

Sequencing of a Cultivated Diploid Cotton Genome - *Gossypium arboreum*

WILKINS Thea A

(Department of Plant and Soil Science, Texas Tech University, Lubbock, TX 79409, USA)

Sequencing the genomes of crop species and model systems contributes significantly to our understanding of the organization, structure and function of plant genomes. In a 'white paper' published in 2007, the cotton community set forth a strategic plan for sequencing the AD genome of cultivated upland cotton that initially targets less complex diploid genomes. This strategy banks on the high degree of conservation between diploid progenitors and AD species that will allow information derived from diploid genomes to be directly applied to the tetraploids. As part of this global plan, Texas Tech University (TTU) has launched a cotton genome sequencing initiative to characterize the gene space and reveal the allelic and transcriptome architecture of the cotton genome by deep sequencing of a cotton (*Gossypium*) progenitor and a cultivated A-genome cotton species. Exploiting the very latest advances in next-generation deep sequencing technology, whole-genome shotgun (WGS) sequencing of the ancestral genome of *Gossypioides kirkii* will be employed to reveal gene content and organization in cotton lineages as a reference to justify sequencing and facilitate the assembly of more complex *Gossypium* genomes. With this in mind, the TTU team elected to target the ancestral genome of *G. kirkii* especially in light of the following considerations: *G. kirkii* is equidistant on the phylogenetic tree to A, D and AD genomes, and has a relatively small genome with minimal repetitive DNA that therefore makes this species eminently suited for this express purpose and approach. Since spinnable fiber first evolved in the A-genome lineage, the TTU team will target WGS of an elite cultivar of *Gossypium arboreum* to identify genes in the gene-space regions of the A-genome that govern important fiber quality traits. The distinct advantage of targeting both *G. arboreum* and *G. kirkii* genomes for WGS will be our ability to glean ancestral traits and genetic organization of the cotton tribe based on the *G. kirkii* genome, which when coupled with the A-genome, will enable identification of fiber genes that evolved during domestication of the A-genome species. This effort will be augmented by the recently completed pilot project funded by the DOE JGI CSP recently completed provides 0.6X genome coverage from the D-genome progenitor (*G. raimondii*). To aid in the assembly of draft genome sequences, there is an urgent need to expand the characterization of the cotton transcriptome as the gene sequences deposited to date in the public sector are not representative of the transcriptome. To address this need, the TTU genome project will launch a gene discovery project to generate ~10 Gb of sequence by whole transcriptome sequencing (WTS) from a wide variety of tissues for digital expression analysis. Preliminary WTS analysis from a fiber library has already yielded thousands of newly identified novel genes. One of the highly anticipated outcomes of the TTU genome project will be a search for allelic diversity in genic regions of fiber genes that will be targeted for development and deployment of DNA markers to molecular breeding programs.