
The U.S. National Cotton Germplasm Collection – Its Contents, Preservation, Characterization, and Evaluation

Richard G. Percy, James E. Frelichowski,
Mark D. Arnold, Todd B. Campbell, Jane K. Dever,
David D. Fang, Lori L. Hinze, Dorrie Main,
Jodi Scheffler, Monica A. Sheehan, Mauricio Ulloa,
Jing Yu and John Yu

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/58386>

1. Introduction

The early history of the cotton industry in the U.S. was, for the most part, a story of importation and adaptation of cottons from Mexico, Guatemala, and the tropics of the Western Hemisphere. By the 19th century there were two distinctive types of upland cotton, with distinctive origins, being grown in the U.S. One type, known as green-seeded cotton, was from southern Mexico. The other dominant type of upland, white-seeded cotton, had its origins in the central plateau of Mexico [1,2]. Although these cottons were grown extensively, no coordinated efforts were made to maintain the original stocks or their progeny. Modern, systematic collection and preservation of cotton in the U.S. only began in response to the outbreak of the boll weevil in the 1880's [3]. Three collection trips between 1902 and 1906 were responsible for the introduction of two cottons, Acala and Kekchi, which would contribute significantly to the development of modern U.S. cultivars [4]. From these initial collecting trips and ensuing trips, federal, university and state experiment stations began to assemble the germplasm collections. Also from these efforts there sprang collaborative efforts that were formalized into U.S. regional projects involving scientists and members from all aspects of the U.S. cotton industry. The first Regional Research Project was S-1 in 1950 (succeeded by many) and it established a priority of acquiring and studying diverse germplasm for cotton improvement [5-7]. Increased organization and centralization of germplasm activities was necessary to manage and increase

cotton germplasm resources. Plant Introduction centers were established with the Agricultural Marketing Act of 1946 and a National Seed Storage Laboratory (now the National Center for Genetic Resources and Preservation or NCGRP) was established in 1958 for long term backup storage of germplasm. Federal funds and staff also were committed to create working collections throughout the U.S. to specialize in crops. In 1960, the cotton sub-collections were deposited at NCGRP with their passport data, and in the early 1980's they were consolidated in College Station, TX, USA. Increased staff, greenhouses, seed processing and storage facilities were added. The curators continued to add germplasm from publicly donated cultivars and germplasm lines, explorations and exchanges with other collections and to record descriptors on new and existing germplasm. The International Board for Plant Genetic Resources (now Bioversity International) established a set of cotton descriptors in 1980 to serve as a guide for data collection for the U.S. and other collections [8]. Databases for the National Plant Germplasm System (NPGS) and the collection are managed online by the Germplasm Resources and Information Network (GRIN).

2. Content, preservation, and distribution

The U.S. National Cotton Germplasm Collection (NCGC) currently has approximately 10,000 accessions covering 45 species of *Gossypium* as described by Fryxell [28]. Representation of species within the collection varies significantly, and to some extent reflects historical opportunities for collecting or exchange with other collections (Table 1.) However, there has been a conscious effort to focus on fiber producing species that readily hybridize with the cultivated cottons. Accessions of the collection are officially recognized by Plant Introduction (PI) numbers in the NPGS, which maintains a permanent long-term backup inventory, or base collection, in the NCGRP in Fort Collins. However the working, or active, collection in College Station, TX further subdivides the collection to reflect historical organization and taxonomic classification. This additional step facilitates easy organization and retrieval of seed from the temperature controlled vault. The working collection is under the care of a full-time curator and supporting staff with the ARS-SPARC in College Station, Texas and it is the source of *Gossypium* germplasm distributed to the public.

The collection is subdivided to seven different sub-collections that consist of: 1) variety sub-collection, primarily *G. hirsutum* (prefixed by SA-), 2) landrace sub-collection, primarily *G. hirsutum* (prefixed by TX-), 3) *G. barbadense* sub-collection, (prefixed by GB-), 4) Asiatic (A-genome species) sub-collection, (prefixed by either A₁-or A₂-for *G. herbaceum* or *G. arboreum* respectively), 5) wild species sub-collection (prefixed by a taxon specific genome letter-number combination [9]), 6) genetic marker sub-collection, and 7) a base sub-collection (i.e. NCGRP) of all materials in sub-collections 1-6 and new plant introductions [3,10]. Sub-collections 1-5 constitute the "working collection", which is routinely seed propagated and distributed by the USDA-ARS at College Station, Texas. Seed of the working collection is maintained at 4°C and 23% relative humidity [11]. Sub-collection 7 is housed at the NCGRP in Fort Collins, CO under conditions more favorable for long-term storage (-20°C, ~5% seed moisture content [12]). The SA sub-collection is short for Stoneville Accession and represents the obsolete cultivar collection originally started at the Delta Branch Experiment Station in Stoneville, MS. The TX

(or TEX) sub-collection was started on the campus of Texas A&M University, College Station, TX, to catalogue *G. hirsutum* landraces and various genetic stocks and is the format continued to the present day, exclusively for *G. hirsutum* photoperiodic landraces or tropically collected material. The GB sub-collection originated from *G. barbadense* accessions curated at Phoenix, AZ, and after consolidation with the main collection, it has continued to expand as new acquisitions have been made.

Gossypium genome/species	ARS					CGRU ²	SC ³	CSRL ⁴	Texas A&M Agrilife ⁵			
	Working Collection ¹								Accessions	Landrace Evaluation	Salt/Drought Evaluation	Characterization, Digital Imaging
	Accessions	Distribution 1987-2013	Seed Increase 2011-2013	Characterization, Digital Imaging 2011-2013	SSR, Seed Oil/Protein Evaluation	CLCV Evaluation	Landrace Evaluation	Salt/Drought Evaluation	Characterization, Digital Imaging	Abiotic Evaluation	Biotic Evaluation	Phenotypic Evaluation
(AD) <i>hirsutum</i> total	6302	42659	2878	3731	1541	631	1380	440+	700+	290	473	400+
(<i>hirs.</i> landraces)	(2522)	-	(1220)	(1174)	-	(510)	(1380)	(440+)	-	-	-	-
<i>barbadense</i>	1584	5502	340	328	430	23	42	20	5	-	25	-
<i>tomentosum</i>	16	98	3	5	2	-	-	-	-	-	3	-
<i>mustelinum</i>	19	125	4	4	7	-	-	-	6	-	6	-
<i>darwinii</i>	138	80	8	4	4	-	-	-	9	-	-	-
(A) <i>herbaceum</i>	194	1487	129	129	49	194	-	20	26	-	2	-
<i>arboresum</i>	1729	4287	1729	1729	145	1729	-	20	24	10	4	-
(B) <i>anomalum</i>	7	95	3	3	5	-	-	-	2	-	-	-
<i>capitis viridis</i>	1	36	1	1	-	-	-	-	-	-	-	-
<i>triphylllum</i>	2	14	0	1	-	-	-	-	1	-	-	-
(C) <i>nandewarense</i>	6	11	1	1	1	-	-	-	-	-	-	-
<i>robinsonii</i>	3	31	0	1	-	-	-	-	1	-	-	-
<i>sturtianum</i>	7	103	1	2	3	-	-	-	-	-	-	-
(D) <i>aridum</i>	14	135	4	2	4	-	-	-	-	-	-	-
<i>armourianum</i>	10	115	1	2	2	-	-	-	-	-	-	-
<i>davidsonii</i>	31	140	4	2	9	-	-	-	2	-	1	-
<i>gossypoides</i>	8	135	4	4	7	-	-	-	-	-	-	-
<i>harknessii</i>	19	72	0	1	-	-	-	-	1	-	1	-
<i>klotzschianum</i>	59	52	1	1	1	-	-	-	2	-	-	-
<i>laxum</i>	2	3	1	1	1	-	-	-	-	-	-	-
<i>lobatum</i>	4	61	1	1	1	-	-	-	-	-	-	-

Gossypium genome/species	ARS					CGRU ²	SC ³	CSRL ⁴	Texas A&M Agrilife ⁵			
	Working Collection ¹								CLCV Evaluation	Landrace Evaluation	Salt/Drought Evaluation	Characterization, Digital Imaging
Accessions	Distribution 1987-2013	Seed Increase 2011-2013	Characterization, Digital Imaging 2011-2013	SSR, Seed Oil/Protein Evaluation								
<i>raimondii</i>	56	315	3	2	4	-	-	-	3	-	-	-
<i>schwendimanii</i>	2	1	1	-	-	-	-	-	-	-	-	-
<i>thurberi</i>	41	268	20	11	9	-	-	-	5	-	1	-
<i>trilobum</i>	11	118	4	3	6	-	-	-	-	-	-	-
(E) <i>areysianum</i>	2	13	1	1	1	-	-	-	1	-	-	-
<i>incanum</i>	4	22	1	1	-	-	-	-	-	-	-	-
<i>somalense</i>	3	65	1	1	1	-	-	-	1	-	-	-
<i>stocksii</i>	4	65	2	2	1	-	-	-	3	-	-	-
(F) <i>longicalyx</i>	4	213	4	4	4	-	-	-	3	-	-	-
(G) <i>australe</i>	11	103	2	2	3	-	-	-	3	-	-	-
<i>bickii</i>	5	82	1	1	4	-	-	-	3	-	-	-
<i>nelsonii</i>	2	11	2	1	1	-	-	-	-	-	-	-
(K) <i>costulatum</i>												
<i>exiguum</i>	1	-	1	-	1	-	-	-	-	-	-	-
<i>marchantii</i>	2	-	2	2	1	-	-	-	-	-	-	-
<i>nobile</i>	2	-	2	2	1	-	-	-	-	-	-	-
<i>populifolium</i>	4	2	1	1	1	-	-	-	-	-	-	-
<i>pulchellum</i>	1	3	1	1	1	-	-	-	-	-	-	-
<i>rotundifolium</i>	1	-	1	-	-	-	-	-	-	-	-	-

¹ Southern Plains Agricultural Research Center, USDA-ARS, College Station, TX

² Crop Genetics Research Unit, USDA-ARS, Stoneville, MS

³ Coastal Plain Soil, Water and Plant Research Center, USDA-ARS, Florence, SC

⁴ Cropping Systems Research Laboratory, USDA-ARS, Lubbock, TX

⁵ Texas A&M Agrilife, Lubbock, TX, data compiled from 2005-2013

Table 1. *Gossypium* germplasm accession holdings, distribution, seed increase, characterization, digital imaging, and evaluation activities of the U.S. National Cotton Germplasm Collection (NCGC) at College Station, USDA-ARS collaborators, and Texas A&M.

To meet the goal of renewing seed of the collection every ten years or less and to ensure the availability of viable seed for distribution, storage, and backup at the NCGRP, accessions of the collection are grown in the Cotton Winter Nursery (CWN) in Tecoman, Mexico, in a local nursery in College Station, or in greenhouses. The majority of the working collection, including photoperiodic and many perennial accessions, are grown, force self-pollinated, and seed increased at the tropical CWN nursery. A large portion of the TX and GB sub-collections are of tropical origin, uncultivated, and are short-day flowering types that require a tropical nursery or a winter greenhouse to induce flowering and reproduction. Wild diploid species (sub-collection 5 of the collection) and recalcitrant accessions with critically low seed numbers must be seed increased in College Station. These high maintenance accessions typically exhibit characteristics such as a perennial growth habit, exerted stigmas, or prerequisite environmental stimuli to induce flowering, to name a few. These accessions are typically grown year round in greenhouses to target seed yields of at least 3000 seeds. Accessions are replenished on a 10 year cycle, as determined by viability studies of cottonseed under storage conditions of the working collection, or as indicated from germination testing on the base collection by the NCGRP.

	2007	2008	2009	2010	2011	2012
U.S.	768/59	1,537/90	2,300/101	1,702/78	3,200/83	4,900/128
Int'l.	919/26	546/18	450/12	65/10	1,254/10	1,100/12
Total	1,687/85	2,083/108	2,750/113	1,767/88	4,454/93	6,000/140

Table 2. Number of accessions/orders distributed from the National Cotton Germplasm Collection (NCGC) in College Station, TX, to requestors from the U.S. and internationally (int'l) in the past six years.

Germplasm users in all countries have equal access to the U.S NCGC, and distributions are limited only by the availability of the germplasm at the time of the request and by proper adherence to customs and phytosanitary laws of the U.S. and the requesting country. The policy of free access has grown out of the belief that germplasm is a world heritage to be freely shared. Seed distribution is provided from the working collection at College Station and shows an upward trend in number of requests and amount of accessions distributed yearly (Table 2). Information regarding materials available can be found online at GRIN-GLOBAL (<http://distribution.grin-global.org/gringlobal/search.aspx>).

Seed samples are supplied in amounts of 25 self-pollinated seed per sample, if sufficient seed is available for distribution (accessions with a minimum of 500 seed). Cooperators requiring more than 25 seeds must arrange for their own seed increase and in some cases are asked to return excess seed generated (if they are acquiring rare or difficult to manage accessions and can increase seed under controlled conditions with careful self-pollination in greenhouses). Some seed requests go unfilled due to critically low seed amounts, repeated germination difficulties with the seed, or because users request materials that are unavailable (e.g. germplasm under patent protection or accessions not present in the collection). Some accessions prove difficult to germinate from original or even increased seed, and require treatments such

as seed scarification or hot water baths to achieve germination. Regular re-inventory of the critical and/or difficult to manage accessions indicate if they need to be re-established from backup seed at the NCGRP, or through a repeated exploration or germplasm exchange with the host country.

3. Collecting and exchange

Although domestic upland (*G. hirsutum*) cotton dominates world production (around 95%), it is generally accepted that its genetic diversity is very low [13-16]. Various bottlenecks and restrictions to genetic variability have occurred during the breeding history of upland cotton that have led to the low diversity currently seen in commercial cotton [17,18]. However, our germplasm collections have historically served as reservoirs of diversity for the breeding community; providing useful variation for disease and insect resistance, fiber quality improvement, and resistance to environmental stress. Collecting and exchange efforts are necessary to maintain and increase the genetic diversity of our collections [18,19]. In addition to early collecting trips that provided the foundation for the germplasm collection, numerous sponsored plant exploration trips and germplasm exchange efforts have occurred since the mid-twentieth century [10,20,21] and continue to occur. In 1982 the sovereign rights of countries over their germplasm was recognized [22,23] and many countries have become signatories to the International Treaty on Plant Genetic Resources for Food and Agriculture, implemented on 29 June 2004. This treaty states as a goal “the conservation and sustainable use of plant genetic resources for food and agriculture and the fair and equitable sharing of the benefits arising out of their use, in harmony with the Convention on Biological Diversity, for sustainable agriculture and food security”. The acquisition of germplasm through collection has become the subject of extensive negotiation with host countries, and often includes the topics of benefit sharing and agreements on proprietary rights of host countries to products originating from germplasm. The latter topic has become a major stumbling block to collecting for the U.S. NCGC. In the past, the collection has only acquired germplasm that could be distributed free and clear of proprietary obligations, and the collection remains committed to this policy. Careful bilateral negotiation with individual nations, assuring mutual benefit, has led to limited germplasm acquisitions by the U.S. national collection in the latter half of the twentieth century. Since 1985 twelve acquisition trips have been made by U.S. scientists to collect germplasm. Ten of the trips were used for *in situ* explorations, while two of the trips were conducted to exchange germplasm with India, China, Russia, and Uzbekistan respectively [24]. Continued collection and germplasm exchange is essential to address gaps in individual collections worldwide and to conserve *ex situ* germplasm threatened by development [25].

The two most recent collecting efforts that resulted in germplasm acquisition by the U.S. collection were to Mexico in 2004-2006 and to Puerto Rico in 2013. Mexico encompasses the species range of several diploid cotton species, and Mexico-Guatemala is the recognized center of origin of *G. hirsutum*, the most widely cultivated species in the world. Mesoamerica is the center of morphological diversity for *G. hirsutum* [26]. Collecting efforts in Mexico during the

last century are not well documented, and published information is limited. A series of collection/exploration trips were made in southern Mexico and Central America in the mid-twentieth century by T.R. Richmond and C. W. Manning in 1946, S. G. Stephens in 1947, and J. O. Ware and C. W. Manning in 1948 [21]. From the germplasm collected in these trips, a series of eight 'landraces' was established that generally corresponded to areas of native production or geographic occurrence. N. Lemeshev (USSR) and Q. Obispo (Mexico) in 1977-1978 made several collection-expeditions covering 16 Mexican states, from the peninsula of Baja California to the state of Yucatan. They collected several landraces and seed of seven *Gossypium* species (*G. armourianum* Kearney, *G. harknessii* Brandegee, *G. trilobum* (DC.) Skovsted, *G. aridum* (Rose & Standley) Skovsted, *G. lobatum* Gentry, *G. laxum* Phillips, and *G. gossypoides* (Ulbrich) Standley) [26]. P.A. Fryxell from 1968 to 1975 made several collection-expeditions in the country, providing a larger number of specimens to herbaria (MEXU and XAL) with clear and precise descriptions of habitat and location of collected accessions [27]. He made the most recent taxonomic classification of *Gossypium* species [28,29]. A. E. Percival and J. M. Stewart of the USDA-ARS, and A. Hernandez and F. de Leon of the Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP) made several collection-expeditions throughout the states of the Yucatan Peninsula and in parts of the states of Tamulipas, Veracruz, Tabasco, Oaxaca and Chiapas in 1984 [10]. Also, A. E. Percival (USDA-ARS), J. M. Stewart (Univ. of Arkansas), E.A. Garcia, and L. Pérez (SAGAR-INIFAP Mexico) made additional collection expeditions in the state of Baja California Sur and parts of Sonora and Sinaloa states in 1990.

The most recent collecting efforts in Mexico for the collection were made by M. Ulloa (USDA), J.M. Stewart (Univ. of Arkansas), and several researchers from Mexico's INIFAP. These collecting efforts covered the Pacific west coast of Mexico, from the state of Sonora (North) to the state of Chiapas (South) and included several trips from 2002-2004 [19,26]. As a result of these efforts, a significant number of additional *Gossypium* accessions from various parts of Mexico are now available for evaluation, including several accessions of each diploid species of the D genome [19,26]. The cooperative collecting efforts of 2002-2006 also led to a significant number of additional *Gossypium* accessions of the subgenus *Houzingenia* from various parts of Mexico being placed in a nursery or botanical garden in Iguala Guerrero Mexico, including several accessions of each of the arborescent species for *ex situ* conservation [19,26]. This botanical garden was supported from 2006 through 2009 by additional funding from Cotton Incorporated (Cary, NC, USA) provided by J. M. Stewart to INIFAP. Today, Mexico maintains this *Gossypium* nursery in Iguala Guerrero (C. Perez-M. *personal communication*).

Since the first collecting trips were made in Mexico, the *in situ* survival of Mexican cotton germplasm has been threatened with increasing human population, modernization of agriculture and urbanization. New roads and population growth continue to increase. At this point, one species (*G. aridum*) of the subsection *Erioxylum* appears not to be threatened, probably because of the great diversity (botanical and geographic) encompassed by this species [26]. However, some of the most recent collected and non-described taxons (e.g., US-72) or ecotypes of the *G. aridum* species [19,26] may be in the process of becoming extinct in the wild. In addition, the D₈ *G. trilobum* species is almost extinct or already extinct. The natural habitat

of this species has been replaced by intense and extensive agricultural production of guava (*Psidium* spp.). We speculate that the crop of guava and *G. trilobum* must have very similar environmental requirements, particularly as influenced by altitude (around 4,000 feet). Whereas in the past *G. trilobum* was considered to be a common, widely distributed species, based on surveys (J.M. Stewart and M. Ulloa, 2004 expedition), we are of the opinion that urbanization and agricultural development have very severely eroded its habitat, and that the species is becoming extinct in the wild (Drs. Stewart and Ulloa *personal communication*). Within the eight races described in *G. hirsutum*, much of the original diversity existing *in situ* appears to have been lost. According to information obtained from local sources, eradication of naturally occurring landrace, feral, and dooryard cottons was attempted in areas of southern Mexico in the 1980's in efforts to remove perceived insect reservoirs. Apparently all attempts at commercial cotton production since then have been abandoned. No commercial fields of cotton were encountered during expeditions between 2002 and 2004 in the central and southern part of Mexico [19,26]. Currently, with the exception of the northern cotton production regions of Mexico, the diversity of the *G. hirsutum* is limited to feral plants that occur opportunistically in waste areas and as occasional home garden plants maintained by rural peoples or village residents [26].

Due to its relative proximity to the U. S. and its status as a U.S. territory, Puerto Rico (PR) was a target of opportunity for collecting efforts in 2013. Puerto Rico was revisited in the most recent germplasm collecting effort by James Frelichowski, the curator of the U.S. collection, Louis Prom, of the USDA-ARS, College Station, TX and by collaborators from the USDA-ARS Tropical Agriculture Research Station (TARS) in Mayaguez, PR. Previous reports on wild cotton in the Caribbean [30,31] and the recent taxonomic identification of a new species in the Dominican Republic [32] justified continued exploration for *Gossypium* in Puerto Rico. The current collection inventory indicates 49 accessions originally collected from Puerto Rico and surrounding islands, but GPS locations and details on habitats are lacking. Judging from the limited passport data, the trend was for wild cotton to occur in areas with drier climates in or near coastal regions. Only four Puerto Rican accessions in the collection were observed to have maturity and productivity ratings similar to cultivar checks in the CWN. The Puerto Rican accessions were more characteristic of landrace cottons, and suggest that cotton in Puerto Rico is likely native or is cotton of tropical origin that has become naturalized.

In the most recent collecting efforts in Puerto Rico, attempts to address gaps in the knowledge of the island's diversity and diversity structure were addressed through tagging of all collection sites and resulting accessions with GPS coordinates and with geographic mapping of native cottons. Seventy-nine cotton plants or populations were photographed and tagged with GPS coordinates to give a detailed map based survey of PR. Numerous other cotton plants or populations were sighted and recorded on a GPS map, but not collected due to time restrictions or safety or access concerns. Southern Puerto Rico was much more populated with cotton than the north, but morphological diversity was seen throughout the island and justified the wide search of the island for wild cotton. Differences in fiber length, quality, seed fuzz, and leaf pubescence were evident during collection, and sometimes individual traits were observed to vary among neighboring plants. This confirms the need to collect from individual

plants and to document their locations carefully. A local botanist, Duane Kolterman and the curator of the Puerto Rico Herbarium, Jeanine Velez Gavilan, indicated that two species inhabit PR, *G. hirsutum* and *G. barbadense* with probable species introgression in discovered specimens. Only one clear cut *G. barbadense* was seen and it was a dooryard cotton, which we were not able to access. Most of the observed cotton plants were similar to collection accessions classified as race 'marie galante' [31], which is typical of the Caribbean and often produces good fiber. However these plants were not as strongly arborescent as the classic 'marie galante' accession. In addition these plants had characters, such as tapering boll shapes, which were typical of *G. barbadense*, and therefore warrant further inspection. In the southwest portion of Puerto Rico, in the Cabo Roja region, the cotton plants were more variable. In the salt flats at the southwest tip of Puerto Rico, the dominant cotton plant had shorter, coarser and darker fibers than that typically observed on the remainder of the island, and were more characteristic of accessions classified as race 'yucatanense' [31]. The southwest portion of PR faces the Dominican Republic where a sixth tetraploid species is proposed, and a priority site for further exploration and collection. Dooryard cottons were in low frequency on the island and may be declining as cotton products and substitutes become more available for purchase. One resident extolled the virtues of natural cotton because of lack of chemicals used on them, and she demonstrated how they were propagated via cuttings, which may be indicative of a hard seed coat frequent in wild cottons. The digital data collected allows us to assemble a vivid description of wild cotton in Puerto Rico in what appears to be an ever changing landscape and justifies exploration to ensure their *ex situ* preservation.

4. Phenotyping and characterization

Phenotypic characterization of the national collection has historically served dual purposes. Until the very recent past (when molecular tools became available), phenotypic descriptors were the primary means of describing the diversity contained in the germplasm collection and rationally classifying that variability. The second role of phenotypic descriptors was to assist breeders and others in identifying germplasm of interest in genetic improvement efforts. However, to be useful, phenotypic descriptors needed to be well defined and universally applied. Throughout the history of the collection, this has been a goal that has been striven for but not completely attained. The convention in the past has been to follow, as much as possible, the guidelines set by Bioversity International (formerly IBPGR) in 1980 [33], revised in 1985 and again in 1995 [34] in creating descriptors. Prior to the universal adoption of computers, the internet, and electronic databases, descriptor data files of the collection's holdings were periodically published, with the last publication occurring in 1987 [10]. Presently, data files of descriptors for accessions of the collection can be found in GRIN-GLOBAL at <http://distribution.grin-global.org/gringlobal/search.aspx> or in the CottonGen database at www.cottongen.org. Descriptors and descriptor sets for phenotyping the collection have not remained static over time, but have evolved as their use has evolved. In 2009, the Cotton Crop Germplasm Committee [14] summarized the need to expand the collection's databases with morphological, agronomic, molecular, and priority trait evaluations to improve user utility of the collection.

Of particular interest to the breeding and crop improvement community was the addition of traits of agronomic value, to supplement traits of a primarily botanical or taxonomic nature adopted from Bioversity International.

Within the collection, a goal has been set to routinely characterize or re-characterize approximately 1,000 accessions (or a tenth of the collection) annually in the CWN, or in local fields or greenhouses at College Station, TX. Although the photoperiodism of a large portion of the collection requires that it be renewed in the tropically located CWN, collecting descriptors at that site has been challenging, due to travel costs and time limitations. With the great diversity of accessions being grown in the CWN, and their variable maturation rates, collecting complete descriptor sets for accessions has been a lengthy, multi-year process. To aid in the description of the collection, digital imaging has been adopted. An immediate concern has been to develop a pictorial reference of the descriptor scores in order to establish clear standards and reduce subjectivity that occurs when scoring descriptors on such a spectrum of diversity in the collection. Mackay and Alercia [35] have described the use of digital imaging to document variation in germplasm, and this approach has been used in several collections [36,37]. Digital imaging can be tailored with software to detect minute differences among accessions and to determine duplicates [38]. Digital photography to document *Gossypium* diversity has been used by collaborators [38,39] and recently by the NCGC staff during descriptor collection. Currently within the collection, digital imaging has been concentrated upon plant structures that are most informative in differentiating and classifying the diversity of the genus. Samples of typical leaves, flowers, and bolls are gathered per accession, tagged, and quickly assembled under high resolution digital cameras. Only mature, fully expanded leaves (below the fifth main stem node) are photographed. Cotton flowers are photographed on the day of anthesis to reveal flower color, petal spots, pollen color, and stigma length. Fully expanded, mature but non-cleft bolls are photographed to reveal shape and locule number. To obtain photographs, several cameras are fixed with tripods, some for an overall view of size of plant tissues and others for a close up of minute leaf hairs, nectaries, and glands. All images are produced with a measured grid background for a consistent size reference. Portable shade is constructed at the CWN or in the local field to stabilize lighting conditions and reduce wind disturbance. Images are identified by photographed tags, and then named according to accession inventory number, year/environment of planting, and a code for the plant tissue or structure photographed.

To date (2011-2013), standardized descriptors and digital images have been collected on over 4,900 cotton producing accessions planted at the CWN and in the field at College Station, TX. These images are in the process of being uploaded into the CottonGen database (www.cottongen.org) for pairing with descriptors for optimal use by the cotton community. The digital library or “virtual herbarium” created by the use of high resolution cameras to produce ‘virtual’ voucher specimens provides easy access to examine the morphological variation within the genus that was previously unattainable with classic herbarium specimens. It is hoped that the digital image library promotes standardization of descriptor data and image creation by cooperating groups and collections, thereby promoting a greater ability to characterize the diversity within *Gossypium*, address gaps in the U.S. and other collections and effectively share and backup germplasm between collections.

Agronomic Traits	Descriptor Name	Method of Observation	Descriptor State
Architecture	Plant Height (cm)	Plant height in centimeters, grown in greenhouse. Measured from pot soil surface to terminal.	ht. (cm)
	Growth Habit	General architectural arrangement of plant type grown in greenhouse.	Compact → spreading Intermediate → bushy
Physiological	Photoperiod	Response of plants flowering habit over a short day to long day growth duration in greenhouse as compared to day neutral upland varieties grown in Lubbock, Texas.	day neutral → strong photo dependent → weak
Retention	Square	Retention of squares that mature into bolls and are not shed due to insect or environmental conditions.	good → weak fair → none
	Boll	Retention of bolls grown to maturity in greenhouse and are not shed due to insect or environment.	good → weak fair → none
	Fiber	Retention of fiber in burr until harvested.	good → weak fair → none
Morphological	Flower Score	Flower score of early, medium or late as determined by the greenhouse planting date. Lubbock, Texas	Early → medium → late
	Flower Duration	Length of time a plant will produce floral buds, determined by photoperiodic response of plants growing in greenhouse during the fall and winter months. Lubbock, Texas	short → medium → long
	1 st Flower Date	Date of first flowers when grown in greenhouse in number of days after planting. Lubbock, Texas	number of days after planting
	Nodes to 1 st Fruiting	Node of first sympodial branch to set fruit on plant growing in greenhouse. Lubbock, Texas	node number (cotyledon = 0)
	Production	Productive-ness	Quality and quantity of mature fiber and seed produced from plants grown in greenhouse. Lubbock, Texas
Seed Index		The weight of 100 fuzzy seed in grams.	WT. (g)

Table 3. Agronomic trait characteristics documented from *Gossypium* species grown in greenhouse environment for phenotypic evaluation at Texas A&M AgriLife Research, Lubbock, TX, USA

Although the U.S. NCGC tries to set standards and methodologies for characterizing the germplasm collection through its internal efforts, it also recognizes and encourages cooperation from the research community in this task. Due to the volume of accessions available for characterization, the finite resources of the collection, and the impact of genotype x environment interaction on many phenotypic traits of interest, a unilateral effort to collect descriptor data is not considered desirable. Collaborative efforts within the research community offer an attractive way to collect relevant data and make it widely available. One such effort has been conducted by Texas A&M Agrilife, Lubbock, TX from 2005 until 2013. Seed of accessions from the U.S. NCGC at College Station, TX were obtained and planted in the greenhouse at Lubbock for seed increase. In conjunction with seed increases, phenotypic descriptors of various *Gossypium* species were recorded and documented using digital photography. Traits selected for digital imaging were based on previous descriptive data recorded using the GRIN guidelines and published taxonomic documents. [27,10] Additional documentation elements were expanded to include descriptor name, state, and method of evaluation in a greenhouse environment (Table 3). Digital images depicting characteristic and state were used as reference standards to help alleviate misinterpretation of the descriptor. This allowed for more decisive, objective, and consistent data collection. Qualitative data has been documented on a nominal, ordinal, or binary scale and quantitative data consist of continuous or discrete numerical values. Over 800 accessions have been regenerated, digitally documented and evaluated (Table 1). The descriptors used by the Texas A&M AgriLife Research and Extension Center did not encompass all aspects of trait documentation, but were intended to give a general observation of an accession's attribute, measurable trait, or characteristic.

5. Evaluation efforts

The screening and evaluation of the national collection for traits such as disease, insect, and environmental stress resistance are beyond the capacity and resources of the national collection to achieve, and necessitate research community participation. Despite this fact, the collection in recent years has tried to move from a passive supplier of germplasm resources to an active participant in germplasm disease and insect resistance screening and evaluation efforts. The collection has a long history of evaluation efforts by the research community for biotic and abiotic stress resistance. In fact the origins of the collection are intertwined with the entry of the boll weevil into the United States in the 1890's and the search for a source of host plant resistance to that pest [4]. Numerous evaluations have been conducted on the collection for phenotypic characteristics of agronomic importance. A sampling of evaluations includes investigations of seed protein [40], seed oil [41], seed gossypol [42], boll weevil resistance [43] and *Cercospora* leaf spot and *Verticillium* wilt resistance [44]. Prior to 2000, over 320 accessions of the collection had been screened for resistance to pink bollworm (*Pectinophora gossypiella* Saunders) [45], and 471 race stocks of *G. hirsutum* were screened for resistance to root knot nematode (*Meloidogyne incognita*) and resistance found in 18 lines [46]. As of 1986, over 200 accessions of the collection had been reported to carry resistance to one or more pests [47].

More recently, in 2009, eleven ongoing evaluations of collection germplasm for disease, insect, and nematode resistance were reported in a status report of U.S. cotton germplasm [14].

In an ongoing project, the Cotton Improvement Program at the Texas A&M AgriLife Research Center, Lubbock, TX began in 2005 to screen accessions from the working collection for resistance to the thrips (Thysanoptera: Thripidae) pest. Thrips were ranked as the number three pest of U. S. cotton in 2012, reducing yield by 0.374 percent and causing a loss of 123,947 bales [48]. In the panhandle and far west Texas, 860,476 acres were treated for pest thrips in 2012 at a cost of \$1.28 per acre [48]. By 2013, 516 accessions from the active collection of the U.S. National Cotton Germplasm Collection had been screened (Table 4). Resistance to thrips was identified in *G. barbadense* accession TX110 (PI 163608) in the first year of screening [49]. A series of studies conducted through much of the 20th century identified, confirmed and characterized thrips resistance in *G. barbadense*. Studies in this progression include: the discovery of resistance in glabrous Egyptian cotton cultivars and the conclusion that resistance is most likely due to a thicker leaf epidermal layer on lower sides of leaves allowing cotton seedlings to tolerate more thrips feeding [50]; confirmation of this work [51]; further characterization of feeding by thrips and measurement of the protrusion of the mandible past the fused maxillae at 11 microns [52] and the identification of a resistant group of *G. barbadense* cottons with a lower leaf epidermal layer of 10-13 microns [53]. This series of studies was summarized by Bowman and McCarty [54] who investigated the heritability of the traits conferring resistance. Interspecific hybridization between resistant *G. barbadense* accession, TX110 and two unreleased elite lines from the Lubbock Texas A&M AgriLife Research Cotton Improvement Program was used to begin a cultivar development project. Selections for thrips resistance and day-neutral flowering habit were made in the segregating F₂ plots, and this process continued for five years. Resistance has been carried to the F₅ and F₆ generations in many individuals, and day-neutral flowering habit and favorable agronomic traits have been improved.

Species	Genome	Accessions tested
<i>Gossypium hirsutum</i>	AD1	473
<i>Gossypium barbadense</i>	AD2	25
<i>Gossypium mustelinum</i>	AD4	3
<i>Gossypium darwinii</i>	AD5	6
<i>Gossypium herbaceum</i>	A1	2
<i>Gossypium arboreum</i>	A2	4
<i>Gossypium thurberi</i>	D1	1
<i>Gossypium harknesii</i>	D2-2	1
<i>Gossypium davidsonii</i>	D3	1

Table 4. By species, number of cotton accessions from U.S. National Cotton Germplasm Collection screened for resistance to the Texas High Plains pest thrips complex at Texas A&M AgriLife Research in Lubbock, Texas, 2005-2013.

In another ongoing project funded by USAID and USDA to identify sources of resistance to Cotton Leaf Curl Virus (CLCuV), germplasm resources were made available through the U.S. NCGC. The ready availability of accessions from the collection, combined with winter nursery seed increase capabilities, GRIN database information and especially the recent addition of standardized descriptor data and digital images, made possible a rapid coordinated CLCuV screening program. CLCuV is a major threat to cotton production in Pakistan and parts of India and has been reported in cotton producing countries in Africa, as well as China and Uzbekistan. This project to identify sources of resistance to CLCuV, helps not only countries such as Pakistan where the virus is already a problem, but also makes resistant germplasm available, should CLCuV become a threat to cotton production in other countries. As part of the project, accessions from the collection were increased at the CWN in Tecoman, Mexico. At the CWN, the collection curator and staff collected descriptor data and digital images for all the accessions. Part of the seed was sent to Pakistan and screened for CLCuV and the rest returned to the collection. In Pakistan, the disease is endemic which allowed replicated screening at three locations. At each location, the screening nurseries included regularly spaced host rows of highly susceptible plants. Any plants identified as resistant were re-tested the following year. All the screening data is being made available for inclusion in the collection database. Previous screening tests had identified *G. arboreum* and *G. herbaceum* as potential sources of resistance, so 1,050 *G. arboreum* and 100 *G. herbaceum* were increased at the CWN and sent for screening. Transferring resistance from diploid *G. arboreum* or *G. herbaceum* into tetraploid *G. hirsutum* is extremely difficult, so the project expanded to evaluate *G. hirsutum*. The NPGS GRIN database was used to identify a subset of 920 *G. hirsutum* accessions with a range of morphologies and originating from diverse geographic regions. CLCuV screening identified 12 resistant *G. hirsutum* accessions originating from northeast Brazil, Central America and the Caribbean. Travel to Pakistan is difficult and the plots could only be evaluated by the USDA researchers once during the field season. However, descriptor data and digital images were e-mailed throughout the season from the Pakistani partners for each of the 12 accessions. These could be compared to the descriptor data and digital images made by the collection curator at the CWN. This Collection information proved essential as 11 of the 12 accessions were photoperiod sensitive and did not flower during the field season in Pakistan. Additional accessions from these areas, as well as other geographic regions, were selected using GRIN and are being screened in Pakistan in 2013. This project has served as a model for a germplasm evaluation effort that serves the germplasm collection as well as the research community. In addition to identifying resistant sources to CLCuV for future cotton improvement efforts, the project made possible the seed renewal of numerous accessions of the collection under controlled conditions and the characterization and digital imaging of these same accessions.

Another current effort involves evaluating germplasm resources to identify lines with physiological and morphological traits that can improve water use efficiency and tolerance to extreme temperature and drought. The decline of water in glaciers, reservoirs and aquifers in many regions, combined with climate change and the unpredictability of precipitation during the growing season, has stimulated efforts to identify germplasm resources that can minimize the elevated production risks associated with crop water deficits. Cultivars with extensive root systems or physiological adaptations to their aerial parts may offer a means to minimize the

impact of periods of short-term or prolonged water stress on yield or quality. Plants vary in their physiological response to stress and their responses vary with the severity and duration of the limited water availability, extreme heat exposure, and other abiotic stresses [55-57]. Diversity in cotton for root architecture has been reported [58], as well as variation in root resistance expressed as the redistribution of water through the soil layer-profile by the root system (hydraulic lift) [59]. Untapped genetic variability for plant and root morphology-architecture types is present in germplasm resources but lacking in modern commercial cultivars [60,61]. However, methods of morphological or phenotypical characterization have progressed slowly in the last 30 years [62]. Currently, accessions from the NCGC are being used by the USDA-ARS Cropping Systems Research Laboratory, Lubbock, TX to initiate this phenotypical characterization. Specifically, drought responses are being examined in accessions of the *Gossypium* Diversity Reference Set, created for an ongoing diversity study being conducted by the USDA-ARS at College Station, Texas. Physiological and biochemical plants responses such as photosynthesis and CO₂ rates, stomata conductance, and osmotic adjustment will be examined and monitored under heat and/or low-temperature stress conditions. Also, a rapid bioassay is being used for monitoring plant stress based on the ability of a "source" leaf to provide sufficient energy for plant growth. Water-deficit stress reduces new growth and lessens the demand on the source leaves. [63,64]. In addition, plant mapping (plant height, nodes, and fruiting position), agronomic (seed cotton yield, seed index, number of bolls per plants, lint percentage, etc.), and fiber quality (length, strength, and fineness) data are used to monitor drought tolerance among cotton entries or accessions. Accessions are evaluated in field replicated plots under well-irrigated and water-deficit stress conditions, using a randomized complete block or incomplete block design with 3–4 replications using five plants as subsamples per replication, or two replications when seed availability from accessions of selected germplasm is limited [65].

In another project to screen the germplasm collection for drought tolerance, 400 accessions have been evaluated for variation in growth parameters by the Cotton Breeding Program at the Texas A&M AgriLife Research Center, Lubbock, TX. Measurements were obtained for the 400 accessions under non-stress conditions by growing plants in granular diatomaceous earth contained in 30 inch tubes made from 3 inch PVC pipe. Adequate water and fertilizer were applied, and day and night temperatures were maintained at 75-85°F and 67-69°F, respectively, to obtain maximum growth rates. After 20 days, plants were removed from the growth medium, washed, and taproot and shoot length were measured. Plants were then separated into roots and shoots, and dried at 140°F for 48 hours. Parameters obtained were: taproot length, shoot length, total root and shoot weights, and shoot to root ratio. All experiments were conducted using a RCB design with five blocks. Recognizing that in many regions declining water levels are associated with increased salinity in irrigation water, the Lubbock AgriLife research group evaluated 290 collection accessions for NaCl tolerance in a hydroponic system. Plants grown in hydroponic medium were evaluated continuously for a variety of salt injury indicators. Seedlings were exposed to increasing increments of salt (NaCl) solution at a consistent rate from 3,000 to 35,000 ppm until plants suffered near 100 percent mortality. Preliminary studies conducted in 2006-2009 show accession, TX 307 (PI 165390) to have

significantly better growth when treated with NaCl using the hydroponic screening system. Seed from a plant selection of TX 307 was increased to use as a control in further salinity studies.

Efforts also are underway to develop remote sensing and high throughput phenotyping methodology to evaluate germplasm variation for response to drought and other environmental stresses. Remote sensing instruments measure and record responses to drought by infrared camera monitoring, sensors and thermometers for leaf and plant canopy measurements. Soil water content is monitored using neutron moisture gages and bucket rain gages. Several research groups are working to develop high throughput field based equipment to phenotype thousands of plots and/or accessions in the minimum amount of time. Equipped with sensors for measuring spectral reflectance, canopy temperature, and plant mapping imaging; a high throughput vehicle has the potential to assess the responses of hundreds of accessions to stress in a fairly short time [62]. Many of the above instruments are developed or under development. Knowledge being acquired will provide acutely needed means of conducting uniform evaluations of germplasm for valuable drought, heat, salt, and other environmental stresses.

6. Genotyping

Molecular tools provide the means to characterize underlying genetic diversity that is not measurable through classical phenotypic descriptors [66]. With the advance of DNA marker technologies, it is now possible to characterize *Gossypium* germplasm not only phenotypically at the levels of whole plants but also genotypically at the levels of whole genomes [67,68]. In general, cotton lags behind other major crops in genomic tools that are available for effective manipulation and exploitation of beneficial genes otherwise buried in *Gossypium* germplasm collections. The first molecular maps for cotton [69-70] were based on the cumbersome and expensive restriction fragment length polymorphism (RFLP) technology that requires large amounts of genomic DNA and generation of radioactive probes physically disseminated to the research community as plasmid or phage clones. Thus, RFLP markers were poorly suited for high-throughput genotyping experiments in germplasm characterization [67]. New DNA marker technologies were needed to coordinate the systematic characterization and the simultaneous comparison among various research efforts involved in cotton diversity analysis and genetic resource preservation.

Microsatellites of simple sequence repeat (SSR) markers, in contrast to RFLP markers, are abundant, co-dominant, and widely distributed throughout the genomes of higher plants. They are amenable to high-throughput assay via multiplex polymerase chain reaction (PCR) bins on automated sequencers [72]. Thus, SSR markers are portable and simple to use for germplasm characterization. Over the last decade, several thousand SSR markers have been developed for the tetraploid genome of cultivated cottons [73-76]. For various applications in cotton, SSR markers have been systematically characterized and genetically mapped in cotton genomes. High-density genetic maps were constructed based on an immortal recombinant inbred line (RIL) population that was developed from a cross between TM-1 and 3-79, the

genetic standards for *Gossypium hirsutum* and *G. barbadense*, respectively [73,75]. Using a balanced diversity panel of 12 cultivated and exotic cotton genotypes representing six *Gossypium* species [77, Table 5], a core set of 105 SSR markers has been developed for coordinated germplasm characterization [72]. This initial set of core SSR markers were carefully selected on the basis of criteria that included reproducible DNA amplification, ability to be multiplexed, reasonable polymorphism information content (PIC), representation of different marker sources, and uniform distribution across the tetraploid cotton genome. The initial core set has only two DNA markers proposed for each chromosome arm of the 26 tetraploid cotton chromosomes [67]. The core markers serve as the standard descriptors to characterize workable sets of *Gossypium* accessions across different gene pools or germplasm sources.

With the development of molecular markers, genetic diversity studies within *Gossypium* have been pursued using accessions from the U.S. NCGC with varying objectives, including understanding the evolutionary process of interspecific gene flow within *G. aridum* [78], the genetic relationships of geographically diverse *G. arboreum* cultivars [79], and the utility of microsatellite primers developed in tetraploid species in the diploid *G. davidsonii* [80], to name a few examples. As cultivars possess the most readily accessible genetic variation for use in breeding programs, a number of diversity investigations have occurred using improved cotton cultivars [81-84]. Typically, many of these diversity studies have been limited in the scope of accessions investigated or the number of markers used [85,79,82]. There has been little standardization in these studies, as different groups of molecular markers were used for each group of accessions. Attempts to characterize larger genomic groups using a standardized marker set have been made in the germplasm collections of Uzbekistan [86,87] and France [88]. Recognizing the need to characterize the genetic diversity within the NCGC and the utility of markers in characterizing and managing collection diversity, the working collection at College Station, TX initiated research to characterize a major portion of the collection using the core SSR marker set described above. Objectives of this characterization included: determining the structure of genetic variability within the collection with the goal of identifying targets for further germplasm collecting and exchange efforts; identification of redundancies, misidentifications, unintended introgression, and gaps within the collection; and developing and validating a core marker set that could be used in comparative studies across collections and over time. The core marker set developed to accomplish this task was used to genotype and analyze a *Gossypium* Diversity Reference Set (GDRS) of 2,254 accessions (approximately 20% of the NCGC). The GDRS represented the range of diversity of *Gossypium*, including nine genomes (eight diploid and one tetraploid genome) and 33 species as represented in the collection (Table 6). DNA profiles of these accessions showed strengths and deficiencies of using the core set of 105 SSR markers.

When applied to the GDRS, the core marker set was most successful at revealing DNA profiles in the *G. hirsutum* and *G. barbadense* species. This was expected, since the SSRs used in the core set were developed from these two species. Many of the remaining species showed incomplete DNA profiles due to lack of PCR amplification or non-informative profiles due to the amplification of monomorphic DNA fragments. The current 105 marker core set was capable of discriminating tetraploid species, and discriminating between the diploid A genome and all

CMD>#	CMD name	Description
CMD01	TM-1	<i>G. hirsutum</i> (AD ₁) genetic standard
CMD02	3-79	<i>G. barbadense</i> (AD ₂) genetic standard
CMD03	Acala Maxxa	California Upland cotton (AD ₁)
CMD04	DPL 458BR	Upland cotton (AD ₁) with significant acreage
CMD05	Paymaster 1218BR	Upland cotton (AD ₁) with significant acreage
CMD06	Fibermax 832	Upland cotton (AD ₁) with significant acreage
CMD07	Stoneville 4892BR	Upland cotton (AD ₁) with significant acreage
CMD08	Pima S-6	Pima (AD ₂) germplasm breeding source
CMD09	<i>G. arboreum</i> (A ₂ -8)	A genome representative
CMD10	<i>G. raimondii</i> (D ₅ -3)	D genome representative
CMD11	<i>G. tomentosum</i> (AD ₃)	Introgression breeding source
CMD12	<i>G. mustelinum</i> (AD ₄)	Introgression breeding source

Table 5. A standardized panel of *Gossypium* germplasm diversity for cotton marker development (CMD) [12]

other diploid genomes. Within *G. hirsutum* and *G. barbadense*, the marker set revealed misclassification, introgression, and accession redundancy and uniqueness, and therefore will be a significant tool in maintaining collection integrity within the primary species. It was determined that detailed characterization efforts would continue to use those core SSR markers that were informative, but further characterization would require the addition of markers that were genome- or species-specific. In this ongoing research effort, diversity reference sets will be identified to represent a minimum of 25% of the accessions available for a given species, genome, or group within the NCGC. Currently, reference sets are being created for the A genome, D genome, and tertiary genomes. These Individual reference sets are being assembled to minimize similarity between accessions within the set, with priority being given to those accessions representing geographic, ecological, and morphological diversity; and differing degrees of human manipulation (i.e., wild species, semi-adapted landraces, or improved cultivars) that would be of interest in a diversity structure analysis. This approach will more readily illustrate the genetic diversity of the 10,000+ accessions in the collection and enhance the usage of this large collection by geneticists and breeders world-wide. The analyses of results from genotyping reference sets will be used to guide the development of smaller core collections for species and genomes. Marker information also will be used to prioritize regeneration efforts, identify redundancy and uniqueness in the collection, and monitor integrity of accessions through regeneration cycles. Coordinated characterization between collections using markers could enhance exchange efforts and allow mutual protection of holdings. Efforts are underway to organize a cooperative international effort to genotype cotton collections with a standard set of markers and a uniform distribution of species and ecotypes.

Genome	Species	Accessions	Accession total
		number (%)	in NCGC
A	<i>G. arboreum</i>	145 (8.4%)	1729
	<i>G. herbaceum</i>	49 (25.3%)	194
AD	<i>G. barbadense</i>	430 (27.1%)	1584
	<i>G. darwinii</i>	4 (2.9%)	138
	<i>G. hirsutum</i>	1541 (24.5%)	6302
	<i>G. mustelinum</i>	7 (36.8%)	19
	<i>G. tomentosum</i>	2 (12.5%)	16
	<i>G. anomalum</i>	5 (71.4%)	7
C	<i>G. nandewarense</i>	1 (16.7%)	6
	<i>G. sturtianum</i>	3 (42.9%)	7
D	<i>G. aridum</i>	4 (28.6%)	14
	<i>G. armourianum</i>	2 (20.0%)	10
	<i>G. davidsonii</i>	9 (29.0%)	31
	<i>G. gossypoides</i>	7 (100.0%)	7
	<i>G. klotzschianum</i>	1 (1.7%)	59
	<i>G. laxum</i>	1 (50.0%)	2
	<i>G. lobatum</i>	1 (25.0%)	4
	<i>G. raimondii</i>	4 (7.1%)	56
	<i>G. thurberi</i>	9 (24.3%)	37
	<i>G. trilobum</i>	6 (54.5%)	11
E	<i>G. areysianum</i>	1 (50.0%)	2
	<i>G. somalense</i>	2 (66.7%)	3
	<i>G. stocksii</i>	2 (50.0%)	4
F	<i>G. longicalyx</i>	4 (100.0%)	4
G	<i>G. australe</i>	3 (27.3%)	11
	<i>G. bickii</i>	4 (80.0%)	5
	<i>G. nelsonii</i>	3 (75.0%)	4
K	<i>G. costulatum</i>	1 (50.0%)	2
	<i>G. exiguum</i>	1 (100.0%)	1
	<i>G. marchantii</i>	1 (100.0%)	1
	<i>G. nobile</i>	1 (100.0%)	1
	<i>G. populifolium</i>	1 (25.0%)	4
	<i>G. pulchellum</i>	1 (100.0%)	1
Overall		2256 (22.0%)	10276

Table 6. Accessions in the *Gossypium* Diversity Reference Set (GDRS), by genome and species.

The low resolution of the 105 core marker set used in the U.S. collection offers a first glimpse at, or general survey of the cotton germplasm collection, but it does not provide detailed characterization of specific genomic regions that may harbor important genes of interest. The size and complexity of the cotton genomes require many more widely applicable DNA markers for more effective germplasm characterization and gene discovery [72-74]. While the growing collection of well characterized SSR markers offers an opportunity of expanding the current set with additional core markers to meet the demand, single nucleotide polymorphism (SNP) markers represent a promising DNA marker system [89]. SNP markers are co-dominant and more abundant in a given genome than SSR markers, and they offer a new opportunity for generating large numbers of intraspecific polymorphisms [90]. Current efforts to develop large numbers of cotton SNP markers are based on reduced genome representation [91] or other limited sequence resources [92]. A targeted genotyping by sequencing (GBS) approach, that is simple and cost-effective, is being explored to simultaneously discover and map SNP markers for traits of interest [93]. With the current efforts to develop whole genome sequence in cotton [94,95] including several members of germplasm standards and mapping parents, hundreds of thousands of SNP markers will be identified in the cotton genomes. These SNP loci will be validated and sorted out in various orthologs, homeologs and paralogs as often encountered in the allotetraploid cotton. Individual sets of core markers specific for each genome, species or land race will be developed from evenly distributed genomic and genic regions, and they will be arrayed in public SNP genotyping chips for high-throughput and more detailed characterization and exploitation of genetic diversity of *Gossypium* germplasm [93,96].

Presently, an investigation has been made of the diversity of upland cotton cultivars using a larger set of available SSR markers. Commercial *G. hirsutum* cultivars being the most accessible sources of genetic variation available to breeding and genetic improvement efforts, a cooperative effort involving the collection was undertaken to determine the variability available and the structure of diversity in this elite group using a broader set of molecular markers. A set of 193 upland cotton cultivars from 26 countries that include most cotton growing regions of the world were sampled from the national collection and elsewhere. These cultivars represent a wide spectrum of the genetic diversity found within the cultivated upland cotton. Due to strong contributions and influence of the U.S. cultivars to early breeding efforts in other countries, a special emphasis was placed upon representing U.S. cultivars in this study. When selecting U.S. cultivars, the following factors were taken into consideration: breeding program, pedigree (if known), geographic region, era of release, planting acreage and breeding value (e.g. as a common parent in breeding programs). The 130 cultivars from the U.S. represented more than 100 years of breeding history. Four hundred forty-eight SSR markers were selected based on their mapping positions in the *G. hirsutum* TM-1x*G. barbadense* 3-79 genetic map [76]. In addition, whenever possible, the physical positions of the selected SSR markers were also located based on the reference sequence of *G. raimondii* genome [95]. The genome-wide linkage disequilibrium (LD) between pairs of SSR marker loci was calculated using the software package JMP Genomics 6.0 (SAS Corporation, Cary, NC). Genetic similarities between all pairs of cultivars were calculated based on the Dice's coefficient [97] using the software package NTSYS-pc [98]. Neighbor-joining (NJ) trees were generated to view the phylogenetic relationship among the cultivars.

The 448 selected SSR markers revealed a total of 1590 alleles belonging to 732 loci that were distributed across the whole cotton genome (Table 7). Of these 732 loci, 523 were polymorphic and 209 were monomorphic. One hundred thirty nine unique alleles were observed in 69 cultivars. Of the 130 U.S. cultivars, 94 had information on the years of their release, ranging from 1899 ('Mebane') to 2010 ('UA48') [99,100]. Using this information, an analysis of unique marker allele numbers contained in the cultivars clearly demonstrated that modern U.S. upland cotton has been gradually losing its genetic diversity during the past century. The pair-wise genetic similarity between all cultivars averaged 0.760 with a range from 0.640 ('Pak 4F CB 4025' and 'Paymaster HS200') to 0.993 ('Rowden' and 'DES 716'). These cultivars were assembled into 15 groups with sub-groups present in groups 4 and 6 based on the phylogenetic tree (Table 8). A detailed phylogenetic tree can be found in Fang et al. 2013 [81]. Molecular marker analysis revealed little relationship between the genetic make-up of cultivars and their countries of origin. Instead, the results from genetic diversity analysis and phylogenetic grouping based on molecular markers were largely congruent with the breeding history and pedigree information of many cultivars [100,101].

The information obtained from this research may benefit cotton geneticists and breeders in a variety of ways. First, breeders could use a subset or a whole set of the markers to genotype their germplasm. Then, they would be able to compare their germplasm to the cultivars analyzed in this study based on the molecular profiles. Second, depending on their breeding objectives, breeders may choose cultivars from the list as the parents to make new crosses, or choose one cultivar to cross with their own favorite cultivar(s) based on their molecular profiles and phylogenetic groups. Third, breeders may use the information to re-evaluate their breeding programs. Fourth, breeders or geneticists may use all or a subset of these cultivars as starting materials to conduct an association mapping study.

Chromosome (A sub-genome)	No. loci	Chromosome (D sub-genome)	No. loci
1	17	14	31
2	18	15	25
3	27	16	20
4	16	17	16
5	39	18	18
6	17	19	39
7	22	20	13
8	21	21	21
9	28	22	20
10	14	23	24
11	21	24	18
12	26	25	19
13	17	26	19
not mapped	166		
Total			732

Table 7. Chromosome distribution of 732 loci within cotton genome for 448 markers used to genotype 193 cultivars.

Group	No. cultivars	No. unique alleles	No. countries of origin	Description
1	5	6	5	Common factor unknown or not clear
2	8	3	3	Cultivars with Coker 100 in their pedigrees
3	10	8	1	Pee Dee germplasm
4	22	34	6	Cultivars from Pakistan, multi-adversity resistance breeding program and Paymaster program in Texas
5	15	4	5	Cultivars from S. America, and those with disease resistance
6	70	29	4	Commercial cultivars developed in the U.S., Australia and China
7	13	4	2	Acala type
8	4	7	3	Common factor unknown or not clear
9	6	1	2	Cultivars from Paymaster program in Texas high plain
10	4	0	3	Common factor unknown or not clear
11	6	14	2	Common factor unknown or not clear
12	3	1	3	Common factor unknown or not clear
13	3	11	2	Common factor unknown or not clear
14	3	1	2	Coker cultivars related to Coker 310
15	19	14	9	Cultivars from Africa, and Coker cultivars with wilt resistance

Table 8. Summary of phylogenetic grouping of 191 cultivars

7. Germplasm databases

In this chapter we have tried to briefly describe the U.S. National Cotton Germplasm Collection and some of the current efforts made to preserve, characterize, and distribute its contents. However, the utility of the germplasm collection and its genetic resources is directly proportional to the knowledge of the genetic diversity within the collection, and the accessibility of that knowledge. Without the ability to disseminate information about the collection broadly to a community with free and easy access, the utility of the collection is greatly diminished. Currently, information about the collection can be accessed at two main database sources: GRIN-GLOBAL and CottonGen. GRIN-GLOBAL (<http://distribution.grin-global.org/gringlobal/search.aspx>) has been adequately described elsewhere and will not be described here. The following is a description of CottonGen and efforts to expand its contents and capabilities.

CottonGen (www.cottongen.org, [102,103]) is a curated and integrated web-based relational database providing centralized access to publicly available genomic, genetic and breeding data for cotton. Initiated in 2011, as the successor to CottonDB [102], CottonGen contains annotated whole genome sequences, genes, transcripts, markers, trait loci, genetic maps, taxonomy, germplasm, publications, data analysis tools, and communication resources for the cotton community. Annotated whole genome sequences of *Gossypium raimondii* are available with

aligned genetic markers, transcripts and protein homologs. These whole genome data can be accessed through genome pages, search tools and the genome browser GBrowse. Most of the published cotton genetic maps can be viewed and compared using the comparative map viewer CMap, and are searchable via map search tools. Search tools also exist for markers, QTLs, germplasm, publications, and trait evaluation data.

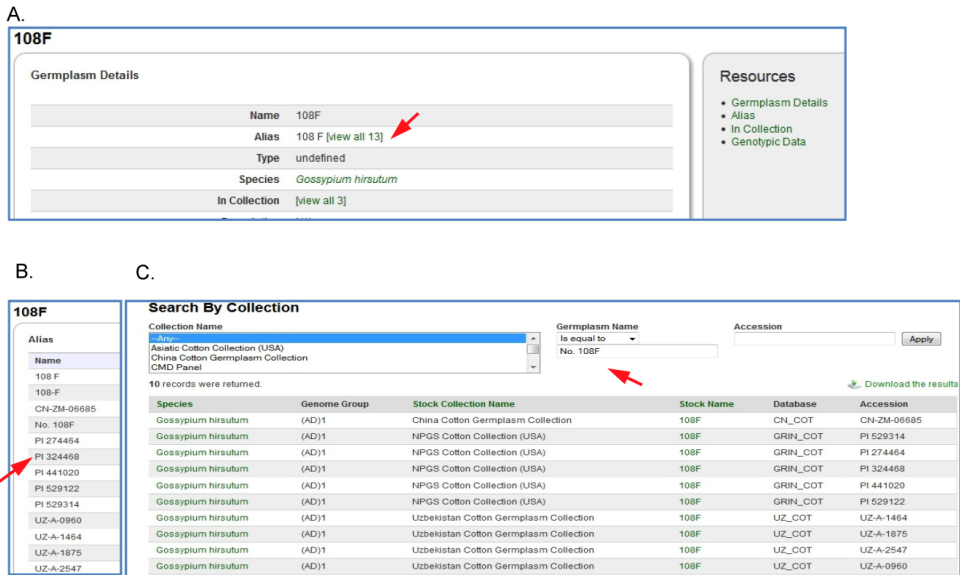


Figure 1. An example of how germplasm names can be used and displayed in CottonGen. A. The details page of germplasm ‘108F’ which shows this germplasm individual has been recorded as 13 different names (aliases). B. lists the 13 names (aliases), some of them are the accessions from various germplasm collections or germplasm groups, such as ‘PI 274464’ which is an accession name used in GRIN. Other names show the others formats used for this germplasm, such ‘108 F’ and ‘No. 108F’. C. Any of the names in this germplasm aliases list can be used in the germplasm name search, such as ‘No. 108F’, the same result table will be displayed if it matches any of the aliases or the standardized name.

A critical function of any community database is to provide data and information that is as accurate, unambiguous and non-redundant as possible. For germplasm pedigree and trait data, this can be particularly challenging as it involves identifying and rationalizing the many aliases used for accession names, and the different descriptors and measurement scales for the same trait. This is often compounded by a lack of standardization within and between collections due to data storage system restrictions, nomenclature differences, and unrecognized errors at the time of data entry. When these issues are resolved by the database curator, the curated data are integrated with other associated data and made accessible to scientists through easy-to-use interfaces to ensure utility to scientists.

The first step in the germplasm curation effort involved creating a validated list of unique cotton germplasm with unique IDs. A good starting point for this list was obtained from the

USDA-ARS GRIN [104] cotton collection PI numbers. In order to resolve the discrepancies, the apparently unique lines that could be identified from the passport data were given an unique “cdbgm” ID in CottonGen, and the accession numbers or names from different collections or germplasm groups (such as the CMD panel), and alternate names cross-referenced to these CottonGen unique IDs. The unique ID, germplasm names and aliases were then expanded to include other germplasm from the GRIN Cotton Crop Science Registration, GRIN Plant Variety Protection (PVP), pedigrees of 642 Upland and 20 Pima cultivars, plus 205 parents representing most of the cotton breeding lines and obsolete cultivars, (kindly provided by Dr. Bowman [100], the cotton germplasm collection of the China Cotton Research Institute, the Chinese Academy of Agricultural Sciences, and the cotton germplasm collection of Uzbekistan Center of Genomics and Bioinformatics, Academy of Sciences of Uzbekistan. As a result of curation efforts, CottonGen currently contains 14,959 assigned unique IDs connected to 53,645 corresponding aliases obtained from 14 collections for 49 *Gossypium* spp. Access and display of germplasm names (standardized and aliases) in CottonGen are shown in Figure 1.

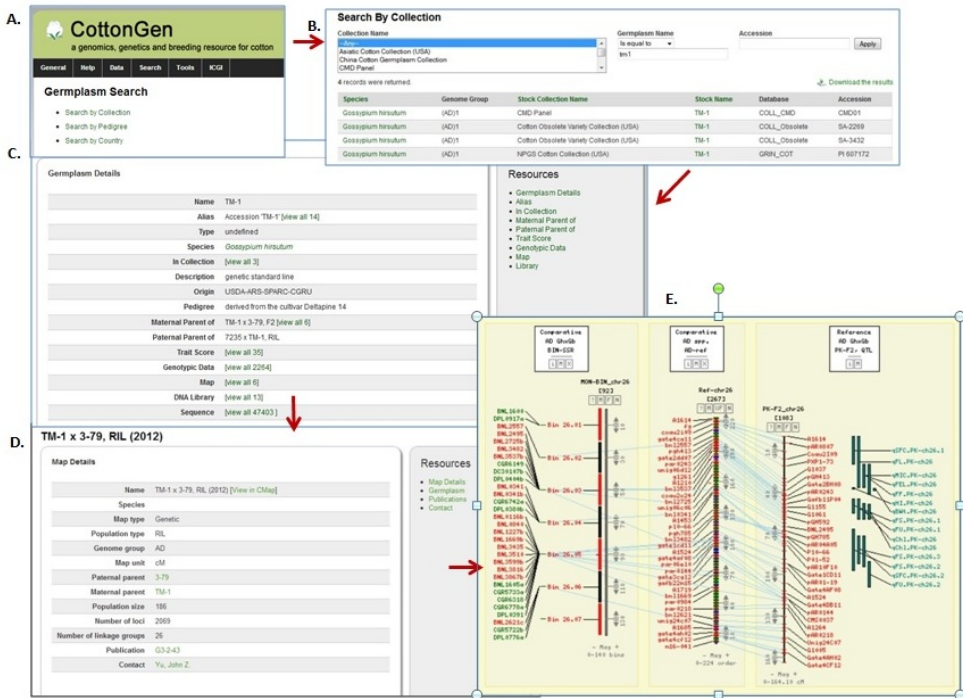


Figure 2. Germplasm search site in CottonGen. A. Multiple germplasm search sites are available based on the type of information users are interested in. B. An example search interface where users can view and search for germplasm and their collection center. C. Germplasm detail page with various tabs to show the detailed information. D. Map tab of the germplasm page shows all the maps for which the germplasm has been used. E. From the map page users can go to CMap for accessed to marker between and within maps, with hyperlinks to the markers detail page.

In addition to germplasm name (and aliases), germplasm data housed in CottonGen includes pedigrees, publicly available passport data, stock collection center information, associated maps, libraries and sequences. The trait evaluation database in CottonGen contains over 118,000 trait scores from 9,000 accessions. The *Gossypium* species summary page (<http://www.cottongen.org/data/species>) provides a list of species along with information such as genome group, haploid chromosome number, geographic origin, and number of accessions, sequences and DNA libraries per species. The species name in the table leads to individual species pages, which show more detail such as common name, images and other additional data such as functional analysis of the genes, both from the National Center for Biotechnology Information (NCBI) and whole genome sequences, which includes Kyoto Encyclopedia of Genes and Genomes (KEGG) [105] and Gene Ontology (GO) [106] analysis reports. The germplasm search page, accessible from <http://www.cottongen.org/search/germplasm> provides access to different types of searchable data (Figure 2). The search by collection page provides a list of germplasm along with stock collection center information. The search can be filtered by collection center name, germplasm name and/or accession name in the stock center. The search by pedigree page provides an interface to search germplasm by pedigree and the search germplasm by country page searches by the country of origin. From the germplasm search page, researchers can go to the germplasm details page, which shows all the detailed information such as pedigree, passport, collection center, image and associated genotypic and phenotypic data (Figure 2). Germplasm can also be searched based on their trait evaluation data. Both the qualitative trait evaluation search and quantitative trait evaluation search site allows the trait values of up to three trait descriptors to be specified to view the germplasm trait data. Data from all the search result sites can be downloaded in Excel files.

Among ongoing efforts of CottonGen is the development of a digital image library to store over 100,000 images provided by the USDA-ARS Research Project: “Genotypic and Phenotypic Analysis and Digital Imaging of Accessions in the US National Cotton Germplasm Collection”. The associated phenotypic data from the same project will be stored in CottonGen as well. Provision of other germplasm data to CottonGen is actively encouraged.

8. Conclusion

The National Cotton Germplasm Collection is a complex amalgamation of several previously existing collections, which present challenges to its continued growth, preservation, characterization, and evaluation. Although habitat loss and international treaties have had significant impact on germplasm collection and exchange efforts, the NCGC continues to grow through mutually beneficial collecting efforts and germplasm exchanges with cooperating countries. Given finite and sometimes constricting resources; efficiency and effectiveness in preserving, characterizing, and evaluating the collection’s contents becomes imperative. One means of increasing the efficiency and effectiveness of the collection has been to enlist the research community in characterizing and evaluating the collection. Currently there are dynamic cooperative efforts to evaluate the collection for drought, heat, and other environmental stresses associated with global climate change. Efforts to find resistance to biotic stresses within

the collection continue, as do efforts to identify positive variation within the collection for agronomic and fiber quality characteristics. The development of genetic marker technology greatly increases the ability to investigate the genetic variation of the collection and offers needed means to manage the collection's contents through identification of redundancy, misclassification, introgression, and sources of unique variability within the collection. Cooperative efforts within the research community to characterize and evaluate the collection, while very effective, could be replicated at an international level with greater impact.

Author details

Richard G. Percy^{1*}, James E. Frelichowski¹, Mark D. Arnold², Todd B. Campbell³, Jane K. Dever², David D. Fang⁴, Lori L. Hinze¹, Dorrie Main⁵, Jodi Scheffler⁶, Monica A. Sheehan², Mauricio Ulloa⁷, Jing Yu⁵ and John Yu¹

*Address all correspondence to: Richard.Percy@ars.usda.gov

1 USDA-ARS, College Station, TX, USA

2 Texas A&M AgriLife Research, Texas A&M Agrilife Research & Extension Center, Lubbock, TX, USA

3 USDA-ARS, Florence, SC, USA

4 USDA-ARS, New Orleans, LA, USA

5 Department of Horticulture, Washington State University, WA, USA

6 USDA-ARS, Stoneville, MS, USA

7 USDA-ARS, Lubbock, TX, USA

References

- [1] Ware JO. Plant Breeding and the Cotton Industry. U.S.D.A. Yearbook; 1936.
- [2] Mauer FM. The Cottons of Mexico, Guatemala, and Columbia. Suppl. 47, Bull. Appl. Bot. Genet. and Pl. Br. 1930.
- [3] Percival AE, Wendel JF, Stewart JM. Taxonomy and Germplasm Resources. In: Smith CW and Cothren JT (eds.) Cotton: Origin, history, technology, and production. New York: John Wiley & Sons; 1999. p. 33-64.

- [4] Smith CW, Cantrell RG, Moser HS, Oakley SR. History of Cultivar Development in the United States. In: Smith CW, Cothren JT (eds.) Cotton: Origin, history, technology, and production. New York: John Wiley & Sons; 1999. p99-172.
- [5] SCSB. Preservation and Utilization of Germplasm in Cotton 1968-1980. Southern Cooperative Series Bulletin No. 256. 1981
- [6] SCSB. Genetics and Cytology of Cotton 1956-67. Southern Cooperative Series Bulletin No. 139. 1968.
- [7] SCSB. Genetics and Cytology of Cotton 1948-55. Southern Cooperative Series Bulletin No. 47. 1956.
- [8] IBPGR. Descriptors for Cotton Species. International Board for Plant Genetic Resources Working Group, Rome, Italy. 1980.
- [9] Beasley JO. Meiotic chromosome behavior in species, species hybrids, haploids and induced polyploids of *Gossypium*. Genetics 1942; 27 25-54.
- [10] Percival AE. The National Collection of *Gossypium* Germplasm. Southern Cooperative Series Bulletin No. 321. 1987.
- [11] Vertucci WE, Roos EE. Theoretical basis of protocols for seed storage. Plant Physiol. 1990;94 1019-1023.
- [12] Cromarty AS, Ellis RH, Roberts EH. The design of seed storage facilities for genetic conservation. International Board for Plant Genetic Resources, Rome, Italy. 1982.
- [13] Abdalla AM, Raddy OUK, El-Zik KM, Pepper AE. Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor. Appl. Genet. 2001;102 222-229.
- [14] Wallace TP, Bowman D, Campbell BT, Chee P, Gutierrez OA, Kohel RJ, McCarty J, Myers G, Percy R, Robinson F, Smith W, Stelly DM, Stewart JM, Thaxton P, Ulloa M, Weaver DB. Status of the USA cotton germplasm collection and crop vulnerability. Genet. Resour. Crop Evol. 2009;56 507-532.
- [15] Xu Q, Zhang X, Nie N, Feng CD. Genetic diversity among cotton cultivars (*G. hirsutum* L.) resistant to Fusarium wilt as estimated by RAPD markers. Scientia Agricultura Sinica 2002;35(3) 272-276.
- [16] Multani DS, and Lyon BR. Genetic fingerprinting of Australian cotton cultivars with RAPD markers. Genome 1995;38 1005-1008.
- [17] Ulloa M., Brubaker C., Chee P. Cotton. In: Kole C. (ed.) Genome Mapping & Molecular Breeding. Vol. 6: Technical Crops. Heidelberg, Berlin, New York, Tokyo: Springer; 2007. p1-49.
- [18] Ulloa M, Percy R, Hutmacher RB, Zhang J. The future of cotton breeding in the Western United States. J. Cot. Sci. 2009;4 246-255.

- [19] Ulloa M, Abdurakhmonov IY, Perez-M C, Percy R, Stewart McDJ. Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers. *Botany* 2013;91 251-259.
- [20] Percival AE, Kohel RJ. Distribution, collection, and evaluation of *Gossypium*. *Adv. Agron.* 1990;44 225-256.
- [21] USDA. The Regional Collection of *Gossypium* Germplasm. United States Department of Agriculture, Agriculture Research Service, ARS-H-2, Oct. 1974.
- [22] Bretting, PK. The U.S. National Plant Germplasm System in an era of shifting international norms for germplasm exchange. In: KE Hummer. (ed.) *Plant Gen. Resources, Proc. XXVII IHC-S1*, International Society of Horticultural Science. Acta. Hort. 760, ISHS. 2007.
- [23] FAO. The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome. 2010. P370. Available from <http://www.fao.org/agriculture/seed/sow2/> (accessed 10 September, 2013).
- [24] Wallace TP, Bowman D, Campbell BT, Chee P, Gutierrez A, Kohel RJ, McCarty J, Meyers G, Percy R, Robinson F, Smith W, Stelly DM, Stewart JM, Thaxton P, Ulloa M, Weaver DB. Status of the USA cotton germplasm collection and crop vulnerability. *Gen. Res. Crop Evol.* 2009;56 507-532.
- [25] Campbell BT, Saha S, Percy R, Frelichowski J, Jenkins J N, Park W, Mayee CD, Gotmare V, Dessauw D, Giband M, Du X, Jia Y, Constable G, Dillon S, Abdurakhmonov I Y, Abdulkarimov A, Rizaeva SM, Abdullaev A, Barroso PAV, Padua JG., Hoffmann LV, Podolnaya L. Status of the global cotton germplasm resources. *Crop Sci.* 2010;50 1161-1179.
- [26] Ulloa M, Stewart JM, Garcia-C EA, Godoy-A A, Gaytán-M A, Acosta-N S. Cotton genetic resources in the western states of Mexico: in situ conservation status and germplasm collection for ex situ preservation. *Genet. Resour. Crop Evol.* 2006;53 (4) 653-668.
- [27] Fryxell PA. *The Natural History of the Cotton Tribe*. College Station and London: Texas A&M University Press; 1979.
- [28] Fryxell PA. A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheede* 1992;2 108-165.
- [29] Fryxell PA, Craven LA, Stewart JM. A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. *Systematic Botany* 1992; 17 91-114.
- [30] Grover CE, Grupp KK, Wanzek RJ, Wendel JF. Assessing the monophyly of polyploidy *Gossypium* species. *Plant Syst. Evol.* 2012;298 1177-1183.
- [31] Schwendiman J, Percival AE, Belot JL. Cotton collecting on Caribbean Islands and South Florida. *FAO/IBPGR Plant Genetic Resources Newsletter.* 1986;66 20-5.

- [32] Krapovickas A, Seijo G. *Gossypium ekmanianum* (Malvaceae), algodón Silvestre de la Republica Dominicana. *Bonpl.* 2008;17(1) 55-63.
- [33] Anonymous. Descriptors for Cotton. AGP:IBPGR. 1980;10.
- [34] Toll JA. Processing of germplasm, associated material and data. In: Guarino L, Ramanatha RV, Reid R, (eds.) *Collecting Plant Genetic Diversity: Technical Guidelines*. Rome, Italy: International Plant Genetic Resources Institute (IPGRI), Plant Production and Protection Division, FAO; 1995. p577-595.
- [35] Mackay M, Alercia A. Processing of germplasm, associated material and data. In: Guarino L, Ramanatha RV, Goldberg E. (eds.) *Collecting Plant Genetic Diversity: Technical Guidelines-2011 Update*. Rome, Italy: Bioversity International; 2011. p1-4.
- [36] Puzone L, Hazekamp T. Characterization and documentation of genetic resources utilizing multimedia databases. *Proc. Workshop, Dec. 19-20, 1996, Naples, Italy: IPGR; 1998. p67.*
- [37] Mincione A, Filippone E, Monti LM. Construction of an image-based germplasm data base for *Phaseolus* spp. *Euphytica*. 1994;75 149-159.
- [38] Smykalova I, Hampel D, Pavelek M. Determination of duplicates of accessions in a germplasm collection of flax/linseed by means of digital image analysis. *Indust. Crops Prod.* 2012;36 177-187.
- [39] Dever JK, Sheehan MA. Update of cotton race-stock screening and phenotypic characterization. *Proc. Beltwide Cotton Conf., 4-7, January 2011, Atlanta, GA. p718-719; 2011.*
- [40] Sheehan M A, Dever JK, Arnold MD. Phenotypic documentation of *Gossypium* species from the US Cotton Germplasm Collection, USDA-ARS, Using Digital Photography. *Proc. Beltwide Cotton Conf., 4-7, January 2010, New Orleans, LA. p753-756; 2010.*
- [41] Kohel RJ, Glueck J, Rooney LW. Comparison of cotton germplasm collections for seed-protein content. *Crop Sci.* 1985;25 961-963.
- [42] Kohel, RJ. Survey of *Gossypium hirsutum* L. germplasm collections for seed-oil percentage and seed characteristics. *Agric. Res. Serv., U.S. Dept. of Agriculture, ARS-S-187; 1978.*
- [43] Dilday RH, Shaver TN. Survey of the regional *Gossypium hirsutum* L., primitive race collection for flowerbud gossypol. *Agr. Res. Serv., U.S. Dept. of Agriculture, Publ. ARS-S-146; 1976.*
- [44] Jenkins JN, Parrott WL. Field evaluation of primitive races of *Gossypium hirsutum* L. for resistance to boll weevil. *Technical Bulletin, Mississippi Agricultural and Forestry Experiment Station, 91; 1978.*

- [45] Jenkins JN, Parrott WL. Field evaluation of primitive races of *Gossypium hirsutum* L. for resistance to *Cercospora* leaf spot and *Verticillium* wilt. Miss. Agric. & For. Exp. Stat. Tech. Bull. 92; 1978.
- [46] USDA. A laboratory evaluation of primitive cotton (*Gossypium hirsutum* L.) races for pink bollworm resistance. USDA. ARS W-30;1975. 9pp.
- [47] Shepherd RL. New sources of resistance to root-knot nematodes among primitive cottons. Crop Sci. 1983;23 999-1002.
- [48] Jenkins, JN. Host plant resistance: advances in cotton. In: Brown JM, Nelson TC. (eds.): proceedings of the Beltwide Cotton Conferences, 4-9 January 1986, Las Vegas, NV, USA. Memphis, TN: National Cotton Council; 1986.
- [49] Williams, MR. Cotton insect losses 2012. In: Boyd S, Huffman M, Robertson B. (eds.): proceedings of the Beltwide Cotton Conferences, 7-10 January 2013, San Antonio, TX, USA. Memphis: National Cotton Council; 2013.
- [50] Arnold MD, Gannaway JR, Sheehan MA. Screening obsolete race stocks and other wild cottons for thrips resistance at the Crops Genetic Research Facility at the Texas A&M Research and Extension Center at Lubbock, Texas. In: Boyd S, Huffman M, Richter D, Robertson B. (eds.): proceedings of the Beltwide Cotton Conferences, 9-12 January 2007, New Orleans, LA. Memphis: National Cotton Council; 2007.
- [51] Wardle R, Simpson R. The biology of Thysanoptera with reference to the cotton plant. Annals of Applied Biology 1927;14 513-528.
- [52] Watts, JG. A study of the biology of the flower thrips *Frankliniella tritici* (Fitch) with special reference to cotton. South Carolina Agricultural Experiment Station Bulletin 306. 1936; 46p.
- [53] Painter RH. Insect Resistance in Crop Plants. New York: MacMillan Co.; 1951.
- [54] Gawaad, AAA, El-Gayar FH, Soliman AS, Zaghlool OA. Studies on *Thrips tabaci* Lindeman, X. Mechanism of resistance to *Thrips tabaci* L. in cotton varieties. Zeitschrift fuer Angewandte Entomologie 1973;73 251-255.
- [55] Bowman DT, McCarty, Jr. JC. Thrips (Thysanoptera: Thripidae) Tolerance in Cotton: Sources and Heritability. Journal of Entomological Science 1997;32:460-471.
- [56] Turnbull MH, Whitehead D, Tissue DT, Schuster WS, Brown KJ, Griffin KL. Responses of leaf respiration to temperature and leaf characteristics in three deciduous tree species vary with site of water availability. Tree Physiol. 2001;21 571-578.
- [57] Chinnusamy V, Schumaker K, Zhu JK. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. J. Exp. Bot. 2004;55 225-236.
- [58] Nixon PJ, Barker M, Boehm M, de Vries R, Komenda J. FtsH-mediated repair of the photosystem II complex in response to light stress. J. Exp. Bot. 2005;56 357-363.
- [59] McMichael BL, Quisenberry JE. Genetic diversity for root-shoot relationships among cotton germplasm. J. Environ. Exptl. Bot. 1991;31 461-470.

- [60] McMichael BL, Lascano RJ. Evaluation of hydraulic lift in cotton (*Gossypium hirsutum* L.) germplasm. *J. Environ. Exptl. Bot.* 2009;68 26-30.
- [61] Ulloa M, Brubaker C, Chee P. Cotton. In: C. Kole (ed.) *Genome Mapping & Molecular Breeding*. Vol. 6. Heidelberg, Berlin, New York, Tokyo: Technical Crops, Springer; 1991. p1-49.
- [62] Ulloa M, Abdurakhmonov IY, Perez-M C, Percy R, Stewart McDJ. Genetic diversity and population structure of cotton (*Gossypium* spp.) of the New World assessed by SSR markers. *Botany* 2013;91 251-259.
- [63] White JW, Andrea-Sanchez P, Gore MA, Broson KF, Coffelt TA, Conley MM, Feldmann KA, French AN, Heun JT, Hunsaker DJ, Jenks MA, Kimball BA, Roth RL, Strand RJ, Thorp KR, Wall GW, Wang G. Field-based phenomics for plant genetics research. *Field Crop Research* 2012;133 101-112.
- [64] Burke JJ. Evaluation of source leaf responses to water-deficit stresses in cotton using a novel stress bioassay. *Plant Physiol.* 2007;42 108-121.
- [65] Burke JJ, Franks C, Burow GB, Xin Z.. Selection System for the Stay-Green Drought Tolerance Trait in Sorghum Germplasm. *Agronomy Journal*. 2010;102(4) 1118-1122.
- [66] Burke JJ, Ulloa M. Initial characterization of a Stoneville 474 x Phytogen 72 recombinant inbred population. In: Boyd S, Huffman M, Robertson B. (eds.): proceedings of the Beltwide Cotton Conferences, 7-10 January 2013, San Antonio, TX, USA. Memphis: National Cotton Council.
- [67] Tanksley SD, McCouch SR. Seed bank and molecular maps: Unlocking genetic potential from the wild. *Science* 1997;277 1063-1066.
- [68] Kohel RJ, Yu JZ. Molecular characterization of *Gossypium* germplasm for cotton improvement, p 67-75. In Engels JMM, Rao VR, Brown AHD, Jackson MT (eds), *Managing Plant Genetic Diversity*, CAB International Publishing, Inc., Wallingford, UK. 2001. 512 pp.
- [69] Abdurakhmonov IY, Buriev ZT, Shermatov SE, Abdullaev AA, Urmonov K, Kushanov F, Egamberdiev SS, Shapulatov U, Abdukarimov A, Saha S, Jenkins JN, Kohel RJ, Yu JZ, Pepper AE, Kumpatala S, Ulloa M. Genetic diversity in *Gossypium* genus. In: Çalişkan M.(ed.) *Genetic Diversity in Plants*. New York, NY: InTech; 2012. p313-338.
- [70] Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH. A detailed RFLP map of cotton, *Gossypium hirsutum* X *Gossypium barbadense*-chromosome organization and evolution in a disomic polyploid genome. *Genetics* 1994;138 829-847.
- [71] Shappley ZW, Jenkins JN, Meredith WR, McCarty JC. An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. *Theoretical and Applied Genetics* 1998;97 756-761.
- [72] Yu JZ, Kohel RJ. Cotton genome research in the United States. In: Jenkins JN, Saha S. (eds.) *Genetic improvement of cotton: Emerging technologies*. Plymouth, UK: The Science Publishers, Inc; 2001. p103-121.

- [73] Yu JZ, Fang DD, Kohel RJ, Ulloa M, Hinze LL, Percy RG, Zhang J, Chee P, Scheffler BE, Jones DC. Development of a core set of SSR markers for the characterization of *Gossypium* germplasm. *Euphytica* 2012;187 203-213.
- [74] Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape J-M, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yellambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D. CMD: A cotton microsatellite database resource for *Gossypium* genomics. *BMC Genomics* 2006;7 132. doi:10.1186/1471-2164-7-132.
- [75] Yu JZ, Kohel RJ, Fang DD, Cho J, Van Deynze A, Ulloa M, Hoffman SM, Pepper AE, Stelly DM, Jenkins JN, Saha S, Kumpatla SP, Shah MR, Hugie WV, Percy RG. A high-density SSR and SNP genetic map of the tetraploid cotton genome. *G3: Genes, Genomes, Genetics* 2012;2 43-58.
- [76] Blenda A, Fang DD, Rami JF, Garsmeur O, Luo F, Lacape JM. A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLoS ONE* 2012;7(9) e45739. doi:10.1371/journal.pone.0045739.
- [77] Fang DD, Yu JZ. Addition of four-hundred fifty-five microsatellite marker loci to the high-density *Gossypium hirsutum* TM-1 x *G. barbadense* 3-79 genetic map. *Journal of Cotton Science*. 2012;16 229-248.
- [78] Yu JZ. A standard panel of *Gossypium* genotypes established for systematic characterization of cotton microsatellite markers. *Plant Breeding News Edition* 2004;148 (1.07).
- [79] Alvarez I, Wendel JF. Cryptic interspecific introgression and genetic differentiation within *Gossypium aridum* (Malvaceae) and its relatives. *Evolution* 2006; 60:505-517.
- [80] Kantartzi SK, Ulloa M, Sacks E, Stewart JM. Assessing Genetic Diversity in *Gossypium arboreum* L. Cultivars using Genomic and EST-derived Microsatellites. *Genetica* 2009; 136:141-147.
- [81] Kuester AP, Nason JD. Microsatellite Loci for *Gossypium davidsonii* (Malvaceae) and Other D-genome, Sonoran Desert Endemic Cotton Species. *American Journal of Botany* 2012; 99 91-93.
- [82] Fang DD, Hinze LL, Percy RG, Li P, Deng D, Thyssen G. A Microsatellite-Based Genome-Wide Analysis of Genetic Diversity and Linkage Disequilibrium in Upland Cotton (*Gossypium hirsutum* L.) Cultivars from Major Cotton-Growing Countries. *Euphytica* 2013; 191 391-401.
- [83] Hinze LL, Dever JK, Percy RG. Molecular Variation Among and Within Improved Cultivars in the U.S. Cotton Germplasm Collection. *Crop Science* 2012; 52 222-230.
- [84] Kalivas A, Xanthopoulos F, Kehagia O, Tsaftaris AS. Agronomic Characterization, Genetic Diversity and Association Analysis of Cotton Cultivars using Simple Sequence Repeat Molecular Markers. *Genetics and Molecular Research* 2011; 10 208-217.

- [85] Maleia MP, Filho PSV, Goncalves-Vidigal MC, Gonela A, Lacanallo GF, Moiana LD, Chamuene A, de Sousa LL, Darben LM. Genetic Divergence Among African and American Cotton (*Gossypium hirsutum* L. race latifolium H.) Cultivars and Inbred Lines Through Random Amplification of Polymorphic DNA (RAPD) Markers. *African Journal of Biotechnology* 2010; 9 8539-8548.
- [86] Kebede H, Burow G, Dani RG, Allen RD. A-genome Cotton as a Source of Genetic Variability for Upland Cotton (*Gossypium hirsutum*). *Genetic Resources and Crop Evolution* 2007; 54 885-895.
- [87] Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev A, Kushanov FN, Salakhutdinov IB, Buriev ZT, Saha S, Scheffler B, Jenkins JN, Abdukarimov A. Molecular Diversity and Association Mapping of Fiber Quality Traits in Exotic *G. hirsutum* L. Germplasm. *Genomics* 2008; 92 478-487.
- [88] Abdurakhmonov IY, Saha S, Jenkins JN, Buriev ZT, Shermatov SE, Scheffler BE, Pepper AE, Yu JZ, Kohel RJ, Abdukarimov A. Linkage Disequilibrium Based Association Mapping of Fiber Quality Traits in *G. hirsutum* L. Variety Germplasm. *Genetica* 2009; 136 401-417.
- [89] Lacape J-M, Dessauw D, Rajab M, Noyer J-L, Hau B. Microsatellite Diversity in Tetraploid *Gossypium* Germplasm: Assembling a Highly Informative Genotyping Set of Cotton SSRs. *Molecular Breeding* 2007; 19 45-58.
- [90] Van Deynze A, Stoffel K, Lee M, Kozik A, Wilkins TA, Cantrell RG, Yu JZ, Kohel RJ, Stelly DM. Sampling nucleotide diversity in cotton. *BMC Plant Biology*. 2009. 9:125.
- [91] Byers RL, Harker DB, Yourstone SM, Maughan PJ, Udall JA. Development and mapping of SNP assays in allotetraploid cotton. *Theoretical and Applied Genetics* 2012; 124 1201-1214. doi: 10.1007/s00122-011-1780-8.
- [92] Verma S, Young CJ, Pepper AE, McCormick R, Percy RG, Yu JZ. De novo SNP discovery and development of an interspecific cotton genome map using a simplified genotyping-by-sequencing (GBS) approach. *Proceedings of the Research Conference on International Cotton Genome Initiative*. 2012. Raleigh, NC.
- [93] Hulse A, Hoegenauer K, Wang F, Stelly D, Ashrafi H, Van Deynze A, Zhang H, Sasaki C, Patterson AH, Schmutz J, Chen ZJ, Udall J, Yu JZ, Jones DC. Localization of allotetraploid *Gossypium* SNPs using physical mapping resources. *Proceedings of the Research Conference on International Cotton Genome Initiative*. 2012. Raleigh, NC.
- [94] Yu JZ, Young CJ, Verma S, Percy RG, McCormick R, Pepper AE. Development of single nucleotide polymorphism markers via sequence-based genotyping in cotton (*Gossypium* spp). In: *Proceedings of the International Conference on Plant and Animal Genomes*, 11-16 January 2013, San Diego, CA.
- [95] Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C, Li Q, Yuan Y, Lu C, Wei H, Gou C, Zheng Z, Yin Y, Zhang X, Liu K, Wang B, Song

- C, Shi N, Kohel RJ, Percy RG, Yu JZ, Zhu Y-X, Wang J, Yu S. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics*. 2012; 44(10) 1098-1104.
- [96] Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, Yoo M, Byers R, Chen W, Faigenboim AD, Duke MV, Gong L, Grimwood J, Grover C, Grupp K, Hu G, Lee T, Li J, Lin L, Liu T, Marler BS, Page JT, Roberts AW, Romanel E, Sanders WS, Szadkowski E, Tan X, Tang H, Xu C, Wang J, Wang Z, Zhang D, Zhang L, Ashrafi H, Bedon F, Bowers JE, Brubaker CL, Chee PW, Das S, Gingle AR, Haigler CH, Harker D, Hoffmann LV, Hovav R, Jones DC, Lemke C, Mansoor S, Rahman M, Rainville LN, Rambani A, Reddy UK, Rong J, Saranga Y, Scheffler BE, Scheffler JA, Stelly DM, Triplett BA, Van Deynze A, Vaslin MFS, Waghmare VN, Walford SA, Wright RJ, Zaki EA, Zhang T, Dennis ES, Mayer KFX, Peterson DG, Rokhsar DS, Wang X, Schmutz J. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 2012; 492 423–427.
- [97] Ganai MW, Durstewitz G, Polley A, Berard A, Buckler IV ES, Charcosset A, Clarke JD, Graner E, McMullen MD, Falque M. A large maize (*Zea Mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One*. 2011. 6(12):e28334. DOI: 10.1371/journal.pone.0028334.
- [98] Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945; 26 297-302.
- [99] Rohlf FJ. NTSYSpc: Numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software. Setauket, NY; 2000.
- [100] Bourland FM, Jones DC. Registration of 'UA48' cotton cultivar. *Journal of Plant Registrations* 2010; 6 15-18.
- [101] Bowman DT, Gutierrez OA, Percy RG, Calhoun DS, May OL. Pedigrees of upland and pima cotton cultivars released between 1970 and 2005. Mississippi State Univ, Mississippi State, MS Mississippi Agricultural and Forestry Experimental Station Bulletin # 1155:<http://msucares.com/pubs/bulletins/b1155.pdf>. 2007.
- [102] May OL, Bowman DT, Calhoun DS. Genetic diversity of US upland cotton cultivars released between 1980 and 1990. *Crop Science* 1995; 35 1570-1574.
- [103] Yu J, Jung S, Cheng C, Ficklin S, Lee T, Zheng P, Jones D, Percy R, Main D. CottonGen: an Integrated Genomics, Genetics and Breeding Database for the Cotton Research Community. In: Proceedings of the International Cotton Genome Initiative Research Conference, 9-12 October 2012, Raleigh, NC, United States. 2012.
- [104] Main D., Yu J, Jung S, Cheng CH, Ficklin SP, Zheng P, Lee T, Jones D, Percy R. Update on CottonGen: An Integrated Genomics, Genetics and Breeding Database for the Cotton Research Community. In: Proceedings of the International Plant and Animal Conference, 12 January 2012, San Diego, CA, United States. 2012.

- [105] Wiersema, J. H. Taxonomic information on cultivated plants in the USDA/ARS germplasm resources information network (GRIN). In: Tukey Jr, HB. (ed). *Acta Horticulturae* 413. International Society for Horticultural Science. Proceedings of the Second International Symposium on the Taxonomy of Cultivated Plants, 10-14 August 1994, Seattle, Washington, USA. Leuven, Belgium; 1995. ISBN 90 6605 967 2. Individual papers from this Acta are available to download from <http://www.actahort.org/books/413/>
- [106] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 2000; 28(1) 27-30.
- [107] Harris M, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* 2004; 32 D258–D261.

