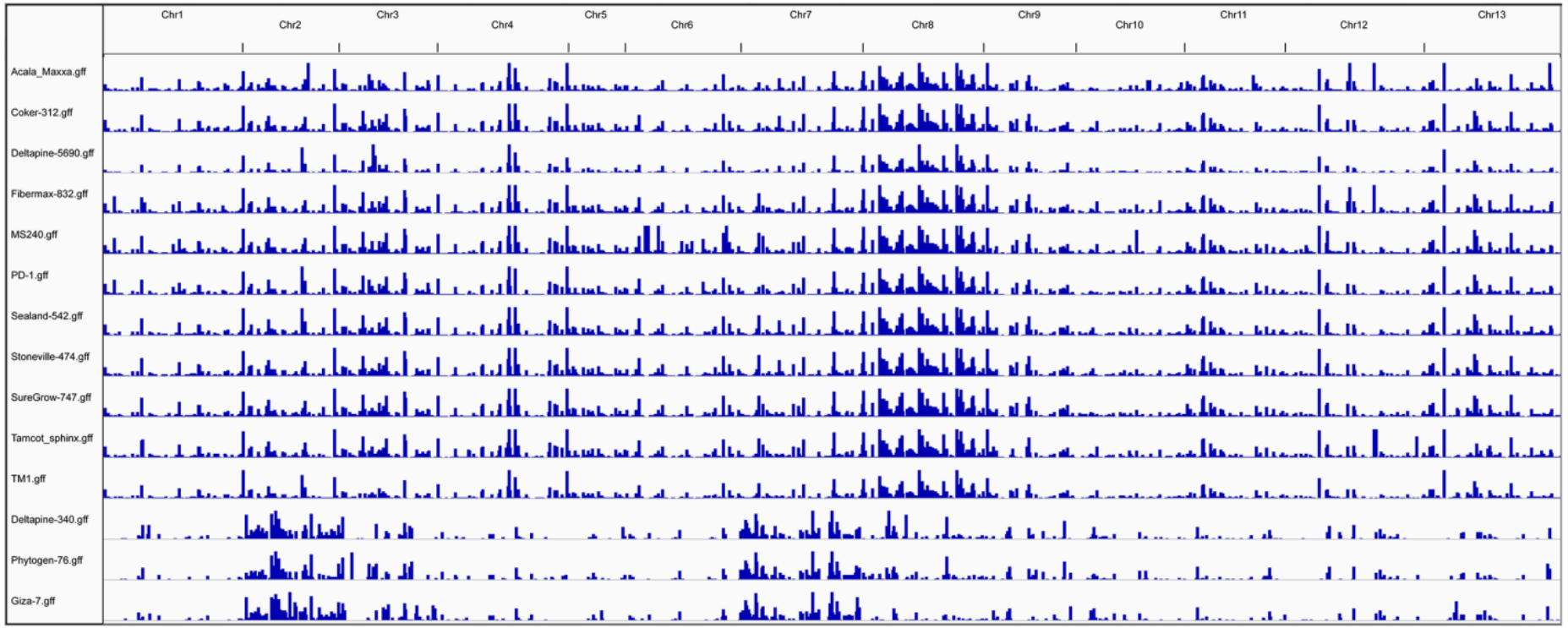
Rational breeding strategy for fiber quality:

Introducing genome segments from *G. barbadense* to *G. hirsutum*

Chaudhary et al., 2009

Interspecies introgression during cotton breeding

Introgression between *G. hirsutum* and *G. barbadense* varies across the At-genome.

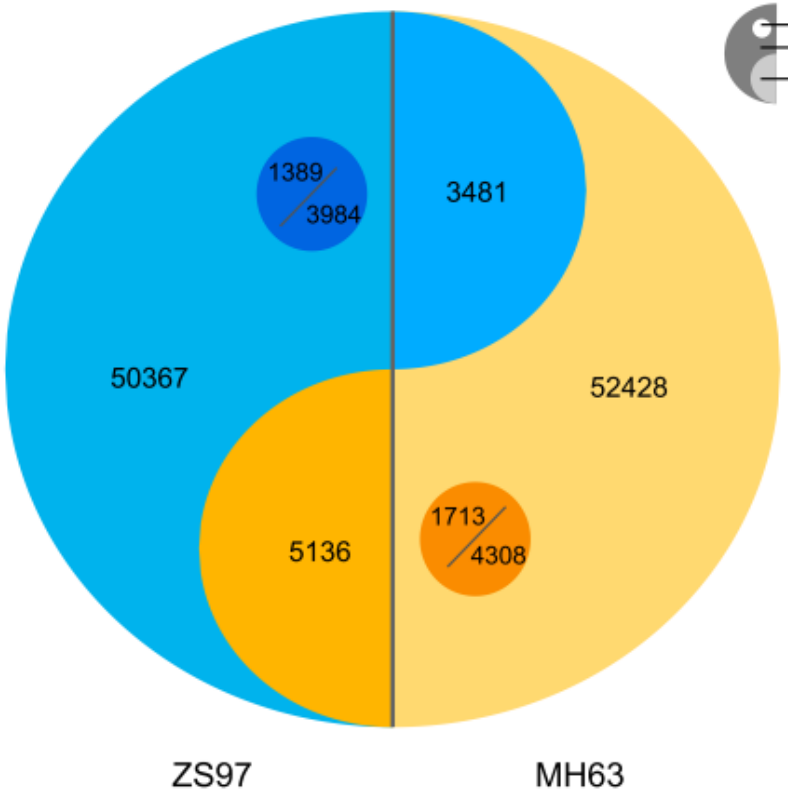


Page et al., PloS Genetics, 2016

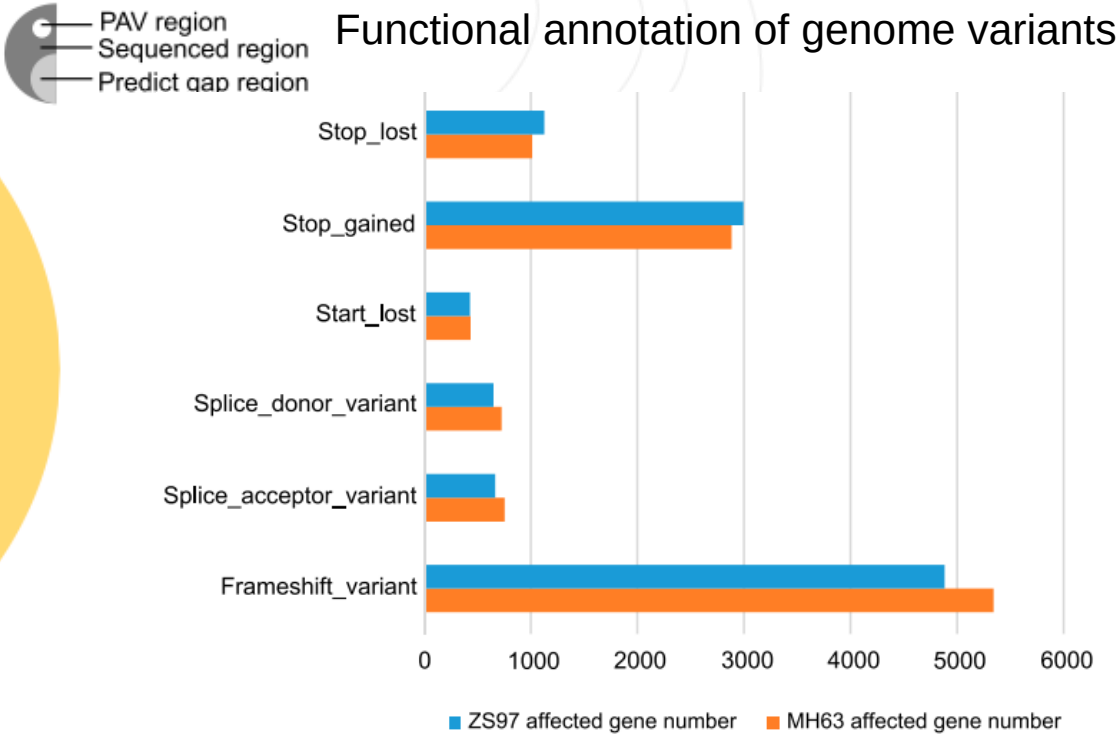
Reference-scale genome sequence should be provided for more efficient genome-based breeding

1. Directly explore genomic differences between two representative species.

The presence/absence variations in two rice species



Functional annotation of genome variants

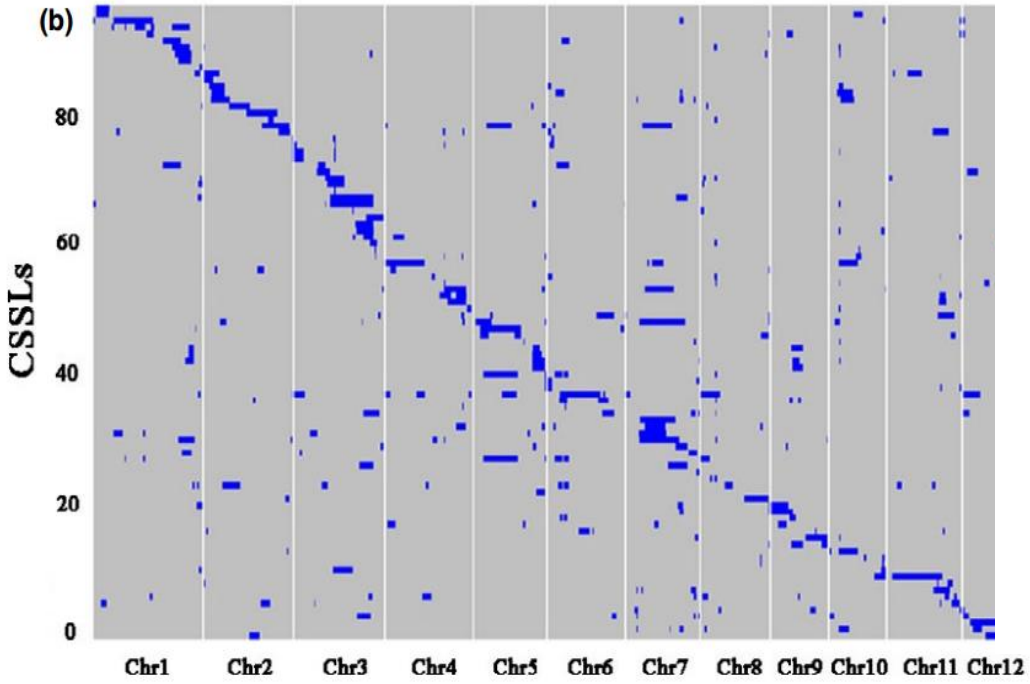


Zhang et al, PNAS, 2016

Reference-scale genome sequence should be provided for more efficient genome-based breeding

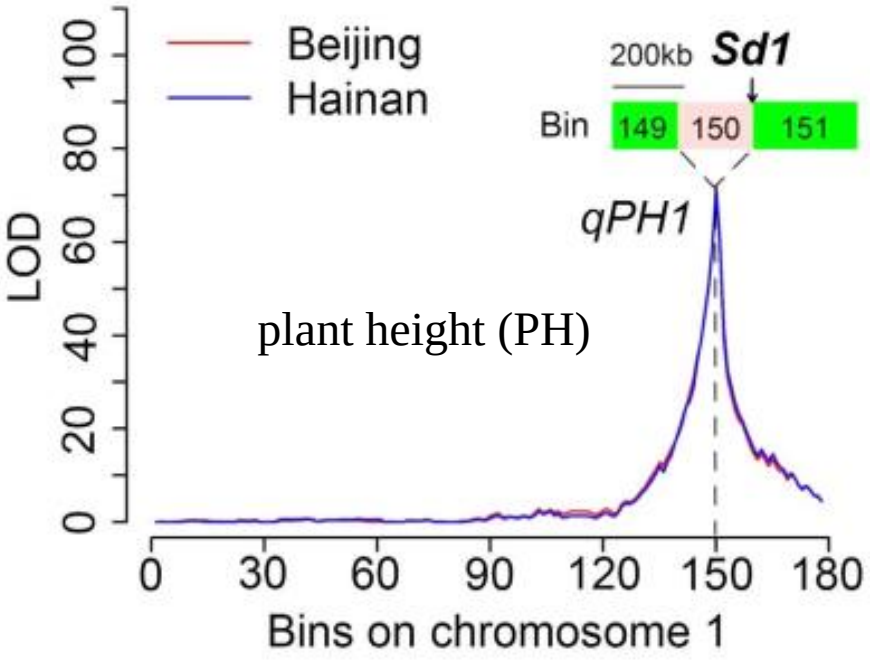
2. Identification of casual variants in introgression lines by QTL mapping and GWAS

Introgression lines between two rice species



Zhu et al., 2015

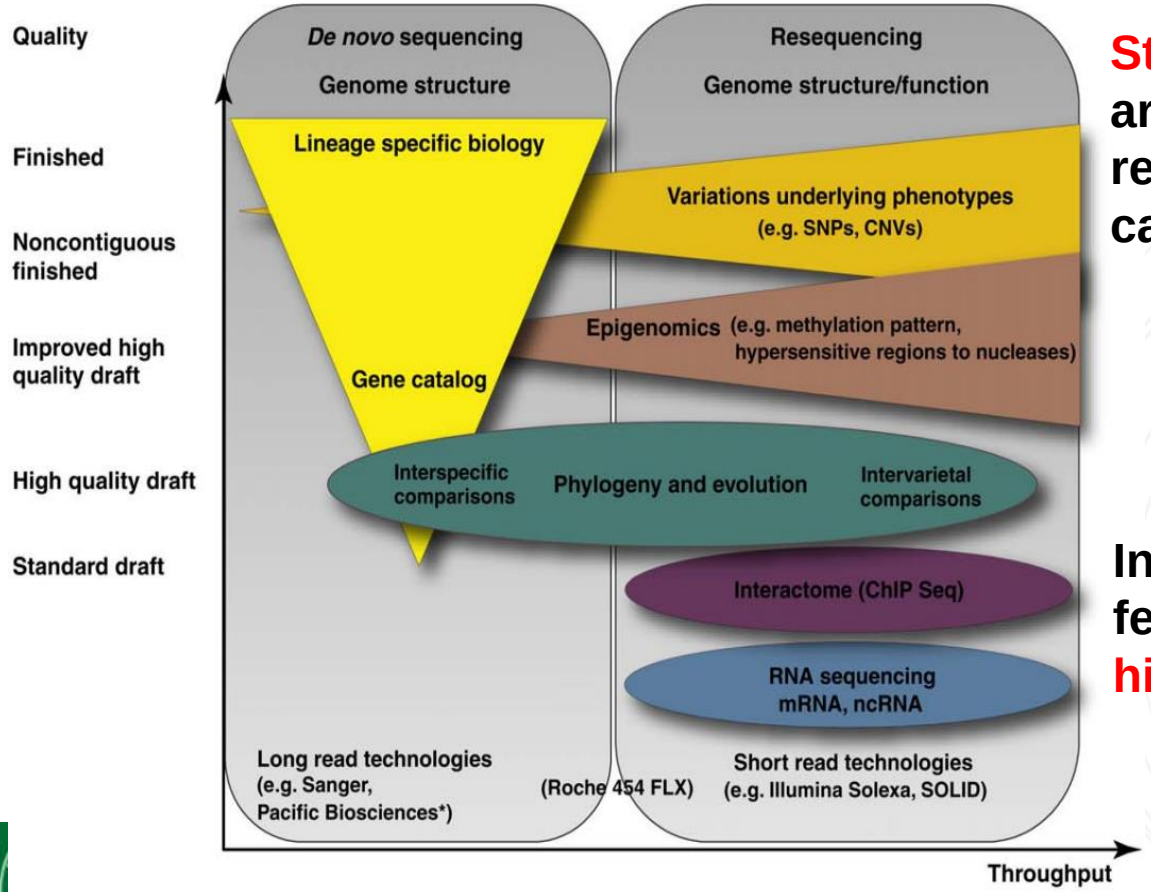
QTL mapping using introgression lines



Ma et al., 2016

Reference-scale genome sequence should be provided for more efficient genome-based breeding

3. Accurate genome sequences facilitate functional gene discovery



Standard draft genome sequences are only suitable for establishing a relatively comprehensive gene catalog.

Inferences about lineage-specific features can only be achieved with high-quality complete sequences.

Feuillet et al., 2011

Acknowledgement

Dr. Lili Tu
Dr. DaojunYuan
Dr. Maojun Wang

Prof. Keith Lindsey
University of Durham

Funding sources:

NSFC

Ministry of Sci & Tech

Ministry of Agriculture

Ministry of Education

Thank you!