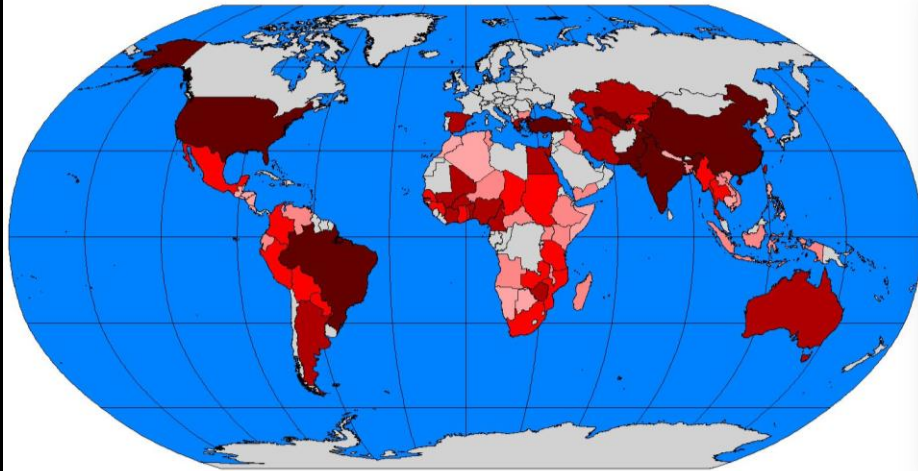


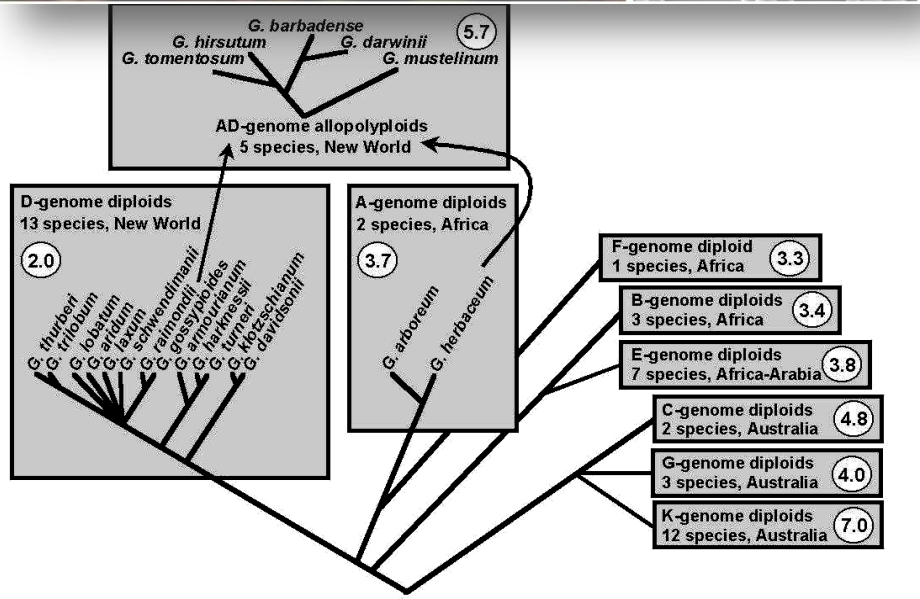
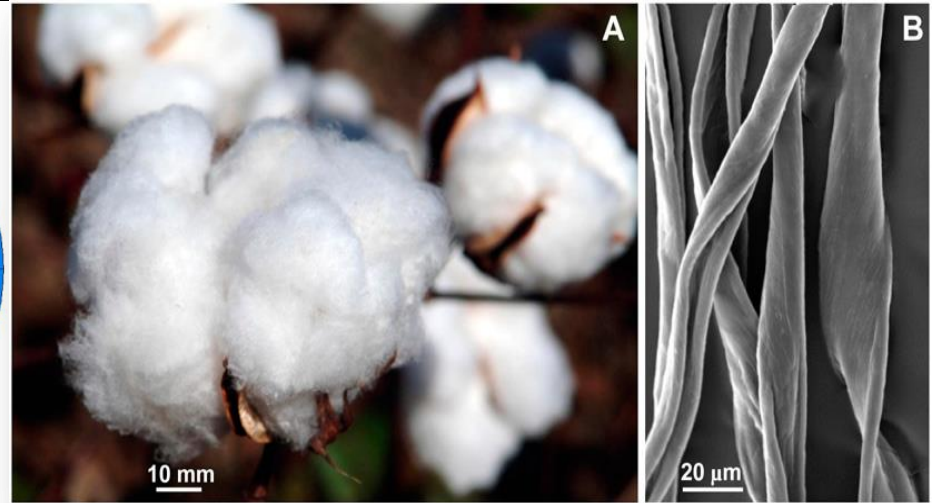
The worldwide economic and scientific importance of cotton warrants eventual DNA sequencing of many *Gossypium* species to answer a wide range of questions



Cotton Cultivation (Metric Tons Lint)

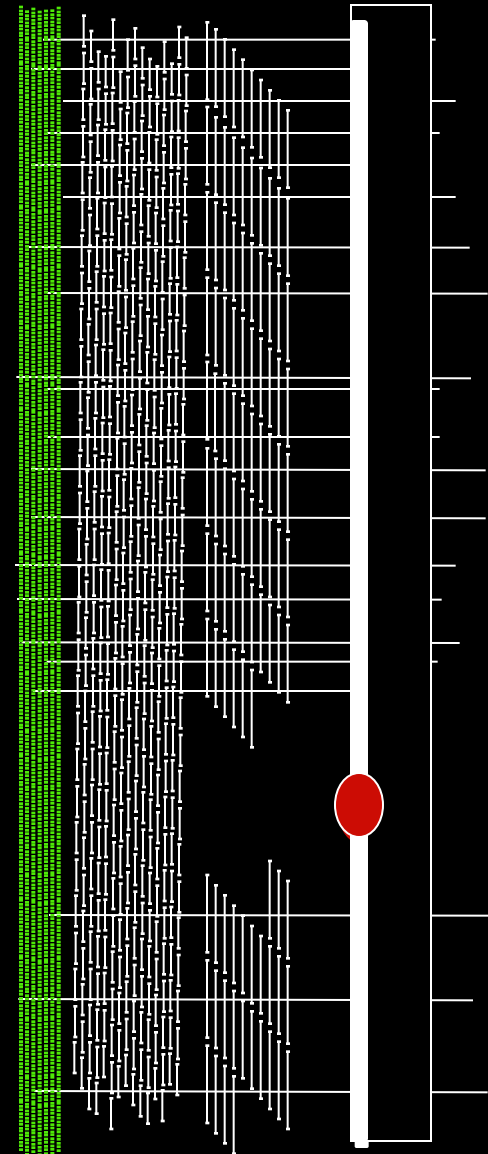
- 1 - 1980
- 2000 - 15000
- 15066 - 97200
- 102000 - 693000
- 872150 - 5323510

Source: FAO



The path to a gold-standard cotton sequence

1. Genetic map of all chromosomes (complete, Rong et al 2004)
2. Physical map, interleaving genetic map with high-coverage BAC library in which each clone has been fingerprinted and end-sequenced to determine relation to one another
 - a. Fingerprinting completed under NSF and CI support, published (Lin et al 2010)
 - b. Paired ends of ~ 9 genome-equivalents of BACs completed.
3. Paired-ends of 15 genome-equiv of fosmids completed (JGI).
4. About 16x genome -eq. 454 sequence completed.
5. Illumina full run (2 x 150 nt), gap-filling and error checking





Cotton
Incorporated

'Gold-Standard' Cotton Genome Sequence Will Advance Fiber, Fuel, and Food Applications

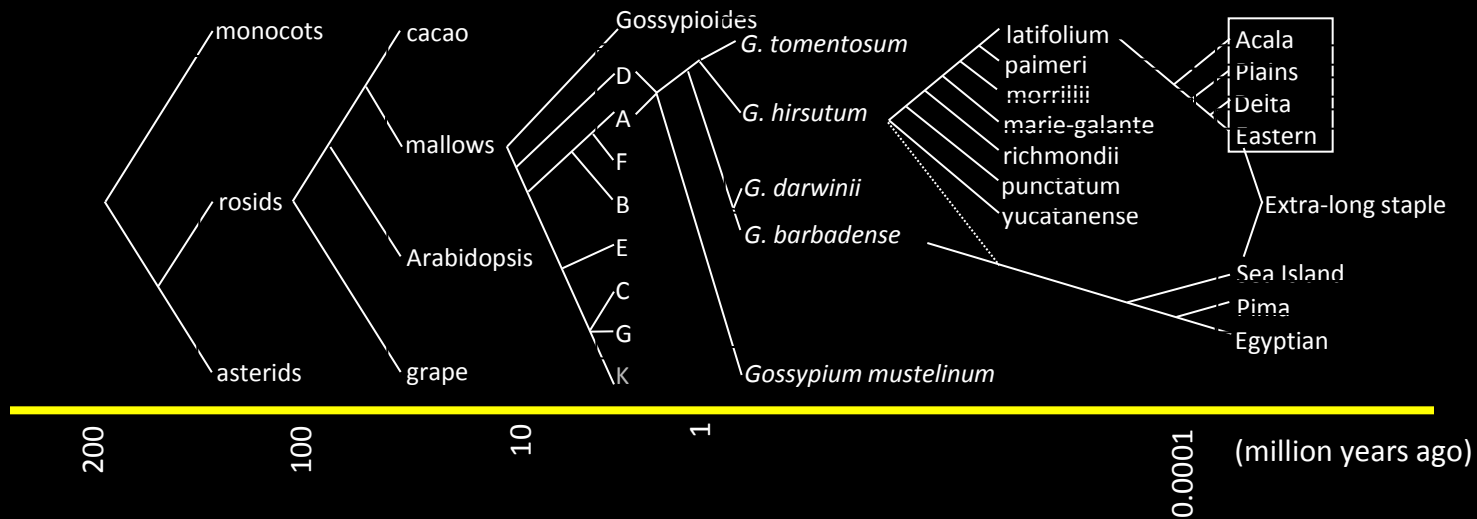
Thursday January 05, 2012

New York, NY

An international consortium, led by Professor Andrew Paterson of the University of Georgia, has made publicly available the first 'gold-standard' genome sequence for cotton. Cotton was among the first plants studied at the molecular level, and the sequence obtained by Paterson and his team is the culmination of a 20-plus year effort in the analysis of cotton genes and genomic DNA. This critical sequence will be invaluable to better understanding and optimizing the production and sustainability of the cotton plant.

We did it!

Some Most of us will enjoy access to sequences of all Gossypium genome types, and also diverse genotypes, during our careers





Cotton
Incorporated

'Gold-Standard' Cotton Genome Sequence Will Advance Fiber, Fuel, and Food Applications

Thursday January 05, 2012

New York, NY

An international consortium, led by Professor Andrew Paterson of the University of Georgia, has made publicly available the first 'gold-standard' genome sequence for cotton. Cotton was among the first plants studied at the molecular level, and the sequence obtained by Paterson and his team is the culmination of a 20-plus year effort in the analysis of cotton genes and genomic DNA. This critical sequence will be invaluable to better understanding and optimizing the production and sustainability of the cotton plant.

We did it!

The draft genome of a diploid cotton *Gossypium raimondii*

Kunbo Wang^{1,6}, Zhiwen Wang^{2,6}, Fuguang Li^{1,6}, Wuwei Ye^{1,6}, Junyi Wang^{2,6}, Guoli Song^{1,6}, Zhen Yue², Lin Cong², Haihong Shang¹, Shilin Zhu², Changsong Zou¹, Qin Li³, Youlu Yuan¹, Cairui Lu¹, Hengling Wei¹, Caiyun Gou², Zequn Zheng², Ye Yin², Xueyan Zhang¹, Kun Liu¹, Bo Wang², Chi Song², Nan Shi², Russell J Kohel⁴, Richard G Percy⁴, John Z Yu⁴, Yu-Xian Zhu³, Jun Wang^{2,5} & Shuxun Yu¹

We have sequenced and assembled a draft genome of *G. raimondii*, whose progenitor is the putative contributor of the D subgenome to the economically important fiber-producing cotton species *Gossypium hirsutum* and *Gossypium barbadense*. Over 73% of the assembled sequences were anchored on 13 *G. raimondii* chromosomes. The genome contains 40,976 protein-coding genes, with 92.2% of these further confirmed by transcriptome data. Evidence of the hexaploidization event shared by the eudicots as well as of a cotton-specific whole-genome duplication approximately 13–20 million years ago was observed. We identified 2,355 syntenic blocks in the *G. raimondii* genome, and we found that approximately 40% of the paralogous genes were present in more than 1 block, which suggests that this genome has undergone substantial chromosome rearrangement during its evolution. Cotton, and probably *Theobroma cacao*, are the only sequenced plant species that possess an authentic *CDN1* gene family for gossypol biosynthesis, as revealed by phylogenetic analysis.

... and more!

"If the only tool you have is a hammer, you tend to see every problem as a nail."

Abraham Maslow, psychologist (1908-1970)

" My father's cardinal lesson was 'Use the right tool for the job.' If you have ever tried to hammer a nail with a can of WD-40, you know what he means. Wisdom ... is the right tool ... "

David Michaelis, biographer (CC '94)

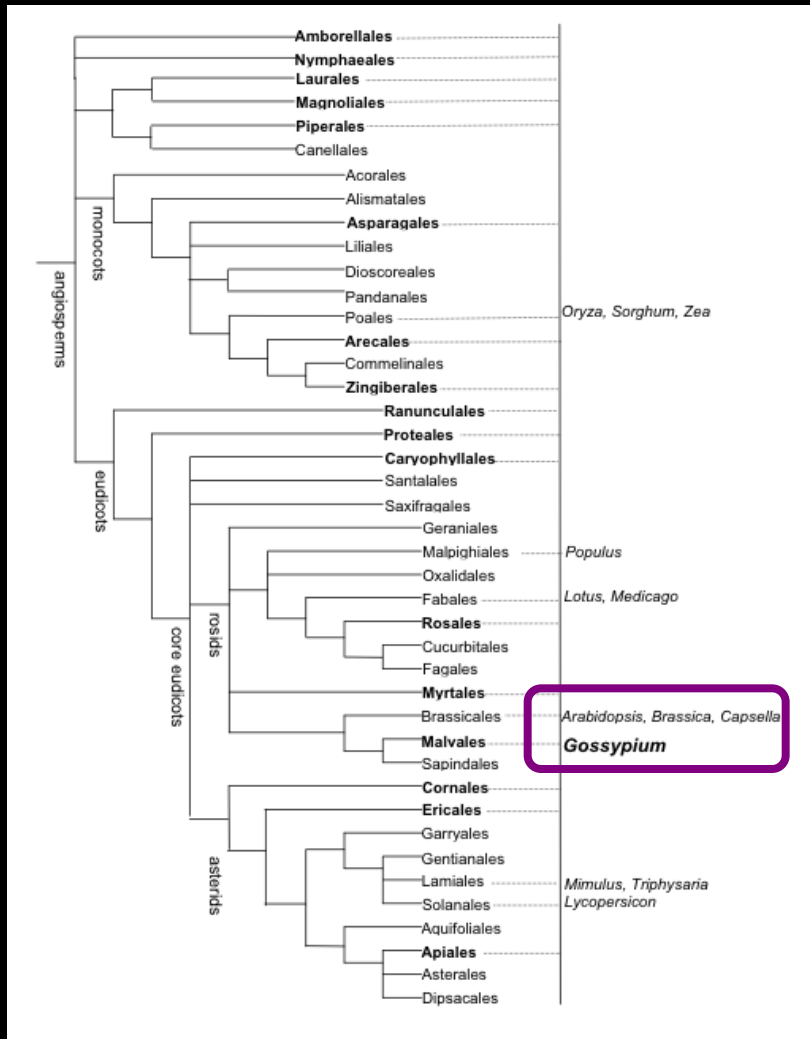
Parameters of different G. raimondii genome assemblies

	<i>Draft</i>	<i>Reference</i>
Scaffold number	4715	1084
N50 (Megabases)	2.3	18.8
Longest	12.8	52.1
Anchored and oriented % genome	52.4%	98.3%
Gene number	40,976	37,505

Draft and reference assemblies are different tools, regarding ...

1. Ability to infer gene orthology and paralogy by robust colinearity-based approaches, especially in large gene families and in angiosperm genomes rich in duplicated genes;
[important to both fundamental (evolutionary) and practical (SNP calling) applications]
2. Ability to infer paleo-evolutionary history, where high contiguity and complete genome coverage are important to intra- and inter-genomic alignments;
[important to 'translational genomics', leveraging information from other taxa in accelerating research progress in cotton]
3. Understanding of the 'genomic landscape,' specifically differences in the nature and properties of euchromatin (which is relatively easy to assemble) and heterochromatin (which is recalcitrant to assembly in many draft sequences)
[important to bridging genomics and the rich history of cotton cytogenetics]

The pot of gold at the end of the genome (so now we have all the answers, right)?



Good news:

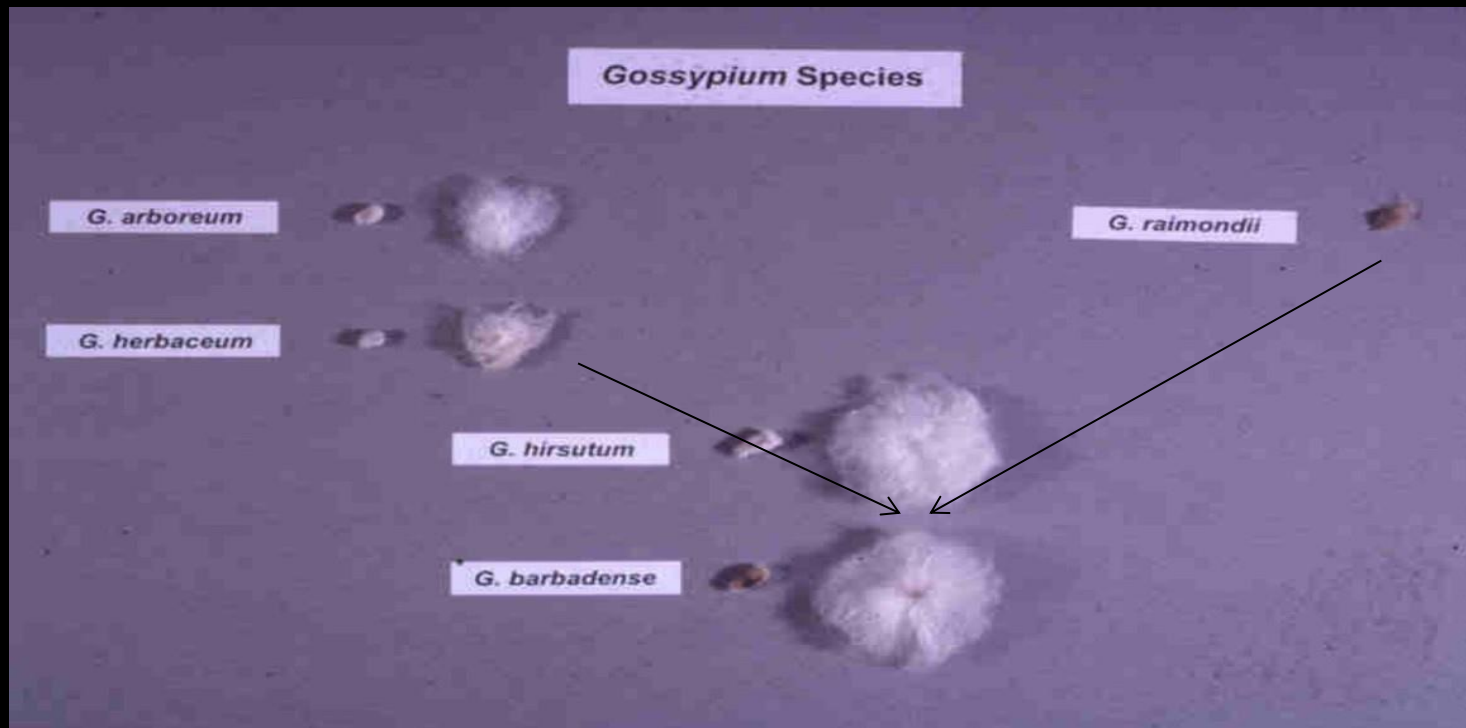
About 55-70% of cotton genes have discernible 'homologs' in *Arabidopsis* ...

permitting us to quickly learn about the many similarities that are fundamental to being a plant.



The pot of gold at the end of the genome?

Bad news: Many cotton genes are not recognizable in *Arabidopsis*, and/or have evolved new functions in the past ~80 million



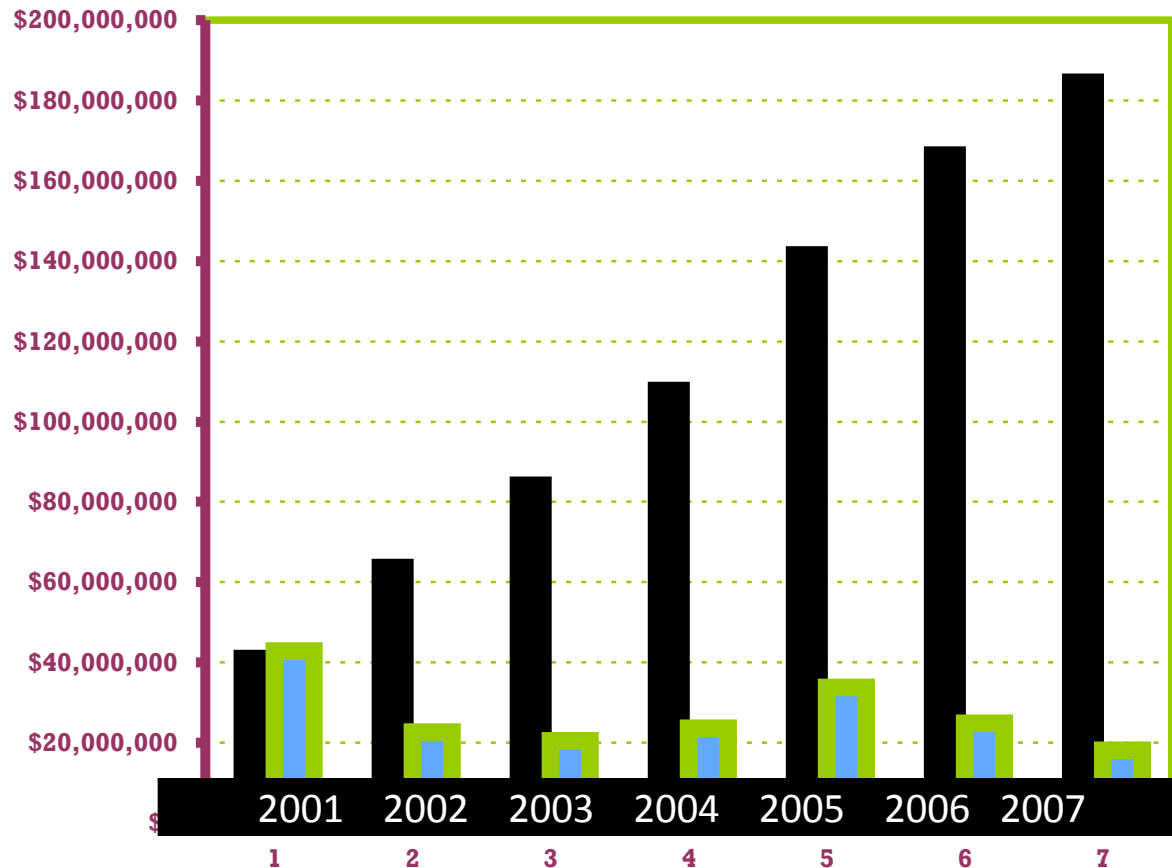
Key changes associated with polyploid formation 1-2 my ago have contributed to the superior phenotypes of the world's elite cottons

“The potential benefits of the post-genomic era in cotton are real and large – improved quality, productivity, and stability; reduced input needs that improve sustainability and environmental stewardship; and value-added features tailored to human needs rather than to self-perpetuation in the wild.”

The greatest challenge facing the cotton community is the conversion of ‘sequence’ to ‘knowledge,’ a challenge that will require investment, creativity, investment, energy, investment, coordination, investment, patience, and investment.”

(Source: Paterson, 2007, Proceedings of the World Cotton Res. Conf)

Arabidopsis got >\$200 million in competitive research funds, and the job is far from done ...



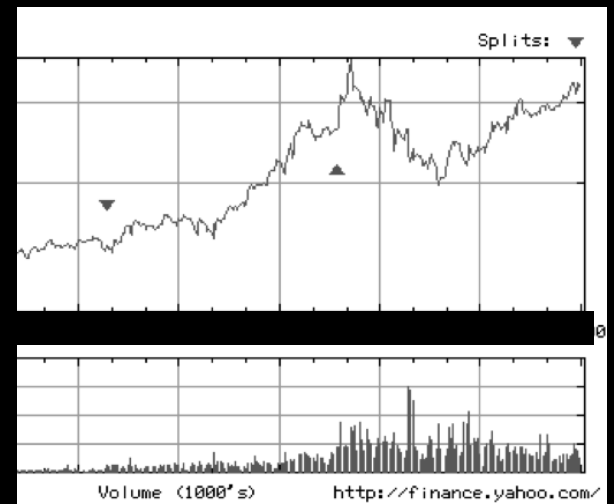
An industrial crop may benefit from a different approach than a botanical model



Utilize public-sector scientists as a virtual R&D resource to add ‘power’ to a unified database of enabling info and resources (with royalty-sharing incentives)...

... linked to ...

A ‘value capture’ mechanism to foster the investments needed to translate discoveries into economic growth.



Collette Abbey
Miguel Arias
Cem Bas
Dr John Bowers
Nino Brown
Rahul Chandnani
Dr. C. B. Chaporkar
Dr Peng Chee
Rosana Compton
Dawn Coplin-Williams
Dr Sayan Das
Dr Laura Decanini
Jian-min Dong
Jamie Estill
Lori Goff
Dr El-sayed Hafez
Dr Chunxiao Jiang
Sameer Khanal
Dr Tien-hung Lan
Lifeng (Phill) Lin
Barry Marler
Lisa Nelson
Dr Richard Noyes
Dr Mark Osterlund
Dr Chan-Hwan Park
Jinesh Patel
Dr Daniel Peterson
Gary Pierce

Dr Debashis Rana
Dr Alesia Reinisch
Dr Junkang Rong
Stefan Schulze
Kim Reischmann
Dr Haibao Tang
Dr Vijay Waghmare
Dr Guo-liang Wang
Dr Xiyin Wang
Dr Zining Wang
Dr Robert Wright
Dr Xinping Zhao
Dr. Dick Auld (TTU)
Dr Randy Boman (TTU)
Dr Steve Brown (UGA)
Dr Charles Cook (USDA-ARS)
Dr Kamal El-Zik (SCSC)
Dr John Gannaway (TAES-LB)
Dr Alan Gingle (UGA)
Dr Ishwarappa Katageri
Dr Don Jones
Dr Russell Kohel (USDA-ARS)
Dr Lloyd May (Deltapine)
Dr Tom McKnight (Biology)
Dr Ed Percival (USDA-ARS)
Dr Jim Price (SCSC)
Dr. Mehboob ur Rahman (Pakistan)
Dr Yehoshua Saranga (Israel)
Dr Wayne Smith (SCSC)
Dr David Stelly (SCSC)
Dr Peggy Thaxton (MS State)
Dr Jonathan Wendel (Iowa St.)
Dr Thea Wilkins (TTU)
Dr Rod Wing (UAZ)
Dr Essam Zaki (Egypt)
Dr Brent Zehr (Mahyco)

Thanks to:

Personnel and collaborators (left);

The 73 co-authors of the reference genome sequence (in review) ;

Our research sponsors (below);

Drs Wendel and Yu, for a stimulating start to this meeting, and;

You, the audience ...in advance, for what I expect to be a constructive and forward-looking dialogue during this meeting and beyond.

Funding:

National Science Foundation Plant Genome Research Program

USDA National Research Initiative and BARD program

Texas and Georgia State Support Committees, Cotton Inc.

Consortium for Plant Biotechnology Research and associated firms

US Dept of Energy Joint Genome Institute Community Sequencing Program