

M.J. Iqbal · O.U.K. Reddy · K.M. El-Zik
A.E. Pepper

A genetic bottleneck in the ‘evolution under domestication’ of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting

Received: 23 May 2000 / Accepted: 23 January 2001

Abstract Reliable information about the evolutionary and genetic relationships of various germplasm resources is essential to the establishment of rational strategies for crop improvement. We used AFLPs to study the genetic relationships of 43 cultivars of *Gossypium hirsutum* representative of the genomic composition of modern ‘Upland’ cotton. The study also included representatives of the related tetraploid species *Gossypium barbadense*, as well as the diploid species *Gossypium raimondii*, *Gossypium incanum*, *Gossypium herbaceum* and *Gossypium arboreum*. We tested 20 primer combinations that resulted in a total of 3,178 fragments. At the species level, and above, genetic similarities based on AFLPs were in agreement with the known taxonomic relationships. Similarity indices ranged from 0.25 to 0.99. Representatives of the *G. hirsutum* germplasm resources utilized in North America, including secondary accessions collected by breeders in Central America (‘Acala’, ‘Tuxtla’, ‘Kekchi’) and the southwestern US (‘Hopi Moencopi’), formed a single cluster with exceedingly limited genetic diversity (with many pairwise similarity indices >0.96). We concluded that these accessions were derived from the same genetic pool. The early maturing or ‘latifolium’ or ‘Mexican Highlands’ cultivars from which these cultivars were derived appear to have had extremely limited genetic diversity, perhaps as a result of a severe ge-

netic bottleneck resulting from the selection pressures of domestication. Outside of the major *G. hirsutum* cluster, well-supported phylogenies were inferred. Inside this cluster, phylogenies were obscured by limited diversity, reticulation and lineage sorting. The implications of these findings for cotton improvement are discussed.

Keywords AFLP · Genetic diversity · Germplasm · Photoperiodism · Selection

Introduction

The ‘evolution under domestication’ (Stephens 1967) of cultivated cotton (*Gossypium* spp.) presents a fascinating story with elements of biogeography, developmental biology, anthropology, and both natural and anthropogenic selection pressures. Only four species of *Gossypium* are presently cultivated, two New World tetraploid species *Gossypium hirsutum* L. and *Gossypium barbadense* L., and two Old World diploid species *Gossypium arboreum* L. and *Gossypium herbaceum* L. The New World tetraploid species arose some 1–2 million years ago through the hybridization of an Old World taxon of the ‘A genome’ cytogenetic group, related to the present-day species *G. herbaceum* and *G. arboreum* L. ($2n = 2x = 26$), with a taxon of the ‘D-genome’ group, related to the new world species *Gossypium raimondii* Ulbrich and *Gossypium gossypoides* L. ($2n = 2x = 26$) (Beasley 1940, 1942; Wendel et al. 1992). After polyploidization, the nascent ‘AD’ disomic tetraploid ($2n = 2x = 52$) gave rise to five extant tetraploid species, including *G. barbadense* and *G. hirsutum* which dominate world cotton production for fiber and seed products.

G. hirsutum was probably first domesticated by pre-Columbian peoples of the Yucatan peninsula (Brubakar and Wendel 1994). The only truly wild *G. hirsutum* variety is ‘yucatanense’, a sprawling perennial shrub of littoral plant communities. Reproductive development of this species is under strict photoperiodic control, with flowering under short days. The primitive domesticated

Communicated by R. Hagemann

M.J. Iqbal and O.U.K. Reddy contributed equally to this paper

MJ Iqbal · KM El-Zik
Department of Soil and Crop Sciences, Texas A & M University,
College Station, TX 77843–2474, USA

O.U.K. Reddy · A.E. Pepper (✉)
Department of Biology, Texas A & M University, College Station,
TX 77843–3258, USA
e-mail: apepper@bio.tamu.edu
Fax: 409–845–2518

Present address:

M.J. Iqbal, Department of Plant, Soil and General Agriculture,
Southern Illinois University, Carbondale, IL 62901, USA

variety 'punctatum' probably arose from within yucatanense. Though still a perennial shrub, many of the early punctatum cultivars had tufted 'green' seeds that were superior for primitive hand-ginning practices. These early domesticated forms of *G. hirsutum* dispersed out of the Yucatan to the rest of Mesoamerica, as well as to northern South America and throughout the Caribbean basin. Ethnobotanical evidence suggests that the important landrace 'latifolium' arose from this germplasm as it was further domesticated in present day southern Mexico and Guatemala (Brubakar et al. 1999). Some accessions classified as latifoliums show photoperiodic flowering, while others are photoperiod independent (J. McCarty, personal communication). According to Cook (1905), the Kekchi indians of Guatemala traditionally intercropped their cotton with peppers (*Capsicum* spp.), and would remove cotton plants as soon as first bolls began to open in order to prevent competition with the developing peppers. This practice would have eliminated late-maturing genotypes. Such a rigorous selection regime for early maturity would have selected for reduced seed dormancy, compact 'annualized' growth habit, and possibly photoperiod-independent flowering, thus creating genotypes suitable for cultivation at higher latitudes with longer summer daylengths. The early maturing latifolium genotypes diffused into the highlands of present-day southern central Mexico prior to the arrival of Europeans, and the resulting cultivars became collectively known as 'Mexican Highlands' (or 'Mexican') varieties.

In colonial North America, cotton production initially consisted of a complex mixture of taxa, including *G. herbaceum* imported to the New World by the British, *G. barbadense* ('Sea Island Cotton') from the Caribbean, and several varieties of *G. hirsutum*, including tufted 'green-seed' types from Mexico. The invention of the cotton gin in 1793 increased the demand for high yield cultivars, eventually leading to the introduction of several non-tufted Mexican Highlands varieties including Burling's Cotton in 1806 and Wyche cotton in 1857 (Smith et al. 1999). *G. hirsutum* of the 'Mexican type' may have been grown in the Stephen F. Austin colony in Texas as early as 1821, and numerous introductions were probably made by soldiers returning from the Mexican-American War (1846–1848). These cultivars were subjected to strong selection to create varieties adapted to local conditions in various cotton-growing regions of North America. Throughout this period, both intentional and unintentional outcrosses occurred between cultivars (Endrizzi et al. 1985). Collectively known as 'American Upland Cotton' the resulting high yielding and adaptable cultivars were dispersed to Europe, Africa and Asia.

In response to the dire boll weevil (*Aethionomus grandis* Boh.) infestation that began at the end of the 19th century, breeders collected new germplasm from the southern Mexican state of Chiapas ('Acala' and 'Tuxtla', 1906), Guatemala ('Kekchi', 1904), and from the Hopi peoples of Arizona ('Hopi Moencopi'). Although not resistant to the weevil, these cultigens were incorporated with germplasm from earlier Mexican Highlands intro-

ductions to provide the genetic foundation for modern Upland *G. hirsutum* cultivation worldwide.

An understanding of the evolutionary and genomic relationships of cotton species and cultivars is critical for further utilization of extant genetic diversity in the development of superior cultivars that will combine the favorable qualities (insect and pathogen resistance, fiber quality and yield traits, tolerance to environmental stresses) conditioned by diverse *G. hirsutum* germplasm resources (El-Zik and Thaxton 1989). Given the status of *G. hirsutum* as the world's most important fiber crop, the genetic composition of various cultivars has not been adequately investigated. Studies using allozymes (Wendel et al. 1992) and RFLPs (Wendel and Brubakar 1993) have been limited by low levels of polymorphism at the intraspecific and interspecific levels. Recently, there have been several reports describing the genetic relationships of several cotton varieties based on PCR-based molecular techniques such as RAPDs (Multani and Lyon 1995; Tatineni et al. 1996; Iqbal et al. 1997) and AFLP (Abdallah et al. 2001). The amplified fragment length polymorphism (AFLP) technique (Zabeau and Vos 1993; Vos et al. 1995) generates a relatively large number of polymorphic amplified fragments compared to the RAPD method. Since each AFLP fragment is considered to be a genetically distinct locus, a large number of loci can be compared between different genotypes in relatively few experiments. We have examined 43 genotypes that are representative of most of the Upland *G. hirsutum* germplasm resources utilized in North America and Asia. Our set of taxa included the Acala, Tuxtla, Kekchi and Hopi Moencopi collections as well as several obsolete cultivars, such as 'Western Stormproof' and 'Wannamaker's Cleveland' which are purported to be direct selections from 19th century germplasm introductions from Mexico (Smith et al. 1999).

Materials and methods

Plant materials and DNA isolation

Three diploid species of *Gossypium*, three *G. barbadense* cultivars, and 43 *G. hirsutum* accessions were used for the study (Table 1). Genomic DNAs were isolated from leaf tissue by the method of Iqbal et al. (1997). DNAs of cultivars Allepo, Nazelli, Cedix, LRA5166, CP 15/2, S-12, CIM-435, CIM-443 and CIM-1100 were provided by Dr. Yusuf Zafar, NIBGE, Faisalabad, Pakistan. After RNase treatment, the concentration of each sample was measured by a fluorimetric assay using Hoechst dye.

AFLP fingerprinting

AFLP fingerprinting (Zabeau and Vos 1993; Vos et al. 1995) was performed using Analysis System I (GIBCO-BRL Life Technologies) according to the manufacturer-supplied protocol with some modifications. Genomic DNA (225 ng) was digested with 0.625 units of *EcoRI* and *MseI* for 2 h at 37°C. *EcoRI/MseI* adapters were ligated to genomic DNA fragments using 0.25 Weiss units of T4 DNA ligase for 2 h at room temperature. Pre-amplification of adaptor-ligated DNA was carried out in a reaction containing 10 µl of pre-amplification primer mix, 1.25 µl of 10 × PCR buffer for