



# Core Collections: Is There Any Value for Cotton Breeding?

Lucy Marie Egan\*, Warren Charles Conaty and Warwick Nigel Stiller

CSIRO Agriculture and Food, Narrabri, NSW, Australia

## OPEN ACCESS

### Edited by:

Linghe Zeng,  
United States Department of  
Agriculture (USDA), United States

### Reviewed by:

Jane Dever,  
Texas A&M AgriLife Research,  
United States  
Jack McCarty,  
USDA/ARS, United States  
Joshua A. Udall,  
United States Department of  
Agriculture (USDA), United States

### \*Correspondence:

Lucy Marie Egan  
lucy.egan@csiro.au

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

Received: 13 March 2022

Accepted: 06 April 2022

Published: 28 April 2022

### Citation:

Egan LM, Conaty WC and  
Stiller WN (2022) Core Collections: Is  
There Any Value for Cotton  
Breeding?  
Front. Plant Sci. 13:895155.  
doi: 10.3389/fpls.2022.895155

Global plant breeding activities are reliant on the available genetic variation held in extant varieties and germplasm collections. Throughout the mid- to late 1900s, germplasm collecting efforts were prioritized for breeding programs to archive precious material before it disappeared and led to the development of the numerous large germplasm resources now available in different countries. In recent decades, however, the maintenance and particularly the expansion of these germplasm resources have come under threat, and there has been a significant decline in investment in further collecting expeditions, an increase in global biosecurity restrictions, and restrictions placed on the open exchange of some commercial germplasm between breeders. The large size of most genebank collections, as well as constraints surrounding the availability and reliability of accurate germplasm passport data and physical or genetic characterization of the accessions in collections, limits germplasm utilization by plant breeders. To overcome these constraints, core collections, defined as a representative subset of the total germplasm collection, have gained popularity. Core collections aim to increase germplasm utilization by containing highly characterized germplasm that attempts to capture the majority of the variation in a whole collection. With the recent availability of many new genetic tools, the potential to unlock the value of these resources can now be realized. The Commonwealth Scientific and Industrial Research Organisation (CSIRO) cotton breeding program supplies 100% of the cotton cultivars grown in Australia. The program is reliant on the use of plant genetic resources for the development of improved cotton varieties to address emerging challenges in pest and disease resistance as well as the global changes occurring in the climate. Currently, the CSIRO germplasm collection is actively maintained but underutilized by plant breeders. This review presents an overview of the Australian cotton germplasm resources and discusses the appropriateness of a core collection for cotton breeding programs.

**Keywords:** core collection, germplasm management, cotton, *Gossypium*, breeding

## INTRODUCTION

By 2050, the world population is estimated to reach 9.1 billion people, and an increase of 70% in agricultural production is needed to feed and clothe the world (FAO, 2009; Tester and Langridge, 2010; Llewellyn, 2018). Land and water availability, climate change, and evolving pest and disease virulence are significant factors that add pressure to already intensive agricultural production systems.

Plant breeding activities will be a critical avenue in addressing these challenges. The selection of traits for future climates such as abiotic and biotic stress tolerance, water use efficiency, and adaptation to new farming environments and systems could develop climate-resilient plant varieties. Target traits could be introduced from the naturally occurring diversity within a crop species, by utilizing the secondary and tertiary gene pools (Maxted et al., 2016) or through induced variation (Holme et al., 2019).

Plant genetic resources (PGR) constitute the foundation of sustainable agriculture and global food security and stability (Bernard et al., 2020). While the need for conservation and use is apparent, the lack of funding is severely impacting conservation efforts across the globe (Anđelković et al., 2020). Inadequate evaluation and characterization of germplasm held in genebanks are key challenges contributing to the underutilization of material and negatively impact future opportunities and funding efforts (Bernard et al., 2020). PGRs are the link between agriculture, environment and trade, so conservation efforts require cooperation from different sectors (Anđelković et al., 2020).

Since the 1970s, the conservation of PGRs has become a large-scale independent activity detached from crop improvement efforts. The germplasm held in genebanks is utilized by plant breeding organizations, often globally, and breeding programs are reliant on the genetic diversity that is available (Rajasekharan, 2015). Elite crop cultivars are generally bred from a narrow genetic base and targeted to high-input-intensive agricultural production, decreasing their ability to adapt to changes in the environment. Efforts need to be increased and directed to utilize germplasm from the genebanks for future climate-resilient breeding targets.

Globally, approximately 7.4 million accessions are preserved in over 1750 genebanks (Upahyaya, 2015), with fiber crops constituting ~104,000 accessions in the collection (FAO, 2010). Fiber crops hold importance in terms of raw materials for the textile industry as well as other industries including building materials, cosmetics, medicine, and biopolymers. Total fiber production is predicted to increase from 50 million tonnes/year in 1999 to 130 million tonnes/year by 2050 (European Commission, 2015). Fiber crops are dominated by cotton (European Commission, 2015), and Brazil, China, India, Russia, the United States of America (USA), and Uzbekistan are the countries that contribute the greatest to cotton production. Although Australia only produces around 3% of the world's cotton, it is a significant player in the global trade of cotton, usually ranking as the third or fourth largest exporter (Khan et al., 2020; Cotton Australia, 2022). Australian cultivars are generally superior to those from many other countries in both seed and fiber qualities (Kilby et al., 2013; Gapare et al., 2017). Approximately, 1% of the global cotton accessions are held in Australian germplasm stocks (Campbell et al., 2010; Abdurakhmonov, 2014).

Cultivated cotton encompasses four species from the cotton genus, *Gossypium*; *G. arboreum* L. (Desi cotton), *G. herbaceum* L. (Levant or Arabian cotton), *G. barbadense* L. (Pima, Egyptian

or Sea Island cotton), and *G. hirsutum* L. (Upland cotton; Wendel et al., 2009; Constable et al., 2015). Globally, *G. hirsutum* is the dominant species used in cotton production systems, making up around 95% of global cotton production (May and Lege, 1999; Zhang et al., 2008). However, the varieties released represent only a small portion of the total variation in the cotton gene pool (Wendel et al., 1992; Shim et al., 2018). This is largely due to the negative effects of linkage drag, deleterious alleles, and the breakdown of co-adapted gene complexes that can be introduced from wild germplasm (Acquaah, 2009) and breeders on the whole shy away from incorporating very unadapted germplasm into their breeding programs except in dire need.

Cotton breeding activities around the world have developed varieties with adaptation to new environments and modern management systems. Despite the past success, the full potential of future Australian varieties is constrained by imminent adverse climates influenced by climate change, changing biotic and abiotic stresses, and legislation and commercial developments that constrains resources used in cotton production (e.g., water, and fertilizers). The continued development of productive cultivars will rely heavily on exploiting the available variation present in conserved *Gossypium* species germplasm.

This paper aims to build on previous reviews on the purpose and relative benefits and challenges of germplasm collections (Fu, 2017; Díez et al., 2018; Walters et al., 2018; Wambugu et al., 2018; Westengen et al., 2018; Gotor et al., 2019; Hanson and Ellis, 2020). The outcome of the review will allow researchers to make an informed decision on the applicability of core collections for their situation. Such informed decisions are vital for plant breeding efforts, and germplasm collections as an effective core collection can provide a workable solution that ensures trait diversity can be harnessed in cultivar development programs.

## CORE COLLECTIONS

A plant germplasm collection is defined as a collection of seeds or breeding material, typically to conserve genetic diversity, but to also act as a source of material to address future challenges in agriculture (van Hintum et al., 2000). Germplasm banks generally hold a wide range of targeted cultivated species and wild relatives. Globally, the key germplasm resources for cotton are based in China (Jia et al., 2014), India (Narayanan et al., 2014), Pakistan (Rahmat et al., 2014), France (Dessauw et al., 2004), Uzbekistan (Sanamyan et al., 2014), Australia (Stiller and Wilson, 2014), and the United States (Percy et al., 2014; Zeng, 2014).

Although genebanks around the world have had great successes in collecting germplasm to conserve crop genetic diversity, the scale of success is often limited by inefficient data and germplasm management. The germplasm collections often become so large that maintenance of the conserved germplasm becomes difficult both from lack of resources and accumulated human errors in processing and handling, and potential contamination during seed regeneration cycles

**Abbreviations:** ATCF, Australian Tropical Crops and Forages Genetic Resource Centre; CSIRO, Commonwealth Scientific and Industrial Research Organisation; CBD, Convention on Biological Diversity; MSE, Managed Stress Environment; PGR, Plant Genetic Resources; USA, United States of America.

(Chen et al., 2016). In addition, when the data associated with germplasm lacks detail and the germplasm is under-characterized and under-evaluated, the germplasm is generally underutilized by breeders (Egan et al., 2019a,b, 2020). The ever-increasing gain that is needed from modern cultivars across all major crops suggests that the genetic variation available through germplasm banks will be even more critical for future breeding activities.

One method to tackle the management of large germplasm collections was suggested by Frankel (1984), through the development of a core collection. It was proposed that a pruned down collection would increase the management efficiency of the accessions and be more attractive to breeders. A core collection is defined as, “a limited set of accessions representing, with a minimum of repetitiveness, the genetic diversity of a crop species and its wild relatives.” The accessions that did not make it into the collection would not be discarded but would be managed in a “reserve collection.” The accessions chosen for the core collection should be genetically, ecologically, and geographically distinct from each other to maximize the diversity represented in the collection (Brown, 1995).

From the definition by Frankel (1984), several operational definitions have followed. For an individual genebank, van Hintum et al. (2000) proposed a core collection “consists of a limited number of accessions from an existing collection, chosen to represent the genetic spectrum in the whole collection”; this is termed the core subset of the collection, or is commonly referred to as a diversity panel. In comparison, a species-specific collection is made up of entries chosen to represent the majority of the genetic diversity of the target species and its wild relatives. It is a comprehensive collection and is often multi-organizational and international. Core collections now exist in chickpea, mungbean (Schafleitner et al., 2015), peanut (Holbrook et al., 1993; Jiang et al., 2014), apple (Liang et al., 2015), safflower (Dwivedi et al., 2005), cowpea (Mahalakshmi et al., 2007), pearl millet (Bhattacharjee et al., 2007), Australian bermudagrass (Jewell et al., 2012), common bean (Kuzay et al., 2020), annual Medicago (Diwan et al., 1994), capsicum (Zewdie et al., 2004), and eggplant (Miyatake et al., 2019). The most notable core collection is the International Barley Core Collection of some 1,600 accessions from around the world (van Hintum et al., 2000).

The development of “subset collections” and mini-core collections has also increased in recent years in response to germplasm management resource limitations (Pande et al., 2006; Sharma et al., 2010, 2012a,b). A subset collection is defined as a core collection for a specific breeding target and would be appropriate for an emergency situation, i.e., a disease decimates a crop, and a source of resistance is urgently needed to initiate a new breeding activity (Mahuku et al., 2003; Silvar et al., 2010). The concept of a mini-core collection is often attractive as the core collections are frequently still too large for the available resources to deeply characterize all of the accessions. Whilst a core collection is generally in the order of 10% of the size of the whole germplasm collection, a mini-core collection aims to subsample the core collection to develop a smaller collection (10% of a core collection) that captures most of the beneficial variation within the crop (Sharma et al., 2012a;

Upadhyaya et al., 2013). Mini-core collections have been developed for chickpea (Upadhyaya and Ortiz, 2001; Pande et al., 2006), sorghum (Sharma et al., 2010), rice (Li et al., 2010), durum wheat (Etminan et al., 2017), mungbean (Sokolkova et al., 2020), cowpea (Fatokun et al., 2018), flax (Diederichsen et al., 2013), and finger millet (Upadhyaya et al., 2010).

In cultivated cotton, there are currently few core collections that exist globally (Xu et al., 2006; Mei et al., 2012; Wang et al., 2013; Ma et al., 2018). However, in Australia, there are as yet no core or mini-core collections for cotton.

## CURRENT GERmplasm UTILIZATION WITHIN AUSTRALIAN COTTON GERmplasm COLLECTIONS

Historically, Australian plant breeders across a range of crops acquired and maintained their own germplasm collections. However, in the early 1980s, the Commonwealth and State governments established a series of genetic resource centers to conserve national germplasm collections. The Australian Tropical Crops and Forages Genetic Resource Centre (ATCF) based in Queensland was the center that included cotton, although that collection has now been transferred to the Australian Grains Genebank in Horsham, Victoria, as part of a national consolidation effort. Currently, the only other two cotton-focused collections in Australia reside at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food in Narrabri, New South Wales and Canberra, Australian Capital Territory.

The CSIRO main germplasm collection at Narrabri is maintained and funded by the active industry breeding program and includes about 2000 accessions. The CSIRO collection in Canberra is largely dedicated to the long-term cold storage of indigenous Australian species that have been collected through germplasm explorations over the last few decades, although nearly all are also entered in the US National Cotton Germplasm Collection. The majority of accessions held in Narrabri are cultivated tetraploid cottons, predominantly cultivars and germplasm stocks of *G. hirsutum* and some *G. barbadense*, with a small number of undomesticated or landrace stocks. The next largest group of accessions belong to the cultivated diploid cottons, with the majority being *G. arboreum* and a smaller number of *G. herbaceum*. A significant number of accessions of the Australian species are also held, particularly *G. australe* L., *G. sturtianum* L., *G. nelsonii* L., and *G. bickii* L. However, these are the result of individual germplasm explorations, so it is unknown what genetic variability exists between the accessions. Small numbers of accessions of most of the other *Gossypium* species are also held (Stiller and Wilson, 2014).

The majority of the successes in utilizing germplasm resources globally have been through the identification of important traits within the cultivated tetraploid species, the primary germplasm pool. From a breeding perspective, it is relatively straightforward to transfer traits from any of the main tetraploid species to *G. hirsutum*, the target of most commercial cotton breeding programs. Significant improvements in fiber quality have been

realized through the introgression of traits from *G. barbadense* into *G. hirsutum* (Niles and Feaster, 1984; Meredith, 1991). In Australia, the most significant successes have been around improvements in disease resistance, particularly against Bacterial Blight, Fusarium Wilt, Verticillium Wilt, and Cotton Bunchy Top, mostly derived from other *G. hirsutum* cultivars.

The cultivated diploid cotton species, the secondary germplasm pool, are also of interest to cotton breeders as a source of novel traits. However, the challenges associated with ploidy differences make successful utilization difficult and a long-term exercise (often more than 20 years). Over the previous few decades, there have been numerous reports of attempts to transfer useful traits from these species (Stewart and Hsu, 1978; Brubaker et al., 1999; Liu et al., 2015a), but documented successes in terms of commercial cultivar releases are limited. However, one prominent example is the introgression of genes from *G. arboreum* and *G. thurberi* into Upland cotton to improve fiber strength (Culp and Harrell, 1974). In more recent times, the focus has shifted to pest and disease resistance and water and heat tolerance traits from the secondary gene pool (Constable, 1998; Singh et al., 2007; Azhar et al., 2009; Cottee et al., 2010, 2014; Trapero et al., 2016; Li et al., 2020; Wilson et al., 2021). This is also relevant to the Australian context where significant resources have been focused on using diploid species to improve resistance to Fusarium wilt and Verticillium wilt, as well as pests such as mites and whitefly.

The tertiary gene pool, including the many indigenous Australian *Gossypium* species, has also held great interest. The characterization of these diploid species has been reasonably well documented (Brown and Brubaker, 2000; Brubaker and Brown, 2003; Becerra Lopez-Lavalle et al., 2011) along with techniques for the development of fertile triple hybrids (Brubaker et al., 1999). Despite extensive research, there are no documented cases of the successful development of a cultivar utilizing these tertiary gene pool resources mainly due to the lack of any recombination of the diploid derived chromosomes with those of tetraploid cotton. However, considerable interest in this area remains and the continued development of advanced breeding techniques such as gene editing may eventually facilitate success in this area.

Historically, genuine germplasm exchange was practiced widely across breeding programs to support genetic gain on a global scale. However, cultivar development in many major field crops, especially corn, soybean, and cotton, is now heavily dominated by multinational corporations who generally adopt the approach of restricting access to their germplasm through various intellectual property protection strategies such as patents and exclusive licensing arrangements. This approach is understandable and has the potential to raise investment in cultivar development as realized economic returns can be reinvested back into research. The downside is an overall reduction in the movement of germplasm between programs and geographies, resulting in some genetic gains not being available to the global breeding community. As many cotton cultivars are only released with genetically modified traits, the germplasm exchange procedure is further complicated. The restricted movement of germplasm may also result in an overall

slowing of genetic gains over time if individual programs become “isolated” and their overall genetic variability is reduced. Thus, it is an imperative that germplasm exchange continues. While there is no simple solution to this, one approach is for breeding programs to partner with these corporate programs or license in germplasm from other programs.

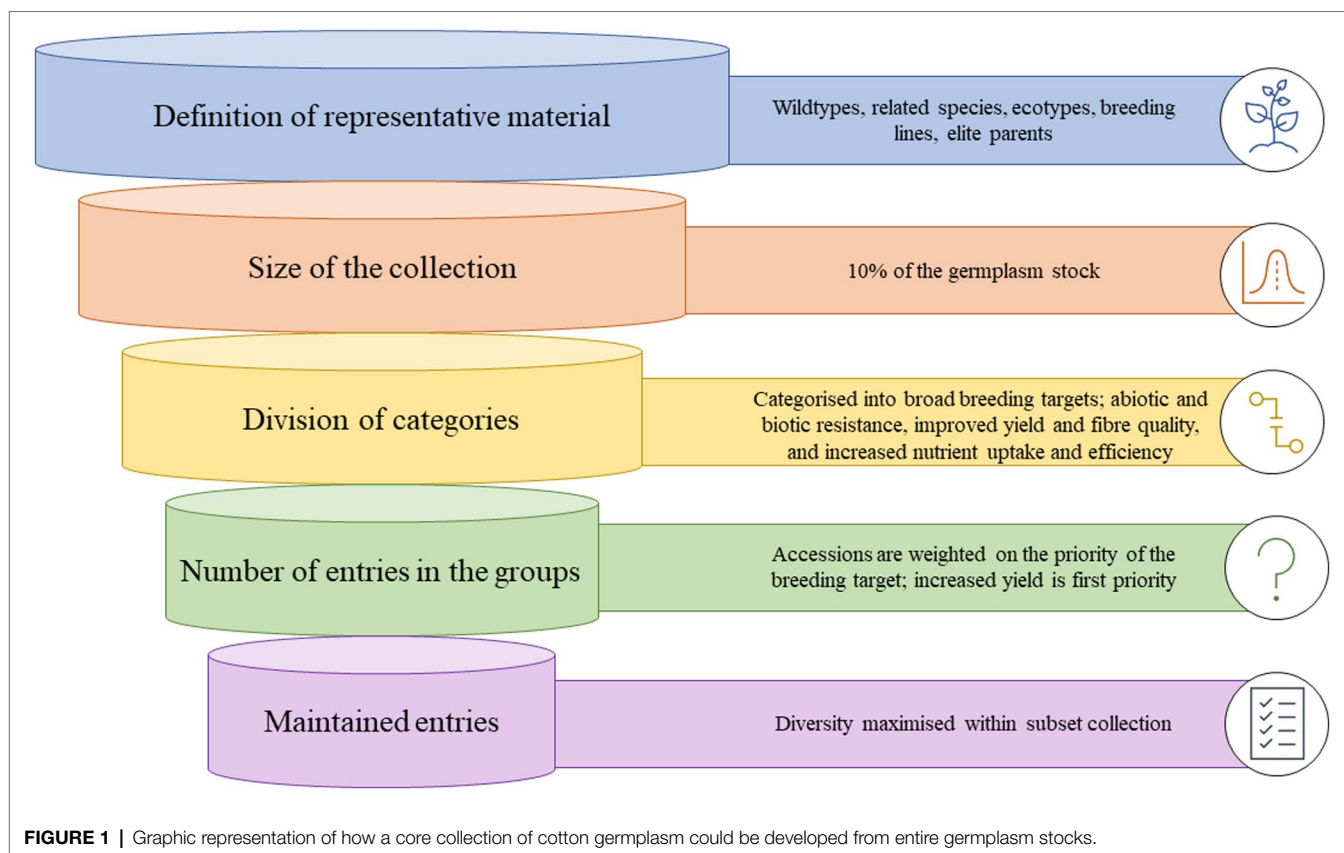
Another potential impediment to the utilization and expansion of germplasm collections as well as germplasm exchange between institutions is the Nagoya Protocol, part of the United Nations Convention on Biological Diversity (CBD). The Protocol is a multilateral international treaty with the purpose of ensuring “fair and equitable sharing of benefits arising out of the utilization of genetic resources,” which covers all organisms, including wild accessions of cotton and its relatives, wherever they are found. Researchers collecting new germplasm from the wild will have to enter into “access and benefit sharing” arrangements with the host country prior to their collection expeditions. These arrangements set out who might profit, and how, from the organisms being used, and stipulate how to distribute the benefits fairly, for example, through co-authorship of publications, or more importantly sharing of profits from products such as newly bred crops or new medicines. Although Australia signed and ratified the CBD in 1993, it is not a signatory to the Nagoya Protocol which was added in 2010 as it already had existing National and State and Territory laws consistent with the Protocol that regulate benefit sharing from any new collections of Australian indigenous biota, including *Gossypium* species. There is still some uncertainty around the interpretation of the Nagoya Protocol in relation to historical collections assembled from before the CBD (Sherman and Henry, 2020), so going forward some care will be needed by breeders in using wild germplasm in breeding, regardless of its origins, unless they can demonstrate Nagoya compliance or at least compliance with local national laws on access and benefit sharing.

## THE DEVELOPMENT OF A CORE COLLECTION

Since the proposal of a core collection, there have been many techniques and sampling methodologies developed. Creating a core collection can be simple or complex and is heavily influenced by resource allocation and information available. Therefore, often a preliminary core collection is developed, that is larger than intended, and then characterized at a deeper level and subset. In general, five generic steps are followed (Figure 1; van Hintum et al., 2000).

**Step 1: Definition of the material that should be represented;** this is termed the domain of the core collection and can range from related species, wild-types, ecotypes, landraces, previous breeding lines or a specific focus area of the diversity of a species. The definition is subjective to the available material, what constitutes a suitable set of material for core development, and the objectives behind the establishment of a core collection.

**Step 2: Decide on the size of the core collection.** Generally, most core collections are 5–20% of the size of the collection from which they were derived, although can be much smaller



if the main collection is very large. A sampling strategy should define both a sampling method and an allocation method, and various sampling strategies and methods have been utilized to develop core collections (Franco-Duran et al., 2019). Based on the neutral allele model, a core collection of 10% of the total collection is likely to contain at least 70% of the variation of the collection, with an ideal collection size of no more than 3,000 accessions (Brown, 1989a,b).

**Step 3: Division of the domain into categories.** This is a methodical process and begins by dividing the material based on broad themes and further characterizing the groups. At every division, the groups should be constructed to maximize variation between the groups.

**Step 4: Decide on the number of entries in the groups of the core collection.** The number of entries of the core collection is heavily influenced by the aim of the core collection, i.e., the breeding targets. The allocation of germplasm across the defined groups is critical to maintaining diversity, and three generic steps have been proposed to determine the number of entries per group: (1) allocate entries in some proportion according to the relative numbers of accessions that occur within each group; (2) if enough material has been characterized with genetic markers, comparisons of marker diversity within the groups and basing the allocation of numbers based on maximizing allelic diversity or richness within each group; (3) subjectively consider the breeders needs and other informal knowledge about the accessions within the groups, i.e., passport data, availability to the collection curator.

**Step 5: Choose the specific entries that will be maintained in the core.** The final process of choosing which actual accessions should be included in the groups, maximize diversity, and serve the purpose of the core collection. However, breeders also need to consider the quality of the documentation of the entries, seed availability, and the role of the accessions in their breeding programs (van Hintum et al., 2000).

## COTTON CORE COLLECTIONS AS A RESOURCE TO DISCOVER NEW SOURCES OF VARIATION FOR BIOTIC AND ABIOTIC STRESS TOLERANCE

Plants are sessile organisms and thus are exposed to environmental conditions and biotic stressors that are detrimental to their growth and development. Improving germplasm to meet abiotic and biotic breeding targets is a long-term process and can involve extended time and resources for successful introgression into elite material. However, there are many examples where plant breeders have effectively utilized well-characterized germplasm from a germplasm center or core collection to introgress tolerance/resistance to abiotic and biotic stresses in crop species (Upadhyaya et al., 2011; Prasanna, 2012; Shivhare and Lata, 2017; Singh et al., 2018; van Zonneveld et al., 2020; Sharma et al., 2021). Climate-proofing future cotton

varieties will heavily involve the incorporation of traits and sources of resistance to biotic and abiotic stresses.

## Biotic Resistance

One of the most important roles for a core collection is when a new disease, or novel variant, threatens the industry and necessitates the rapid establishment of a breeding project to confer host plant resistance. However, biotic stresses are complex due to the high interaction of the pest or disease with the plant, the environment, and evolving pest and disease virulence. Currently, the diseases of Fusarium wilt, Black Root Rot, and Verticillium wilt, and the increase in secondary pests with the introduction of genetically modified cotton resistant to the major Lepidopteran pests are the greatest biotic threats to the Australian cotton industry. A good example of the utilization of germplasm to address a rising disease threat is our breeding for Fusarium wilt resistance in cotton.

Fusarium wilt is caused by the pathogen *Fusarium oxysporum* f sp. *vasinfectum* and was first recognized in Australia in 1993 (Kochman, 1995) when most of the commercially grown cultivars were highly susceptible. Production losses of up to 100% were reported from fields with high levels of inoculum. It is thought that the indigenous and genetically unique Australian Fusarium wilt (Wang et al., 2004) pathogen became prominent in response to wide-scale planting of highly susceptible cotton cultivars and then spread to nearly all cotton-growing regions via infected soil, flood waters, or on cotton trash. An initial screen from the main CSIRO germplasm collection of *G. hirsutum* genotypes from the United States, South America, Africa, Asia, and Europe revealed that most were susceptible, but a few genotypes showed field survival up to 50% higher than the best extant Australian cultivars. Over the next years, over 200 genotypes from the collection were targeted for evaluation based on material from the countries/regions/programs that showed promise in the early screening. One line from India MCU-5 (Lopez-Lavalle et al., 2012) possessed significantly higher levels of resistance and became the initial source of resistance in a breeding process with moderately resistant local elite varieties. The introgression of MCU-5 germplasm ultimately resulted in the release of Sicot F-1 in 2004, which had at least twice the resistance to Fusarium wilt of the best cultivar in 1994 (Allen et al., 2004).

Screening our germplasm collection has also identified genotypes for breeding purposes with increased resistance to Cotton Bunchy Top disease (Ellis et al., 2016) and Bacterial Blight (Rungis et al., 2002), and more recently Verticillium wilt.

## Abiotic Resistance

The major abiotic stressors affecting the Australian cotton industry include temperature extremes, water availability, salinity/sodicity, mineral toxicity, and UV radiation. Ultimately all abiotic stressors adversely alter morphological, biochemical, and physiological plant mechanisms, resulting in reduced plant growth and productivity, preventing crops from reaching their genetic yield potential. Water stress is the major environmental factor influencing crop productivity, where approximately 40% of the global land area is affected by drought (Trenberth et al.,

2014). Thus, significant research has been dedicated to the development of drought-tolerant cotton germplasm and due to its complexity is an interesting case study for the development of core collections.

Like most core collections, passport data (such as the climate of origin and/or target environments of breeding lines) provide a good basis for the initial development of a drought-tolerant cotton mini-core collection. In addition, landrace accessions and cotton wild relatives are essential for the inclusion as they may have drought traits that have been lost from cultivated varieties due to selection pressure for yield potential under optimal conditions (Dempewolf et al., 2017). The yield performance of germplasm under controlled field conditions provides additional confidence in the development of a drought-focused core collection. Under water-limited conditions, the genetic potential of a genotype may not be expressed, which limits the ability to resolve statistical genotype differences. This is particularly important in variable rainfall environments and can be mitigated through the implementation of a managed stress environment (MSE) protocol. An MSE protocol ensures water is applied to “rainfed” germplasm trials when crop yield is likely to fall below the threshold for resolving genetic differences with confidence (Conaty et al., 2018). An MSE protocol has been used to identify germplasm for our drought tolerance research using a germplasm panel that could be utilized as a mini-core collection. In addition, germplasm evaluations are conducted in multiple rainfed and irrigated environments, allowing for paired comparisons and the development of drought resistance indices used in selection, e.g., Fischer and Maurer (1978) and Mwadzingeni et al. (2017).

Several institutions have developed core collections in other crop species with increased tolerance to abiotic stresses: chickpea (Upadhyaya et al., 2013), sorghum (Upadhyaya et al., 2009), barley (Muñoz-Amatrián et al., 2014), bermudagrass (Anderson, 2005), and peanut (Upadhyaya et al., 2014). These activities have had promising results and should encourage similar efforts in cotton.

## Variation in the Cotton Genome

Before the advent of molecular genetics, plants were assessed on their diversity in morphological and phenotypic traits as well as their pedigree and geographical distribution. Viewed molecularly, cotton diversity is the variation of the genes within each species. Recent advances in technology have resulted in ever-decreasing costs of high-throughput sequencing that has enabled the assembly of allopolyploid cotton reference genome sequences in a restricted set of “standard” genetic lines (Li et al., 2015; Liu et al., 2015b). These standard lines provide only a snapshot of cotton genome structure and so it is only recently with the rise of third-generation sequencing (long-read sequencing), that the true diversity within each species can be viewed through the completion of pan-genome sequences (Bayer et al., 2020). Pan-genomes represent the genomic diversity of a species that includes core genes, found in all individuals, as well as variable genes, which are present-absent in some individuals and are likely the result of local adaptations. A simple comparison of two *G. hirsutum* cultivars TM-1 (the genetic standard for

*G. hirsutum*) and zhongmiansuo24, a commercial cultivar (Yang et al., 2019) revealed extensive gene order and gene structural diversity, and in more extensive comparisons in other crop species, there has been large variable gene content (15–40% gene presence and absence variation) between individuals. Although pan-genomes are still in their infancy for many crop species, a pan-genome approach should enable an informed choice for the selection of core collection accessions based on both gene sequence and gene presence/absence diversity between individuals or accessions. For disease-resistant plants, for example, NOD-like receptor (including disease resistance like genes) pan-genome diversity (Barragan and Weigel, 2020) could be used to prioritize lines for a core collection. Molecular surveys of diversity, therefore, could enable the selection of core lines that encompass species diversity without the requirement for extensive phenotype analysis.

## THE CHALLENGES OF DEVELOPING A CORE COLLECTION

Although core collections have the potential to address the underutilization of germplasm that is a frequent challenge for genebanks, there are concerns from breeders about certain aspects of the development stage. Three key areas highlighted by Brown and Spillane (1999) are the evaluation criteria used, completeness of data, and the disposal of germplasm not in the core collection. First, the choice of traits to be incorporated into the core collection needs to be flexible. Many breeders approach genebank curators with specific traits in mind that they want to incorporate into their program, and if the core collection does not contain the desired traits then germplasm utilization will remain stagnant (Egan et al., 2019a). The collection also needs to have the ability to pivot direction if new breeding targets and trends emerge (Brown and Spillane, 1999). Second, the core collection may lack crucial accessions if it was formed using incomplete information (Brown, 1995). If the collection was developed using misleading or incomplete information, then it could lead to a false assumption that the core collection is a broader sample than it actually is (Brown and Spillane, 1999). Finally, there are concerns about the potential disposal of accessions that are not included in the core. The intention of a core collection is not to replace the main collection but to increase its utilization, so curators should make every effort to maintain underutilized accessions in longer-term storage as part of the main collection, perhaps with more restricted access by users, or to transfer them to other institutions with greater resources rather than dispose of any accessions due to resource constraints (Brown, 1989b; Brown and Spillane, 1999). Therefore, there is a stronger case to develop a core collection to utilize some of the germplasm rather than have none of it used at all. Overall, when a germplasm center decides to develop a core collection, the risk to the whole collection is low, at least for the major crops (Hamon et al., 1995).

While the concept of identifying germplasm that is resistant to abiotic or biotic stressors to include in the core collection is hopeful, there are significant challenges associated with

identifying said germplasm. The two major challenges are: (1) identifying a source of resistance to the stress and (2) developing an effective screening method to characterize the level of resistance. The development of a core collection for resistance to an abiotic or biotic stress is limited by the variation available in germplasm. In some circumstances, particularly for biotic stresses, there may not be any resistant germplasm in either the main or core collection and the pest or pathogen will be better controlled by management strategies. Therefore, a wide screen of the germplasm available using an appropriate phenotyping method is critical before deciding if a core collection is an effective solution. Once a source of resistance is identified, a high-throughput screening assay that is reflective of field resistance needs to be developed. However, this is often a lengthy process and is complicated by several factors. First, the susceptibility of plants to the stress varies with respect to timing, duration, and intensity and often under field conditions, plants encounter specific abiotic and biotic stresses in combination with other stressors. Second, controlled environment and pot-based experiments are generally not well correlated with field performance. Finally, resistance to abiotic and biotic stressors is often polygenic, including a complex of gene networks involved in stress sensing, signal transduction, and expression of stress-responsive genes. Therefore, the development of core collections to assess a wide range of germplasm rapidly and effectively for abiotic and biotic stress resistance can be an ambitious goal.

While core collections for drought tolerance have been assembled, gains in limited water and rainfed cotton productivity have largely been incremental and reflective of improvements in yield potential under non-stressed conditions. Thus, future genetic gain in abiotic stress resistance will require a combination of traditional plant breeding and new breeding methods such as genomic selection, as well as the integration of panomics (Weckwerth et al., 2020), novel field-based phenomics (White et al., 2012; Zhao et al., 2019), and the application of machine learning and artificial intelligence to breeding for complex plant traits (Niazian and Niedbała, 2020; Nabwire et al., 2021). However, these methods are currently prohibitively expensive and/or will require a core collection of diverse germplasm to efficiently assess whether they will be effective in cotton breeding.

The major bottleneck for the development of core collections remains the required level of investment. Developing and maintaining a germplasm collection is time-consuming and expensive, and there is a clear reluctance to investment due to the large cost of germplasm maintenance and the continual characterization of germplasm for numerous traits. Adding to the expense is the high likelihood that novel resistance traits will often lie in related cotton species, rather than older varieties or landraces, and characterizing and introgressing resistance from those more distant sources is a lengthy and costly process. To encourage funds to be allocated to germplasm centers, the development of a core collection must show a clear return on investment and have an applied and impactful outcome to commercial breeding programs.

## WILL CORE COLLECTIONS PROVIDE BENEFIT TO AUSTRALIAN COTTON BREEDING ACTIVITIES?

The challenges facing the Australian cotton industry include increasing production costs, pests, and diseases, and new environmental conditions due to climate change. While these challenges are both numerous and diverse, a core collection would only be useful to the CSIRO breeding program if it was a single collection that contained deeply characterized accessions for multiple traits. This is because the fundamental focus of the breeding program is the production of commercial cotton varieties, not germplasm banking. Thus, the investment required for the core collection must match the desired outcome and provide impact in terms of cultivar development. However, it should be acknowledged that the approach used for trait screening and ultimately cultivar development over the last few decades by CSIRO has utilized the central principles involved in developing a core collection. The screening and characterization of approximately 200 diverse accessions for *Fusarium* wilt resistance (see “Biotic Resistance,” above) are a good example of how the principles of developing a core collection can be employed in breeding.

Although yield will always be the main priority for the CSIRO cotton breeding program, the core collection would not directly target high-yielding accessions. Rather, the traits targeted would include those with less complex genetic backgrounds, such as biotic stress resistance and the fiber quality parameters of length, strength, and elongation. It is an ambitious goal that abiotic stress tolerance would be included as a characterization parameter due to the complexities in phenotyping for that type of tolerance. It is probable that the categories across the core collection would contain an unbalanced number of accessions as target traits may only be present in low frequencies throughout our entire germplasm collection. Conversely, for traits that are common, the highest performing accessions would be selected for inclusion in the core collection. Importantly, adequate representation from related cotton species must be included, historically sources of host plant resistance to biotic stressors have been identified in landrace and crop wild relatives in the CSIRO cotton breeding program (Allen et al., 2004; Miyazaki et al., 2012; Wilson et al., 2021). We conclude that CSIRO should pursue the development of a core collection through a greater characterization of our main collection at both a molecular and phenotypic level and augmenting it where necessary with new material from global collections to fill gaps and ensure a high level of genetic diversity across several traits types of interest, but of a size that we can economically maintain through regular regeneration of our stocks.

## CONCLUSION

Conserving germplasm is, in the long term, the most efficient and inexpensive method of genetic conservation for agriculturally important plants. However, the conservation of germplasm quickly becomes unmanageable, and accessions will be underutilized if a cost-effective management system is not implemented. Core collections are one strategy to increase the utilization of germplasm

and have moved from a period of experimentation to one of increasing popularity and endorsement.

The development of core collections presents great opportunities for cotton and provides a resource use-efficient strategy enabling the identification of the most accessible source for a given trait (Brown, 1989b). Core collections that represent the majority of the diversity available in cotton would demonstrate significant value for germplasm conservation and breeding activities. Future-proofing the global cotton industry will rely heavily on the introgression of germplasm for novel traits, particularly focused on resistance to abiotic and biotic stresses.

Nevertheless, the significant challenges of developing a core collection must be clearly acknowledged. To be of real value, the core collection should target traits specific to that required by the industry that a breeding program is supporting. This will be different for different countries, regions, and production systems. Significant investment from the program would need to be allocated to develop and maintain the core collection. Therefore, the target traits must be a continuing important breeding goal to the relevant industry. Due to the vastly different cotton production systems across the globe, it is unlikely that a single global cotton core collection would provide merit to all breeding programs; rather, a country- and trait-specific core collection model would be more appropriate.

Finally, to fully exploit the potential that core collections can bring to a breeding program, coordinated efforts between the private and public sectors are required. The goal of these efforts is to ensure that meaningful germplasm access and exchange can occur, resulting in core collections that better represent the diversity of the focus traits of that specific core collection. However, it must be acknowledged that germplasm protection is fundamental for all commercial breeding programs. Therefore, access to the full range of germplasm is unlikely. This can in part be mitigated through commercial agreements providing access to recently released cultivars, and private sector breeding programs releasing commercially obsolete germplasm to publicly available germplasm libraries.

## AUTHOR CONTRIBUTIONS

LE conceived and drafted the review. LE, WC, and WS designed and wrote the review. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by Cotton Breeding Australia, a joint venture between Cotton Seed Distributors Ltd. Australia and CSIRO Agriculture and Food.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of Dr. Iain Wilson and Dr. Danny Llewellyn.



## REFERENCES

- Abdurakhmonov, I.Y. (2014). *World Cotton Germplasm Resources*. Germany: Books on Demand.
- Acquaah, G. (2009). *Principles of Plant Genetics and Breeding*. New Jersey: John Wiley & Sons.
- Allen, S., McNamara, G., Kochman, J., Constable, G., Reid, P., and Stiller, W. (2004). Sicot F-1: A variety with increased resistance to Fusarium Wilt. In "Proceedings of the 12th Australian Cotton Conference, Australian Cotton Growers Research Association," Wee Waa, Australia.
- Andelković, V., Cvejić, S., Jocić, S., Kondić-Špika, A., Jeromela, A. M., Mikić, S., et al. (2020). Use of plant genetic resources in crop improvement – example of Serbia. *Genet. Resour. Crop. Evol.* 67, 1935–1948. doi: 10.1007/s10722-020-01029-9
- Anderson, W. F. (2005). Development of a forage bermudagrass (*Cynodon* sp.) core collection. *Grassl. Sci.* 51, 305–308. doi: 10.1111/j.1744-697X.2005.00040.x
- Azhar, F., Ali, Z., Akhtar, M., Khan, A., and Trethowan, R. (2009). Genetic variability of heat tolerance, and its effect on yield and fibre quality traits in upland cotton (*Gossypium hirsutum* L.). *Plant Breed.* 128, 356–362. doi: 10.1111/j.1439-0523.2008.01574.x
- Cotton Australia (2022). *Where Is Cotton Grown?* [Online]. Available: <https://cottonaustralia.com.au/where-is-cotton-grown> [Accessed February 21, 2022].
- Barragan, A. C., and Weigel, D. (2020). Plant NLR diversity: the known unknowns of Pan-NLRomes. *Plant Cell* 33, 814–831. doi: 10.1093/plcell/koaa002
- Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., and Edwards, D. (2020). Plant pan-genomes are the new reference. *Nat. Plants* 6, 914–920. doi: 10.1038/s41477-020-0733-0
- Becerra Lopez-Lavalle, L., Matheson, B., and Brubaker, C. (2011). A genetic map of an Australian wild *Gossypium* C genome and assignment of homoeologies with tetraploid cultivated cotton. *Genome* 54, 779–794. doi: 10.1139/g11-037
- Bernard, A., Barreneche, T., Donkpegan, A., Lheureux, F., and Dirlwanger, E. (2020). Comparison of structure analyses and core collections for the management of walnut genetic resources. *Tree Genet. Genomes* 16, 1–14. doi: 10.1007/s11295-020-01469-5
- Bhattacharjee, R., Khairwal, I., Bramel, P. J., and Reddy, K. (2007). Establishment of a pearl millet [*Pennisetum glaucum* (L.) R. Br.] core collection based on geographical distribution and quantitative traits. *Euphytica* 155, 35–45. doi: 10.1007/s10681-006-9298-x
- Brown, A. (1989a). "The case for core collections," in *The Use of Plant Genetic Resources*. eds. A. Brown, O. Frankel, D. Marshall and J. Williams (Cambridge: Cambridge University Press), 136–156.
- Brown, A. (1989b). Core collections: a practical approach to genetic resources management. *Genome* 31, 818–824. doi: 10.1139/g89-144
- Brown, A. (1995). "The core collection at the crossroads," in *Core Collections of Plant Genetic Resources*, eds. T. Hodgkin, A. H. D. Brown, T. J. L. van Hintum and E. A. V. Morales. (Italy: International Plant Genetic Resources Institute (IPGRI)), 3–19.
- Brown, A., and Brubaker, C. (2000). Genetics and the conservation and use of Australian wild relatives of crops. *Aust. J. Bot.* 48, 297–303. doi: 10.1071/BT98079
- Brown, A., and Spillane, C. (1999). "Implementing core collections - principles, procedures, progress, problems and promise," in *Core Collections for Today and Tomorrow*. eds. R. Johnson and T. Hodgkin (Rome, Italy: IPGRI), 1–9.
- Brubaker, C. L., and Brown, A. H. (2003). The use of multiple alien chromosome addition aneuploids facilitates genetic linkage mapping of the *Gossypium* G genome. *Genome* 46, 774–791. doi: 10.1139/g03-063
- Brubaker, C., Brown, A., Stewart, J. M., Kilby, M., and Grace, J. (1999). Production of fertile hybrid germplasm with diploid Australian *Gossypium* species for cotton improvement. *Euphytica* 108, 199–214. doi: 10.1023/A:1003641217653
- Campbell, B., Saha, S., Percy, R., Frelchowski, J., Jenkins, J. N., Park, W., et al. (2010). Status of the global cotton germplasm resources. *Crop Sci.* 50, 1161–1179. doi: 10.2135/cropsci2009.09.0551
- Chen, J., Zavala, C., Ortega, N., Petrolis, C., Franco, J., Burgueño, J., et al. (2016). The development of quality control genotyping approaches: a case study using elite maize lines. *PLoS One* 11:e0157236. doi: 10.1371/journal.pone.0157236
- Conaty, W. C., Johnston, D. B., Thompson, A. J., Liu, S., Stiller, W. N., and Constable, G. A. (2018). Use of a managed stress environment in breeding cotton for a variable rainfall environment. *Field Crop Res.* 221, 265–276. doi: 10.1016/j.fcr.2017.10.012
- Constable, G. (1998). Breeding and cultivar development of cotton for specific cropping systems. In "Proceeding of the World Cotton Research Conference – 2", New Frontiers in Cotton Research, Athens, 3–9.
- Constable, G., Llewellyn, D., Walford, S. A., and Clement, J. D. (2015). "Cotton breeding for fiber quality improvement" in *Industrial Crops. Handbook of Plant Breeding*. eds. V. M. V. Cruz and D. A. Dierig (New York: Springer), 191–232.
- Cottee, N., Tan, D., Bange, M., Cothren, J., and Campbell, L. (2010). Multi-level determination of heat tolerance in cotton (*Gossypium hirsutum* L.) under field conditions. *Crop Sci.* 50, 2553–2564. doi: 10.2135/cropsci2010.03.0182
- Cottee, N. S., Wilson, I. W., Tan, D. K., and Bange, M. P. (2014). Understanding the molecular events underpinning cultivar differences in the physiological performance and heat tolerance of cotton (*Gossypium hirsutum*). *Funct. Plant Biol.* 41, 56–67. doi: 10.1071/FP13140
- Culp, T., and Harrell, D. (1974). Breeding methods for improving yield and fiber quality of upland cotton (*Gossypium hirsutum* L.). *Crop Sci.* 13, 686–689.
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., and Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082. doi: 10.2135/cropsci2016.10.0885
- Dessauw, D., Nguyen, T.-B., Lacape, J.-M., and Hau, B. (2004). Cotton genetic resources at CIRAD. *4th International Cotton Genome Initiative ICGI-2004 Workshop*.
- Diederichsen, A., Kusters, P. M., Kessler, D., Baines, Z., and Gugel, R. K. (2013). Assembling a core collection from the flax world collection maintained by plant gene resources of Canada. *Genet. Resour. Crop. Evol.* 60, 1479–1485. doi: 10.1007/s10722-012-9936-1
- Díez, M. J., De la Rosa, L., Martín, I., Guasch, L., Cartea, M. E., Mallor, C., et al. (2018). Plant genebanks: present situation and proposals for their improvement. The case of the Spanish network. *Frontiers. Plant Sci.* 9:1794. doi: 10.3389/fpls.2018.01794
- Diwan, N., Bauchan, G. R., and McIntosh, M. S. (1994). A core collection for the United States annual *Medicago* germplasm collection. *Crop Sci.* 34, 279–285. doi: 10.2135/cropsci1994.0011183X003400010051x
- Dwivedi, S. L., Upadhyaya, H. D., and Hegde, D. M. (2005). Development of core collection using geographic information and morphological descriptors in safflower (*Carthamus tinctorius* L.) germplasm. *Genet. Resour. Crop. Evol.* 52, 821–830. doi: 10.1007/s10722-003-6111-8
- Egan, L. M., Hofmann, R. W., Barrett, B. A., Ghamkhar, K., and Hoyos-Villegas, V. (2019a). Identification of founding accessions and patterns of relatedness and inbreeding derived from historical pedigree data in a white clover germplasm collection in New Zealand. *Crop Sci.* 59, 2087–2099. doi: 10.2135/cropsci2018.11.0688
- Egan, L. M., Hofmann, R. W., Ghamkhar, K., and Hoyos-Villegas, V. (2019b). Identification of founding accessions and patterns of relatedness and inbreeding derived from historical pedigree data in a red clover germplasm collection in New Zealand. *Crop Sci.* 59, 2100–2108. doi: 10.2135/cropsci2019.01.0045
- Egan, L. M., Hofmann, R., Seguin, P., Ghamkhar, K., and Hoyos-Villegas, V. (2020). Pedigree analysis of pre-breeding efforts in *Trifolium* spp. germplasm in New Zealand. *BMC Genet.* 21, 104–113. doi: 10.1186/s12863-020-00912-9
- Ellis, M. H., Stiller, W. N., Phongkham, T., Tate, W. A., Gillespie, V. J., Gapare, W. J., et al. (2016). Molecular mapping of bunchy top disease resistance in *Gossypium hirsutum* L. *Euphytica* 210, 135–142. doi: 10.1007/s10681-016-1713-3
- Etmian, A., Pour-Aboughadareh, A., Mohammadi, R., Ahmadi-Rad, A., Moradi, Z., Mahdavian, Z., et al. (2017). Evaluation of genetic diversity in a mini core collection of Iranian durum wheat germplasms. *J. Ani. Plant Sci.* 27, 1582–1587.
- European Commission (2015). *Fibre crops as a sustainable source of bio-based materials for industrial products in Europe and China*.
- FAO (2009). *Global Agriculture towards 2050. How to Feed the World 2050: High-Level Expert Forum*. Available at: [https://www.fao.org/fileadmin/templates/wsfs/docs/Issues\\_papers/HLEF2050\\_Global\\_Agriculture.pdf](https://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf)

- FAO (2010). "The state of *ex situ* conservation," in *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome, Italy: FAO), 55–90.
- Fatokun, C., Girma, G., Abberton, M., Gedil, M., Nwachukwu, N., Oyatom, O., et al. (2018). Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. *Sci. Rep.* 8, 16035–16010. doi: 10.1038/s41598-018-34555-9
- Fischer, R., and Maurer, R. (1978). Drought resistance in spring wheat cultivars. I. Grain yield responses. *Aust. J. Agric. Res.* 29, 897–912. doi: 10.1071/AR9780897
- Franco-Duran, J., Crossa, J., Chen, J., and Hearne, S. J. (2019). The impact of sample selection strategies on genetic diversity and representativeness in germplasm bank collections. *BMC Plant Biol.* 19, 520–517. doi: 10.1186/s12870-019-2142-y
- Frankel, O. H. (1984). "Genetic Perspectives of Germplasm Conservation," in *Genetic Manipulation: Impact on Man and Society*. eds. W. Arber, K. Llimensee, W. J. Peacock and P. Starlinger (Cambridge, UK: Cambridge University Press).
- Fu, Y. B. (2017). The vulnerability of plant genetic resources conserved *ex situ*. *Crop Sci.* 57, 2314–2328. doi: 10.2135/cropsci2017.01.0014
- Gapare, W., Conaty, W., Zhu, Q.-H., Liu, S., Stiller, W., Llewellyn, D., et al. (2017). Genome-wide association study of yield components and fibre quality traits in a cotton germplasm diversity panel. *Euphytica* 213:66. doi: 10.1007/s10681-017-1855-y
- Gotor, E., Caracciolo, F., and Scafetti, F. (2019). On the international flow of plant genetic resources: forecasting the impacts of an evolving legal framework on CGIAR genebanks. *Biodiversity International*.
- Hamon, S., Dussert, S., Noirot, M., Anthony, F., and Hodgkin, T. (1995). Core collections - accomplishments and challenges. *Plant Breed. Abst.* 65, 1125–1133.
- Hanson, J., and Ellis, R. H. (2020). Progress and challenges in *ex situ* conservation of forage germplasm: grasses, herbaceous legumes and fodder trees. *Plan. Theory* 9:446. doi: 10.3390/plants9040446
- Holbrook, C. C., Anderson, W. F., and Pittman, R. N. (1993). Selection of a core collection from the US germplasm collection of peanut. *Crop Sci.* 33, 859–861. doi: 10.2135/cropsci1993.0011183X003300040044x
- Holme, I. B., Gregersen, P. L., and Brinch-Pedersen, H. (2019). Induced genetic variation in crop plants by random or targeted mutagenesis: convergence and differences. *Front. Plant Sci.* 10:1468. doi: 10.3389/fpls.2019.01468
- Jewell, M. C., Zhou, Y., Loch, D. S., Godwin, I. D., and Lambrides, C. J. (2012). Maximizing genetic, morphological, and geographic diversity in a core collection of Australian bermudagrass. *Crop Sci.* 52, 879–889. doi: 10.2135/cropsci2011.09.0497
- Jia, Y., Sun, J., and Du, X. (2014). "Cotton germplasm resources in China," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (United Kingdom: IntechOpen), 35–53.
- Jiang, H., Huang, L., Ren, X., Chen, Y., Zhou, X., Xia, Y., et al. (2014). Diversity characterization and association analysis of agronomic traits in a Chinese peanut (*Arachis hypogaea* L.) mini-core collection. *J. Integr. Plant Biol.* 56, 159–169. doi: 10.1111/jipb.12132
- Khan, M. A., Wahid, A., Ahmad, M., Tahir, M. T., Ahmed, M., Ahmad, S., et al. (2020). "World cotton production and consumption: An overview" in *Cotton Production and Uses*. eds. A. Shakeel and H. Mirza (Singapore: Springer), 1–7.
- Kilby, C. R., Tan, D. K., Duggan, B., and Constable, G. (2013). Yield components of high-yielding Australian cotton cultivars. *Cotton Res. J.* 5, 117–130.
- Kochman, J. (1995). Fusarium wilt in cotton: A new record in Australia. *Australas. Plant Pathol.* 24:74. doi: 10.1071/APP9950074
- Kuzay, S., Hamilton-Conaty, P., Palkovic, A., and Gepts, P. (2020). Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity? *Crop Sci.* 60, 1398–1414. doi: 10.1002/csc2.20032
- Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., et al. (2015). Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 33, 524–530. doi: 10.1038/nbt.3208
- Li, X., Shi, W., Broughton, K., Smith, R., Sharwood, R., Payton, P., et al. (2020). Impacts of growth temperature, water deficit and heatwaves on carbon assimilation and growth of cotton plants (*Gossypium hirsutum* L.). *Environ. Exp. Bot.* 179:104204. doi: 10.1016/j.envexpbot.2020.104204
- Li, X., Yan, W., Agrama, H., Hu, B., Jia, L., Jia, M., et al. (2010). Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. *Genetica* 138, 1221–1230. doi: 10.1007/s10709-010-9521-5
- Liang, W., Dondini, L., De Franceschi, P., Paris, R., Sansavini, S., and Tartarini, S. (2015). Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. *Plant Mol. Biol. Report.* 33, 458–473. doi: 10.1007/s11105-014-0754-9
- Liu, Q., Chen, Y., Wang, Y., Chen, J., Zhang, T., and Zhou, B. (2015a). A new synthetic allotetraploid ( $A_1A_1G_2G_2$ ) between *Gossypium herbaceum* and *G. australe*: bridging for simultaneously transferring favorable genes from these two diploid species into upland cotton. *PLoS One* 10:e0123209. doi: 10.1371/journal.pone.0123209
- Liu, X., Zhao, B., Zheng, H.-J., Hu, Y., Lu, G., Yang, C.-Q., et al. (2015b). *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Sci. Rep.* 5, 1–14. doi: 10.1038/srep14139
- Llewellyn, D. (2018). Does global agriculture need another green revolution? *Engineering* 4, 449–451. doi: 10.1016/j.eng.2018.07.017
- Lopez-Lavalle, L. A. B., Gillespie, V. J., Tate, W. A., Ellis, M. H., Stiller, W. N., Llewellyn, D. L., et al. (2012). Molecular mapping of a new source of Fusarium wilt resistance in tetraploid cotton (*Gossypium hirsutum* L.). *Mol. Breed.* 30, 1181–1191. doi: 10.1007/s11032-012-9705-z
- Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., et al. (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50, 803–813. doi: 10.1038/s41588-018-0119-7
- Mahalakshmi, V., Ng, Q., Lawson, M., and Ortiz, R. (2007). Cowpea [*Vigna unguiculata* (L.) Walp.] core collection defined by geographical, agronomical and botanical descriptors. *Plant genetic. Resources* 5, 113–119. doi: 10.1017/S1479262107837166
- Mahuku, G. S., Jara, C., Cajiao, C., and Beebe, S. (2003). Sources of resistance to angular leaf spot (*Phaeoisariopsis griseola*) in common bean core collection, wild *Phaseolus vulgaris* and secondary gene pool. *Euphytica* 130, 303–313. doi: 10.1023/A:1023095531683
- Maxted, N., Dulloo, M. E., and Ford-Lloyd, B. V. (2016). "how the focused identification of germplasm strategy (FIGS) is used to mine plant genetic resources collections for adaptive traits," in *Enhancing Crop GenePool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement* (England: CABI).
- May, O. L., and Lege, K. E. (1999). "Development of the world cotton industry," in *Cotton: Origin, History, Technology and Production*. eds. W. C. Smith and J. T. Cothren (New York: John Wiley & Sons), 65–97.
- Mei, Y., Zhou, J., Xu, H., and Zhu, S. (2012). Development of sea island cotton (*Gossypium barbadense* L.) core collection using genotypic values. *Aust. J. Crop Sci.* 6:673.
- Meredith, W. (1991). "Contributions of introductions to cotton improvement," in *Use of Plant Introductions in Cultivar Development: Part 1*. eds. H. Shands and L. Wiesner (Madison, WI: Crop Science Society of America).
- Miyatake, K., Shinmura, Y., Matsunaga, H., Fukuoka, H., and Saito, T. (2019). Construction of a core collection of eggplant (*Solanum melongena* L.) based on genome-wide SNP and SSR genotypes. *Breed. Sci.* 69, 498–502. doi: 10.1270/jsbbs.18202
- Miyazaki, J., Stiller, W. N., and Wilson, L. J. (2012). Novel cotton germplasm with host plant resistance to twospotted spider mite. *Field Crop Res.* 134, 114–121. doi: 10.1016/j.fcr.2012.05.006
- Muñoz-Amatriain, M., Cuesta-Marcos, A., Endelman, J. B., Comadran, J., Bonman, J. M., Bockelman, H. E., et al. (2014). The USDA barley core collection: genetic diversity, population structure, and potential for genome-wide association studies. *PLoS One* 9:e94688. doi: 10.1371/journal.pone.0094688
- Mwadingeni, L., Figlan, S., Shimelis, H., Mondal, S., and Tsilo, T. J. (2017). Genetic resources and breeding methodologies for improving drought tolerance in wheat. *J. Crop Improv.* 31, 648–672. doi: 10.1080/15427528.2017.1345816
- Nabwire, S., Suh, H.-K., Kim, M. S., Baek, I., and Cho, B.-K. (2021). Application of artificial intelligence in phenomics. *Sensors* 21:19. doi: 10.3390/s21134363
- Narayanan, S., Vidyasagar, P., and Babu, K. S. (2014). "Cotton germplasm in India - new trends," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (United Kingdom: IntechOpen).

- Niazian, M., and Niedbala, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture-Basel* 10:436. doi: 10.3390/agriculture10100436
- Niles, G., and Feaster, C. (1984). "Breeding," in *Cotton*. eds. R. Kohel and C. Lewis (Madison, WI: Crop Science Society of America), 201–231.
- Pande, S., Kishore, G. K., Upadhyaya, H., and Rao, J. N. (2006). Identification of sources of multiple disease resistance in mini-core collection of chickpea. *Plant Dis.* 90, 1214–1218. doi: 10.1094/PD-90-1214
- Percy, R., Frelichowski, J., Arnold, M., Campbell, B., Dever, J., Fang, D., et al. (2014). "The US national cotton germplasm collection—its contents, preservation, characterization, and evaluation," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (United Kingdom: IntechOpen), 167–201.
- Prasanna, B. (2012). Diversity in global maize germplasm: characterization and utilization. *J. Biosci.* 37, 843–855. doi: 10.1007/s12038-012-9227-1
- Rahmat, Z., Mahmood, A., Abdullah, K., and Zafar, Y. (2014). "Cotton Germplasm of Pakistan," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (United Kingdom: IntechOpen).
- Rajasekharan, P. (2015). "Gene banking for ex situ conservation of plant genetic resources," in *Plant Biology and Biotechnology*. eds. B. Bahadur, R. M. Venkat and L. Sahijram (New Delhi: Springer), 445–459.
- Rungis, D., Llewellyn, D., Dennis, E., and Lyon, B. (2002). Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium hirsutum* L.) cultivar. *Aust. J. Agric. Res.* 53, 551–560. doi: 10.1071/AR01121
- Sanamyan, M., Petlyakova, J., Rakhmatullina, E., and Sharipova, E. (2014). "Cytogenetic collection of Uzbekistan," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (United Kingdom: IntechOpen), 247–287.
- Schafleitner, R., Nair, R. M., Rathore, A., Wang, Y. W., Lin, C. Y., Chu, S. H., et al. (2015). The AVRDC—the world vegetable center mungbean (*Vigna radiata*) core and mini core collections. *BMC Genomics* 16, 1–11. doi: 10.1186/s12864-015-1556-7
- Sharma, R., Rao, V., Upadhyaya, H., Reddy, V. G., and Thakur, R. (2010). Resistance to grain mold and downy mildew in a mini-core collection of sorghum germplasm. *Plant Dis.* 94, 439–444. doi: 10.1094/PDIS-94-4-0439
- Sharma, M., Rathore, A., Mangala, U. N., Ghosh, R., Sharma, S., Upadhyay, H., et al. (2012a). New sources of resistance to Fusarium wilt and sterility mosaic disease in a mini-core collection of pigeonpea germplasm. *Eur. J. Plant Pathol.* 133, 707–714. doi: 10.1007/s10658-012-9949-9
- Sharma, S., Sharma, R., Govindaraj, M., Mahala, R. S., Satyavathi, C. T., Srivastava, R. K., et al. (2021). Harnessing wild relatives of pearl millet for germplasm enhancement: challenges and opportunities. *Crop Sci.* 61, 177–200. doi: 10.1002/csc2.20343
- Sharma, R., Upadhyaya, H., Manjunatha, S., Rao, V., and Thakur, R. (2012b). Resistance to foliar diseases in a mini-core collection of sorghum germplasm. *Plant Dis.* 96, 1629–1633. doi: 10.1094/PDIS-10-11-0875-RE
- Sherman, B., and Henry, R. J. (2020). The Nagoya protocol and historical collections of plants. *Nature Plants* 6, 430–432. doi: 10.1038/s41477-020-0657-8
- Shim, J., Mangat, P. K., and Angeles-Shim, R. (2018). Natural variation in wild *Gossypium* species as a tool to broaden the genetic base of cultivated cotton. *Journal of plant science. Current Res.* 2, 1–9. doi: 10.24966/PSCR-3743/100005
- Shivhare, R., and Lata, C. (2017). Exploration of genetic and genomic resources for abiotic and biotic stress tolerance in pearl millet. *Front. Plant Sci.* 7:2069. doi: 10.3389/fpls.2016.02069
- Silvar, C., Casas, A., Kopahnke, D., Habekuß, A., Schweizer, G., Gracia, M., et al. (2010). Screening the Spanish barley core collection for disease resistance. *Plant Breed.* 129, 45–52. doi: 10.1111/j.1439-0523.2009.01700.x
- Singh, R. P., Prasad, P. V., Sunita, K., Giri, S., and Reddy, K. R. (2007). Influence of high temperature and breeding for heat tolerance in cotton: a review. *Adv. Agron.* 93, 313–385. doi: 10.1016/S0065-2113(06)93006-5
- Singh, S., Vikram, P., Sehgal, D., Burgueño, J., Sharma, A., Singh, S. K., et al. (2018). Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Sci. Rep.* 8, 12527–12511. doi: 10.1038/s41598-018-30667-4
- Sokolikova, A., Burlyaeva, M., Valiannikova, T., Vishnyakova, M., Schafleitner, R., Lee, C.-R., et al. (2020). Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the world vegetable Gene Bank (Taiwan). *BMC Plant Biol.* 20, 363–369. doi: 10.1186/s12870-020-02579-x
- Stewart, J. M., and Hsu, C. L. (1978). Hybridization of diploid and tetraploid cottons through in-ovulo embryo culture. *J. Hered.* 69, 404–408. doi: 10.1093/oxfordjournals.jhered.a108982
- Stiller, W. N., and Wilson, I. W. (2014). "Australian cotton germplasm resources," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (Croatia: IntechOpen), 1–34.
- Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.1126/science.1183700
- Trapero, C., Wilson, I. W., Stiller, W. N., and Wilson, L. J. (2016). Enhancing integrated pest management in GM cotton systems using host plant resistance. *Front. Plant Sci.* 7:500. doi: 10.3389/fpls.2016.00500
- Trenberth, K. E., Dai, A., Van Der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., et al. (2014). Global warming and changes in drought. *Nat. Clim. Chang.* 4, 17–22. doi: 10.1038/nclimate2067
- Upadhyaya, H., Dronavalli, N., Dwivedi, S., Kashiwagi, J., Krishnamurthy, L., Pande, S., et al. (2013). Mini core collection as a resource to identify new sources of variation. *Crop Sci.* 53, 2506–2517. doi: 10.2135/cropsci2013.04.0259
- Upadhyaya, H., Dwivedi, S., Vadez, V., Hamidou, F., Singh, S., Varshney, R., et al. (2014). Multiple resistant and nutritionally dense germplasm identified from mini core collection in peanut. *Crop Sci.* 54, 679–693. doi: 10.2135/cropsci2013.07.0493
- Upadhyaya, H., and Ortiz, R. (2001). A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298. doi: 10.1007/s00122-001-0556-y
- Upadhyaya, H., Pundir, R., Dwivedi, S., Gowda, C., Reddy, V. G., and Singh, S. (2009). Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Sci.* 49, 1769–1780. doi: 10.2135/cropsci2009.01.0014
- Upadhyaya, H., Sarma, N., Ravishankar, C., Albrecht, T., Narasimhudu, Y., Singh, S., et al. (2010). Developing a mini-core collection in finger millet using multilocation data. *Crop Sci.* 50, 1924–1931. doi: 10.2135/cropsci2009.11.0689
- Upadhyaya, H. D., Thudi, M., Dronavalli, N., Gujaria, N., Singh, S., Sharma, S., et al. (2011). Genomic tools and germplasm diversity for chickpea improvement. *Plant Gene Res.* 9, 45–58. doi: 10.1017/S1479262110000468
- Upadhyaya, H. (2015). Establishing core collections for enhanced use of germplasm in crop improvement. *Ekin J. Crop Breed. Gene.* 1, 1–12.
- van Hintum, T. J., Brown, A., Spillane, C., and Hodgkin, T. (2000). *Core Collections of Plant Genetic Resources*. Rome, Italy: International Plant Genetic Resources Institute.
- van Zonneveld, M., Rakha, M., Tan, S., Chou, Y. Y., Chang, C. H., Yen, J. Y., et al. (2020). Mapping patterns of abiotic and biotic stress resilience uncovers conservation gaps and breeding potential of *Vigna* wild relatives. *Sci. Rep.* 10:2111. doi: 10.1038/s41598-020-58646-8
- Walters, C., Richards, C. M., and Volk, G. M. (2018). "Genebank conservation of germplasm collected from wild species," in *North American Crop Wild Relatives. Vol. 1*. eds. S. L. Greene, K. A. Williams, C. K. Houry, M. B. Kantar and L. F. Marek (Cham: Springer International Publishing), 245–280.
- Wambugu, P. W., Ndjiondjop, M.-N., and Henry, R. J. (2018). Role of genomics in promoting the utilization of plant genetic resources in genebanks. *Brief. Funct. Genomics* 17, 198–206. doi: 10.1093/bfpg/ely014
- Wang, B., Brubaker, C. L., and Burdon, J. J. (2004). Fusarium species and Fusarium wilt pathogens associated with native *Gossypium* populations in Australia. *Mycol. Res.* 108, 35–44. doi: 10.1017/S0953756203008803
- Wang, J.-C., Hu, J., Guan, Y.-J., and Zhu, Y.-F. (2013). Effect of the scale of quantitative trait data on the representativeness of a cotton germplasm sub-core collection. *J Zhejiang Univ Sci B* 14, 162–170. doi: 10.1631/jzus.B1200075
- Weckwerth, W., Ghatak, A., Bellaire, A., Chaturvedi, P., and Varshney, R. K. (2020). PANOMICS meets germplasm. *Plant Biotechnol. J.* 18, 1507–1525. doi: 10.1111/pbi.13372
- Wendel, J. F., Brubaker, C., Alvarez, I., Cronn, R., and Stewart, J. M. (2009). "evolution and natural history of the cotton genus, in genetics and genomics of cotton," in *Plant Genetics and Genomics: Crops and Models*. ed. A. H. Paterson (New York, NY: Springer), 3–22.
- Wendel, J. F., Brubaker, C. L., and Percival, A. E. (1992). Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* 79, 1291–1310. doi: 10.1002/j.1537-2197.1992.tb13734.x

- Westengen, O. T., Skarbø, K., Mulesa, T. H., and Berg, T. (2018). Access to genes: Linkages between genebanks and farmers' seed systems. *Food Sec.* 10, 9–25. doi: 10.1007/s12571-017-0751-6
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crop Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003
- Wilson, I. W., Moncuquet, P., Ellis, M., White, R. G., Zhu, Q.-H., Stiller, W., et al. (2021). Characterization and genetic mapping of black root rot resistance in *Gossypium arboreum* L. *Int. J. Mol. Sci.* 22:2642. doi: 10.3390/ijms22052642
- Xu, H., Mei, Y., Hu, J., Zhu, J., and Gong, P. (2006). Sampling a core collection of island cotton (*Gossypium barbadense* L.) based on the genotypic values of fiber traits. *Genet. Resour. Crop. Evol.* 53, 515–521. doi: 10.1007/s10722-004-2032-4
- Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., et al. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* 10:2989. doi: 10.1038/s41467-019-10820-x
- Zeng, L. (2014). "Broadening the genetic base of upland cotton in US cultivars: genetic variation for lint yield and fiber quality in germplasm resources," in *World Cotton Germplasm Resources*. ed. I. Y. Abdurakhmonov (United Kingdom: IntechOpen), 231–246.
- Zewdie, Y., Tong, N., and Bosland, P. (2004). Establishing a core collection of capsicum using a cluster analysis with enlightened selection of accessions. *Genet. Resour. Crop. Evol.* 51, 147–151. doi: 10.1023/B:GRES.0000020858.96226.38
- Zhang, H.-B., Li, Y., Wang, B., and Chee, P. W. (2008). Recent advances in cotton genomics. *Int. J. Plant Gen.* 2008, 1–20. doi: 10.1155/2008/742304
- Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: current status and perspectives. *Front. Plant Sci.* 10:714. doi: 10.3389/fpls.2019.00714

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Egan, Conaty and Stiller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.