

1 Cotton

Mauricio Ulloa¹, Curt Brubaker², and Peng Chee³

¹ WICS, Research Unit, Cotton Enhancement Program, ASDA-ARS, 17053 N. Shafter Avenue, Shafter, CA 93263, USA
e-mail: mulloa@pw.ars.usda.gov

² Centre for Plant Biodiversity Research, CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia

³ Dep Crop & Soil Science, University of Georgia, Coastal Plain Experiment Station, 2356 Rainwater Road, Tifton, GA 31793-0748, USA

1.1 Introduction

The most important renewable natural textile fiber worldwide and the world's sixth largest source of vegetable oil is cotton. After losing some ground to synthetic fibers in the past, the demand for cotton has been steadily growing. Between 1998 and 2003, there were five consecutive records for cotton consumption with a net gain of 2.6 million tons of cotton finding its way to textile mills (Valderrama 2004). The other side of this coin is that the competitiveness of the supply market has also increased. Production is also increasing, as is yield per hectare, and the price of cotton relative to that of other textile fibers is declining (Valderrama 2004; Anonymous 2005). The cotton growers who survive through the next decade will be those who can produce the most cotton per unit input and produce cotton whose characteristics attract price premiums. This is the message being delivered to cotton breeders around the world. While perennial concerns regarding pest and disease resistance and yield remain, water-use efficiency/drought tolerance, fiber quality and uniformity are high priorities for the future of cotton improvement. The challenge for breeders is that the high-priority breeding objectives for the next decade involve phenotypically complex traits controlled by many interacting genes. Making real breeding gains will require an unprecedented understanding of the molecular genetics of these complex traits. The purpose of this chapter is to provide the reader with an understanding of recent research advances in cotton molecular DNA technologies and genetic linkage mapping. These are the primary tools breeders will use to reach their objectives in the next decade.

While much attention has been focused on genetic linkage maps and quantitative trait locus (QTL) analyses in the recent literature, the fundamental analytical process, genetic linkage analysis, dates back to Mendel. The sophisticated genetic linkage maps of the past decade are the result of our ability to look at DNA sequence differences directly with molecular markers and to use powerful computers that analyze genetic linkage at scales that were impossible to do by hand. Unable to view genes directly, a geneticist could, for a small number of genes controlling a trait with a discrete and variable phenotype, infer the number of genes involved and their mode of action and interaction with other genes. However, the application of this information to cotton breeding was limited because the number of agronomically useful traits that are phenotypically discrete and variable is low, too low for genomewide sampling of genetic linkages. The critical agronomic traits (e.g., yield, quality, disease resistance) were controlled by complex interactions of many genes distributed across the genome that produced a phenotype that varied continuously between two extremes.

The event that allowed the modern cotton breeder to dissect complex agronomic traits genetically was the discovery that DNA-modifying enzymes could be isolated from living organisms (e.g., restriction enzymes, DNA and RNA polymerases, kinases, ligases) and used to manipulate and characterize the cotton genome. The new generation of cotton breeders, the molecular breeders, could use these molecular tools to locate and characterize the sequences of DNA-encoding proteins and the DNA – protein binding regions that controlled their expression, developmentally and temporally. Even when the inferred gene is yet to be located, sequenced, and characterized, molecular breeders could use these natural DNA modifying enzymes to identify sequence differ-

ences among breeding lines and then apply traditional genetic linkage analyses to identify sequence differences (e.g., randomly amplified polymorphic DNAs, RAPDs, restriction fragment length polymorphisms, RFLPs, amplified fragment length polymorphisms, AFLPs, simple sequence repeats, SSRs, sequence-characterized amplified regions, SCARs, single nucleotide polymorphisms, SNPs) that could be used as surrogates for the inferred genes (marker-assisted selection, MAS). Molecular markers are being continuously developed, which allow cotton geneticists to sample all regions of the cotton genome. When this genome saturation is combined with quantitative genetic analyses, cotton breeders finally will have the tools they need to identify the location of the genes (QTLs) conditioning the expression of complex agronomic traits, such as yield (Ulloa and Meredith 2000; Paterson et al. 2003), water tolerance, and water-use efficiency (Ulloa et al. 2000; Saranga et al. 2001).

The fundamental goal of genetic linkage mapping is, and always has been, gene discovery, or when the gene itself is yet to be located, the identification of a linked surrogate that can be used in its stead. Identifying the genes conditioning a trait, their mode of expression, and how they interact gives the cotton breeder the greatest leverage in manipulating the trait to the desired effect. Traditionally this was a process of combining the most effective alleles, from what was available in the germplasm into a single genotype. The advent of transgenic methodologies and the discovery of which genes are interacting to condition a critical trait allows cotton breeders to look beyond traditional germplasm for useful genes, but the efficacy of transgenic technology is entirely dependent on gene discovery. Over the past decade new gene discovery tools have emerged that do not depend on genetic linkage analyses (e.g., bacterial artificial chromosome, BAC, physical mapping, microarrays, and differential display). While these new tools provide exciting new avenues for gene discovery that circumvent some of the logistical limitations of genetic linkage mapping, genetic linkage mapping will continue to be a critical tool for gene discovery in cotton breeding. So this is a particularly appropriate time to review how genetic linkage mapping can be effectively integrated into future gene discovery methodologies.

To understand the complexities of genetic linkage mapping in cotton, we must first understand the evolution and structure of the cotton genome. Understanding the cotton genome is complex because cotton, the lint used in textiles, is derived from the

seed trichomes (hairs) of not one, but four *Gossypium* species: two Old World diploid species, *G. arboreum* L. and *G. herbaceum* L., and two New World allotetraploid species, *G. barbadense* L. and *G. hirsutum* L. Seed trichomes are common in the plant world, and while seed trichomes of the lengths we see in the four *Gossypium* cotton species are certainly uncommon, it is not the length of fiber that makes cotton unique. The singular feature of cotton fibers that sets them apart from other plant seed trichomes and other non-cotton *Gossypium* species is the fact that the cellulose microfibrils are laid down in a spiral arrangement. When the boll (a capsule in botanical terms) opens and the seed trichomes desiccate, they collapse into convoluted ribbons; this is what makes cotton fibers “spinnable.” While this chapter will focus primarily on the most widely cultivated of the four species, *G. hirsutum*, the other three cannot be ignored in any discussion of cotton genome mapping. *G. barbadense* contributes a significant proportion of the extra-long staple lint required for a significant niche of the cotton market and, importantly, is typically used as a parent to generate genetic populations and/or segregating families for genetic linkage mapping. *G. arboreum* and *G. herbaceum*, which are still cultivated in Africa and Asia, are important because they are the only extant representatives of the lineage in which the convoluted seed trichome first evolved (Saunders 1961), and are therefore critical experimental models for gene discovery. Also important are the New World D-genome diploid species, which contributed the second genome to the tetraploid cottons and are an important source of genes for lint quality.

1.1.1

Origin and Domestication of Cotton

The cotton genus, *Gossypium* (Malvaceae) comprises 45 diploid and five allotetraploid species (Fryxell 1979, 1992; Percival et al. 1999; Ulloa et al. 2006). *Gossypium* species are found in tropical and subtropical regions of Australia, Africa, Arabia, Asia, and the Americas. Cotton breeders typically group the *Gossypium* species into one of eight diploid genomes (A – G and K) and one allotetraploid genome (AD) (Beasley 1940, 1942; Phillips and Strickland 1966; Edwards and Mirza 1979; reviewed by Endrizzi et al. 1985; Stewart 1995; Percival et al. 1999). These genomic designations are functional groups based on the similarities in chromosome size and structure and the success of inter-

specific crosses. In general, within genomes hybrids are fertile and their chromosomes recombine during meiosis; however, intergenomic hybrids are generally infertile and there are few stable bivalents at meiosis (Stewart 1995). *G. barbadense* and *G. hirsutum* are AD-genome tetraploids; *G. arboreum* and *G. herbaceum* are A-genome diploids.

The diploid *Gossypium* species have 13 pairs of chromosomes, while the tetraploid species have 26 chromosome pairs. Haploid nuclear DNA amounts (1C) range from 980 to 3,425 Mbp (Edwards et al. 1974; Kadir 1976; Bennett et al. 1982; Endrizzi et al. 1985; Michaelson et al. 1991; Stewart 1995; Grover et al. 2004). The length of the A-genome nuclear DNA is approximately 1,860 Mbp. The D genome is smaller with a haploid length of 980 Mbp. Genome size in the AD tetraploids is roughly additive: 2,835 Mbp.

1.1.2

Evolution of Cotton

Evolution of the Old World Cottons

The Old World A-genome diploid cottons belong to an early *Gossypium* lineage that comprises the 14 African and Arabian species (genomes A, B, E, F). Within this group the evidence suggests that each of the genomes represents a monophyletic group of species (Wendel and Albert 1992; Seelanan et al. 1997). Because wild populations of *G. arboreum* have never been identified, it has been suggested that that *G. arboreum* is simply a recent segregate of domesticated forms of *G. herbaceum* (Hutchinson et al. 1947; Hutchinson 1954, 1959). As discussed later, cytogenetic, genetic, and phylogenetic studies, however, clearly indicate that the two A-genome species diverged prior to the independent domestication of each species (Wendel et al. 1989; Liu et al. 2001).

Evolution of the New World Cottons

The New World tetraploid cottons are allotetraploids that combine one genome derived from an A-genome ancestor and a second genome from a D-genome ancestor (Endrizzi et al. 1985; Galau and Wilkins 1989; Wendel 1989; Brubaker et al. 1999; Liu et al. 2001; Senchina et al. 2003; Wendel and Cronn 2003), two lineages that diverged about 6.0–7.3 million years ago (Senchina et al. 2003). How this occurred remains a mystery. All the known AD tetraploid species are indigenous to the New World, but there is no evi-

dence of any A-genome species in the New World, and there are no D-genome species outside the New World. Nonetheless, 1.3–1.7 million years ago an A-genome plant was pollinated by a D-genome plant and produced an AD diploid hybrid, which, in a single polyploidization event, gave rise to an AD allotetraploid with an A-genome cytoplasm (Wendel 1989; Cronn et al. 1996; Small et al. 1998; Senchina et al. 2003; Wendel and Cronn 2003). After the original AD tetraploid entity arose, it subsequently diverged into three distinct lineages (Seelanan et al. 1997; Small et al. 1998; Cronn et al. 2002). One lineage is represented by *G. mustelinum*, which is found only in a small region of northeast Brazil. The two remaining lineages contain a domesticated species and wild island endemic species: *G. barbadense*/*G. darwinii* (Galapagos Islands) and *G. hirsutum*/*G. tomentosum* (Hawaii).

Of the extant A- and D-genome species, *G. arboreum* and *G. herbaceum* are the best diploid models of the tetraploid At subgenome and *G. raimondii* is the best model of the Dt subgenome (Wendel and Cronn 2003). The gross structural arrangement of the *G. arboreum* genome differs from that of the At subgenome by only two translocations, while the *G. herbaceum* genome differs from the At subgenome by three translocations (Brown and Menzel 1950; Gerstel 1953; Menzel and Brown 1954; Endrizzi et al. 1985; Brubaker et al. 1999), but the phylogenetic data, however, clearly indicate that these two species diverged after the polyploidization event and therefore both of the extant A-genome species are phylogenetically equidistant from the At subgenome and equally appropriate living models of the At subgenome. There has been considerable speculation over the years as to which extant D-genome species is the best model of the Dt subgenome (Endrizzi et al. 1985; Wendel 1989; Wendel et al. 1995), but the most recent and convincing data clearly indicate that *G. raimondii* is the closest living relative of the ancestor of the Dt subgenome and therefore its best living model (reviewed by Wendel and Cronn 2003).

1.1.3

Morphology and Taxonomy of the Cotton Species

The Old World Diploid Cotton Species

The two A-genome cotton species, *G. arboreum* and *G. herbaceum*, are shrubs or subshrubs less than 2-m tall with palmately lobed leaves and yellow flowers (also in *G. arboreum* white to red to purple) that are



<http://www.springer.com/978-3-540-34537-4>

Technical Crops

Kole, C. (Ed.)

2007, XIX, 204 p., Hardcover

ISBN: 978-3-540-34537-4