

The International Cotton Genome Initiative

Roy G. Cantrell¹, O. Lloyd May², Curt Brubaker³ and Tian-zhen Zhang⁴

¹Cotton Incorporated, Agricultural Research, Cary NC USA

²University of Georgia, Crop & Soil Science, Tifton GA USA

³CSIRO Plant Industry, Centre for Plant Biodiversity Research, Canberra AUSTRALIA

⁴Nanjing Agricultural University, National Key Laboratory of Crop Genetics & Germplasm Enhancement, Nanjing CHINA

Correspondence author RCantrell@cottoninc.com

ABSTRACT

Genome technologies present an unprecedented opportunity for the scientific community to make significant genetic improvements in cotton. The International Cotton Genomics Initiative (ICGI) at the fundamental level is an effort to develop a framework for collaboration and cooperation on cotton genomics research and its application. The last decade has seen unprecedented advances in the use of DNA technology to unravel the genetic secrets of plants and animals. International collaborative efforts are underway to map and characterize the genomes of many important organisms. The recent publication of the sequencing effort for *Arabidopsis* and *Oryza* illustrates the scope and power of the technology. The study of the complex allopolyploid cotton genome is scientifically very challenging and requires a coordinated multidisciplinary research effort. Many challenges face cotton production at the present and in the future. Concerns over water use and pesticide inputs abound. In some important growing regions of the world, a production plateau has been reached and stability of production is a serious concern and intense competition from man-made fibers jeopardizes future profitability of cotton production. The ICGI was created at a meeting of cotton researchers in Canberra, Australia in February 2000. Since then, productive workshops have been held in Montpellier, France in 2001 and Nanjing, China in 2002. The objectives of ICGI are to: 1) reduce redundancy of research effort and maximize rate of progress in research to understand the cotton genome by providing a forum for international researchers, 2) foster tool development to begin integrating genetic and physical maps, 3) accelerate development of consensus cotton linkage map comprised of framework markers that are transferable from lab to lab, 4) foster rapid application of new genomic tools to cotton improvement, 5) develop comprehensive forum for exchange and communication within cotton scientific community and with the *Arabidopsis* model genome community and 6) develop standardized nomenclature for DNA markers, maps and etc. Progress has been made in many of these areas the past three years yet much of the results have not made it out of the individual labs working many times in isolation from each other. Barriers still exist that prevent meaningful collaboration that would enable real gains in cotton genomics that would

lead to sustained genetic improvements in cotton germplasm. ICGI is the only forum that can facilitate the necessary multidisciplinary research efforts on a global scale to address some of these issues.

Introduction

Cotton plays a critical role as a sustainable fiber product compared to man-made synthetic fibers that require significant inputs from petroleum-based products. Cotton has for many years been viewed as the world's most important cash crop, and sustains the agricultural economies of many nations. Unfortunately, many challenges face cotton production at the present and in the future. Concerns over water use and low profitability abound. Cotton breeders and biotechnologists have tried to respond to these challenges yet a sustained production plateau has been observed for the last decade in the United States (Anonymous, 1999). This is very similar to that reported on a worldwide basis for grain crops (Mann, 1999). Conventional plant breeding to simultaneously enhance production and fiber quality (value-added fiber) has been hindered by complex antagonistic genetic relationships (Green and Culp, 1990). To a certain extent, biotechnology approaches focusing on a few single transgene events may face the same obstacle unless comprehensive genomic information is available for genes involved in cotton fiber development. Knowledge of possible preferred sites for insertion of transgenes in cotton genetic engineering can be generated from genomic studies (Kohel *et al.*, 2000). Genomic information of this type will be essential for future registration and approval of genetically engineered cotton as well as information on the efficacy and stability of the transgene event in the cotton genome. The major biotechnological initiatives in cotton (*G. hirsutum* L.) are centered in the private sector with focus on single-event transgenic technology to provide various forms of plant protection and improve the inherent value of the fiber. Genomics will permit dramatic genetic modification of fiber development to improve productivity and fiber quality. Powerful genomic tools for fundamental research of cotton fiber development include recently developed fiber-specific ESTs and large-insert cotton BAC resources.

The last decade has seen unprecedented advances in the use of DNA technology to unravel the genetic secrets of plants and animals. International collaborative efforts are underway to map and characterize the genomes of many important organisms. The most remarkable advances have been made in model genomes and diploid organisms. It is estimated that over 70% of Angiosperms including important crops (e.g. canola, coffee, cotton, oats, sugarcane, and wheat) are polyploid. The study of the complex allopolyploid cotton genome is scientifically very challenging and requires a coordinated multi-disciplinary research effort. To significantly advance genomic research in cot-

ton, the International Cotton Genome Initiative (ICGI) was formed with the following objectives:

- A. Reduce redundancy of similar work by genome scientists and maximize rate of progress in research efforts to understand the cotton genome.
- B. Accelerate development of a portable consensus genetic linkage map of cotton through distribution of PCR-based framework markers and linkage maps containing those markers.
- C. Foster tool development and collaborations to begin integrating genetic and physical map.
- D. Foster rapid application of new genomic tools and information to:
 1. Polyploid evolution of *Gossypium* spp.
 2. Cytogenetics of *Gossypium* spp.
 3. *Gossypium* germplasm resources and enhancement
- E. Develop comprehensive forum for communication with *Arabidopsis* model genome community to foster application of model genome tools to the cotton genome.

Active international consortia are in place in many plant and animal systems. For example: The International Triticeae Mapping Initiative (ITMI), International Grass Genome Initiative, International Sugarcane Biotechnology Consortium, International Grape Genome Initiative, International Rice Genome Sequencing Project, and *Musa* Genomics Consortium. Many of these provide a key role in planning and conducting international genomic workshops for collaborations and scientific exchange and documentation.

Cotton (*Gossypium hirsutum* L. $2n=4x=52$, AADD) is an allopolyploid, and is characterized by a large, complex genome. To date about 1,500 DNA markers have collectively been developed for the cotton genome (Reinisch *et al.*, 1994; Paterson and Smith 1999). Several DNA marker linkage maps for cotton are published (Reinisch *et al.*, 1994; Shapley *et al.*, 1998a; Shapley *et al.*, 1996). These maps have provided a valuable tool for cotton gene and QTL identification, mapping and positional cloning. However, none of these linkage maps have the numbers of linkage groups as the haploid chromosome number of cotton ($n = 26$). Additional markers are needed to complete a saturated genetic map of cotton. Given that the genome size of cotton is 2,200 Mb/1C (Arumuganathan and Earle, 1991) covering 5,000 cM (Reinisch *et al.*, 1994), on average marker density it is about 1,500 Kb per DNA marker. This is far sparser than the DNA marker densities of other crops having comparable agronomic importance such as maize and soybean. DNA microsatellites or simple sequence repeats (SSR) have been developed for the cotton genome and are being anchored to cotton chromosomes and optimized for high-throughput genotyping (Liu *et al.*, 2000a; 2000b). As these are placed on the various cotton linkage maps worldwide, the opportunity exists for consensus map construction. This is a necessity because there are many traits of interest to cotton scientists, i.e.

yield, fiber quality, disease and insect resistance. Of course, no single mapping population is segregating for all of the alleles of genes controlling these traits. Much of the genetic marker data exists at various research centers for computation of a consensus map, yet concerted merging of the necessary data is lacking.

Fortunately, the physical DNA content of 400kb per 1cM in tetraploid *G. hirsutum* is less than that of organisms such as humans and tomato, and not substantially larger than that of *Arabidopsis thaliana* for which genes have been successfully cloned by "chromosome walking" (Paterson and Smith, 1999). Large amounts of moderately and highly repetitive DNA elements (Zhao *et al.*, 1998) in the cotton genome will complicate this approach of gene discovery. In addition, map-based cloning will be hindered in the polyploid genome by the presence of homoeologous duplicate loci that will yield ambiguous results on cotton BACs or YACs. Paterson and Smith (1999) suggested the exploitation of the diploids for physical mapping and map-based cloning to circumvent this difficulty.

The recent availability of large-insert BAC libraries of tetraploid and diploid *Gossypium* enable researchers to begin physical mapping of the cotton genome (Peterson *et al.*, 2000). The financial resources are unlikely to be available for large-scale physical mapping of the complete genome, however, targeted regions or genome islands are amendable to integrated genetic and physical mapping. Regions where quantitative trait loci reside (Jiang *et al.*, 2000; Ulloa *et al.*, 2000; Shapley *et al.*, 1998b) or genomic sites at or surrounding important transgene insertion events (Kohel *et al.*, 2000) are logical target regions. Physical mapping will have to be integrated with genetic mapping to resolve the ambiguities and complexities of gene duplication and homoeologies in the tetraploid genome. All of this will take a strategic coordinated effort to advance this aspect of cotton genomics and fully utilize the powerful genomic tools available.

The present status of cotton genomics and DNA markers was reviewed at length by a small group of scientists at a workshop held in Canberra Australia 16-17 February 2000. It became very obvious to the participants that an international collaborative effort was necessary to coordinate future cotton genomics research. Cotton genome research was presented as plagued by gaps in knowledge, redundancy of effort and lack of coordination. The announcement of the ICGI was first published in the *Journal of Cotton Science* as a letter to the cotton science community (<http://www.jcotsci.org/2000/issue02/toc.html>). It was also obvious at the workshop that funds were needed to sponsor and support ICGI activities for it to be successful. The international steering committee of the newly formed ICGI was charged with securing funds to continue the efforts of the ICGI. In early 2001, funds were solicited from diverse sectors of the cotton community with an interest in ICGI. The response from the private

sector for start-up support for ICGI was significant. Approximately 52 000 USD was raised for this purpose and deposited in an ICGI Foundation account. Donors include Monsanto (USA), Delta and Pineland Co. (USA), Stoneville Pedigreed Seed Co. (USA), CIRAD (France), Cotton Research and Development Corp. (AUS), Cotton Incorporated (USA), Cotton Foundation (USA), Syngenta (USA), Aventis (USA) and Dow AgroSciences (USA). This level of support from the global cotton industry attests to the timeliness and critical nature of ICGI to the cotton industry. The initial start-up support for ICGI covered the cost of the ICGI Workshop in Montpellier, France 5-7 June 2001. The official brochure for this meeting and the conference proceedings can be found at the ICGI web page (<http://icgi.tamu.edu>). There were 35 participants at this workshop representing researchers in the public and private sectors. Most of the France 2001 meeting was devoted to ICGI organizational discussions. The planning and coordination of this meeting was handled effectively by Drs. Jean-Marc LaCape and Marc Giband of CIRAD.

The Second ICGI Workshop was held in Nanjing, China on June 3-6, 2002. This was a remarkable event with the presentation of over 100 oral or poster papers on all aspects of cotton genetics and genomics. The success of this workshop is due to the efforts of Dr. Tian-zhen Zhang and coordination by Nanjing Agricultural University and the Chinese Academy of Science. This meeting serves as a model for future ICGI technical meetings. The abstracts of all presentations are published in a special issue of *Cotton Science* 2002 vol. 14.

The breadth of ICGI requires that careful organization structure be in place to handle the diverse subject areas effectively. Working Groups (WG) were created to provide a forum and mechanism for researchers of similar interest to organize effectively carry out the mission of ICGI. The WG designated for ICGI are: 1) structural genomics and mapping, 2) functional genomics, 3) germplasm and genetic stocks, 4) evolutionary and comparative genomics, and 5) bioinformatics. The individual WG are best positioned to define long-term research and collaborative goals. WG's can document resource gaps and organize informative meetings. An electronic election for Chairperson of each WG is scheduled for Spring 2003 and each WG Chair will serve on the ICGI Steering Committee. The Chair can communicate activities of the WG to the ICGI as a whole. The WG structure allows for individuals within ICGI to function in a smaller setting and evolve at a pace not dependent on the whole of ICGI.

The determined and focused ICGI-WG's can begin to make real scientific advances in understanding the cotton genome. Framework DNA marker development is a priority of the structural genomics WG. Contemporary PCR-based DNA markers have great util-

ity in bridging the gap between genetic mapping and genome sequencing (physical mapping) (Schuler 1998). They can be easily converted to sequence-tagged sites (STS) to facilitate the construction of high-resolution genetic maps as demonstrated in human (Dib *et al.*, 1996), mouse (Dietrich *et al.*, 1996) and rice (Chen *et al.*, 1997). STS markers can thus serve as a framework for saturation and expansion of genetic maps with additional DNA markers to achieve genome-wide coverage of the cotton genome (Liu *et al.*, 2000). This will facilitate integration of the genetic and physical maps (Schuler 1998). Framework DNA markers are especially useful in a disomic tetraploid, such as cotton (*Gossypium hirsutum* L., $2n=4x=52$, AADD) where RFLP maps have been constructed from both interspecific (Reinisch *et al.*, 1994; Lacape *et al.*, 2003) and intraspecific (Shapple *et al.*, 1998; Ulloa *et al.*, 2002) mapping populations. Genomic map development in a complex polyploid involves a stepwise process that builds upon previous genetic and cytogenetic information (Stelly, 1993). *G. hirsutum* L. aneuploid stocks are employed to locate markers to individual chromosomes and identify linkage groups to chromosomes. DNA markers are being used in combination with cytogenetic stocks to create framework maps and further localize markers to chromosome regions (Liu *et al.*, 2000; LaCape *et al.*, 2003). Genome-wide cotton genetic maps can assist in determining the correct homologous relationships of chromosomes in a disomic tetraploid like cotton. A total of 11 of the 13 expected homologous pairs of chromosomes in tetraploid cotton have been identified (Paterson and Smith, 1999). Yet, dense marker coverage of all cotton chromosomes and assignment to linkage groups on existing genetic maps is lacking.

DNA markers have tremendous potential in tagging genes (QTLs) controlling complex traits (Ulloa *et al.*, 2000; Ulloa and Meredith 2000; Jiang *et al.*, 2000). These complex traits often have significant genotype X environment interactions. As more portable cotton genetic populations are developed (i.e. recombinant inbred lines), extensive evaluation and QTL mapping would be possible through ICGI-WG. The verification of putative QTL's from population to population can also be achieved through the sharing of genetic data through ICGI-WG collaboration. This will ultimately contribute to tagging of major traits with DNA markers suitable for marker-assisted selection in cotton genetic improvement.

The functional genomics ICGI-WG collaborations can provide a valuable "conduit" for the flow of genetic information in complex genomes, such as cotton. Recent phylogenetic evidence based on parsimony analyses of two plastid genes (*rbcl* and *atpB*) and a nuclear gene (18S rDNA) place *Malvales* and *Brassicales* both in the "eurosoid II" large group hinting at the potential for comparative genome analysis of *Gossypium* spp. and *Arabidopsis* (Soltis *et al.*, 1999). At the functional genomic level, *Arabidopsis* is a pow-

erful research tool. The genetic resources to study functional genomics in *Gossypium* spp. has grown dramatically in recent years (<http://cfg.ucdavis.edu>).

The success of ICGI in the long-term depends on broad international participation and interest. This will lead to active Working Groups that achieve the mission of ICGI. Interested scientists are encouraged to sign up electronically at <http://icgi.tamu.edu>. The list of current participants is listed at that web site. This opportunity to forge productive international collaborative research efforts should not be missed. The sustained genetic improvement of the productivity and profitability of cotton globally depends on these modern technology advances. In Winter 2003, the new Chair of ICGI was elected, Dr. David Stelly (stelly@tamu.edu). The full and newly elected ICGI Steering Committee should be in place by Summer 2003.

References

- Anonymous, (1999). Report of the American Cotton Producers yield committee. (http://www.cotton.org/ncc/acp_yield_report.htm).
- Arumuganathan, K. and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, **9**: 208-218.
- Chen X, Temnykh S, Xu, Y., Cho, Y.G. and McCough, S.R. (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, **95**: 553-567.
- Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damronboles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., Evans, C.A., DeAngelis, M.M., Levinson, D.M., Kruglyak, L., Goodman, N., Copeland, N.G., Jenkins, N.A., Hawkins, T.L., Stein, L., Page, D.C. and Lander, E.S. (1996). A comprehensive genetic map of the mouse genome. *Nature*, **380**: 149-152.
- Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Wesssenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380**: 152-154.
- Green, C.C. and Culp, T.W. (1990). Simultaneous improvement of yield, fiber quality, and yarn strength. *Crop Science*, **30**: 66-69.
- Jiang, C., Wright, R., Woo, S., Del Monte, T. and Paterson, A. (2000). QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theoretical and Applied Genetics*, **100**: 409-418.
- Kohel, R.J., Quisenberry, J.E., Cartwright, G. and Yu, J. (2000). Linkage analysis of transgenes inserted into cotton via *Agrobacterium tumefaciens* transformation. *Journal of Cotton Science*, **4**: 65-69. (<http://www.jcotsci.org>).
- Lacape, J.M., Nguyen, T.B., Thibivilliers, S., Courtois, B., Cantrell, R.G., Burr, B. and Hau, B. (2003). A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome (in press)*.
- Liu, S., Saha, S., Stelly, D. Burr, B. and Cantrell, R.G. (2000a). Chromosomal assignment of microsatellite loci in cotton. *Journal of Heredity*, **91**: 326-332.
- Liu, S., Cantrell, R.G., McCarty, J.C. Jr. and Stewart, J. McD. (2000b). SSR-based assessment of genetic diversity in cotton race stock accessions. *Crop Science*, **40**: 1459-1469.
- Mann, C.C. (1999). Crop scientists seek a new revolution. *Science*, **283**: 310-314
- Paterson, A.H. and Smith, R.H. (1999). Future horizons: Biotechnology for cotton improvement. p.415-432. In C.W. Smith and J.T. Cothren (eds.) *Cotton: Origin, history, technology and production*. John Wiley & Sons, Inc, New York, NY.
- Peterson, D.G., Tomkins, J.P., Frisch, D.A., Wing R.A. and Paterson, A.H. (2000). Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide. *Journal of Agricultural Genomics*, **5** (<http://www.ncgr.org/research/jag/>).
- Schuler, G.D. (1998). Electronic PCR: bridging the gap between genome mapping and genome sequencing. *Trends in Biotech.*, **16**: 456-459.
- Shappelley, Z.W., Jenkins, J.N., Watson, C.E., Kahler A.L. and Meredith, W.R. (1996). Establishment of molecular markers and linkage groups in two F2 populations of Upland cotton. *Theoretical and Applied Genetics*, **92**: 915-919.
- Shappelley, Z.W., Jenkins, J.N., Meredith, W.R. and McCarty, J.C. Jr. (1998a). An RFLP linkage map of upland cotton, *Gossypium hirsutum* L. *Theoretical and Applied Genetics*, **97**: 756-761.
- Shappelley, Z.W., J Jenkins, N., Zhu, J. and McCarty, J.C. (1998b). Quantitative trait loci associated with agronomic and fiber traits of upland cotton. *Journal of Cotton Science*, **4**: 153-163 (<http://www.jcotsci.org>).
- Soltis, P.S., Soltis, D.E. and Chase, M.W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature*, **402**: 402-403.
- Stelly, D.M. (1993). Interfacing cytogenetics with the cotton genome mapping effort. In Beltwide Cotton Conference, New Orleans, LA, 10-14 January (Herber DJ and Richter DA, eds). Memphis, TN: National Cotton Council of America; 1545-1550.
- Ulloa, M., Cantrell, R., Percy, R., Zeiger, E. and Lu, Z. (2000). QTL Analysis of stomatal conductance and relationship to lint yield in an interspecific cotton. *Journal of Cotton Science* **4**:10-18 (<http://www.jcotsci.org>).
- Ulloa, M. and Meredith, W.R. Jr. (2000). Genetic linkage map and QTL analysis of agronomic and fiber quality traits in intraspecific populations. *Journal of Cotton Science*, **4**: 161-170 (<http://www.jcotsci.org>).
- Ulloa, M. and Meredith, W.R., Jr., Shappelley, Z.W.

- and Khaler, A.L. (2002). RFLP genetic linkage maps from four $F_{2,3}$ populations and a joinmap of *Gossypium hirsutum* L. *Theoretical and Applied Genetics*, **104**: 200-208.
- Zhao, X., Ji, Y., Ding, X., Stelly, D.M. and Paterson, A.H. (1998). Macromolecular organization and genetic mapping of a rapidly evolving chromosome-specific tandem repeat family (B77) in cotton (*Gossypium*). *Plant Molecular Biology*, **38**: 1031-1042.