## ABSTRACT

PETTIT, NICOLE ANNETTE. Molecular Genetic Resources and Trait Analyses for Improved Stress Resistance in Upland Cotton (*Gossypium hirsutum* L). (Under the direction of Vasu Kuraparthy).

The production of cotton is heavily impacted by a number of diseases and abiotic stresses like drought and temperature. In order to aid breeders in addressing these concerns in their breeding programs, the projects in this thesis worked to identify and create a number of resources and new lines of inquiry for disease resistance and stress tolerance. The cotton species *Gossypium* displays variation in pollen color both interspecifically and intraspecifically. The most predominantly grown species in the U.S., Upland cotton (*Gossypium hirsutum* L.), is typically cream with some exceptions showing yellow pollen. Pima cotton (*Gossypium barbadense* L.) pollen is a deep yellow or orange. To identify the genetic mechanisms behind pollen pigmentation in Upland cotton, a Genome-wide association study (GWAS) of 136 Upland cotton accessions phenotyped for pollen color and genotyped with the cottonSNP63K array was performed. Three significant marker trait associations were detected within a 226 Kb region on the short arm of chromosome A05. Segregation ratios of two biparental recombinant inbred line (RIL) mapping populations developed by crossing parental accessions DES 56 (white pollen) and Sealand #2 and Southland M1 (yellow pollen) and manually phenotyped in the $F_1$ and $F_5$ generations showed yellow pollen color in Upland cotton is under simple genetic control and is dominant to cream. Consensus mapping of the two $F_5$ RIL populations mapped the pollen color locus *Gh-$P_{A1}$* to a 5.36 cM region (112.83 Kb) between the flanking markers i12529Gh and i12525Gh corresponding with the larger 226 Kb region identified in GWAS. Four putative gene sequences were identified in this genomic region on the *G. hirsutum* cv. TM1 CRI v1 reference genome of which *phytoene synthase* (PSY) is the most probable candidate gene due to its role in a rate limiting step in the carotenoid biosynthetic pathway. Understanding the genetic controls and molecular mechanisms behind pollen color could aid in understanding the roles of pigment antioxidants in cotton pollen physiology.

Cotton bacterial blight (CBB) caused by the pathogen *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) is one of the most devastating diseases in cotton around the world, causing significant damage to cotton production. A broad-spectrum resistance locus *BB-13* was previously mapped to the long arm of the D02 chromosome. In the current study we made allele

specific molecular markers of 10 of the linked single nucleotide polymorphisms (SNPs) previously identified for *BB-13* using KASP/PACE primers. The markers were ran on 72 Upland cotton accessions previously phenotyped with *Xcm* race 18 to determine the accuracy of the markers. Additionally, a total of 145 lines from two $F_6$ RIL populations of Acala Maxxa by Arkot 8102 and CABD3CABCH-1-89 were phenotyped and used for marker validation. The marker NCBB-KASP4 developed from the CottonSNP63K array-based flanking SNP i25755Gh consistently had the lowest error rate of 4.17% in the diversity panel and 0.95% and 2.17% in the two RIL populations as compared to the known phenotypes. Using the SNP of NCBB-KASP4, i25755Gh, phenotypic predictions were made in 1,128 *Gossypium* accessions previously genotyped with the cottonSNP63K array. Of these lines, 253 had been manually phenotyped and errors in predictions were able to be identified. These analysis validated NCBB-KASP4 and its associated SNP i25755Gh to be a reliable molecular marker for detection and selection for CBB resistance in cotton making them the first publically available molecular marker for CBB resistance.

# Molecular Genetic Resources and Trait Analyses for Improved Stress Resistance in Upland Cotton (*Gossypium hirsutum* L)

by

Nicole Pettit

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Master of Science

Crop Science

Raleigh, North Carolina

2023

APPROVED BY:

| | |
|---|---|
| _____ | _____ |
| Vasu Kuraparthy<br>Chair of Advisory Committee | Gina Brown-Guerdira |
| _____ | |
| Terri Long | |

## DEDICATION

To my loving family who have guided and supported me in all my endeavors. To my mother who has always been my greatest ally and stood with me to face all of life's experiences. To my father who is my biggest encouragement and who pushes me to keep improving. And to my grandparents whose love and care have guided me along the way.

Without their support I would never have made it to where I am and where I am going. With them I have no limits. Thank you.

## BIOGRAPHY

Nicole Pettit was born on April 6th, 1998 in Cary, North Carolina to Audrey Pettit and Hugh Pettit. She graduated from Panther's Creek High School in Cary, NC in 2015. She went on to obtain her B.S. in Plant Biology from North Carolina State University in 2021 with a minor in Marine Science. In her undergraduate studies she spent time researching cotton fiber physiology in Dr. Candace Haigler's lab where she fell in love with crop science and research. With a focus on plant breeding, she joined Dr. Vasu Kuraparthy's cotton breeding program as a Master's Student in Augest of 2021.

## ACKNOWLEDGMENTS

I would like to thank Dr. Vasu Kuraparthy for the opportunity to continue my education in crop science and the guidance he has given me in my time in my Master's program. I would also like to thank Dr. Anjan Gowda for his mentorship and friendship. Thank you to Dr. Candace Haigler and my committee members Dr. Terri Long and Dr. Gina Brown-Guerdira. Thank you as well to Dr. Gloria Muday and her lab members who have helped in the pollen biochemical liquid chromatography-mass spectrometry and various aspects of this project. I would also like to thank Cotton Incorporated and Cotton Growers for providing funding throughout my Master's program.

Finally, thank you to my family including my parents Audrey Pettit and Hugh Pettit, my older brother Peter, and my grandparents Lavonne and Donald Eckstrom and Annette and Richard Pettit.

## TABLE OF CONTENTS

**Chapter 1: General Review of Literature……...1**

*1.1 Cotton Origins and Domestication……..1*

*1.2 Cotton Production……2*

*1.3 Challenges in Cotton Production……3*

*1.4 Abiotic Stresses……5*

*1.4.1 Heat stress in cotton……...6*

*1.4.2 Molecular biology of heat stress response in plants……8*

*1.5 ROS Scavenging……9*

*1.6 Introduction to Flavonoids……10*

*1.6.1 Flavonoids in cotton……11*

*1.7 Introduction to Carotenoids……12*

*1.7.1 Carotenoids in plants……...15*

*1.7.2 Carotenoids in abiotic stress tolerance……16*

*1.7.3 Carotenoids in cotton……..17*

*References……19*

**Chapter 2: Identification and molecular genetic mapping of major effect pollen pigmentation locus *Gh-P*$_{A1}$ in Upland cotton (*Gossypium hirsutum* L.)……..50**

*Key message……50*

*2.1 Abstract……...50*

*Keywords……...51*

*Abbreviations……51*

*2.2 Introduction……51*

*2.3 Material and Methods……..54*

*2.3.1 Plant material……...54*

*2.3.2 Phenotyping……..54*

*2.3.3 DNA isolation and genotyping*……..*56*

*2.3.4 Genome-wide association studies (GWAS)*……..*57*

*2.3.5 Linkage mapping*……..*57*

*2.3.6 Candidate gene analysis*……..*58*

*2.3.7 Data analysis*……*59*

*2.4 Results*……*59*

*2.4.1 Phenotypic analysis*……*59*

*2.4.2 LC-MS metabolite quantification*……..*60*

*2.4.3 Genome-wide association studies*……*61*

*2.4.4 Linkage mapping*……..*65*

*2.4.5 Candidate gene identification and allele combinations.*……..*71*

*2.5 Discussion*……..*71*

*2.5.1 Linkage mapping*……..*72*

*2.5.2 Candidate gene analysis and PSY gene*……...*73*

*2.5.3 Historic translocation event gave rise to Gh-$P_{A1}$*……..*73*

*2.5.4 Origins of yellow pollen in Upland cotton*……..*74*

*Acknowledgements*……..*76*

*Conflicts of interest*……..*77*

*Funding*……..*77*

*Supplementary material*……*77*

*References*……*79*

**Chapter 3: Development and validation of Kompetitive Allele-Specific PCR (KASP) markers for bacterial blight resistant locus *BB-13* in Upland cotton (*Gossypium hirsutum* L.)……...92**

*Abbreviations*……92

*Abstract*……92

*Keywords*……93

*Core ideas*……93

*3.1 Introduction*……94

*3.2 Material and Methods*……96

*3.2.1 Plant material*……96

*3.2.2 Phenotyping of the diversity panel and RIL populations* ……97

*3.2.3 DNA isolation*……97

*3.2.4 Primer design*……98

*3.2.5 KASP/PACE genotyping assay*……98

*3.2.6 Data analysis*……99

*3.3 Results*……100

*3.3.1 Phenotyping of the diversity panel and RIL populations*……100

*3.3.2 Validation of KASP markers in the diversity panel*……101

*3.3.3 Evaluation of KASP markers in the RIL populations*……108

*3.3.4 Phenotypic predictions based on allelic results*……109

*3.4 Discussion*……110

*Author contributions*……112

*Acknowledgements*……113

*Conflicts of interest*……..*113*

*Funding information*……*113*

*References*……*114*

*Appendix*……*121*

# LIST OF TABLES

## Chapter 2

**Table 1:** Phenotypic segregation ratios of pollen color in the two $F_5$ RIL populations…59

**Table 2:** Putative genes identified by combined GWAS and linkage analyses of *Gh-P$_{A1}$* in the Upland cotton.……..70

**Supplementary Table 1:** List of the 99 accessions used from the elite diversity panel and 37 accessions from the tropical landrace panel used in GWAS analysis of pollen color in *G. hirsutum*.……...….*See attached file*

**Supplementary Table 2:** SNP density and coverage statistics of 19,349 polymorphic SNP identified in 136 *G. hirsutum* accessions used for GWAS…..*See attached file*

**Supplementary Table 3:** Percentage of phenotypic variance explained by the SNP markers associated with *Gh-P$_{A1}$* identified by GWAS analysis…...*See attached file*

**Supplementary Table 4:** Chromosomal marker information of the 40D linkage map of DES 56 × Sealand #2.……..*See attached file*

**Supplementary Table 5:** Chromosomal marker information from the 42D linkage map of DES 56 × Southland M1.…….....*See attached file*

**Supplementary Table 6:** Coverage statistics of the 40D linkage map of DES 56 × Sealand #2 compared to the sequence-based physical map……*See attached file*

**Supplementary Table 7:** Coverage statistics of the 42D linkage map of DES 56 × Southland M1 compared to the sequence-based physical map…...*See attached file*

**Supplementary Table 8:** Consensus linkage map information of the *Gh-P$_{A1}$* locus on the short arm of chromosome A05.……...*See attached file*

**Supplementary Table 9:** PSY and associated genes found in *G. hirsutum*…*See attached file*

## Chapter 3

**Table 1:** Allele pattern of eight KASP markers for cotton bacterial blight (CBB) resistance locus *BB-13* validated in the elite diversity (DIV) panel……..*104*

**Table 2:** Primer sequence information of the six best performing KASP markers for *BB-13* in Upland cotton.……*109*

**Supplementary Table 1:** List of the complete 382 *Gossypium* accessions in the elite diversity (DIV) panel……..*See attached file*

**Supplementary Table 2:** Primer sequence information of the ten KASP markers utilized in the current study to identify a diagnostic marker for *BB-13* locus…*See attached file*

**Supplementary Table 3:** Phenotypic segregation ratios in the RIL populations evaluated for resistance to race 18 of CBB.…….....*See attached file*

**Supplementary Table 4:** Error rates for KASP markers in the biparental RIL populations.……..*See attached file*

**Supplementary Table 5:** Error rates for the SNP i25755Gh in 253 *Gossypium hirsutum* accessions of DIV panel.……..*See attached file*

**Supplementary Table 6:** Predictions of CBB response phenotypes of 875 *Gossypium* germplasm lines based on i25755Gh SNP allele calls……....*See attached file*

# LIST OF FIGURES

## Chapter 1

**Figure 1:** Caroeneoid biosynethesis pathway as deswcribed in Figure 3 from Tanaka & Ohmiya, 2008.……*14*

## Chapter 2

**Figure 1:** Pollen color phenotypes of Upland cotton (*Gossypium hirsutum*) **(a)** parental accessions and **(b)** $F_1$s of four RIL populations.……*55*

**Figure 2:** Flavonol quantification using liquid chromatography-mass spectrometry in **(a)** parental accessions and **(b)** $F_5$ RIL populations 40D and 42D……*60*

**Figure 3:** Genome wide association study for pollen color in GWAS panel of *G. hirsutum* **(a)** Manhattan plot detecting significant signals associated with pollen color and **(b)** Quantile-quantile plot showing deviations from expected p-values for pollen color……*62*

**Figure 4:** Marker allele combinations of pollen color in the GWAS panel **(a)** Regional Manhattan plot of the short arm of chromosome A05 and **(b)** Five marker allele combinations were observerd using the three significant markers.……*63*

**Figure 5:** Consensus mapping of the short arm of chromosome A05 made from combining two linkage maps developed from $F_5$ RIL populations (DES 56 × Sealand #2 and DES 56 × Southland M1)……*68*

**Supplementary figure 1:** SNP marker density and distribution for 26 Upland cotton chromosomes in linkage mapping and the genome wide association study……*122*

**Supplementary figure 2: (a-b)** heat map for 40D (DES 56 × Sealand #2) and 42D (DES 56 × Southland M1) population based linkage map and **(c-d)** comparison of collinearity between the genetically mapped linkage groups and the sequence based physical chromosome maps……*124*

**Supplementary figure 3:** Single marker analysis of pollen color for all 26 Upland cotton chromosomes for **(a)** 40D $F_5$ RIL population (DES 56 × Sealand #2) and **(b)** 42D $F_5$ RIL population (DES 56 × Southland M1).……..*126*

**Chapter 3**

**Figure 1:** Phenotypic response of Upland cotton accessions to race 18 of *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) at seedling stage.……*101*

**Figure 2:** Allele discrimination Cartesian plots showing FAM and HEX fluorescence determining resistant and susceptible marker allele pattern in *Gossypium hirsutum* lines.…… …….*102*

# CHAPTER 1: General Review of Literature

## 1.1 Cotton Origins and Domestication

Cotton, *Gossypium* spp., is one of the world's leading sources of natural fibers and is grown in all continents except Antarctica. These perennials express a wide range of variations in growth structure from bushes to small trees. The *Gossypium* genus evolved from the Malvaceae family 5 to 7 MYA (Chen *et al.*, 2020; Huang *et al.*, 2020; Hu *et al.*, 2019; Li *et al.*, 2015; Viot & Wendel, 2023). The Malvaceae family, also known as the mallow family, contains around 4,225 different species of plants (Wang *et al.*, 2021). The genus *Gossypium* itself contains over 50 species with four subgenres, *Gossypium*, *Houzingenia*, *Sturtia*, and *Karpas* (Viot & Wendel, 2023). There are eight different diploid genomes (2n=2x=26) that are denoted A to G and K found in the *Gossypium* genus (Hu *et al.*, 2021; Seelanan *et al.*, 1999; Viot & Wendel, 2023; Wendel & Albert, 1992; Wendel & Grover, 2015) as well as one AD allopolyploid genome (2n=4x=52). The AD genome tetraploids arose due to a natural hybridization of the African-Asian A genome with the American D genome (Hu *et al.*, 2021; Seelanan *et al.*, 1999; Viot & Wendel, 2023; Wendel & Albert, 1992; Wendel & Grover, 2015). It is thought that an A genome seed parent, that was equally related to the two extant A genome species, hybridized 0.7 to 1 MYA with a D genome ancestor that was similar to *G. raimondi* which then underwent chromosome doubling (Beasley, 1940; Chen *et al.*, 2017; Hu *et al.*, 2021; Li *et al.*, 2015; Senchina *et al.*, 2003; Skovsted, 1934; Viot & Wendel, 2023; Wendel, 1989; Wendel & Grover, 2015). Of the four cultivated *Gossypium* species, two species are diploids with an A genome (*G. herbaceum* L. and *G. arboreum* L.) and two species are allopolyploids (*G. hirsutum* L. and *G. barbadense* L.) (Viot & Wendel, 2023). *Gossypium barbadense*, or Pima cotton, is thought to originate in South America. Fossil records suggest that humans may have begun cultivating cotton in the middle of the 7th millennium BP and evidence of domestication was found in Real Alto from 6,350 YBP (Damp & Pearsall, 1994; Moulherat *et al.*, 2002; Viot & Wendel, 2023). The origin of domestication for *Gossypium hirsutum*, or Upland cotton, is more difficult to specify due to its prolific use worldwide. Available evidence suggest that domestication likely began in the Yucatan peninsula in Mexico 5.5 to 6 thousand YBP (Alavez *et al.*, 2021; Smith & Stephens, 1971; Viot & Wendel, 2023; Wendel *et al.*, 1992; Yuan *et al.*, 2021). These independent origins show that the species domesticated and evolved parallel to one another. The traits selected for during early domestication for both species consisted of longer, more abundant

fibers that are easier to detach and bigger seeds with thinner seed coats to reduce dormancy and increase germination (Brubaker *et al.*, 1999; Fryxell, 1979). Now breeders select from traits including fiber quality, yield, plant architecture, photoperiod-insensitivity, resistance, and stress tolerance (Gross & Strasburg, 2010).

### 1.2 Cotton Production

Cotton is a fiber, feed, and oil seed crop most known for producing 81% of the natural fiber produced globally (FAO, 2021). Besides the lint on the seed's epidermis, the plant also produces seed oil and meal for livestock due to its oleoproteaginous seeds (Bellaloui *et al.*, 2015). Cotton is usually planted from March to June but can be planted as early as February depending on the species, location, and weather (NCCA, 2013; USDA ERS, 2022). Harvesting depends on the planting time, growing region, as well as the variety, ranging from August to December (USDA ERS, 2022). It takes at least 40 to 50 days for fibers to mature on the seeds (Hake & Silvertooth, 1990). In the US, cotton is predominantly grown in the Cotton Belt comprising 17 states in the southern half of the country ranging from California to Virginia (USDA ERS, 2022). The top-producing states are Georgia, Mississippi, Arkansas, Texas, and North Carolina with Texas producing around 40% of the US cotton (USDA ERS, 2022).

The US is one of the world's leading exporters of cotton (USDA ERS, 2022). In the US, 97% of the cotton produced is upland cotton while only 3% is Pima cotton (USDA ERS, 2022). Upland cotton plants have increased production and adaptability as compared to Pima cotton. Pima cotton fibers, however, are longer, thinner, and stronger than Upland fibers, making them of higher quality and used in the production of premium fabrics. In 2022, the US planted 12.48 million acres of cotton, 97% of which were transgenic varieties (NCCA, 2013; NCCA, 2022). Of this cotton, on average around 70% was exported, resulting in over 3 million metric tons in exports valued at 9.02 billion dollars (NCCA, 2013; USDA FAS, 2022). Around 2.4 million bales of the remaining cotton per year on average are used in domestic US textile mills (NCCA, 2013). The downstream economic impacts of the US cotton industry make it the number one value-added crop in the country accounting for 420 thousand jobs and approximately $120 billion in total economic activity annually (NCCAb, accessed 2023). The income generated from cotton production is an integral part of food security on smallholder farms across the globe, where the majority of cotton production occurs (Ferreira, 2019). In addition to its importance as

a fiber crop for the textile industry, cotton seed is an important source of protein-rich feed for livestock and oil for cooking purposes (NCCAb, accessed 2023), making cotton one of the most diversely valuable crops domesticated. As such, the biotic and environmental stresses that have wide-spread, negative effects on cotton production are devastating for rural communities as well as the national and global economy.

### 1.3 Challenges in Cotton Production

With such an important crop, research on factors affecting cotton production is conducted continuously. Despite an extensive diversity in existing germplasm, through years of interspecific and intraspecific breeding the varieties used in commercial production have limited genetic diversity (Tyagi *et al.*, 2014; Van Esbroeck *et al.*, 1999; Wallace *et al.*, 2009). Due to a select few lines being regularly used in their development, current commercially used upland cultivars only consist of 1% of the total genetic diversity available (Bowman *et al.*, 1996; Brubaker *et al.*, 1999; May *et al.*, 1995; Wallace *et al.*, 2009; Wendel *et al.*, 1992). This is particularly true in transgenic cotton as all available varieties originate from the same original line, Coker 312 and few inhouse germplasm lines used in the breeding programs (Wallace *et al.*, 2009). Limited genetic diversity and similar ancestors can lead to genetic bottlenecks, linkage blocks, and limited allelic diversity that are difficult to overcome with traditional breeding methods (Brown, 1983). Traditional breeding methods rely heavily on crossing and re-selections using a limited set of elite lines which could further decrease genetic diversity (Bowman *et al.*, 1996; Brubaker *et al.*, 1999; May *et al.*, 1995; Wendel *et al.*, 1992).

In addition to the limited genetic diversity challenges, cotton is susceptible to a number of biotic and abiotic stresses that can lead to a decrease in plant fitness, yield, and quality of the plant fibers. Cotton growth and yield also depend heavily on adequate weed control, especially in the first eight weeks of planting (Cotton inc a, accessed 2023). Insects can cause damage to many of the plant structures and can act as a transmission source for a number of diseases. A number of nematode species are found in the US Cotton Belt including the root-knot nematode and the reniform nematode (*Meloidogyne* spp. and *Rotylenchulus reniformis,* respectively) (Ahumada & Gorny, 2019; Ciancio, 2021; Uppalapati *et al.*, 2010). Nematodes will feed on the root system of host plants. This damage can lead to galls of swelled tissues as well as chlorosis and stunting under high severity, causing severe loss (Ahumada & Gorny, 2019; Lawrence *et al.*, 2018).

Nematodes can also act as a carrier for certain bacterial diseases including some species of Fusarium wilt (da Silva *et al.*, 2019).

Cotton Seedling diseases caused 378.9 thousand bales of cotton to be lost in 2017 (Lawrence *et al.*, 2018). These seedling diseases cover Root Rot, Black Root, Damping Off, and a number of early diseases (Ahumada, 2023; NCCAa, accessed 2023; Sinclair, 1965). *Pythium* spp. are the causal agents of Root Rot which infects the seed, emerging radical, and hypocotyl which can lead to chlorosis and stunting (Ahumada, 2023; NCCAa, accessed 2023). Black Root is caused by the pathogen *Thielaviopsis basicola* during the seedling stage (Ahumada, 2023; NCCAa, accessed 2023). This can cause the below-ground structures to turn black and rot in young plants or form collars in more developed plants (NCCAa, accessed 2023). Damping Off can be caused by a number of different pathogens either pre- or post-emergence (NCCAa, accessed 2023). *Rhizoctonia* species, including *Rhizoctonia solani*, will infect the seedling at the soil level, damaging the hypocotyl and resulting in post-emergence Damping Off (Ahumada, 2023; NCCAa, accessed 2023). Other pathogens that can affect seedlings include the fungus *Fusarium* spp. (Ahumada, 2023). *Fusarium oxysporum* f. sp. *vasinfectum* Race 4 (FOV4) causes Fusarium wilt, one of the most damaging diseases in US cotton (Cotton inc b, accessed 2023; Lawrence *et al.*, 2018; Zhu *et al.*, 2021). Race 4 of FOV is a highly virulent race that infects the vascular systems through the roots of the plant (Cianchetta & Davis, 2015; Cotton inc b, accessed 2023; Zhu *et al.*, 2021). Infection leads to vascular damage causing staining of the vascular tissues, wilting, chlorosis, necrosis, stunting, defoliation, and death (Cotton inc b, accessed 2023; Zhu *et al.*, 2021). Another cause of wilt is Verticillium wilt caused by the fungus *Verticillium dahliae* (Khleb) which is also one of the most devastating diseases in cotton (El-Zik, 1985; Lawrence *et al.*, 2018; Meena *et al.*, 2019). Symptoms of this disease are chlorotic spots on leaves leading to necrosis, wilting, defoliation, staining and damage of the vascular tissues, and death (El-Zik, 1985; Meena *et al.*, 2019).

Blight diseases like Cotton Bacterial Leaf Blight (CBB) and Ascochyta Blight of Cotton also cause heavy losses to the crop. First reported in the US in 1891, CBB is an extremely damaging disease around the world (Atkinson, 1892). CBB is caused by the bacterial pathogen *Xanthomonas* citri subsp. *malvacearum* (*Xcm*) and can infect and damage all parts of the cotton plant. There are over 22 identified races of *Xcm* with race 18 being the most prevalent and virulent race in the US (Delannoy *et al.*, 2005; El-Zik & Thaxton, 1994; Zhang *et al.*, 2020).

Infection spreads through any openings on the plant's surface causing localized water-soaked lesions to form (Innes, 1983; Kemerait *et al.*, 2017; Verma, 1986). As the disease progresses the lesions grow, the leaves may fray, and defoliation may occur (Innes, 1983; Kemerait *et al.*, 2017; Phillips *et al.*, 2017; Ridgway *et al.*, 1984; Verma, 1986). The infected bolls can lead to decreases in the yield and quality of the fiber (El-Zik & Thaxton, 1994; Kemerait *et al.*, 2017; Phillips *et al.*, 2017). The removal of the residual lint on the seeds through acid delinting and the planting of resistant cultivars had successfully managed the disease in the US for over 50 years until the disease reemerged in 2011 (Allen, 2011; Xiao *et al.*, 2009; Zhang *et al.*, 2020). Ascochyta Blight is caused by infection of the pathogen *Ascochyta gossypii* (Cotton inc c; Meena *et al.*, 2019). Infections are particularly severe during particularly wet and cool weather. Infection causes circular lesions in the plant's shoots and can damage or discolor the fibers (Cotton inc c; Meena *et al.*, 2019).

Cotton is also susceptible to a number of follicular leaf spot diseases such as Target Leaf Spot (*Corynespora casiicola*), Cercospora Leaf Spot (*Cercospora gossypina* and Mycosphaerella gossypina), Alternaria Leaf Spot (*Alternaria macrospora* and *Alternaria alternata*), and Stemphylium Leaf Spot (*Stemphylium solani*) (Cotton inc c; Meena *et al.*, 2019). Leaf spot diseases cause colored lesions on infected leaf surfaces and can lead to wilting, necrotic tissue, and defoliation (Cotton inc c; Meena *et al.*, 2019). Other minor diseases that affect cotton production in the US include Grey Mildew (*Ramularia areola* (Atk.), Areolate Mildew (*Ramularia gossypii*), Anthracnose of cotton (*Colletotrichum gossypii*), Cotton leafroll dwarf disease (cotton leafroll dwarf virus), and Cotton leaf curl (begomovirus) (Chen *et al.*, 2019; Cotton inc c; Edula *et al.*, 2023; Meena *et al.*, 2019).

### 1.4 Abiotic Stresses

In addition to these biotic factors, abiotic stresses can also cause drastic effects on cotton growth and development and economic viability of cotton production in the US (Azhar *et al.*, 2020; Demirel *et al.*, 2014). Abiotic stresses can cause detrimental effects on the plant's photosynthetic ability (Cothren, 1999; DeRidder & Salvucci, 2007) and plant signaling (Digonnet *et al.*, 1997). In cotton, the main abiotic stresses affecting production are temperature, drought, and salinity stresses (Allen & Aleman, 2011). The effects of these stresses often exacerbate each other. For instance, the damage caused by temperature stress will be greater

when the plant is also experiencing drought stress due to a lack of water that is needed for the plants internal cooling systems. The temperature of the plant canopy is shown to regularly be cooler than the surrounding air temperatures though the exact difference varies inversely with the vapor pressure of the area (Brown, 2008; Brown & Zeiher, 1997; Brown & Zeiher, 1998a; Hake & Silvertooth, 1990; Idso *et al.*, 1986; Zeiher *et al.*, 1994; Zeiher *et al.*, 1995). Vapor pressure and thus canopy temperature can be modeled using meteorological conditions including atmospheric humidity/ dew point and air temperature (Brown & Zeiher, 1998a). Higher levels of humidity can reduce radiative cooling and transpiration during the day and night respectively resulting in higher canopy temperatures (Brown, 2008; Hake & Silvertooth, 1990).

### 1.4.1 Heat stress in cotton

The optimal temperature for *G. hirsutum* plants is around 24-30°C in the plant canopy during the day (Azhar *et al.*, 2020; Brown, 2008; Radin *et al.*, 1994). Heat stress can adversely affect a number of different processes in the crop including its reproduction, metabolism, productivity, and yield (Azhar *et al.*, 2020; Demirel *et al.*, 2014). Heat events can rapidly affect the plant and the impacts vary depending on the extent and duration of the high temperatures (Azhar *et al.*, 2020; Brown, 2008; Hasanuzzaman *et al.*, 2013). Studies found that under high heat (above 28°C) square and boll retention was reduced, particularly 3-5 days after anthesis (Brown, 2008; Hodges *et al.*, 1993; Reddy *et al.*, 1992). This reduction in boll retention causes a poor reproductive performance (Brown, 2008; Hodges *et al.*, 1993; Reddy *et al.*, 1992). Squares are susceptible to heat stress as early as 15 days before flowering where exposure to high temperatures above 30°C can cause floral structure abnormalities including smaller flowers that may not fully open, shortened stamen causing apparent elongated stigmas, lack of dehiscence in the anthers, and asynchronicity in the production of male and female organs (Brown, 2001; Brown, 2008; Brown *et al.*, 1995; Brown & Zeiher, 1997; Brown & Zeiher, 1998b; Brown & Zeiher, 2008; Zeiher *et al.*, 1994; Zeiher *et al.*, 1995). These abnormalities were reported to result in aborted bolls during the 3-5 DPA (Days Post Anthesis) time period. The bolls will still abort during this period even if temperatures return to normal, reinforcing that this boll loss is due to the initial heat stress event (Brown & Zeiher, 2008). Surviving bolls themselves can also appear smaller, malformed, or misshapen like parrot beaked bolls as well as have fewer seeds due to excessive consumption of carbohydrates during increased respiration from high

temperatures (Arevalo *et al.*, 2008; Brown & Zeiher, 2008; Hake & Silvertooth, 1990; Liu *et al.*, 2006). As a result, heat stress may reduce the crop's yield with studies showing a decline in 100-245 lb/a in years with high heat stress in Arizona (Brown, 2001; Brown, 2008; Zhao *et al.*, 2005).

High temperatures can affect the quality of the fibers as well. Generally, the effect is not as significant as yield due to the plant balancing the boll load through boll dropping or a reduction in boll size to ensure that the plant has enough carbohydrates to supply the number of maturing bolls (Hake & Silvertooth, 1990). High temperature effects may cause the cellulose rings developing on the growing fibers to be thicker, increasing the fiber width and micronaire (Hake & Silvertooth, 1990). Additionally, the fiber's length and strength may be reduced especially during sudden high temperature events (Hake & Silvertooth, 1990).

Exposure to above-optimal temperatures in *in-vitro* studies resulted in a reduction in pollen germination starting at 28°C and pollen tube development starting at 34°C (Brown, 2008; Burke *et al.*, 2004; Song *et al.*, 2015). This reduction continued until temperatures reached 37°C where pollen germination was dramatically reduced and at 43°C pollen tubes would no longer be able to elongate. Exposure to high temperatures during a temperature-sensitive stage of pollen development in early squares negatively affects meiosis (Meyer, 1966), the pollen may become sterile, and the anthers may fail to dehisce (Azhar *et al.*, 2020; Baloch *et al.*, 2000; Hake & Silvertooth, 1990; Hofmann *et al.*, 1988; Liu *et al.*, 2006; Taha *et al.*, 1981; Zhang *et al.*, 2010). Lower quantities of pollen grains was shown to increase the time for pollen tubes to fully develop, further exacerbating the reduction in fertilization rates (Ter-Avanesian, 1978; Zahid *et al.*, 2016).

Heat stress is managed through methods to limit the exposure of the plant to high air temperatures and humidity as well as through the selection of different varieties (Azhar *et al.*, 2020; Brown, 2008; Hake & Silvertooth, 1990). One way to avoid exposing the cotton plants to high air temperatures during peak blooming season is to plant the cotton during the optimal planting period for the area (Brown, 2008). By planting the cotton on time, growers are planting early enough that peak blooming time does not coincide with peak air temperatures or storm periods for the area. This is why it is important to look at recommended planting times for the local climate based on the number of heat units (HU), or high heat events, that have occurred that year (Brown, 2008). Some planting season models also take into account wet seasons to limit the

plants' exposure to high humidity at peak flowering season. Another way to avoid humidity is through proper irrigation methods of frequent and light waterings to keep the soil moisture high without oversaturating it (Brown, 2008; Hake & Silvertooth, 1990). Inadequate irrigation may cause the plants to experience drought stress which can exacerbate heat stress (Azhar *et al.*, 2020; Brown, 2008; Rizhsky *et al.*, 2004). Drought stress can cause the stomata of the plant to close in order to reduce evaporation, which results in higher canopy temperatures (Brown, 2008). Further, proper plant population density (PPD) is important for growers to keep in mind during planting to maintain proper airflow and ensure high photosynthetic capacity (Yao *et al.*, 2016).

Selecting varieties with appropriate maturation to plant in each growing region is another way for growers to limit heat stress effects in their crops. By selecting short-season varieties, growers can potentially avoid high temperature or high humidity periods due to a quickened flowering period (Brown, 2008). Growers can also choose heat-tolerant varieties that may reduce the effects caused by heat stress. There are a number of different molecular methods to account for limiting heat stress. These methods include selecting plants with differently controlled transpiration methods using stomatal conductance (Cornish *et al.*, 1991; Lu *et al.*, 1998; Monteith, 1995; Ur-Rahman, 2005; Ulloa *et al.*, 2000) or changes to the lipid membrane composition (Ur-Rahman *et al.*, 2004).

### 1.4.2 Molecular biology of heat stress response in plants

Although the genetic basis of heat stress tolerance is less understood in plants, a multitude of studies were conducted to understand the molecular and biochemical changes associated with heat stress response in model plants as well as in crop plants. It was well reported that plants work to limit the damage caused by heat stress through molecular regulation of membrane stability maintainence, production of compatible solutes, epigenetic regulation, the triggering of signaling pathways, and the control of certain protective proteins through transcriptional and post transcriptional methods (Azhar *et al.*, 2020). Heat shock proteins (HSP) are proteins that accumulate in plants under heat stress to protect themselves from adverse effects, allowing for a rapid response after exposure (Abrol & Ingram, 1996; Burke *et al.*, 1985; Cottee *et al.*, 2013; Nakamoto & Hiyama, 1999; Schöffl *et al.*, 1999). HSPs are able to mitigate these effects through their roles in the matenence of cell integrity and photosynthesis by protecting proteins from denaturing (Barua *et al.*, 2003; Bowen *et al.*, 2002; Vierling, 1991). There are a number of

different studies identifying the HSPs and other proteins in cotton that are expressed under high temperatures (Azhar *et al.*, 2020; Burke & Chen, 2015; Burke *et al.*, 1985; Demirel *et al.*, 2014). Plants response to abiotic stress happens rapidly through transcription and post transcriptional molecular modifications through transcriptional factors and microRNA (miRNA) (Yamaguchi-Shinozaki & Shinozaki, 2006; Zahid *et al.*, 2016). MiRNAs are a post transcriptional control that are differentially expressed under abiotic stresses that negatively affect many biological processes in plants, though the specific methods in cotton are less understood (Sun, 2012; Yin *et al.*, 2012). Abscisic acid (ABA) is a plant hormone involved in plant signaling under stress conditions and is shown to be involved in overexpressing or under expressing certain miRNAs (Christmann *et al.*, 2005; Cutler *et al.*, 2010; Reyes & Chua, 2007; Zahid *et al.*, 2016). The overexpression of miRNAs that are known to be involved in specific stress responses in a plant can be used to create tolerant varieties (Zahid *et al.*, 2016). While ABA is often linked to drought stress, the phytohormone's involvement in stomatal conductance through the calcium in the guard cells closing can also exacerbate the effect of the plant under high temperatures (Macrobbie, 1997; Schmidt *et al.*, 1995). Calcium is also used in a plant's stress response through the control of the calcium channels in the plasma membrane, enabling many signaling pathways (Zahid *et al.*, 2016). Of these signaling pathways, an excessive accumulation of reactive oxygen species (ROS) can be detrimental in plants, leading to cell death (Blanvillain *et al.*, 2011). Therefore, another method of heat tolerance is through the production of antioxidants like flavonoids, carotenoids, and other secondary metabolites due to their role in ROS scavenging (Azhar *et al.*, 2020; Baker, 2008; Latowski *et al.*, 2011; Zahid *et al.*, 2016).

### 1.5 ROS Scavenging

ROS are highly reactive oxygen-containing signaling molecules that are found in both plant and animal tissues. The different species of ROS have different production sites and reactivities making them able to be used in a number of different signaling pathways (Chapman *et al.*, 2019; Mittler, 2017). Superoxide anions, hydrogen peroxide ($H_2O_2$), hydroxyl radicals, and singlet oxygen are some of the more common species of ROS (Foyer & Noctor, 2020; Juan *et al.*, 2021). These ROS species may further oxidize themselves or other proteins creating a cascade of free radical production (Martin *et al.*, 2022). ROS production can be induced in certain conditions, like when the plant is under stress (Chapman *et al.*, 2019; Martin *et al.*, 2022). This process is

called ROS bursts. Small changes in the abundance of ROS through these bursts are used as a signal for the plant for plant growth and development as well as for stress responses (Baxter *et al.*, 2014; Willems *et al.*, 2016). ROS may target specific enzymes or transcription factors to change their activity to aid in signaling (Martin *et al.*, 2022; Mase & Tsukagoshi, 2021). When targeting specific polymers in the cell wall, ROS can trigger the loosening or stiffening of the cell wall (Martin *et al.*, 2022). ROS signaling is involved in root development and elongation, guard cell signaling, reproduction, and seed development (Chapman *et al.*, 2019). Certain ROS species are known to be involved in the development of mature pollen and the pollen tube at several stages (Chapman *et al.*, 2019; Hsieh & Huang, 2007; Xie *et al.*, 2014). However due to the highly reactive nature of ROS, high amounts can cause oxidative damage and become toxic. ROS causes oxidative damage by stealing electrons from other molecules and forming more free radicals (Chapman *et al.*, 2019). As such, it is important for plants to be able to manage ROS levels in the cells. There are a number of different methods both enzymatic and nonenzymatic to scavenge the excessive ROS in plant cells, some of which are the same as in human cells. Enzymes like catalase and peroxidase can convert $H_2O_2$ into $H_2O$ and $O_2$, preventing the cascade of free radical production (Chapman *et al.*, 2019). The antioxidant properties of certain secondary metabolites like flavonoids and carotenoids work to lower the levels of ROS in the plant, thus reducing the damage occurring due to stress events. These secondary metabolites include flavonoids, carotenoids, ascorbate, and tocochromanols (vitamin E) (Chapman *et al.*, 2019). The production of antioxidants themselves can change under stress conditions or through the production of ROS (Lee *et al.*, 2023; Shi *et al.*, 2018; Xu *et al.*, 2015; Yang *et al.*, 2018; Zhang *et al.*, 2017). These antioxidants are produced in different pathways, at different locations, and have different functions, protecting the plants from oxidative stress.

### 1.6 Introduction to Flavonoids

Flavonoids are a classification of secondary metabolites with over 9,000 compounds presently known in plants (Corradini *et al.*, 2011; Mierziak *et al.*, 2014). They have a base molecular structure of 2-phenylbenzopyranone with C6-C3-C6 carbon rings (Corradini *et al.*, 2011). These compounds can be further categorized into different classes of flavonoids including flavones, isoflavones, flavonols, anthocyanidins, flavanones, flavanols, chalcones, and aurones based on their molecular structure and stage in the flavonoid biosynthetic process (Chapman *et*

*al.*, 2019; Mierziak *et al.*, 2014; Nix *et al.*, 2017). They are produced in all parts of the plants in the endoplasmic reticulum and have a variety of functions due to their diverse structures (Nix *et al.*, 2017). Flavonoids are known to act as pigments, fragrances, and flavoring agents to appeal to potential pollinators or seed dispersers (Koes *et al.*, 1994; Wink, 2003; Wink, 2016) as well as both promoting and deterring herbivory and oviposition (Feeny *et al.*, 1988; Nuessly *et al.*, 2007; Ohsugi *et al.*, 1985; Tabashnik, 1987; War *et al.*, 2012). Thus, these flavonoids can protect plants from biotic stresses by affecting the attraction, flavor, nutritional content, and toxicity of specific structures (Harbone & Williams, 2000; Gould & Lister, 2006; Mallikarjuna *et al.*, 2004; Simmonds, 2003; Simmonds & Stevenson, 2001; Sosa *et al.*, 2004). Flavonoids also play an important role in the signaling pathway of auxin (Jacobs & Rubery, 1988; Mierziak *et al.*, 2014; Muday & Murphy, 2002; Peer & Murphy, 2007). Additionally, flavonoids can act as a general antibacterial, antifungal, and antiviral agent due to their antioxidant properties in scavenging ROS that build up during infection or by disrupting plant-pathogen interactions (Chan *et al.*, 1978; Hanny, 1980; Hedin *et al.*, 1992; Mace *et al.*, 1978; Rani & Pratyusha, 2013; Shaver & Lukefahr, 1969).

Flavonoids protect from abiotic stresses through maintenance of the redox state of the cells (Mierziak *et al.*, 2014; Zhang *et al.*, 2017). All flavonoids are antioxidants due to their ring structure and conjugated double bonds that gives a low redox potential (Amić *et al.*, 2003; Buettner, 1993; Rice-Evans *et al.*, 1996; Seyoum *et al.*, 2006). The molecular structure of flavonoids allows for single-electron oxidation of free radicals like ROS due to acting as a hydrogen donor by transferring a proton from the A or B ring of the flavonoid (Heim *et al.*, 2002; Mierziak *et al.*, 2014). Modifications of either of these rings and their hydroxyl groups can impact the efficiency of the electron transfer from ROS (Cos *et al.*, 1998; Heim *et al.*, 2002).

### 1.6.1 Flavonoids in cotton

The flavonoid pathway has been extensively studied in plants (Chapman *et al.*, 2019; Gayomba *et al.*, 2016; Winkel-Shirley, 2002). The exact composition of flavonoids varies in different plant species (Tohge *et al.*, 2018). Over 50 flavonoids have been identified in cotton (Ismailov *et al.*, 1994; Nix *et al.*, 2017). Flavonols are the most common class in *G. hirsutum* though no flavones, chalcones, or aurones have been reported (Nix *et al.*, 2017). Floral flavonoid compositions have been the most commonly studied and they contain the majority of the

flavonoids found in cotton. These flavonoid floral profiles have been used to differentiate between *Gossypium* species (Parks, 1965a; Parks 1965b; Parks, 1967). Investigations into leaf reddening have been shown to be due to the accumulation of anthocyanin pigments in order to protect the plant from oxidative stresses (Edreva *et al.*, 2006; Kangatharalingam *et al.*, 2002; Nix *et al.*, 2017). The flavonoid naringenin has been shown to reduce fiber development (Khan *et al.*, 2019; Tan *et al.*, 2013) and flavonoids have been shown to be involved in fiber pigmentation in brown colored cottons (Feng *et al.*, 2013; Lu *et al.*, 2017; Xiao *et al.*, 2014; Yan *et al.*, 2018). The role of flavonoids in plant defense have been regularly studied against a variety of insect pests and fungal pathogens (Chan *et al.*, 1978; Hanny, 1980; Hedin *et al.*, 1992; Mace *et al.*, 1978; Rani & Pratyusha, 2013; Shaver & Lukefahr, 1969).

Investigations into the two colorations of *G. hirsutum* anthers by Hanney looked at the gossypol, flavonoid, and condensed tannin profiles for five cultivars (Hanny, 1980). This paper identified 13 flavonoids found in cotton anthers with gossypetin and quercetin being the major flavonoids regardless of the anther color. Hanney found only minor quantitative differences in the flavonoids between the five lines but reported a higher gossypol percentage in yellow anthers.

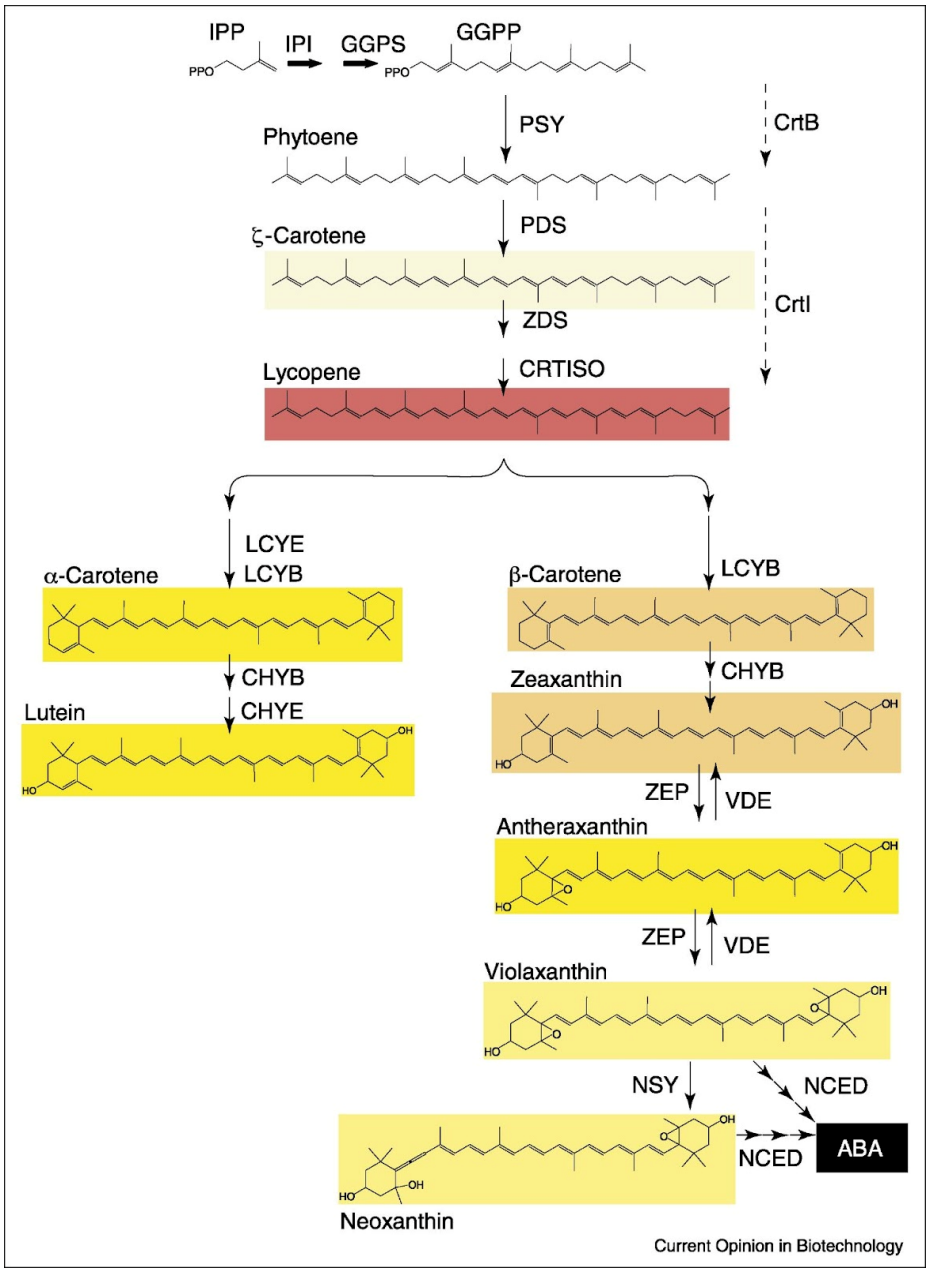
### 1.7 Introduction to Carotenoids

Carotenoids are a group of tetraterpenoid yellow, orange, and red pigments found in plants, animals, and bacteria. Carotenoids possess a double-bond conjugated system called chromophores, which gives the secondary metabolites their pigmentation and stabilization properties (Gruszecki & Strzałka, 2005). Carotenoids can be characterized into two categories, the xanthophylls that contain oxygen functional groups, and carotenes which are pure hydrocarbon molecules (Llorente *et al.*, 2017; Sun *et al.*, 2018). Both groups are synthesized and stored in plastids in plants which act as a sink in the biosynthetic process (Sun *et al.*, 2018). The carotenoid biosynthesis pathway has been heavily studied (Figure 1) (Cazzonelli & Pogson, 2010; Fiore *et al.*, 2006; Llorente *et al.*, 2017; Pankratov *et al.*, 2016; Sun *et al.*, 2018). The first carotenoid, phytoene is a colorless pigment formed through the condensation of two geranylgeranyl diphosphate (GGPP) by phytoene synthase (PSY) through the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway (Cazzonelli & Pogson, 2010; Rodriguez-Concepcion, 2010). This step is reported to be the major rate-limiting step in the carotenoid biosynthesis

pathway. Multiple PSY genes have been found in the genome of some plants with alleles expressed in a tissue specific manner (Welsch *et al.*, 2008). Some PSY genes have been reported to be differentially expressed under abiotic stress (Meier *et al.*, 2011; Qin *et al.*, 2011) and under light (Toledo-Ortiz & Rodríguez-Concepción, 2010).

**Figure 1:** Caroeneoid biosynethesis pathway as described in Figure 3 from Tanaka & Ohmiya, 2008. Labled "Carotenoid biosynthesis pathway. Enzymes are abbreviated as follows: CrtB, bacterial phytoene synthase; CrtI, bacterial phytoene desaturase; DXPS, 1-deoxy-d-xylulose-5-phosphate synthase; DXR, 1-deoxy-d-xylulose-5-phosphate reductoisomerase; IPI, isopentenyl pyrophosphate isomerase; GGDP, geranylgeranyl diphosphate synthase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ζ-carotene desaturase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; CHYB, β-ring hydroxylase; CHYE, ε-ring hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; CRTISO, carotenoid isomerase; NSY, neoxanthin synthase; NCED, 9-cis-epoxycarotenoid dioxygenase."

**Figure 1** (continued).




### 1.7.1 Carotenoids in plants

Carotenoids play a variety of important roles in plant development and signaling. They act as pigments and aromatic compounds to aid in pollinator attraction and seed dispersal (Bradshaw & Schemske, 2003; Simkin *et al.*, 2004; Walter *et al.*, 2010). They are also used as precursors in the synthesis of signaling molecules including ABA and strigolactone (Al-Babili & Bouwmeester, 2015; Du *et al.*, 2010; Rani *et al.*, 2008; Schwartz *et al.*, 1997). Some carotenoids

like xanthophylls have been shown to be rate limiting to ABA synthesis and thus play a role in the plant's tolerance to abiotic stresses like drought and high temperatures as well as other ABA developmental processes like seed dormancy and maturation (Chinnusamy *et al*., 2008; Du *et al.*, 2010). Strigolactone in plants can influence bud growth, shoot branching, and seed germination (Gomez-Roldan *et al.*, 2008; Matusova *et al.*, 2005; Umehara *et al.*, 2008). Carotenoid derived signaling molecules also are involved in mycorrhizal symbiotic relationships in the plants root structure (Akiyama *et al.*, 2005; Walter *et al.*, 2010). Carotenoids also act as precursors to vitamin A making them an important area of study in plants grown for human food, medicine, and cosmetics (Johnson, 2002; Krinsky & Johnson 2005). Many crops have been engineered to increase carotenoid content to increase vitamin A content in human foods to combat vitamin A deficiency including the well known Golden Rice (Datta *et al.*, 2006; Schaub *et al.*, 2005; Ye *et al.*, 2000). Carotenoids are an important part of the light harvesting complexes in the photosystems, absorbing light from a range in the spectral region (Pogson *et al.*, 2005; Polívka & Frank, 2010). The most well known function of carotenoids is their role in the photosystem machinery and photoprotection through ROS scavenging (Hashimoto *et al.*, 2016; Liu *et al.*, 2004).

### 1.7.2 Carotenoids in abiotic stress tolerance

It was reported that under high light intensity, singlet oxygen ROS can be produced when $O_2$ interacts with chlorophyll that has been excited to a triplet state (3 Chl) (Pospíšil, 2012). Other ROS species that can accumulate under overexcitation conditions are superoxides, hydrogen peroxide, and hydroxyl radicals (Osmond, 1994; Wise & Naylor, 1987). The carotenoids xanthophylls and zeaxanthin have been shown to reduce singlet oxygen formation through 3 Chl quenching and non-photochemical quenching (NPQ) (Havaux *et al.*, 2000; Krinsky, 1979; Latowski *et al.*, 2011; Pospíšil, 2012; Rutherford *et al.*, 1981). Temperature and light intensity stress often occur simultaneously, with both potentially damaging photosystem II and reducing the photosynthetic efficiency of the plant (Baker, 2008). This suggests that carotenoids play an important role in managing photo-oxidative damage due to abiotic stress. Lipid degradation can be decreased by the presence of carotenoids in the thylakoid membrane in cells (Havaux *et al.*, 1991; Sarry *et al.*, 1994; Woodall *et al.*, 1997). Studies have shown that under temperature stress, carotenoid content can increase, showing a decrease in the chlorophyll;

carotenoid ratio which recovered after the stressor was removed (Lee *et al.*, 2023). In sweet potatoes (*Ipomoea batatas*) a transgenic line with an Orange (Or) overexpression increased the carotenoid accumulation of up to 23-fold in cali and were associated with increased salt and heat stress tolerance (Kim *et al.*, 2019; Kim *et al.*, 2021). Carotenoids are able to scavenge ROS due to their antioxidant nature, limiting the oxidative damage to the plant's membranes, proteins, and DNA during abiotic stress in plants (Chapman *et al.*, 2019; Jørgensen & Skibsted, 1993; Latowski *et al.*, 2011; Liguori *et al.*, 2017).

### 1.7.3 Carotenoids in cotton

Recent carotenoid studies in cotton have shown an association of increased carotenoid content with salinity stress (Cai *et al.*, 2021; Zafar *et al.*, 2022), high temperature stress (Manan *et al.*, 2021), low temperature stress (Snider *et al.*, 2022), boron deficiency (de Souza *et al.*, 2023), and drought stress (Cai *et al.*, 2021; Ni *et al.*, 2022). A study by Cai *et al*. (2021) found 60 carotenoid cleavage oxygenases (CCOs) in three cotton species. CCOs are used in the cleavage of carotenoids and are needed in the biosynthesis of ABA (Gavassi *et al.*, 2021; Schwartz *et al.*, 2004; Vishwakarma *et al.*, 2017). Cai *et al.* (2021) showed an association with some of the identified CCOs in drought and salt stress. Much of current cotton research is focusing on how carotenoids affect fiber development and elongation. Transgenic cotton lines were developed with a positive regulator of carotenoid accumulation $GhOR1^{Del}$ and a fiber specific promoter which showed an increase in carotenoids as well as ABA and ethylene in the elongating fibers (Zeng *et al.*, 2023). ABA is known to play a role in controlling the elongation of cotton fibers (Beasley & Ting, 1973; Gokani *et al.*, 1998; Pierce *et al.*, 2019; Zeng *et al.*, 2023). Zeng *et al.* (2023) supported this in their transgenic lines showing an increase in fiber length, strength, and fitness as compared to the non-transgenic control. This suggests that an increase in carotenoid content in fibers can increase the ABA and aid in cotton fiber elongation and quality through promotion of ethylene synthesis (Zeng *et al.*, 2023). To increase the provitamin A levels in cotton seeds and thus cotton seed oil, transgenic lines have been developed by manipulating carotenoid biosynthesis in developing cottonseeds through PSY (Yao *et al.*, 2018). These transgenic lines have shown a 6-fold increase in carotenoid content, specifically β-carotene, in extracted cottonseed oil. Combining this with previous transgenic lines showing a significant reduction in the toxic gossypol in seeds (Gao *et al.*, 2022) could produce cotton seed meal that is

high in provitamin A and safe for human consumption. Cotton carotenoid profiles in different tissues have been found (Thompson *et al.*, 1968) though little has been done in cotton anthers or pollen specifically. De Souza *et al.* (2023) noted a 30% increase in carotenoid content in anther samples when experiencing boron deficiency suggesting carotenoid content in male reproductive tissues can be influenced by abiotic stresses. Further studies on the flavonoid and carotenoid profiles of yellow and cream pigments in only the pollen grains have not been performed. Knowledge of the antioxidant properties of pigmented pollen grains may reveal a specialized method of heat tolerance on cotton reproductive structures. This study aims to identify the genetic mechanisms behind yellow pollen pigmentation in *G. hirsutum*. As floral organs are one of the first to show signs of heat damage in cotton and are also intrinsically important to the production of fibers, antioxidants in pollen grains may offer a novel method of temperature tolerance. Breeders may incorporate this easily visible trait into their breeding programs to add an additional level of abiotic stress tolerance in cotton cultivars.

## REFERENCES

Abrol, Y. P., & Ingram, K. T. (1996). Effects of higher day and night temperatures on growth and yields of some crop plants. In: *Global climate change and agricultural production: Direct and indirect effects of changing hydrological, pedological and plant physiological processes.* eds F. A. Bazzaz and W. G. Sombroek *Wiley, Chichester*, 124-140.

Ahumada, D. (2023). Cotton Seedling Diseases. https://content.ces.ncsu.edu/cotton-seedling-diseases. Accessed July 2023.

Ahumada, D., & Gorny, A. (2019). Root-Knot Nematode of Cotton. https://content.ces.ncsu.edu/cotton-root-knot-nematodes. Accesses July 2023.

Akiyama, K., Matsuzaki, K. I., & Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature*, *435*(7043), 824-827. https://doi.org/10.1038/nature03608

Alavez, V., Cuervo-Robayo, Á. P., Martínez-Meyer, E., & Wegier, A. (2021). Eco-geography of feral cotton: a missing piece in the puzzle of gene flow dynamics among members of *Gossypium hirsutum* primary gene pool. *Frontiers in Ecology and Evolution*, *9*, 653271. https://doi.org/10.3389/fevo.2021.653271

Al-Babili, S., & Bouwmeester, H. J. (2015). Strigolactones, a novel carotenoid-derived plant hormone. *Annual review of plant biology*, *66*, 161-186. https://doi.org/10.1146/annurev-arplant-043014-114759

Allen, R. D., & Aleman, L. (2011). Abiotic stress and cotton fiber development. *Stress physiology in cotton*, 149.

Allen, T. (2011). ALERT: Bacterial blight of cotton. Mississippi Crops. http://www.mississippi-crops.com/2011/07/15/alert-bacterial-blight-of-cotton/

Amić, D., Davidović-Amić, D., Bešlo, D., & Trinajstić, N. (2003). Structure-radical scavenging activity relationships of flavonoids. *Croatica chemica acta, 76*(1), 55-61.

Atkinson, G. F. (1892). Some diseases of cotton. Alabama Agricultural Experiment Station. http://hdl.handle.net/11200/2696

Azhar, M. T., Wani, S. H., Chaudhary, M. T., Jameel, T., Kaur, P., & Du, X. (2020). Heat tolerance in cotton: morphological, physiological, and genetic perspectives. In *Heat Stress Tolerance in Plants: Physiological, Molecular and Genetic Perspectives*, 1-22. https://doi.org/10.1002/9781119432401.ch1

Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annual Review of Plant Biology, 59*, 89-113. https://doi.org/10.1146/annurev.arplant.59.032607.092759

Baloch, M. J., Lakho, A. R., Bhutto, H., Memon, A. M., Panhwar, G. N., & Soomro, A. H. (2000). Estimates of combining ability and genetic parameters for yield and fibre traits in Upland cotton. *Pakistan Journal of Biological Sciences (3)*1183-1186.

Barua, D., Downs, C. A., & Heckathorn, S. A. (2003). Variation in chloroplast small heat-shock protein function is a major determinant of variation in thermotolerance of photosynthetic electron transport among ecotypes of *Chenopodium album*. *Functional Plant Biology*, *30*(10), 1071-1079. https://doi.org/10.1071/FP03106

Baxter, A., Mittler, R., & Suzuki, N. (2014). ROS as key players in plant stress signaling. *Journal of experimental botany*, *65*(5), 1229-1240. https://doi.org/10.1093/jxb/ert375

Beasley, C. A., & Ting, I. P. (1973). The effects of plant growth substances on in vitro fiber development from fertilized cotton ovules. *American Journal of Botany*, *60*(2), 130-139. https://doi.org/10.1002/j.1537-2197.1973.tb10209.x

Beasley, J. O. (1940). The origin of American tetraploid *Gossypium* species. *The American Naturalist*, *74*(752), 285-286. https://doi.org/10.1086/280895

Bellaloui, N., Stetina, S. R., & Turley, R. B. (2015). Cottonseed protein, oil, and mineral status in near-isogenic *Gossypium hirsutum* cotton lines expressing fuzzy/linted and fuzzless/linted seed phenotypes under field conditions. *Frontiers in plant science*, *6*, 137. https://doi.org/10.3389/fpls.2015.00137

Blanvillain, R., Young, B., Cai, Y. M., Hecht, V., Varoquaux, F., Delorme, V., Lancelin, J. M., Delseny, M., & Gallois, P. (2011). The *Arabidopsis* peptide kiss of death is an inducer of

programmed cell death. *The EMBO journal*, *30*(6), 1173-1183. https://doi.org/10.1038/emboj.2011.14

Bowen, J., Lay-Yee, M., Plummer, K. I. M., & Ferguson, I. A. N. (2002). The heat shock response is involved in thermotolerance in suspension-cultured apple fruit cells. *Journal of Plant Physiology*, *159*(6), 599-606. https://doi.org/10.1078/0176-1617-0752

Bowman, D. T., May, O. L., & Calhoun, D. S. (1996). Genetic base of upland cotton cultivars released between 1970 and 1990. *Crop Science*, *36*(3), 577-581. https://doi.org/10.2135/cropsci1996.0011183X003600030008x

Bradshaw Jr, H. D., & Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, *426*(6963), 176-178. https://doi.org/10.1038/nature02106

Brown P.W. (2001). Heat stress and cotton yields in Arizona. In: *Cotton: A College of Agriculture Report.* AZ1224. College of Agriculture, Univ. of Arizona, Tucson, AZ.

Brown, P. W. (2008). Cotton Heat Stress. Arizona Cooperative Extension, Tucson, AZ. https://extension.arizona.edu/sites/extension.arizona.edu/files/pubs/az1448.pdf

Brown, P.W. & C.A. Zeiher. (1997). Cotton heat stress. In: *Cotton: A College of Agriculture Report.* Series P-108. College of Agriculture, Univ. of Arizona, Tucson, AZ. p. 91-104.

Brown, P.W. & C.A. Zeiher. (1998a). A model to estimate canopy temperature in the desert southwest. In: *Proceedings of 1998 Beltwide Cotton Conferences*, San Diego, CA. Natl. Cotton Council. Memphis, TN. Vol. 2: 1734.

Brown, P.W. & C.A. Zeiher. (1998b). *Development of an effective screen for identifying cotton cultivars tolerant to elevated night temperatures during the monsoon*. Final Report: Project 96-342AZ. Cotton Inc., Raleigh, NC.

Brown, P.W. & C.A. Zeiher. 2008. Impact of short duration heat stress on the reproductive performance of Upland cotton. Submitted to: *Cotton Science*.

Brown, P.W., C.A. Zeiher & J. Silvertooth. 1995. Response of upland cotton to elevated night temperature I. In: *Proceedings of 1995 Beltwide Cotton Conferences*, San Antonio, TX. National Cotton Council, Memphis, TN. Vol. 2: 1129.

Brown, W. L. (1983). Genetic diversity and genetic vulnerability—an appraisal. *Economic botany*, *37*, 4-12. https://doi.org/10.1007/BF02859301

Brubaker, C.L., Bourland, F.M., & Wendel, J.F. 1999. The origin and domestication of cotton. In: *Smith, C.W., and J.T. Cothren, Eds., Cotton: origin, history, technology and production*. New York: John Wiley & Sons, pp. 23–32.

Buettner, G. R. (1993). The pecking order of free radicals and antioxidants: lipid peroxidation, α-tocopherol, and ascorbate. *Archives of biochemistry and biophysics*, *300*(2), 535-543. https://doi.org/10.1006/abbi.1993.1074

Burke, J. J., & Chen, J. (2015). Enhancement of reproductive heat tolerance in plants. *PLoS One*, *10*(4), e0122933. https://doi.org/10.1371/journal.pone.0122933

Burke, J. J., Hatfield, J. L., Klein, R. R., & Mullet, J. E. (1985). Accumulation of heat shock proteins in field-grown cotton. *Plant physiology*, *78*(2), 394-398. https://doi.org/10.1104/pp.78.2.394

Burke, J. J., Velten, J., & Oliver, M. J. (2004). In vitro analysis of cotton pollen germination. *Agronomy journal*, *96*(2), 359-368. https://doi.org/10.2134/agronj2004.3590

Cai, X., Jiang, Z., Tang, L., Zhang, S., Li, X., Wang, H., Liu, C., Chi, J., Zhang, X., & Zhang, J. (2021). Genome-wide characterization of carotenoid oxygenase gene family in three cotton species and functional identification of *GaNCED3* in drought and salt stress. *Journal of Applied Genetics*, *62*(4), 527-543. https://doi.org/10.1007/s13353-021-00634-3

Cazzonelli, C. I., & Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends in plant science*, *15*(5), 266-274. https://doi.org/10.1016/j.tplants.2010.02.003

Chan, B. G., Waiss Jr, A. C., Binder, R. G., & Elliger, C. A. (1978). Inhibition of lepidopterous larval growth by cotton constituents. *Entomologia Experimentalis et Applicata*, *24*(3), 294-300. https://doi.org/10.1111/j.1570-7458.1978.tb02785.x

Chapman, J. M., Muhlemann, J. K., Gayomba, S. R., & Muday, G. K. (2019). RBOH-dependent ROS synthesis and ROS scavenging by plant specialized metabolites to modulate plant development and stress responses. *Chemical research in toxicology*, *32*(3), 370-396. https://doi.org/10.1021/acs.chemrestox.9b00028

Chen, T., Saeed, Q., He, Z., & Lu, L. (2019). Transmission efficiency of *Cotton leaf curl Multan virus* by three cryptic species of *Bemisia tabaci* complex in cotton cultivars. *PeerJ*, *7*, e7788. https://doi.org/10.7717/peerj.7788

Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., Ding, M., Ye, W., Kirkbride, R.C., Jenkins, J., Plott, C., Lovell, J., Lin, Y. M., Vaughn, R., Liu, B., Simpson, S., Scheffler, B. E., Wen, L., Saski, C. A., Grover, C. E., Hu, G., Conover, J. L., Carlson, J. W., Shu, S., Boston, L. B., Williams, M., Peterson, D. G., McGee, K., Jones, D. C., Wendel, J. F., Stelly, D. M., Grimwood, J., & Schmutz, J. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nature genetics*, *52*(5), 525-533. https://doi.org/10.1038/s41588-020-0614-5

Chen, Z., Zhao, N., Li, S., Grover, C. E., Nie, H., Wendel, J. F., & Hua, J. (2017). Plant mitochondrial genome evolution and cytoplasmic male sterility. *Critical reviews in plant sciences*, *36*(1), 55-69. https://doi.org/10.1080/07352689.2017.1327762

Chinnusamy, V., Gong, Z., & Zhu, J. K. (2008). Abscisic acid-mediated epigenetic processes in plant development and stress responses. *Journal of integrative plant biology*, *50*(10), 1187-1195. https://doi.org/10.1111/j.1744-7909.2008.00727.x

Christmann, A., Hoffmann, T., Teplova, I., Grill, E., & Müller, A. (2005). Generation of active pools of abscisic acid revealed by in vivo imaging of water-stressed Arabidopsis. *Plant physiology*, *137*(1), 209-219. https://doi.org/10.1104/pp.104.053082

Cianchetta, A. N., & Davis, R. M. (2015). Fusarium wilt of cotton: Management strategies. *Crop Protection*, *73*, 40-44. https://doi.org/10.1016/j.cropro.2015.01.014

Ciancio, A. (2021). Observations on a Novel Bacterial Pathogen of Root-Knot Nematodes (*Meloidogyne* spp.). *Pathogens*, *10*(10), 1226. https://doi.org/10.3390/pathogens10101226

Cornish, K., Radin, J. W., Turcotte, E. L., Lu, Z., & Zeiger, E. (1991). Enhanced photosynthesis and stomatal conductance of Pima cotton (*Gossypium barbadense* L.) bred for increased yield. *Plant Physiology*, *97*(2), 484-489. https://doi.org/10.1104/pp.97.2.484

Corradini, E., Foglia, P., Giansanti, P., Gubbiotti, R., Samperi, R., & Laganà, A. (2011). Flavonoids: chemical properties and analytical methodologies of identification and quantitation in foods and plants. *Natural product research*, *25*(5), 469-495. https://doi.org/10.1080/14786419.2010.482054

Cos, P., Ying, L., Calomme, M., Hu, J. P., Cimanga, K., Van Poel, B., Pieters, L., Vlietinck, A. J., & Berghe, D. V. (1998). Structure− activity relationship and classification of flavonoids as inhibitors of xanthine oxidase and superoxide scavengers. *Journal of natural products*, *61*(1), 71-76. https://doi.org/10.1021/np970237h

Cothren, J.T. (ed.) (1999). Physiology of the cotton plant. In: *Cotton: Origin, History, Technology, and Production*, 207–268. Wiley Blackwell.

Cottee, N. S., Wilson, I. W., Tan, D. K., & Bange, M. P. (2013). Understanding the molecular events underpinning cultivar differences in the physiological performance and heat tolerance of cotton (*Gossypium hirsutum*). *Functional Plant Biology*, *41*(1), 56-67. https://doi.org/10.1071/FP13140

Cotton incorporated a. Cotton Weed Management. https://www.cottoninc.com/cotton-production/ag-research/weed-management/. Accessed July 2023.

Cotton incorporated b. Identification and Management of Fusarium Wilt Race 4. https://www.cottoninc.com/cotton-production/ag-research/plant-pathology/identification-and-management-of-fusarium-wilt-race-4/. Accessed July 2023.

Cotton incorporated c. Plant Pathology Research. https://www.cottoninc.com/cotton-production/ag-research/plant-pathology/. Accessed July 2023.

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., & Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *Annual review of plant biology*, *61*, 651-679. https://doi.org/10.1146/annurev-arplant-042809-112122

da Silva, M. B., Davis, R. F., Doan, H. K., Nichols, R. L., Kemerait, R. C., Halpern, H. C., Brewer, M. T., Jagdale, G., & Chee, P. W. (2019). Fusarium wilt of cotton may commonly result from the interaction of *Fusarium oxysporum* f. sp. *vasinfectum* with *Belonolaimus longicaudatus*. *Journal of Nematology*, *51*(1), 1-10. https://doi.org/10.21307/jofnem-2019-015

Damp, J. E., & Pearsall, D. M. (1994). Early cotton from coastal Ecuador. *Economic Botany*, *48*, 163-165. https://doi.org/10.1007/BF02908209

Datta, K., Rai, M., Parkhi, V., Oliva, N., Tan, J., & Datta, S. K. (2006). Improved 'golden' indica rice and post-transgeneration enhancement of metabolic target products of carotenoids (β-carotene) in transgenic elite cultivars (IR64 and BR29). *Current Science*, 935-939. http://www.jstor.org/stable/24094294

de Souza, J. P., de Mello Prado, R., Campos, C. N. S., Junior, G. D. S. S., Costa, M. G., de Pádua Teixeira, S., & Gratão, P. L. (2023). Silicon modulate the non-enzymatic antioxidant defense system and oxidative stress in a similar way as boron in boron-deficient cotton flowers. *Plant Physiology and Biochemistry*, *197*, 107594. https://doi.org/10.1016/j.plaphy.2023.02.024

Delannoy, E., Lyon, B. R., Marmey, P., Jalloul, A., Daniel, J. F., Montillet, J. L., Essenberg, M., & Nicole, M. (2005). Resistance of cotton towards *Xanthomonas campestris* pv. *malvacearum*. *Annual Review of Phytopathology*, *43*, 63-82. https://doi.org/10.1146/annurev.phyto.43.040204.140251

Demirel, U., Gür, A., Can, N., & Memon, A. R. (2014). Identification of heat responsive genes in cotton. *Biologia plantarum*, *58*(3), 515-523. https://doi.org/10.1007/s10535-014-0414-9

DeRidder, B. P., & Salvucci, M. E. (2007). Modulation of Rubisco activase gene expression during heat stress in cotton (*Gossypium hirsutum* L.) involves post-transcriptional mechanisms. *Plant Science, 172*(2), 246-254. https://doi.org/10.1016/j.plantsci.2006.08.014

Digonnet, C., Aldon, D., Leduc, N., Dumas, C., & Rougier, M. (1997). First evidence of a calcium transient in flowering plants at fertilization. *Development, 124*(15), 2867-2874. https://doi.org/10.1242/dev.124.15.2867

Du, H., Wang, N., Cui, F., Li, X., Xiao, J., & Xiong, L. (2010). Characterization of the β-carotene hydroxylase gene *DSM2* conferring drought and oxidative stress resistance by increasing xanthophylls and abscisic acid synthesis in rice. *Plant physiology, 154*(3), 1304-1318. https://doi.org/10.1104/pp.110.163741

Edreva, A., Gürel, A., Gesheva, E., & Hakerlerler, H. (2002). Reddening of cotton (*Gossypium hirsutum* L.) leaves. *Biologia plantarum, 45*, 303-306. https://doi.org/10.1023/A:1015121428714

Edula, S. R., Bag, S., Milner, H., Kumar, M., Suassuna, N. D., Chee, P. W., Kemerait, R. C., Hand, L. C., Snider, J. L., Srinivasan, R., & Roberts, P. M. (2023). Cotton leafroll dwarf disease: An enigmatic viral disease in cotton. *Molecular Plant Pathology*. https://doi.org/10.1111/mpp.13335

El-Zik, K. M. (1985). Integrated control of Verticillium wilt of cotton. *Plant Disease, 69*(12), 1025-1032. https://doi.org/10.1094/PD-69-1025

El-Zik, K. M., & Thaxton, P. M. (1994). Breeding for resistance to bacterial blight of cotton in relation to races of the pathogen. In: *Challenging the future: World Cotton Research Conference-1*, Melbourne.

FAO 2021. Recent trends and prospects in the world cotton market and policy developments. Food and Agriculture Organization of the United Nations, Markets and Trade Division - Economic and Social Development Stream Markets. https://www.fao.org/3/cb4589en/cb4589en.pdf

Feeny, P., Sachdev, K., Rosenberry, L., & Carter, M. (1988). Luteolin 7-O-(6 ″-O-malonyl)-β-d-glucoside and *trans*-chlorogenic acid: Oviposition stimulants for the black swallowtail butterfly. *Phytochemistry*, *27*(11), 3439-3448. https://doi.org/10.1016/0031-9422(88)80745-3

Feng, H., Tian, X., Liu, Y., Li, Y., Zhang, X., Jones, B. J., Sun, Y., & Sun, J. (2013). Analysis of flavonoids and the flavonoid structural genes in brown fiber of upland cotton. *PLoS One*, *8*(3). https://doi.org/10.1371/journal.pone.0058820

Ferreira, I., (2019). Five ways cotton production can help achieve zero hunger. https://medium.com/world-food-programme-insight/five-ways-cotton-production-can-help-achieve-zero-hunger-4ece9d3614ae. Accessed October 2023.

Fiore, A., Dall'Osto, L., Fraser, P. D., Bassi, R., & Giuliano, G. (2006). Elucidation of the β-carotene hydroxylation pathway in *Arabidopsis thaliana*. *FEBS letters*, *580*(19), 4718-4722. https://doi.org/10.1016/j.febslet.2006.07.055

Foyer, C. H., & Noctor, G. (2020). Redox homeostasis and signaling in a higher-CO2 world. *Annual Review of Plant Biology*, *71*, 157-182. https://doi.org/10.1146/annurev-arplant-050718-095955

Fryxell, P. A. (1979). *The natural history of the cotton tribe (Malvaceae, tribe Gossypieae)*. Texas A & M University Press.

Gao, W., Zhu, X., Ding, L., Xu, B., Gao, Y., Cheng, Y., Dai, F., Liu, B., Si, Z., Fang, L., Guan, X., Zhu, S., & Hu, Y. (2022). Development of the engineered "glanded plant and glandless seed" cotton. *Food Chemistry: Molecular Sciences*, *5*, 100130. https://doi.org/10.1016/j.fochms.2022.100130

Gavassi, M. A., Silva, G. S., da Silva, C. D. M. S., Thompson, A. J., Macleod, K., Oliveira, P. M. R., Cavalheiro, M. F., Domingues, D. S., & Habermann, G. (2021). NCED expression is related to increased ABA biosynthesis and stomatal closure under aluminum stress. *Environmental and Experimental Botany*, *185*, 104404. https://doi.org/10.1016/j.envexpbot.2021.104404

Gayomba, S. R., Watkins, J. M., & Muday, G. K. (2016) Flavonols Regulate Plant Growth and Development through Regulation of Auxin Transport and Cellular Redox Status. In: *Recent Advances in Polyphenol Research* (Yoshida, K., Cheynier, V., and Quideau, S., Eds.) John Wiley & Sons: Chichester, U.K.

Gokani, S. J., Kumar, R., & Thaker, V. S. (1998). Potential role of abscisic acid in cotton fiber and ovule development. *Journal of plant growth regulation*, *17*, 1-5. https://doi.org/10.1007/PL00007005

Gomez-Roldan, V., Fermas, S., Brewer, P. B., Puech-Pagès, V., Dun, E. A., Pillot, J. P., Letisse, F., Matusova, R., Danoun, S., Portais, J. C., Bouwmeester, H., Bécard, G., Beveridge, C. A., Rameau, C., & Rochange, S. F. (2008). Strigolactone inhibition of shoot branching. *Nature*, *455*(7210), 189-194. https://doi.org/10.1038/nature07271

Gould, K. S., & Lister, C. (2006). Flavonoid functions in plants. In *Flavonoids: Chemistry, biochemistry and applications*, 397-441.

Gross, B. L., & Strasburg, J. L. (2010). Cotton domestication: dramatic changes in a single cell. *BMC biology*, *8*(1), 1-3. https://doi.org/10.1186/1741-7007-8-137

Gruszecki, W. I., & Strzałka, K. (2005). Carotenoids as modulators of lipid membrane physical properties. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1740*(2), 108-115. https://doi.org/10.1016/j.bbadis.2004.11.015

Hake, K., & Silvertooth, J. (1990). High Temperature Effects on Cotton. https://www.cotton.org/tech/physiology/cpt/plantphysiology/upload/CPT-July90-REPOP.pdf. Accessed July 2023.

Hanny, B. W. (1980). Gossypol, flavonoid, and condensed tannin content of cream and yellow anthers of five cotton (*Gossypium hirsutum* L.) cultivars. *Journal of Agricultural and Food Chemistry*, *28*(3), 504-506. https://doi.org/10.1021/jf60229a021

Harborne, J. B., & Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochemistry*, *55*(6), 481-504. https://doi.org/10.1016/S0031-9422(00)00235-1

Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., & Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in

plants. *International journal of molecular sciences*, *14*(5), 9643-9684. https://doi.org/10.3390/ijms14059643

Hashimoto, H., Uragami, C., & Cogdell, R. J. (2016). Carotenoids and photosynthesis. *Carotenoids in nature: Biosynthesis, regulation and function*, 111-139. https://doi.org/10.1007/978-3-319-39126-7_4

Havaux, M., Bonfils, J. P., Lutz, C., & Niyogi, K. K. (2000). Photodamage of the photosynthetic apparatus and its dependence on the leaf developmental stage in the *npq1* Arabidopsis mutant deficient in the xanthophyll cycle enzyme violaxanthin de-epoxidase. *Plant Physiology*, *124*(1), 273-284. https://doi.org/10.1104/pp.124.1.273

Havaux, M., Gruszecki, W. I., Dupont, I., & Leblanc, R. M. (1991). Increased heat emission and its relationship to the xanthophyll cycle in pea leaves exposed to strong light stress. *Journal of Photochemistry and Photobiology B: Biology*, *8*(4), 361-370. https://doi.org/10.1016/1011-1344(91)80112-U

Hedin, P. A., Jenkins, J. N., & Parrott, W. L. (1992). Evaluation of flavonoids in *Gossypium arboreum* (L.) cottons as potential source of resistance to tobacco budworm. *Journal of chemical ecology*, *18*, 105-114. https://doi.org/10.1007/BF00993746

Heim, K. E., Tagliaferro, A. R., & Bobilya, D. J. (2002). Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. *The Journal of nutritional biochemistry*, *13*(10), 572-584. https://doi.org/10.1016/S0955-2863(02)00208-5

Hodges, H. F., Reddy K. R., McKinion J. M., & Reddy V. R. (1993). *Temperature effects on cotton.* Mississippi State University Experimental Station Bulletin.

Hofmann, W. C., Kittock, D. L., & Alemayehu, M. (1988). Planting seed density in relation to cotton emergence and yield. *Agronomy Journal*, *80*(5), 834-836. https://doi.org/10.2134/agronj1988.00021962008000050030x

Hsieh, K., & Huang, A. H. (2007). Tapetosomes in *Brassica* tapetum accumulate endoplasmic reticulum–derived flavonoids and alkanes for delivery to the pollen surface. *The Plant Cell*, *19*(2), 582-596. https://doi.org/10.1105/tpc.106.049049

Hu, G., Grover, C. E., Jareczek, J., Yuan, D., Dong, Y., Miller, E., Conover, J.L., & Wendel, J. F. (2021). Evolution and diversity of the cotton genome. In *Cotton precision breeding* (pp. 25-78). Cham: Springer International Publishing. pp. 25–78. https://doi.org/10.1007/978-3-030-64504-5_2

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., Baruch, K., Rang, D., Liu, X., Ruan, Y. L., Rahman, M. U., Han, J., Wang, K., Wang, Q., Wu, H., Mei, G., Zang, Y., Han, Z., Xu, C., Shen, W., Yang, D., Si, Z., Dai, F., Zou, L., Huang, F., Bai, Y., Zhang, Y., Brodt, A., Ben-Hamo, H., Zhu, X., Zhou, B., Guan, X., Zhu, S., Chen, X., & Zhang, T. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature genetics*, *51*(4), 739-748. https://doi.org/10.1038/s41588-019-0371-5

Huang, G., Wu, Z., Percy, R. G., Bai, M., Li, Y., Frelichowski, J. E., Hu, J., Wang, K., Yu, J.Z., & Zhu, Y. (2020). Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nature genetics*, *52*(5), 516-524. https://doi.org/10.1038/s41588-020-0607-4

Idso, S. B., Clawson, K. L., & Anderson, M. G. (1986). Foliage temperature: Effects of environmental factors with implications for plant water stress assessment and the CO2/climate connection. *Water Resources Research*, *22*(12), 1702-1716. https://doi.org/10.1029/WR022i012p01702

Innes, N. L. (1983). Bacterial blight of cotton. Biological Reviews, 58(1), 157–176. https://doi.org/10.1111/j.1469-185X.1983.tb00383.x

Ismailov, A. I., Karimdzhanov, A. K., Islambekov, S. Y., & Rakhimkhanov, Z. B. (1994). Flavonoids of the cotton plant and plants close to it. *Chemistry of Natural Compounds*, *30*, 1-14. https://doi.org/10.1007/BF00638411

Jacobs, M., & Rubery, P. H. (1988). Naturally occurring auxin transport regulators. *Science*, *241*(4863), 346-349. https://doi.org/10.1126/science.241.4863.346

Johnson, E. J. (2002). The role of carotenoids in human health. *Nutrition in clinical care*, *5*(2), 56-65. https://doi.org/10.1046/j.1523-5408.2002.00004.x

Jørgensen, K., & Skibsted, L. H. (1993). Carotenoid scavenging of radicals. Effect of carotenoid structure and oxygen partial pressure on antioxidative activity. *Zeitschrift fur Lebensmittel-untersuchung und-forschung*, *196*(5), 423-429. https://doi.org/10.1007/bf01190806

Juan, C. A., Pérez de la Lastra, J. M., Plou, F. J., & Pérez-Lebeña, E. (2021). The chemistry of reactive oxygen species (ROS) revisited: outlining their role in biological macromolecules (DNA, lipids and proteins) and induced pathologies. *International Journal of Molecular Sciences*, *22*(9), 4642. https://doi.org/10.3390/ijms22094642

Kangatharalingam, N., Pierce, M. L., Bayles, M. B., & Essenberg, M. (2002). Epidermal anthocyanin production as an indicator of bacterial blight resistance in cotton. *Physiological and Molecular Plant Pathology*, *61*(3), 189-195. https://doi.org/10.1006/pmpp.2002.0434

Kemerait, B., Allen, T., Lu, S., Rothrock, C., Faske, T., Woorward, J., Wheeler, T., Isakeit, T., Bart, R., Phillips, A., Lawrence, K., Hagan, A., Price, P., Mehl, H., Dufault, N., Kelly, H., & Nichols, R. (2017). Identification and management of bacterial blight of cotton. Cotton Incorporated. https://www.cottoninc.com/cotton-production/ag-research/plant-pathology/management-bacterial-blight-cotton/

Khan, A. Q., Li, Z., Ahmed, M. M., Wang, P., Zhang, X., & Tu, L. (2019). Eriodictyol can modulate cellular auxin gradients to efficiently promote in vitro cotton fibre development. *BMC plant biology*, *19*(1), 1-15. https://doi.org/10.1186/s12870-019-2054-x

Kim, S. E., Kim, H. S., Wang, Z., Ke, Q., Lee, C. J., Park, S. U., Ahn, M. J., & Kwak, S. S. (2019). A single amino acid change at position 96 (Arg to His) of the sweetpotato Orange protein leads to carotenoid overaccumulation. *Plant Cell Reports*, *38*, 1393-1402. https://doi.org/10.1007/s00299-019-02448-4

Kim, S. E., Lee, C. J., Park, S. U., Lim, Y. H., Park, W. S., Kim, H. J., Ahn, M. J., Kwak, S.S., & Kim, H. S. (2021). Overexpression of the golden SNP-carrying *Orange* gene enhances carotenoid accumulation and heat stress tolerance in sweetpotato plants. *Antioxidants*, *10*(1), 51. https://doi.org/10.3390/antiox10010051

Koes, R. E., Quattrocchio, F., & Mol, J. N. (1994). The flavonoid biosynthetic pathway in plants: function and evolution. *BioEssays*, *16*(2), 123-132. https://doi.org/10.1002/bies.950160209

Krinsky, N. I. (1979). Carotenoid protection against oxidation. *Carotenoids°C 5*, 649-660. https://doi.org/10.1016/B978-0-08-022359-9.50018-7

Krinsky, N. I., & Johnson, E. J. (2005). Carotenoid actions and their relation to health and disease. *Molecular aspects of medicine*, *26*(6), 459-516. https://doi.org/10.1016/j.mam.2005.10.001

Latowski, D., Kuczyńska, P., & Strzałka, K. (2011). Xanthophyll cycle–a mechanism protecting plants against oxidative stress. *Redox Report*, *16*(2), 78-90. https://doi.org/10.1179/174329211X13020951739938

Lawrence, K. S., Hagan. A., Norton, R., Hu, J., Faske, T., Hutmacher, R., Muller, J., Small, I., Grabau, Z., Kemerait, R., Overstreet, C., Price, P., Lawrence, G., Allen, T., Atwell, S., Idowu, J., Boman, R., Goodson, J., Kelly, H., Woodward, J. E., Wheeler, T. A., & Mehl, H. L. (2018). Cotton disease estimate committee reports, 2017. In: *2018 Beltwide Cotton Conferences*, San Antonio, TX. https://www.cotton.org/beltwide/proceedings/2005-2022/data/conferences/2018/papers/18194.pdf

Lee, C., Harvey, J. T., Qin, K., & Leskovar, D. I. (2023). Physio-biochemical responses of grafted tomatoes differing in thermotolerance to heat stress and recovery. *Scientia Horticulturae*, *308*, 111546. https://doi.org/10.1016/j.scienta.2022.111546

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J. Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R. G., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y.,Ye, W., Liu, X., Zhang, X., Liu, W., Wei, H., Wei, S., Huang, G., Zhang, X., Zhu, S., Zhang, H., Sun, F., Wang, X., Liang, J., Wang, J., He, Q., Huang, L., Wang, J., Cui, J., Song, G., Wang, K., Xu, X., Yu, J. Z., Zhu, Y., & Yu, S. (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature biotechnology*, *33*(5), 524-530. https://doi.org/10.1038/nbt.3208

Liguori, N., Xu, P., Van Stokkum, I. H., Van Oort, B., Lu, Y., Karcher, D., Bock, R., & Croce, R. (2017). Different carotenoid conformations have distinct functions in light-harvesting regulation in plants. *Nature communications*, *8*(1), 1994. https://doi.org/10.1038/s41467-017-02239-z

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., & Chang, W. (2004). Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature*, *428*(6980), 287-292. https://doi.org/10.1038/nature02373

Liu, Z., Yuan, Y. L., Liu, S. Q., Yu, X. N., & Rao, L. Q. (2006). Screening for high-temperature tolerant cotton cultivars by testing in vitro pollen germination, pollen tube growth and boll retention. *Journal of Integrative Plant Biology*, *48*(6), 706-714. https://doi.org/10.1111/j.1744-7909.2006.00276.x

Llorente, B., Martinez-Garcia, J. F., Stange, C., & Rodriguez-Concepcion, M. (2017). Illuminating colors: regulation of carotenoid biosynthesis and accumulation by light. *Current Opinion in Plant Biology*, *37*, 49-55. https://doi.org/10.1016/j.pbi.2017.03.011

Lu, N., Roldan, M., & Dixon, R. A. (2017). Characterization of two TT2-type MYB transcription factors regulating proanthocyanidin biosynthesis in tetraploid cotton, *Gossypium hirsutum*. *Planta*, *246*, 323-335. https://doi.org/10.1007/s00425-017-2682-z

Lu, Z., Percy, R. G., Qualset, C. O., & Zeiger, E. (1998). Stomatal conductance predicts yields in irrigated Pima cotton and bread wheat grown at high temperatures. *Journal of Experimental Botany*, 453-460. https://www.jstor.org/stable/23695978

Mace, M. E., Bell, A. A., & Stipanovic, R. D. (1978). Histochemistry and identification of flavanols in *Verticillium* wilt-resistant and-susceptible cottons. *Physiological Plant Pathology*, *13*(2), 143-149. https://doi.org/10.1016/0048-4059(78)90027-9

MacRobbie, E. A. C. (1997). Signaling in guard cells and regulation of ion channel activity. *Journal of Experimental Botany*, 515-528. https://www.jstor.org/stable/23695629

Mallikarjuna, N., Kranthi, K. R., Jadhav, D. R., Kranthi, S., & Chandra, S. (2004). Influence of foliar chemical compounds on the development of *Spodoptera litura* (Fab.) in interspecific derivatives of groundnut. *Journal of Applied Entomology*, *128*(5), 321-328. https://doi.org/10.1111/j.1439-0418.2004.00834.x

Manan, A., Zafar, M. M., Ren, M., Khurshid, M., Sahar, A., Rehman, A., Firdous, H., Youlu, Y., Razzaq, A., & Shakeel, A. (2022). Genetic analysis of biochemical, fiber yield and quality traits of upland cotton under high-temperature. *Plant Production Science*, *25*(1), 105-119. https://doi.org/10.1080/1343943X.2021.1972013

Martin, R. E., Postiglione, A. E., & Muday, G. K. (2022). Reactive oxygen species function as signaling molecules in controlling plant development and hormonal responses. *Current Opinion in Plant Biology*, *69*, 102293. https://doi.org/10.1016/j.pbi.2022.102293

Mase, K., & Tsukagoshi, H. (2021). Reactive oxygen species link gene regulatory networks during *Arabidopsis* root development. *Frontiers in Plant Science*, *12*, 660274. https://doi.org/10.3389/fpls.2021.660274

Matusova, R., Rani, K., Verstappen, F. W., Franssen, M. C., Beale, M. H., & Bouwmeester, H. J. (2005). The strigolactone germination stimulants of the plant-parasitic *Striga* and *Orobanche* spp. are derived from the carotenoid pathway. *Plant physiology*, *139*(2), 920-934. https://doi.org/10.1104/pp.105.061382

May, O. L., Bowman, D. T., & Calhoun, D. S. (1995). Genetic diversity of US upland cotton cultivars released between 1980 and 1990. *Crop Science*, *35*(6), 1570-1574. https://doi.org/10.2135/cropsci1995.0011183X003500060009x

Meena, P. K., Meena, L., & Verma, A. K. (2019). A review on biological management of diseases of cotton (*Gossypium* sp.) L. *International Journal of Advanced Scientific Research and Management*, *4*(2).

Meier, S., Tzfadia, O., Vallabhaneni, R., Gehring, C., & Wurtzel, E. T. (2011). A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. *BMC Systems Biology*, *5*, 1-19. https://doi.org/10.1186/1752-0509-5-77

Meyer, V. G. (1966). Environmental effects on the differentiation of abnormal cotton flowers. *American Journal of Botany*, *53*(10), 976-980. https://doi.org/10.1002/j.1537-2197.1966.tb06861.x

Mierziak, J., Kostyn, K., & Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules*, *19*(10), 16240-16265. https://doi.org/10.3390/molecules191016240

Mittler, R. (2017). ROS are good. *Trends in plant science*, *22*(1), 11-19. https://doi.org/10.1016/j.tplants.2016.08.002

Monteith, J. L. (1995). A reinterpretation of stomatal responses to humidity. *Plant, Cell & Environment*, *18*(4), 357-364. https://doi.org/10.1111/j.1365-3040.1995.tb00371.x

Moulherat, C., Tengberg, M., Haquet, J. F., & Mille, B. (2002). First evidence of cotton at Neolithic Mehrgarh, Pakistan: analysis of mineralized fibres from a copper bead. *Journal of Archaeological Science*, *29*(12), 1393-1401. https://doi.org/10.1006/jasc.2001.0779

Muday, G. K., & Murphy, A. S. (2002). An emerging model of auxin transport regulation. *The Plant Cell*, *14*(2), 293-299. https://doi.org/10.1105/tpc.140230

Nakamoto, H. & Hiyama, T. (1999). Heat-shock proteins and temperature stress. In: *Handbook of Plant and Crop Stress* (ed. M. Pessarakli), *Marcel Dekker, New York*, 399-416.

National Cotton Council of America (NCCA). (2013) World of Cotton (http://www.cotton.org/econ/world/) and Frequently Asked Questions (http://www.cotton.org/edu/faq/index.cfm) Accessed July 2023.

National Cotton Council of America (NCCA). (2022). 2022 Cotton Acreage. https://www.cotton.org/news/releases/2022/cplnt.cfm. Accessed July 2023.

National Cotton Council of America (NCCAa). Cotton Seedling Disease Identification. https://www.cotton.org/tech/pest/seedling/identification.cfm. Accessed July 2023.

National Cotton Council of America (NCCAb). Economics of Cotton. https://www.cotton.org/pubs/cottoncounts/fieldtofabric/economics.cfm. Accessed July 2023.

Ni, K., Dai, M., Lu, X., Zhang, Y., Fan, Y., Xu, N., Feng, X., Huang, H., Wang, J., Rui, C., Zhang, H., Lei, Y., Liu, X., Jiang, T., Han, M., Sun, L., Chen, X., Wang, D., Wang, J., Wang, S., Chen, C., Guo, L., Zhao, L., & Ye, W. (2022) Pretreatment of NaCl enhances the drought resistance of cotton by regulating the biosynthesis of carotenoids and abscisic acid. *Frontiers in Environmental Science, 10,* 998141. https://doi.org/10.3389/fenvs.2022.998141

Nix, A., Paull, C., & Colgrave, M. (2017). Flavonoid profile of the cotton plant, *Gossypium hirsutum*: A review. *Plants, 6*(4), 43. https://doi.org/10.3390/plants6040043

Nuessly, G. S., Scully, B. T., Hentz, M. G., Beiriger, R., Snook, M. E., & Widstrom, N. W. (2007). Resistance to *Spodoptera frugiperda* (Lepidoptera: Noctuidae) and *Euxesta stigmatias* (Diptera: Ulidiidae) in sweet corn derived from exogenous and endogenous genetic systems. *Journal of Economic Entomology*, *100*(6), 1887-1895. https://doi.org/10.1093/jee/100.6.1887

Ohsugi, T., Nishida, R., & Fukami, H. (1985). Oviposition stimulant of *Papilio xuthus*, a *citrus*-feeding-Swallowtail butterfly. *Agricultural and biological chemistry*, *49*(6), 1897-1900. https://doi.org/10.1080/00021369.1985.10867003

Osmond, C.B. What is photoinhibition? Some insights from comparisons of shade and sun plants., In: *Photoinhibition of photosynthesis: from molecular mechanisms to the field.* Baker NR, Bowyer JR (eds) Oxford: Bios Science Publishers; 1994. pp. 1–24.

Pankratov, I., McQuinn, R., Schwartz, J., Bar, E., Fei, Z., Lewinsohn, E., Zamin, D., Giovannoni, J. J., & Hirschberg, J. (2016). Fruit carotenoid-deficient mutants in tomato reveal a function of the plastidial isopentenyl diphosphate isomerase (IDI 1) in carotenoid biosynthesis. *The Plant Journal*, *88*(1), 82-94. https://doi.org/10.1111/tpj.13232

Parks, C. R. (1965a). Floral pigmentation studies in the genus *Gossypium*. I. Species specific pigmentation patterns. *American Journal of Botany*, *52*(3), 309-316. https://doi.org/10.1002/j.1537-2197.1965.tb06791.x

Parks, C. R. (1965b). Floral pigmentation studies in the genus *Gossypium* II chemotaxonomic analysis of diploid *Gossypium* species. *American Journal of Botany*, *52*(8), 849-856. https://doi.org/10.1002/j.1537-2197.1965.tb07257.x

Parks, C. R. (1967). Floral pigmentation studies in the genus *Gossypium*. III. Qualitative analysis of total flavonol content for taxonomic studies. *American Journal of Botany*, *54*(3), 306-315. https://doi.org/10.1002/j.1537-2197.1967.tb06925.x

Peer, W. A., & Murphy, A. S. (2007). Flavonoids and auxin transport: modulators or regulators?. *Trends in plant science*, *12*(12), 556-563. https://doi.org/10.1016/j.tplants.2007.10.003

Phillips, A. Z., Berry, J. C., Wilson, M. C., Vijayaraghavan, A., Burke, J., Bunn, J. I., Allen, T. W., Wheeler, T., & Bart, R. S. (2017). Genomics-enabled analysis of the emergent disease cotton bacterial blight. *Public Library of Science Genetics*, *13*(9), e1007003. https://doi.org/10.1371/journal.pgen.1007003

Pierce, E. T., Graham, B. P., Stiff, M. R., Osborne, J. A., & Haigler, C. H. (2019). Cultures of *Gossypium barbadense* cotton ovules offer insights into the microtubule-mediated control of fiber cell expansion. *Planta*, *249*, 1551-1563. https://doi.org/10.1007/s00425-019-03106-5

Pogson, B. J., Rissler, H. M., Frank, H. A. (2005) The roles of carotenoids in photosystem II of higher plants. In *'Photosystem II: the light-driven water: plastoquinone oxidoreductase'*. (Eds T Wydrzynski, K Satoh) pp. 515–537. Springer-Verlag: Dordrecht

Polívka, T., & Frank, H. A. (2010). Molecular factors controlling photosynthetic light harvesting by carotenoids. *Accounts of chemical research*, *43*(8), 1125-1134. https://doi.org/10.1021/ar100030m

Pospíšil, P. (2012). Molecular mechanisms of production and scavenging of reactive oxygen species by photosystem II. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1817*(1), 218-231. https://doi.org/10.1016/j.bbabio.2011.05.017

Qin, X., Coku, A., Inoue, K., & Tian, L. (2011). Expression, subcellular localization, and *cis*-regulatory structure of duplicated phytoene synthase genes in melon (*Cucumis melo* L.). *Planta, 234*, 737-748. https://doi.org/10.1007/s00425-011-1442-8

Radin, J. W., Lu, Z., Percy, R. G., & Zeiger, E. (1994). Genetic variability for stomatal conductance in Pima cotton and its relation to improvements of heat adaptation. *Proceedings of the National Academy of Sciences*, *91*(15), 7217-7221. https://doi.org/10.1073/pnas.91.15.7217

Rani, K., Zwanenburg, B., Sugimoto, Y., Yoneyama, K., & Bouwmeester, H. J. (2008). Biosynthetic considerations could assist the structure elucidation of host plant produced rhizosphere signalling compounds (strigolactones) for arbuscular mycorrhizal fungi and parasitic plants. *Plant Physiology and Biochemistry*, *46*(7), 617-626. https://doi.org/10.1016/j.plaphy.2008.04.012

Rani, P. U., & Pratyusha, S. (2013). Defensive role of *Gossypium hirsutum* L. anti-oxidative enzymes and phenolic acids in response to *Spodoptera litura* F. feeding. *Journal of Asia-Pacific Entomology*, *16*(2), 131-136. https://doi.org/10.1016/j.aspen.2013.01.001

Reddy, K. R., Hodges, H. F., & Reddy, V. R. (1992). Temperature effects on cotton fruit retention. *Agronomy journal*, *84*(1), 26-30. https://doi.org/10.2134/agronj1992.00021962008400010006x

Reyes, J. L., & Chua, N. H. (2007). ABA induction of miR159 controls transcript levels of two MYB factors during Arabidopsis seed germination. *The Plant Journal*, *49*(4), 592-606. https://doi.org/10.1111/j.1365-313X.2006.02980.x

Rice-Evans, C. A., Miller, N. J., & Paganga, G. (1996). Structure-antioxidant activity relationships of flavonoids and phenolic acids. *Free radical biology and medicine*, *20*(7), 933-956. https://doi.org/10.1016/0891-5849(95)02227-9

Ridgway, R. L., Bell, A. A., Veech, J. A., & Chandler, J. M. (1984). Cotton protection practices in the USA and world. In *Cotton* (Vol. 24). R. J. Kohel & C. F. Lewis (Eds.), pp. 265–365 American Society of Agronomy. https://doi.org/10.2134/agronmonogr24.c9

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S., & Mittler, R. (2004). When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant physiology*, *134*(4), 1683-1696. https://doi.org/10.1104/pp.103.033431

Rodríguez-Concepción, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. *Archives of Biochemistry and Biophysics*, *504*(1), 118-122. https://doi.org/10.1016/j.abb.2010.06.016

Rutherford, A. W., Paterson, D. R., & Mullet, J. E. (1981). A light-induced spin-polarized triplet detected by EPR in photosystem II reaction centers. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *635*(2), 205-214. https://doi.org/10.1016/0005-2728(81)90020-7

Sarry, J. E., Montillet, J. L., Sauvaire, Y., & Havaux, M. (1994). The protective function of the xanthophyll cycle in photosynthesis. *FEBS letters*, *353*(2), 147-150. https://doi.org/10.1016/0014-5793(94)01028-5

Schaub, P., Al-Babili, S., Drake, R., & Beyer, P. (2005). Why is golden rice golden (yellow) instead of red?. *Plant Physiology*, *138*(1), 441-450. https://doi.org/10.1104/pp.104.057927

Schmidt, C., Schelle, I., Liao, Y. J., & Schroeder, J. I. (1995). Strong regulation of slow anion channels and abscisic acid signaling in guard cells by phosphorylation and dephosphorylation events. *Proceedings of the National Academy of Sciences*, *92*(21), 9535-9539. https://doi.org/10.1073/pnas.92.21.953

Schöffl, F., Prandl, R., and Reindl, A. (1999). Molecular responses to heat stress. In: *Molecular Responses to Cold, Drought, Heat and Salt Stress in Higher Plants* (eds. K. Shinozaki and K. Yamaguchi-Shinozaki), 81–98. Austin, Texas: R.G. Landes Co.

Schwartz, S. H., Qin, X., & Loewen, M. C. (2004). The biochemical characterization of two carotenoid cleavage enzymes from Arabidopsis indicates that a carotenoid-derived compound inhibits lateral branching. *Journal of Biological Chemistry*, *279*(45), 46940-46945. https://doi.org/10.1074/jbc.M409004200

Schwartz, S. H., Tan, B. C., Gage, D. A., Zeevaart, J. A., & McCarty, D. R. (1997). Specific oxidative cleavage of carotenoids by VP14 of maize. *Science*, *276*(5320), 1872-1874. https://doi.org/10.1126/science.276.5320.1872

Seelanan, T., Brubaker, C. L., McD, J., Craven, L. A., & Wendel, J. F. (1999). Molecular systematics of Australian *Gossypium* section *grandicalyx* (Malvaceae). *Systematic Botany*, 183-208. https://doi.org/10.2307/2419548

Senchina, D. S., Alvarez, I., Cronn, R. C., Liu, B., Rong, J., Noyes, R. D., Paterson, A.H., Wing, R.A., Wilkins, T.A., & Wendel, J. F. (2003). Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Molecular biology and evolution*, *20*(4), 633-643. https://doi.org/10.1093/molbev/msg065

Seyoum, A., Asres, K., & El-Fiky, F. K. (2006). Structure–radical scavenging activity relationships of flavonoids. *Phytochemistry*, *67*(18), 2058-2070. https://doi.org/10.1016/j.phytochem.2006.07.002

Shaver, T. N., & Lukefahr, M. J. (1969). Effect of flavonoid pigments and gossypol on growth and development of the bollworm, tobacco budworm, and pink bollworm. *Journal of Economic Entomology*, *62*(3), 643-646. https://doi.org/10.1093/jee/62.3.643

Shi, H., Liu, G., Wei, Y., & Chan, Z. (2018). The zinc-finger transcription factor ZAT6 is essential for hydrogen peroxide induction of anthocyanin synthesis in *Arabidopsis*. *Plant molecular biology*, *97*, 165-176. https://doi.org/10.1007/s11103-018-0730-0

Simkin, A. J., Schwartz, S. H., Auldridge, M., Taylor, M. G., & Klee, H. J. (2004). The tomato *carotenoid cleavage dioxygenase 1* genes contribute to the formation of the flavor volatiles β-ionone, pseudoionone, and geranylacetone. *The Plant Journal*, *40*(6), 882-892. https://doi.org/10.1111/j.1365-313X.2004.02263.x

Simmonds, M. S. (2003). Flavonoid–insect interactions: recent advances in our knowledge. *Phytochemistry*, *64*(1), 21-30. https://doi.org/10.1016/S0031-9422(03)00293-0

Simmonds, M. S., & Stevenson, P. C. (2001). Effects of isoflavonoids from Cicer on larvae of Heliocoverpa armigera. *Journal of Chemical Ecology*, *27*, 965-977. https://doi.org/10.1023/A:1010339104206

Sinclair, J. B. (1965). *Cotton seedling diseases and their control*. https://repository.lsu.edu/agexp/683

Skovsted, A. (1934). Cytological studies in cotton: II. Two interspecific hybrids between Asiatic and new world cottons. *Journal of Genetics*, *28*, 407-424. https://doi.org/10.1007/BF02981763

Smith, C. E., & Stephens, S. G. (1971). Critical identification of Mexican archaeological cotton remains. *Economic botany*, *25*, 160-168. https://doi.org/10.1007/BF02860076

Snider, J. L., Pilon, C., Hu, W., Wang, H. M., Tishchenko, V., Slaton, W., Chastain, D., & Parkash, V. (2022). Net photosynthesis acclimates to low growth temperature in cotton seedlings by shifting temperature thresholds for photosynthetic component processes and respiration. *Environmental and Experimental Botany*, *196*, 104816. https://doi.org/10.1016/j.envexpbot.2022.104816

Soliz, L. M. A., Oosterhuis, D. M., Coker, D. L., & Brown, R. S. (2008). Physiological response of cotton to high night temperature. *The Americas Journal of Plant Science and Biotechnology*, *2*(2), 63-68.

Song, G., Jiang, C., Ge, X., Chen, Q., & Tang, C. (2015). Pollen thermotolerance of upland cotton related to anther structure and *HSP* expression. *Agronomy Journal*, *107*(4), 1269-1279. https://doi.org/10.2134/agronj14.0458

Sosa, T., Chaves, N., Alias, J. C., Escudero, J. C., Henao, F., & Gutiérrez-Merino, C. (2004). Inhibition of mouth skeletal muscle relaxation by flavonoids of *Cistus ladanifer* L.: a plant defense mechanism against herbivores. *Journal of chemical ecology*, *30*, 1087-1101. https://doi.org/10.1023/B:JOEC.0000030265.45127.08

Sun, G. (2012). MicroRNAs and their diverse functions in plants. *Plant molecular biology*, *80*, 17-36. https://doi.org/10.1007/s11103-011-9817-6

Sun, T., Yuan, H., Cao, H., Yazdani, M., Tadmor, Y., & Li, L. (2018). Carotenoid metabolism in plants: the role of plastids. *Molecular plant*, *11*(1), 58-74. https://doi.org/10.1016/j.molp.2017.09.010

Tabashnik, B. E. (1987). Plant secondary compounds as oviposition deterrents for cabbage butterfly, *Pieris rapae* (Lepidoptera: Pieridae). *Journal of chemical ecology*, *13*, 309-316. https://doi.org/10.1007/BF01025890

Taha, M. A., Malik, M. N. A., Chaudhry, F. I., & Makhdum, M. I. (1981). Heat-induced sterility in cotton sown during early April in West Punjab. *Experimental Agriculture*, *17*(2), 189-194. https://doi.org/10.1017/S0014479700011443

Tan, J., Tu, L., Deng, F., Hu, H., Nie, Y., & Zhang, X. (2013). A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant physiology*, *162*(1), 86-95. https://doi.org/10.1104/pp.112.212142

Tanaka, Y., & Ohmiya, A. (2008). Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. *Current opinion in biotechnology*, *19*(2), 190-197. https://doi.org/10.1016/j.copbio.2008.02.015

Ter-Avanesian, D. V. (1978). The effect of varying the number of pollen grains used in fertilization. *Theoretical and Applied Genetics*, *52*, 77-79. https://doi.org/10.1007/BF00281320

Thompson, A. C., Henson, R. D., Hedin, P. A., & Minyard, J. P. (1968). Constituents of the cotton bud: XII. The carotenoids in buds, seeds and other tissue. *Lipids*, *3*(6), 495-497. https://doi.org/10.1007/BF02530892

Tohge, T., Perez de Souza, L., & Fernie, A. R. (2018). On the natural diversity of phenylacylated-flavonoid and their in planta function under conditions of stress. *Phytochemistry Reviews*, *17*, 279-290. https://doi.org/10.1007/s11101-017-9531-3

Toledo-Ortiz, G., Huq, E., & Rodríguez-Concepción, M. (2010). Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proceedings of the National Academy of Sciences*, *107*(25), 11626-11631. https://doi.org/10.1073/pnas.0914428107

Tyagi, P., Gore, M. A., Bowman, D. T., Campbell, B. T., Udall, J. A., & Kuraparthy, V. (2014). Genetic diversity and population structure in the US upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics, 127*(2), 283–295. https://doi.org/10.1007/s00122-013-2217-3

Ulloa, M., Cantrell, R. G., Percy, R. G., Zeiger, E., & Lu, Z. (2000). QTL analysis of stomatal conductance and relationship to lint yield in an interspecific cotton. *Journal of Cotton Science, 4*(1), 10-18.

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., Magome, H., Kamiya, Y., Shirasu, K., Yoneyama, K., Kyozuka, J., & Yamaguchi, S. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature*, *455*(7210), 195-200. https://doi.org/10.1038/nature07272

Uppalapati, S. R., Young, C. A., Marek, S. M., & Mysore, K. S. (2010). Phymatotrichum (cotton) root rot caused by *Phymatotrichopsis omnivora*: retrospects and prospects. *Molecular plant pathology*, *11*(3), 325-334. https://doi.org/10.1111/j.1364-3703.2010.00616.x

Ur-Rahman, H. (2005). Genetic analysis of stomatal conductance in upland cotton (*Gossypium hirsutum* L.) under contrasting temperature regimes. *The Journal of Agricultural Science, 143*(2-3), 161-168. https://doi.org/10.1017/S0021859605005186

Ur-Rahman, H., Malik, S. A., & Saleem, M. (2004). Heat tolerance of upland cotton during the fruiting stage evaluated using cellular membrane thermostability. *Field Crops Research*, *85*(2-3), 149-158. https://doi.org/10.1016/S0378-4290(03)00159-X

USDA ERS. 2022. Cotton Sector at a Glance. https://www.ers.usda.gov/topics/crops/cotton-and-wool/cotton-sector-at-a-glance/#. Accessed July 2023.

USDA FAS. (2022). U.S. Cotton Exports in 2022. https://www.fas.usda.gov/data/commodities/cotton. Accessed July 2023.

Van Esbroeck, G. A., Bowman, D. T., May, O. L., & Calhoun, D. S. (1999). Genetic similarity indices for ancestral cotton cultivars and their impact on genetic diversity estimates of

modern cultivars. *Crop Science*, *39*(2), 323-328. https://doi.org/10.2135/cropsci1999.0011183X003900020003x

Verma, J. P. (1986). *Bacterial blight of cotton.* CRC Press.

Vierling, E. (1991). The roles of heat shock proteins in plants. *Annual review of plant biology*, *42*(1), 579-620. https://doi.org/10.1146/annurev.pp.42.060191.003051

Viot, C. R., & Wendel, J. F. (2023). Evolution of the cotton genus, *Gossypium*, and its domestication in the Americas. *Critical Reviews in Plant Sciences*, *42*(1), 1-33. https://doi.org/10.1080/07352689.2022.2156061

Vishwakarma, K., Upadhyay, N., Kumar, N., Yadav, G., Singh, J., Mishra, R. K., Kumar, V., Verma, R., Upadhyay, R. G., Pandey, M., & Sharma, S. (2017). Abscisic acid signaling and abiotic stress tolerance in plants: a review on current knowledge and future prospects. *Frontiers in plant science*, *8*, 161. https://doi.org/10.3389/fpls.2017.00161

Wallace, T. P., Bowman, D., Campbell, B. T., Chee, P., Gutierrez, O. A., Kohel, R. J., McCarty, J., Myers, G., Percy, R., Robinson, F., Smith, W., Stelly, D. M., Stewart, J. M., Thaxton, P., Ulloa, M., & Weaver, D. B. (2009). Status of the USA cotton germplasm collection and crop vulnerability. *Genetic resources and crop evolution*, *56*, 507-532. https://doi.org/10.1007/s10722-008-9382-2

Walter, M. H., Floss, D. S., & Strack, D. (2010). Apocarotenoids: hormones, mycorrhizal metabolites and aroma volatiles. *Planta*, *232*, 1-17. https://doi.org/10.1007/s00425-010-1156-3

Wang, J. H., Moore, M. J., Wang, H., Zhu, Z. X., & Wang, H. F. (2021). Plastome evolution and phylogenetic relationships among Malvaceae subfamilies. *Gene*, *765*, 145103. https://doi.org/10.1016/j.gene.2020.145103

War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., & Sharma, H. C. (2012). Mechanisms of plant defense against insect herbivores. *Plant signaling & behavior*, *7*(10), 1306-1320. https://doi.org/10.4161/psb.21663

Welsch, R., Wust, F., Bar, C., Al-Babili, S., & Beyer, P. (2008). A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional

diversification of phytoene synthase genes. *Plant Physiology*, *147*(1), 367-380. https://doi.org/10.1104/pp.108.117028

Wendel, J. F. (1989). New World tetraploid cottons contain Old World cytoplasm. *Proceedings of the National Academy of Sciences*, *86*(11), 4132-4136. https://doi.org/10.1073/pnas.86.11.4132

Wendel, J. F., & Albert, V. A. (1992). Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany*, 115-143. https://doi.org/10.2307/2419069

Wendel, J. F., Brubaker, C. L., & Percival, A. E. (1992). Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *American Journal of Botany*, *79*(11), 1291-1310. https://doi.org/10.1002/j.1537-2197.1992.tb13734.x

Wendel, J.F., and Grover, C.E. 2015. Taxonomy and evolution of the cotton genus, *Gossypium* In: *Cotton*. 2nd. ed. Vol 57. Fang, D.D., and R.G. Percy, Eds., Madison, WI, USA: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., pp. 25–44. https://doi.org/10.2134/agronmonogr57.2013.0020

Willems, P., Mhamdi, A., Stael, S., Storme, V., Kerchev, P., Noctor, G., Gevaert, K., & Van Breusegem, F. (2016). The ROS wheel: refining ROS transcriptional footprints. *Plant Physiology*, *171*(3), 1720-1733. https://doi.org/10.1104/pp.16.00420

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, *64*(1), 3-19. https://doi.org/10.1016/S0031-9422(03)00300-5

Wink, M. (2016). Evolution of secondary plant metabolism. *eLS*, 1-11. https://doi.org/10.1002/9780470015902.a0001922.pub3

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Current opinion in plant biology*, *5*(3), 218-223. https://doi.org/10.1016/S1369-5266(02)00256-X

Wise, R. R., & Naylor, A. W. (1987). Chilling-enhanced photooxidation: evidence for the role of singlet oxygen and superoxide in the breakdown of pigments and endogenous antioxidants. *Plant Physiology*, *83*(2), 278-282. https://doi.org/10.1104/pp.83.2.278

Woodall, A. A., Britton, G., & Jackson, M. J. (1997). Carotenoids and protection of phospholipids in solution or in liposomes against oxidation by peroxyl radicals: relationship between carotenoid structure and protective ability. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1336*(3), 575-586. https://doi.org/10.1016/S0304-4165(97)00007-X

Xiao, J., Fang, D. D., Bhatti, M., Hendrix, B., & Cantrell, R. (2009). A SNP haplotype associated with a gene resistant to *Xanthomonas axonopodis* pv. *malvacearum* in upland cotton (*Gossypium hirsutum L.*). *Molecular Breeding*, *25*, 593–602. https://doi.org/10.1007/s11032-009-9355-y

Xiao, Y. H., Yan, Q., Ding, H., Luo, M., Hou, L., Zhang, M., Yao, D., Liu, H. S., Li, X., Zhao, J., & Pei, Y. (2014). Transcriptome and biochemical analyses revealed a detailed proanthocyanidin biosynthesis pathway in brown cotton fiber. *PLoS One*, *9*(1), e86344. https://doi.org/10.1371/journal.pone.0086344

Xie, H. T., Wan, Z. Y., Li, S., & Zhang, Y. (2014). Spatiotemporal production of reactive oxygen species by NADPH oxidase is critical for tapetal programmed cell death and pollen development in *Arabidopsis*. *The Plant Cell*, *26*(5), 2007-2023. https://doi.org/10.1105/tpc.114.125427

Xu, W., Dubos, C., & Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends in plant science*, *20*(3), 176-185. https://doi.org/10.1016/j.tplants.2014.12.001

Yamaguchi-Shinozaki, K., & Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual Review of Plant Biology*, *57*, 781-803. https://doi.org/10.1146/annurev.arplant.57.032905.105444

Yan, Q., Wang, Y., Li, Q., Zhang, Z., Ding, H., Zhang, Y., Liu, H., Luo, M., Liu, D., Song, W., Liu, H., Yao, D., Ouyang, X., Li, Y., Li, X., Pei, Y., & Xiao, Y. (2018). Up-regulation of

*GhTT 2-3A* in cotton fibres during secondary wall thickening results in brown fibres with improved quality. *Plant Biotechnology Journal*, *16*(10), 1735-1747. https://doi.org/10.1111/pbi.12910

Yang, L., Wen, K. S., Ruan, X., Zhao, Y. X., Wei, F., & Wang, Q. (2018). Response of plant secondary metabolites to environmental factors. *Molecules, 23*(4), 762. https://doi.org/10.3390/molecules23040762

Yao, D., Wang, Y., Li, Q., Ouyang, X., Li, Y., Wang, C., Ding, L., Hou, L., Luo, M., & Xiao, Y. (2018). Specific upregulation of a cotton phytoene synthase gene produces golden cottonseeds with enhanced provitamin A. *Scientific reports, 8*(1), 1348. https://doi.org/10.1038/s41598-018-19866-1

Yao, H., Zhang, Y., Yi, X., Zhang, X., & Zhang, W. (2016). Cotton responds to different plant population densities by adjusting specific leaf area to optimize canopy photosynthetic use efficiency of light and nitrogen. *Field Crops Research, 188*, 10-16. https://doi.org/10.1016/j.fcr.2016.01.012

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P., & Potrykus, I. (2000). Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science, 287*(5451), 303-305. https://doi.org/10.1126/science.287.5451.303

Yin, Z., Li, Y., Yu, J., Liu, Y., Li, C., Han, X., & Shen, F. (2012). Difference in miRNA expression profiles between two cotton cultivars with distinct salt sensitivity. *Molecular biology reports*, *39*, 4961-4970. https://doi.org/10.1007/s11033-011-1292-2

Yuan, D., Grover, C. E., Hu, G., Pan, M., Miller, E. R., Conover, J. L., Hunt, S.P., Udall, J.A., & Wendel, J. F. (2021). Parallel and intertwining threads of domestication in allopolyploid cotton. *Advanced Science*, *8*(10), 2003634. https://doi.org/10.1002/advs.202003634

Zafar, M. M., Shakeel, A., Haroon, M., Manan, A., Sahar, A., Shoukat, A., Mo, H., Farooq, M. A., & Ren, M. (2022). Effects of salinity stress on some growth, physiological, and biochemical parameters in cotton (*Gossypium hirsutum* L.) germplasm. *Journal of Natural Fibers, 19*(14), 8854-8886. https://doi.org/10.1080/15440478.2021.1975596

Zahid, K. R., Ali, F., Shah, F., Younas, M., Shah, T., Shahwar, D., Hassan, W., Ahmad, Z., Qi, C., Lu, Y., Iqbal, A. & Wu, W. (2016). Response and tolerance mechanism of cotton *Gossypium hirsutum* L. to elevated temperature stress: a review. *Frontiers in plant science*, *7*, 937. https://doi.org/10.3389/fpls.2016.00937

Zeiher, C.A., N. Matumba, P. Brown and J. Silvertooth. 1995. Response of Upland cotton to elevated night temperature II. In: *Proceedings of 1995 Beltwide Cotton Conferences*, 4-7 Jan. 1995, San Antonio, TX. National Cotton Council, Memphis, TN. Vol. 2: 1129.

Zeiher, C.A., P.W. Brown, J.C. Silvertooth, N. Matumba and N. Mitton. 1994. The effect of night temperature on cotton reproductive development. In: *Cotton: A College of Agriculture Report. Series P-96*. University of Arizona, Tucson, AZ. p. 26-32.

Zeng, J., Yao, D., Luo, M., Ding, L., Wang, Y., Yan, X., Ye, S., Wang, C., Wu, Y., Zhang, J., Li, Y., Ran, L., Dai, Y., Chen, Y., Wang, F., Lai, H., Liu, N., Fang, N., Pei, Y., & Xiao, Y. (2023). Fiber-specific increase of carotenoid content promotes cotton fiber elongation by increasing abscisic acid and ethylene biosynthesis. *The Crop Journal*. 11(3), 774-784. https://doi.org/10.1016/j.cj.2023.02.010

Zhang, H., Liang, W., Yang, X., Luo, X., Jiang, N., Ma, H., & Zhang, D. (2010). *Carbon starved anther* encodes a MYB domain protein that regulates sugar partitioning required for rice pollen development. *The Plant Cell*, *22*(3), 672-689. https://doi.org/10.1105/tpc.109.073668

Zhang, J., Bourland, F., Wheeler, T., & Wallace, T. (2020). Bacterial blight resistance in cotton: genetic basis and molecular mapping. *Euphytica*, *216*, 1-19. https://doi.org/10.1007/s10681-020-02630-w

Zhang, Q., Liu, M., & Ruan, J. (2017). Metabolomics analysis reveals the metabolic and functional roles of flavonoids in light-sensitive tea leaves. *BMC plant biology*, *17*(1), 1-10. https://doi.org/10.1186/s12870-017-1012-8

Zhao, D., Reddy, K. R., Kakani, V. G., Koti, S., & Gao, W. (2005). Physiological causes of cotton fruit abscission under conditions of high temperature and enhanced ultraviolet-B

radiation. *Physiologia Plantarum*, *124*(2), 189-199. https://doi.org/10.1111/j.1399-3054.2005.00491.x

Zhu, Y., Abdelraheem, A., Lujan, P., Idowu, J., Sullivan, P., Nichols, R., Wedegaertner, T., & Zhang, J. (2021). Detection and characterization of Fusarium Wilt (*Fusarium oxysporum* f. sp. *vasinfectum*) race 4 causing fusarium wilt of cotton seedlings in New Mexico. *Plant Disease, 105*(11), 3353-3367. https://doi.org/10.1094/PDIS-10-20-2174-RE

# CHAPTER 2: Identification and molecular genetic mapping of major effect pollen pigmentation locus *Gh-$P_{A1}$* in Upland cotton (*Gossypium hirsutum* L.)

As prepared for submission in Journal of Experimental Botany

Nicole Pettit[1], S Anjan Gowda[1], Dr. Muday's lab[2], Vasu Kuraparthy[1]
(Author list and order is yet to be finalized)

[1] Crop Sciences Department, North Carolina State University, Raleigh, North Carolina, USA, 27607
[2] Department of Biology and the Center for Molecular Signaling, Wake Forest University, Winston Salem, North Carolina, USA 27109.

**Key message**: GWAS and linkage mapping identified a major locus *Gh-$P_{A1}$* with a candidate gene of carotenoid biosynthesis pathway controlling pollen color variation in upland cotton which could give insight into cotton pollen physiology.

## 2.1 Abstract

Cotton genus *Gossypium* shows wide variation for pollen color. Pima cotton (*Gossypium barbadense* L.), known for premium fiber quality, is characterized by yellow/orange colored pollen, while the widely cultivated Upland cotton, *G. hirsutum* L, usually has cream colored pollen. Genome-wide association study (GWAS) using 136 accessions of Upland cotton phenotyped for pollen color and genotyped with cottonSNP63K array, identified three significant marker trait associations within a 226 Kb region on chromosome A05. Phenotypic analyses in two recombinant inbred line (RIL) mapping populations in the $F_1$ and $F_5$ generations, derived from parental accessions DES 56 (white pollen), and Sealand #2 and Southland M1 (yellow pollen), showed the yellow pollen color was controlled by a single gene with yellow being dominant to cream. Further, molecular mapping using the RIL populations genotyped with SNP markers localized the *Gh-$P_{A1}$* locus to a 5.36 cM region chromosome A05. The flanking markers (i12529Gh and i12525Gh) of the *Gh-$P_{A1}$* locus in the linkage map corresponded to a 112.83 Kb region in the sequence based physical map. This genomic region consisted of four putative genes of which the *phytoene synthase* (PSY) involved in the carotenoid biosynthetic pathway is the most probable candidate gene for

pollen color in cotton. Biochemical analysis of the phytochemicals in pollen grains confirmed that the pollen color differences in upland cotton were not due to flavonoids. Deciphering the molecular genetic architecture of *Gh-$P_{A1}$* could aid in the understanding of the role of antioxidants in cotton pollen physiology and adaptation.

**Keywords:** Cotton, pollen color, anther color, carotenoid, phytoene synthase (PSY), linkage mapping, pigments, GWAS

**Abbreviations**: ABA, abscisic acid; cM, centimorgan; DIV, diversity panel; GWAS, genome wide association studies; LC-MS, liquid chromatography-mass spectrometry; MTA, marker trait association; PSY, phytoene synthase; ROS, reactive oxygen species; SDR, segregated distortion region; SMA, single marker analysis; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; TLA, tropical landrace accession

## 2.2 Introduction

Grown worldwide, cotton (*Gossypium* spp.) is the main source of natural fiber and one of the leading sources of plant oil and animal protein in the world (USDA-ERS, 2023). Due to its commercial and economic significance, cotton is grown in every continent except for Antarctica. The genus *Gossypium* comprises over 50 different species of which four are commercially grown. Two of these species are grown in the United States, *Gossypium barbadense* L. and *Gossypium hirsutum* L. both being allotetraploids (2n = 4x = 52). Ninety seven percent of U.S. grown cotton is *G. hirsutum*, also called Upland cotton (USDA-ERS, 2022). Upland cotton has high productivity and wide adaptability. On the other hand, *G. barbadense*, also known as Pima cotton, produces fibers that are longer, stronger, and finer than Upland cotton though it produces less yield (Holladay *et al*., 2021). Pima cotton is also characterized by having a high pollen load. Pollen and anther color in cotton plants can range from a buff cream to a deep golden yellow or orange (Harland, 1929; Stephens, 1954). Anther color is said to be inseparable from pollen color in cotton (Harland, 1929). Pima cotton is characterized by having either golden yellow or orange pollen while Upland cotton is almost always cream. Turcotte and Feaster (1966) used segregation ratios of Pima cotton to determine that yellow or orange pollen in Pima is controlled by two complementary pollen color genes ($P_1$ and $P_2$ genes). Cream colored pollen is thought to have the

genotype $p_1p_1P_2P_2$ or $p_1p_1p_2p_2$, orange pollen would have the genotype $P_1P_1p_2p_2$, and yellow pollen would have the genotype $P_1P_1P_2P_2$. Upland cotton is characterized by predominantly cream colored pollen while significantly less number of accessions have yellow colored pollen (USDA-GRIN, 2023). Genetic control of pollen color in Upland cotton is unknown.

There are a number of pigment types in plants with chlorophylls, carotenoids, flavonoids, and betalains being the major pigment found in flowers (Narbona *et al*., 2021). Of these, flavonoids are the most diverse. Chlorophylls, the major photosynthetic pigments, absorb light in the red or blue wavelengths, carotenoids absorb mostly blue light, while betalains and flavonoids can absorb a range of colors. Chlorophylls are not common in floral organs and betalains are not naturally produced in cotton (Li *et al.*, 2023; Narbona *et al*., 2021). As such, the pigmentation found in cotton pollen is likely due to either flavonoids or carotenoids. Anthocyanins, a class of flavonoids, reflect light in the shades of purple, blue, red, and pink, and certain other classes can produce tawny to yellow colors. Other classes of flavonoids reflect and absorb ultraviolet light, making them non-visible to the human eye (Narbona *et al.*, 2021). Carotenoids reflect light that is yellow, orange, and sometimes red.

Flavonoids and carotenoids have a variety of functions in plants in addition to their roles as pigments and aromatic components (Bradshaw & Schemske, 2003; Koes *et al.*, 1994; Walter *et al*., 2010; Wink, 2003; Wink, 2016). Most notably, both carotenoids and flavonoids work as antioxidants in plants, protecting the plants from a buildup of reactive oxygen species (ROS) under stress conditions by managing the redox state of cells (Baker, 2008; Latowski *et al*., 2011; Mierziak *et al*., 2014; Zhang *et al.*, 2017). While it can be an important part of cell signaling, excessive ROS can lead to oxidative damage and cell death. Therefore, managing ROS levels is a major concern in crop plants and antioxidant pigments' role in stress tolerance is being heavily studied (Azhar *et al.*, 2020; Baker, 2008; Latowski *et al.*, 2011; Zahid *et al.*, 2016).

Flavonoids and carotenoids are of particular interest in cotton research due to their impact on cotton fiber quality. Tuttle *et al*. (2015) found that oxidative stress was lower in the Pima line *G. barbadense* cv Phytogen 800 as compared to the Upland line *G. hirsutum* cv Deltapine 90 due to better ROS management which they link to the antioxidant ascorbate (Vitamin C). ROS are known to play a role in fiber initiation and early elongation but can be detrimental in excessive levels (Li *et al.*, 2007; Mei *et al*., 2009; Zhang *et al*., 2010). Fiber development has been shown to be reduced in the presence of the flavonol naringenin (Khan *et al*., 2019; Tan *et al*., 2013).

Carotenoids act as a precursor to abscisic acid (ABA) which is known to play a role as a signaling molecule under stress conditions (Chinnusamy *et al*., 2008; Du *et al*., 2010). ABA is particularly important in cotton due to its involvement in fiber elongation (Beasley & Ting, 1973; Gokani *et al*., 1998; Pierce *et al*., 2019; Zeng *et al*., 2023). Zeng *et al.* (2023) created transgenic lines using a fiber specific promoter and a positive regulator of carotenoid accumulation, GhOR1$^{Del}$. These lines had stronger and longer fibers with a higher fitness which they attributed to the increase in carotenoid content increasing the synthesis in ABA and thus promoting ethylene synthesis. Flavonoids have also been studied due to their effects on fiber pigmentation in brown fibers (Feng *et al*., 2013; Lu *et al*., 2017; Xiao *et al*., 2014; Yan *et al*., 2018). As such, flavonoids and carotenoids play an important role in overall plant health and cotton fiber quality.

Yellow pollen has been found to suppress Tobacco Budworm's growth, a common insect pest in cotton, suggesting a possible benefit for the darker colored pollen plants (Hanny *et al*., 1979). A later study attributed this property to the increase in the concentration of the toxic compound gossypol in yellow anthers (Hanny, 1980). Hanny (1980) found yellow anthers to have significantly higher tannin content in addition to the increased gossypol abundance as compared to cream anthers in Upland cotton. Pollen coloration and pigments are associated in a variety of roles in plants. An increase in antioxidant activity mediating ROS levels in pollen grains may lead to an increase in abiotic stress tolerance in pollen (Muhlemann *et al.*, 2018; Rezanejad, 2009; Rutley *et al.*, 2021). Considering Pima cotton is characterized by yellow/orange colored pollen and recent studies implicating carotenoids in light protection (Baker, 2008; Hashimoto *et al*., 2016; Liu *et al.*, 2004), photosynthesis (Pogson *et al*., 2005; Polívka & Frank, 2010), and improved fiber quality (Beasley & Ting, 1973; Gokani *et al.*, 1998; Zeng *et al.*, 2023) it important to understand the biological significance of pollen color variation in tetraploid cotton. Deciphering the molecular genetic basis of pollen color variation is a prerequisite to study the role of pollen color in plant biology and to utilize the variation for cotton improvement. The objective of the current study was to study the genetics of the yellow pollen color trait and identify and map the gene(s) controlling the major pollen color phenotypic variation in Upland cotton.

## 2.3 Material and methods

### *2.3.1 Plant material*

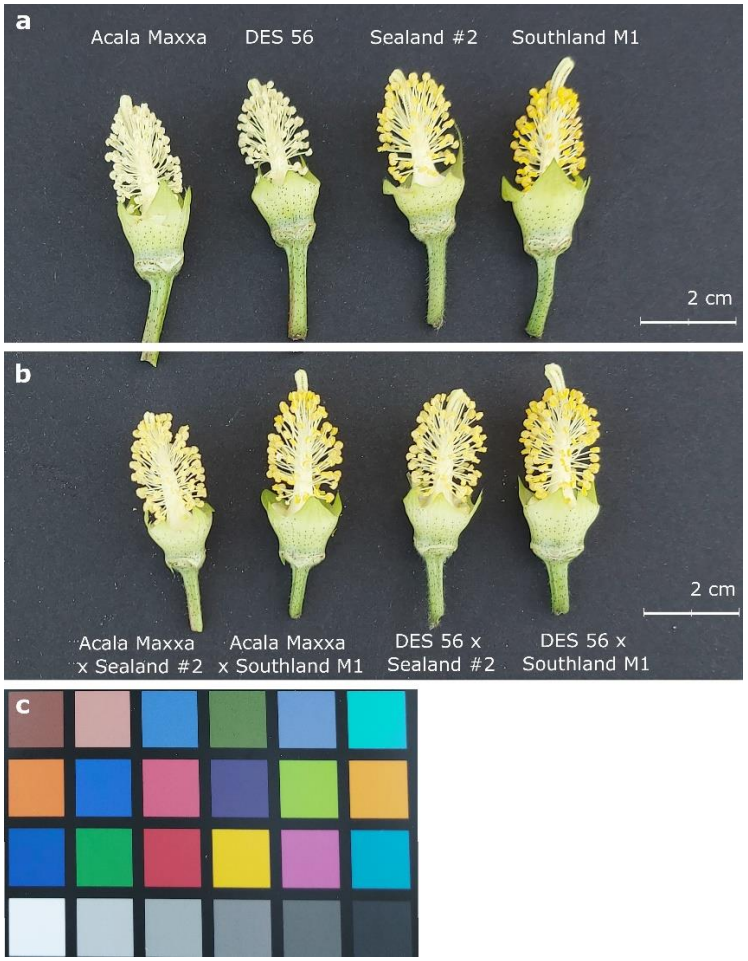
Ninety-nine accessions from the elite diversity (DIV) panel of *G. hirsutum* assembled by Tyagi *et al.* (2014) and 37 accessions from the flowering tropical landrace (TLA) panel of *G. hirsutum* assembled by Kaur *et al*. (2017) were used in for genome wide association studies (GWAS) of *Gh-$P_{A1}$* (Supplementary Table 1). The DIV and TLA panels (together called the GWAS panel) are representative of the genetic diversity found in the elite cultivars of U.S. Upland cotton and the tropical landraces, respectively. Following GWAS, two $F_5$ recombinant inbred line (RIL) populations were used in linkage analysis for further specification of the genomic region of *Gh-$P_{A1}$* locus. The four RIL populations, developed as a part of Nested Association Mapping (NAM) populations, were made by crossing DES 56 (PI 529520; DIV-126) (Bridge & Chism, 1978) or Acalla Maxxa (PI 540885; DIV-17) as the maternal cream pollen colored parent. The paternal yellow pollen colored parents were Sealand #2 (PI 528872; DIV-306) and Southland M1 (PI 601652; DIV-316). The two RIL populations used in linkage analysis were from the crosses of DES 56 × Sealand #2 (designated as population 40D) and DES 56 × Southland M1 (designated as population 42D). Plants for each generation were manually selfed and generations were advanced using the single seed descent method until the $F_5$ generation. The 40D population consisted of 122 individual RILs while the 42D population consisted of 111 individual RILs. LC-MS analysis was performed on $F_1$ and $F_6$ lines of the two RIL populations as well as $F_1$ plants from the two RIL populations derived from crossing Acala Maxxa as the common female to male parental accessions Sealand #2 and Southland M1. LC-MS was also performed on all four parents of the four RIL populations.

### *2.3.2 Phenotyping*

Ninety-nine elite DIV accessions, 37 TLA accessions, and the four RIL populations along with their parents and $F_1$ plants were manually phenotyped for pollen color. For each line/accession, plants were planted at Central Crops Research Station, Clayton, NC in May. Observations were taken on 2-6 flowers between 9:00 AM and 12:00 PM following anther dehiscence. Observations were taken on multiple plants per line whenever possible to detect any segregation in pollen color. For the RIL populations, pollen color was classified as either cream colored, like parent 1, or

yellow colored, like parent 2 (Figure 1). The association mapping GWAS panel accessions were classified as either cream or yellow pollen. Classifications were done in direct sunlight and observations taken during cloudy or rainy days were avoided.

**Figure 1:** Pollen color phenotypes of Upland cotton (*Gossypium hirsutum*) **(a)** parental accessions and **(b)** $F_1$s of four RIL populations used for molecular genetic mapping of pollen color in Upland cotton. **(c)** Color chart to ensure accurate color comparisons.




*Phytochemical estimation*: Pollen samples were collected from flowers collected from the field between 9:00 and 10:00 AM. For phenotypic bulk segregant analysis, yellow and cream pollen samples were individually bulked from a single flower of 11 to 15 lines of yellow and cream colored lines, respectively, from each $F_6$ RIL population. Pollen samples were collected from $F_1$s of both RIL populations and from two additional crosses of Acala Maxxa as the female cream

parent by Sealand #2 and Southland M1 as the two male parents. For all four RIL populations and the four parental lines, pollen from two to three flowers were collected. After pollen was removed from the flowers, the samples were weighed to get a dry pollen weight and immediately frozen in liquid nitrogen. These pollen samples were stored in a -80°C freezer.

Flavonols were extracted from thawed pollen samples by first treating with a 1.05mM Formononetin extraction buffer and incubating at 4°C for five minutes. Samples were then treated with 2N HCL and incubated at 75°C for 45 minutes. Ethyl acetate was then added and samples were spun at 16,100g for five minutes to separate the samples into phases. The first phase was separated into a new tube and another round of ethyl acetate treatment was performed. The first phase solution was then dried down and resuspended in 50 uL of acetonitrile. A blank and internal standard of extraction buffer were also prepared.

Flavonol HP-LC was performed on a TSQ Triple Quadrupole LC-MS system (ThermoFisher, Waltham, MA, USA) with the Accela system autosampler (ThermoFisher, Waltham, MA, USA) using a Luna 3 µm C18(2) 100 Å, LC Column 150 x 3 mm (Phenomenex, Torrance, CA, USA). The Thermo Xcalibur program of the TSQ machine was used to input the sequence information and obtain the area under the peaks for each flavonol. Standards of each flavonol were obtained through Sigma-Aldrich Fine Chemicals and INDOFINE chemicals and run to calculate standard curves. The concentrations of the pollen samples were calculated using the dry weight of the pollen.

*2.3.3 DNA isolation and genotyping*

Leaf tissue from the second or third true leaves were collected and immediately frozen in liquid nitrogen from three to four week old seedlings of the RIL populations and their parental lines. Tissue samples were homogenized using an YG5 3422011 macerator (BTLab systems, St. Louis, MO, USA) and stored at -80°C. The DNeasy Plant Mini-Kit (Qiagen USA, Germantown, MD, USA) and a modified DNA extraction protocol were used to isolate DNA from the ground leaf tissue. A Nanodrop$^{TM}$ spectrophotometer ND1000 was then used to evaluate the concentration and quality of the isolated DNA (260/230 > 1.5 and 260/280 = 1.8 to 2.0). The isolated DNA was genotyped using the cottonSNP63K array at Texas A&M Institute for Genome Science and Society as described in Hulse-Kemp *et al.* (2015).

*2.3.4 Genome-wide association studies (GWAS)*

The genotypic data of GWAS panel from the cottonSNP63K array was filtered with the following criteria to eliminate poor quality markers (a) SNPs with minor allele frequency less than 0.05 were eliminated (b) The markers with missing and heterozygous calls >30% were eliminated. The probes sequences of the target SNP were anchored on to specific chromosome positions on the *G. hirsutum* cv. TM1 CRI v1 (Yang *et al.*, 2019) reference genome through a local BLAST database. The grouping data from six upland cotton linkage maps was also used anchoring positions of SNP on the sub genomes (Gowda *et al.*, 2022; Shrestha *et al.*, 2023; Zhang *et al.*, 2019; Zhu *et al.*, 2021). The missing SNP calls in the filtered and anchored SNP dataset was imputed with Beagle software version 5.4 (Browning *et al.*, 2018; Browning *et al.*, 2021). The pollen color phenotype was coded as "1" for yellow pollen, "0" for cream pollen. The phenotypes in the binary format and high quality imputed SNP dataset were used for GWAS. The mixed linear model (MLM) with Kinship (K) and principal components (PCA) as covariates was utilized for GWAS analysis in *rMVP* package (Buckler *et al.*, 2006; Yin *et al.*, 2021). The Bonferroni correction threshold was used for identifying significant marker trait associations (MTAs) (Hochberg, 1988). The regional Manhattan plot was drawn using the *qqman* R package (Turner, 2018). The proportion of phenotypic variance explained (PVE %) by MTAs was calculated using a method as described by Shim *et al.* (2015). The region in the vicinity of significant MTAs were identified through Haploview software (Barrett *et al.*, 2005).

*2.3.5 Linkage mapping*

The cottonSNP63K array provided genotypic information for 63,058 candidate SNP markers. These markers were filtered to remove low quality markers by removing monomorphic and heterozygous markers in addition to markers with missing parental information as described by Tyagi *et al.* (2014). A minor allele frequency of greater than 0.05 and a maximum missing percentage of 30% were used to ensure only high quality markers were considered. Linkage mapping and analysis were performed using JoinMap 5.0 software (Van Ooijen, 2018). Following the formatting of this software, phenotype data was coded as "a" for cream colored pollen lines, "b" for yellow colored pollen lines, and "u" for any lines with missing or unknown phenotypes. For genotypic information, the lines that matched the alleles for DES 56 were coded as "a", the lines that matched the alleles of the respective yellow parent were coded as "b", heterozygous

alleles were coded as “h”, and any missing information was coded as “u”. A logarithm of the odds threshold (LOD) value of 10.0 and 9.0 were used for populations 40D and 42D respectively. The calculated segregation distortion was obtained using a Chi-square test with an alpha value of 0.05. Segregation distortion regions (SDR) were obtained by identifying areas that had at least three adjacent SNP markers showing segregation distortion. Redundant markers were removed by calculating the similarity of loci in JoinMap. Markers were considered redundant and removed if they had a similarity of 1. All calculations were made using the Kosambi mapping function. Linkage maps were constructed by combining linkage groups and marker ordering was completed using the Maximum Likelihood Method in Joinmap.

After linkage maps were constructed, the maps were evaluated with pairwise recombination matrixes, collinearity between linkage positions and physical positions, and single marker analysis. The R/QTL package (Broman *et al.,* 2003) in RStudio was used to obtain the pairwise recombination matrix and LOD score plots. Physical positions of the SNP markers on the reference genome *G. hirsutum* cv. TM1 CRI v1 (Yang *et al.*, 2019) were obtained through the CottonGen Blast+ online tool (https://www.cottongen.org/blast). SNP marker order was compared to the physical map through the ClicO FS online tool (http://103.47.253.210:3000/) (Cheong *et al.*, 2015). Single marker analysis was performed using Windows QTL Cartographer V2.5 (Wang *et al.*, 2012). LOD and chromosome graphs were reconstructed using the RStudio package, tidyverse (Wickham at al., 2019) with a LOD threshold level of 2.5.

The arm of the chromosome was identified based on the physical positions of anchor markers identified by Shan *et al.* (2016). Linkage maps of the identified chromosomal arm were reconstructed using the RECORD software (Van *et al.*, 2005). In this program, the iteration-based “Recommendation counting and ordering” algorithm was performed with 1,000 steps using the Kosambi mapping function. The linkage maps of the two populations’ chromosome arms were combined into a single consensus linkage map through the RStudio datapack, LPmerge (Endelman *et al.*, 2014). The MapChart version 2.32 software (Voorrips, 2002) was used to draw the consensus linkage map and the physical map.

### *2.3.6 Candidate gene analysis*

Once the target genomic region of *Gh-P*$_{A1}$ was identified, the *G. hirsutum* cv. TM1 CRI v1 (Yang *et al.*, 2019) reference genome was scanned for candidate genes between the identified flanking

SNP markers using the cotton functional genomics database (https://cottonfgd.net/browse/) (Zhu *et al.*, 2017). The Arabidopsis gene identifiers for each candidate gene were obtained through the CottonGen Tripal Megasearch tool (https://www.cottongen.org/tripal_megasearch). These gene identifiers were then analyzed in the Arabidopsis Tair database (https://www.arabidopsis.org/) for more information on the putative gene expression and biological function.

*2.3.7 Data analysis*

Segregation ratios of pollen colors for the $F_5$ RIL populations were calculated using a chi-square analysis under the assumption that the pollen color is controlled by a single dominant gene.

LC-MS quantifications were calculated using a single factor ANOVA test to identify any differences between the groups. If any significant difference were found, T-tests were performed between each combination of samples using a Bonferroni corrected alpha value. Statistics were run with the Analysis ToolPak add-in for Excel.

## 2.4 Results

*2.4.1 Phenotypic analyses*

One hundred and twenty three lines were phenotyped for the 40D $F_5$ RIL population. Of those 123 lines, 51 of them showed cream colored pollen (41.46%), 56 showed yellow pollen (45.53%), and 16 (13.01%) lines unknown or heterozygous. One hundred and eleven $F_5$ lines were phenotype for the 42D RIL population. Of those, 52 (46.85%) were cream, 54 (48.65%) were yellow, and 5 (4.50%) lines were unknown or heterozygous. Chi square analysis of both populations showed a 1:1 ratio of cream to yellow pollen phenotypes (p-value 0.63 and 0.85 for 40D and 42D respectively) (Table 1). This suggests that the gene(s) controlling the correlation in pollen in these populations is a under simple genetic control in Upland cotton. These findings are supportive of previous pollen color segregation studies (Balls, 1912; Stephens, 1954; Turcotte & Feaster, 1966; Zhang *et al.*, 2016).

**Table 1:** Phenotypic segregation ratios of pollen color in the two $F_5$ RIL populations in *G. hirsutum*. Phenotypic observations were taken during bright sunny days between 9:00AM and 12:00PM. Both populations showed a 1:1 ratio of cream to yellow pollen.

**Table 1** (continued).

| Cross | Population number | Number of RIL lines | | χ2 test statistic | P-value |
|---|---|---|---|---|---|
| | | Cream | Yellow | | |
| DES 56 × Sealand #2 | 40D | 51 | 56 | 0.234 | 0.629 |
| DES 56 × Southland M1 | 42D | 52 | 54 | 0.038 | 0.846 |
| Acala Maxxa × Sealand #2 | 40A | 51 | 37 | 2.227 | 0.136 |
| Acala Maxxa × Southland M1 | 42A | 32 | 47 | 2.848 | 0.091 |

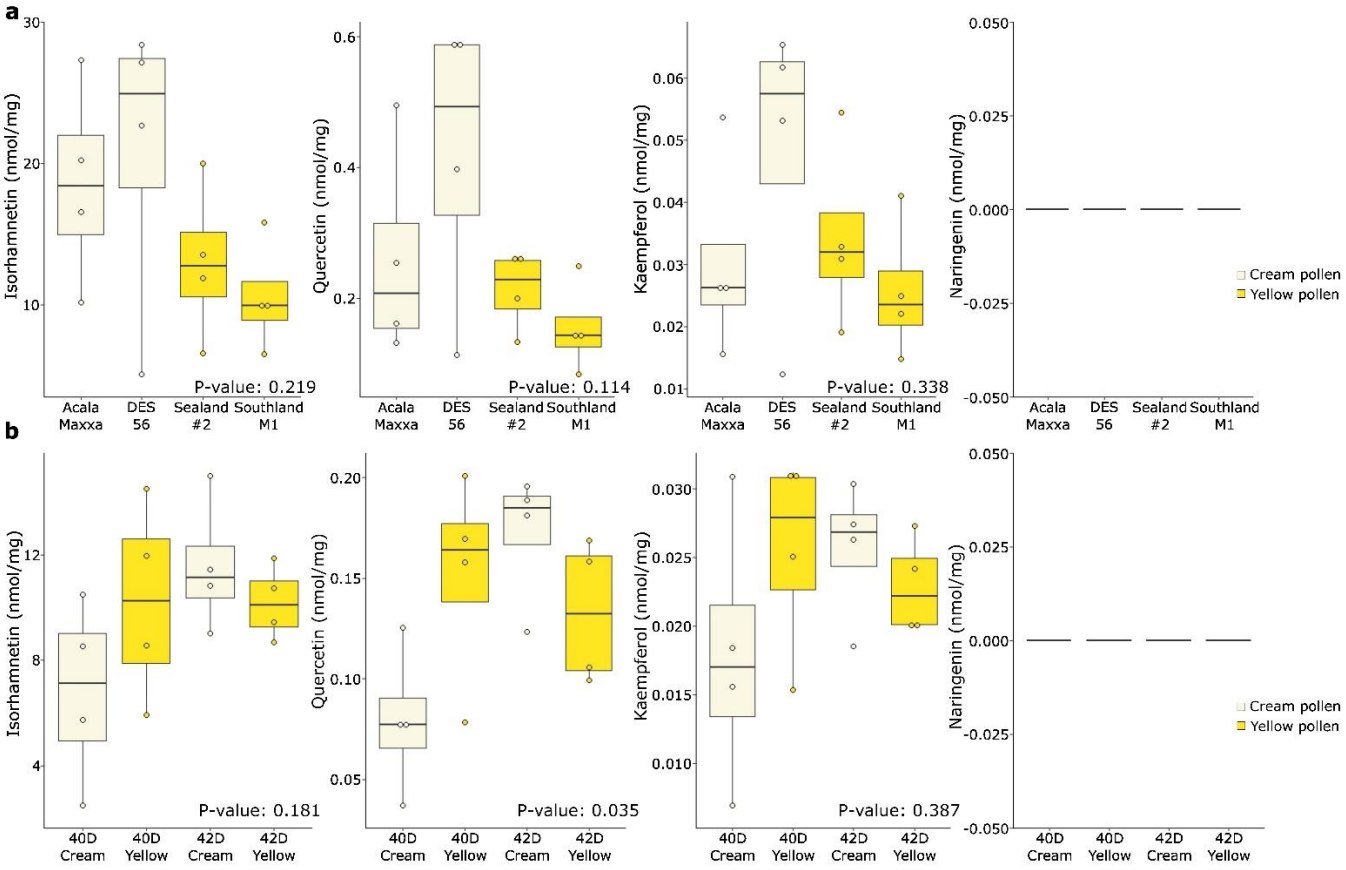
Of the 136 accessions phenotyped for the GWAS analysis, 114 (83.82%) had cream colored pollen and only 22 (16.18%) had yellow pollen. Of those lines, 24.32% (9 lines) and 13.13% (13 lines) were yellow pollen in the DIV and TLA panels respectively.

*2.4.2 LC-MS metabolite quantification*

Flavonol quantification on pollen grains through LC-MS were performed to study the pigment profiles in the yellow and cream colored plants. The major flavonols quercetin, kaempferol, isorhamnetin, and naringenin were analyzed (Figure 2). Naringenin was not detectable in any of the pollen samples. Quercetin, kaempferol, and isorhamnetin were all present though no differences were detected between the cream and yellow colored parents (p-value 0.11, 0.34, and 0.22 respectively). Flavonol estimation was also performed using the individual pollen color bulks for each color type of the RIL populations to study the association of the phytochemicals with pollen color. As with the parents, kaempferol and isorhamnetin were present though no differences were determined (P-value 0.39 and 0.18 respectively). Single factor ANOVA found a significant difference between the groups (p-value 0.03) for quercetin though no significant differences were observed when performing individual T-tests with a Bonferroni-corrected alpha value (Figure 2). Isorhamnetin showed a much higher relative abundance in all samples compared to the other flavonols (Figure 2).

**Figure 2:** Flavonol quantification using liquid chromatography-mass spectrometry. **(a)** Individual flavonols quantified in parental accessions DES 56 (cream), Acala Maxxa (cream), Sealand #2 (yellow) and Southland M1 (yellow). **(b)** Individual flavonols quantified in bulk

samples of cream pollen and yellow color RILs. RIL populations 40D and 42D derived from DES 56 × Sealand #2 and DES 56 × Southland M1 cross combinations, respectively. Significance among groups was calculated using single factor ANOVA and individual T-tests using Bonferroni correction to determine any differences between groups.




*2.4.3 Genome-wide association studies*

Association mapping GWAS panel of 136 upland cotton accessions involving 99 core set accessions of elite cotton DIV panel and 37 photoperiod insensitive accession of TLA panel with their phenotype on pollen color was used for GWAS analysis. A total of 19,349 polymorphic SNPs were identified in the Association mapping panel. These markers spanned a total length of 2,216.5 Mb of *G. hirsutum* cv. TM1 CRI v1 reference assembly (99.5%) (Yang *et al.*, 2019) (Supplementary Table 2; Supplementary figure 1). The marker density varied from 71.58 Kb/SNP in the D-sub genome to 235.05 Kb/SNP in the A-sub genome. Overall the high SNP density of

153.31 Kb/SNP was found in the association mapping panel. The MLM model (K+PCA) of GWAS identified three significant MTAs above Bonferroni threshold *p*-value (5.59) for the pollen color phenotype (Supplementary Table 3; Figure 3). All the three MTAs were identified on the short arm of A05 chromosome (103.57-103.75 Mb) using the SSR anchor marker JESPR-50 (Shan *et al.*, 2016). The significant MTAs in the study, i12529Gh, i12525Gh and i12524Gh were found with PVE of 32.37, 24.78, 26.17 percent, respectively (Supplementary Table 3). These three MTAs were all found within a single short region of 226 Kb (103.54-103.77 Mb) (Figure 4).

**Figure 3:** Genome wide association study for pollen color in GWAS panel of *G. hirsutum* (consisting of 99 diversity panel and 37 flowering tropical landrace panel accessions). **(a)** Manhattan plot of all 26 Upland chromosomes detected three significant signals associated with pollen color above the Bonferroni corrected threshold ($-Log_{10}$ p-value = 5.59) through mixed linear model (MLM) method. **(b)** Quantile-quantile plot showing observed p-values and association signals deviating from expected p-values for pollen color.

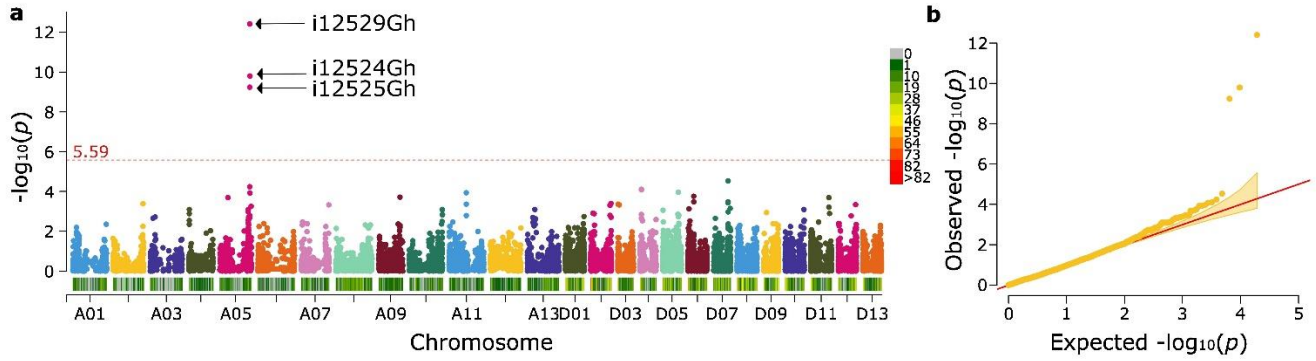
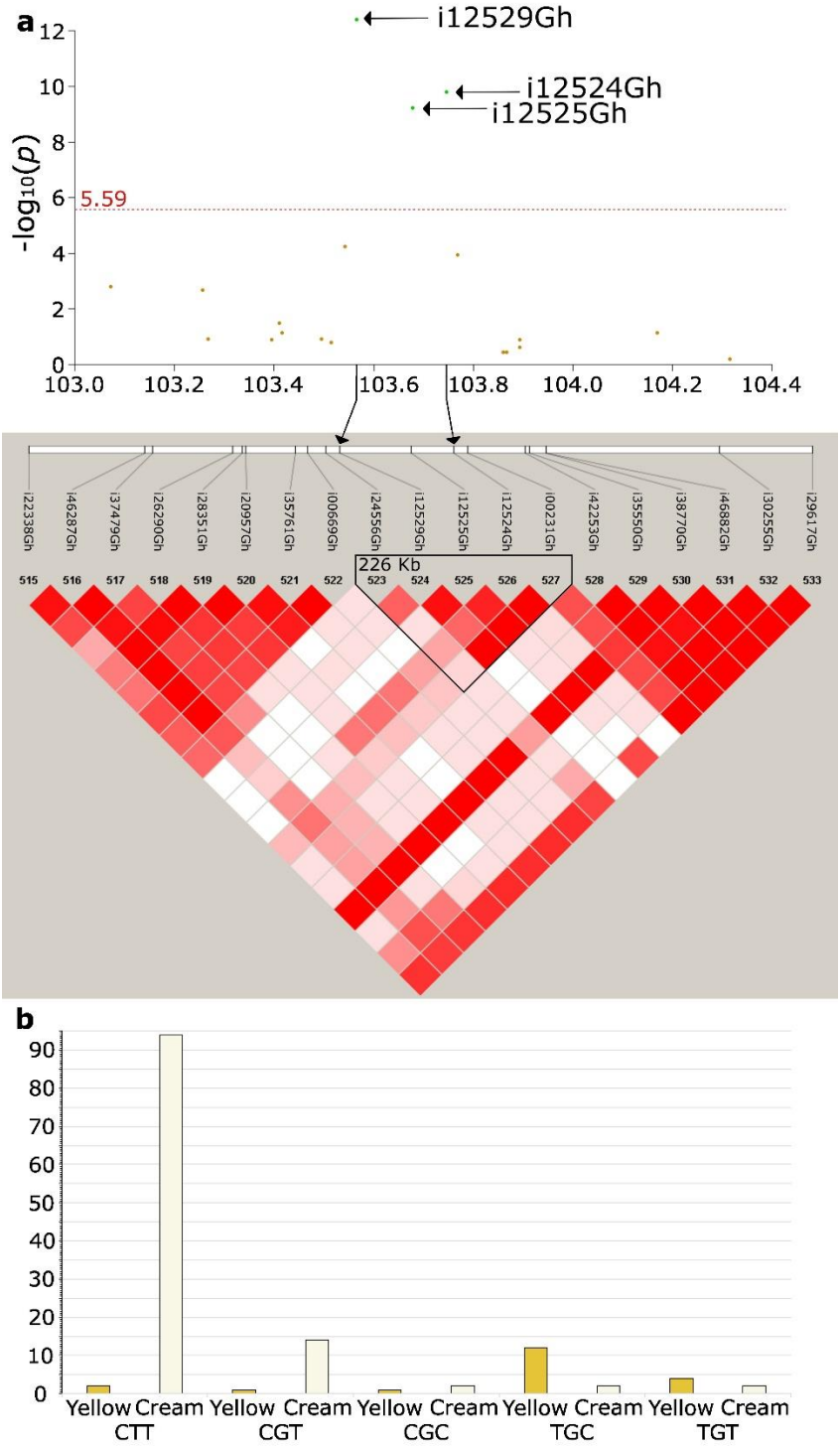


**Figure 4:** Marker allele combinations of pollen color in the GWAS panel. 136 Upland accessions consisting of 99 diversity panel and 37 flowering tropical landrace panel accessions. **(a)** Regional Manhattan plot of the short arm of chromosome A05 identified three marker trait associations over the Bonferroni corrected threshold value ($-Log_{10}$ p-value = 5.59). The three significant signals were found in a 226 Kb region (103.54-103.77 Mb). **(b)** Five marker allele combinations were observed using the three significant markers.

**Figure 4** (continued).



*2.4.4 Linkage mapping*

To further analyze the genetics of pollen color variation through segregating populations and to delineate the genomic region of *Gh-$P_{A1}$*, linkage mapping was performed with the two biparental $F_5$ RIL populations. Analysis of the 63,058 SNP markers in the 40D population showed 8,502 (13.48%) were polymorphic between the parental lines DES 56 and Sealand #2. After filtering for quality markers, 3,454 markers were used in the construction of the 40D linkage map. Using a LOD score of 10.0, a total of 74 linkage groups were calculated. Of those, 65 linkage groups were combined to construct 26 final linkage groups for each chromosome, with a total of 3,264 markers used. Nine linkage groups and a total of 190 SNP markers were not able to be assigned to a chromosome. These compiled 26 linkage groups had an average marker density of 1.67 centimorgan (cM) and spanned 4,941.42 cM in total (Supplementary Table 4; Supplementary figure 1). The A genome spanned a total of 2,478.39 cM with 1,387 markers. The D genome spanned 2,463.02 cM with 1,877 markers. The lowest marker density was found in chromosome A09 with 3.83 cM and the chromosome with the highest marker density was chromosome D07 with 0.80 cM. Two hundred and nine markers (6.40%) showed segregation distortion (Supplementary Table 4). The A genome had more distortion with 115 distorted loci and 9 SDRs as compared to the D genome with 94 distorted loci and only 7 SDRs.

For the 42D population, 63,058 SNP markers were analyzed showing 8,896 markers (14.11%) that were polymorphic between the parental lines of DES 56 and Southland M1. After filtering the markers, 4,337 markers were used to construct the 42D linkage map. For this map, a LOD score of 9.0 identified 74 linkage groups. Sixty one linkage groups (3,913 markers) were used to construct the 26 chromosomal linkage groups. Thirteen linkage groups and a total of 424 markers were unable to be assigned. The final linkage map spanned 5,280.43 cM with an average marker density of 1.52 cM (Supplementary Table 5; Supplementary figure 1). The A genome consisted of 1,741 markers over 2,515.99 cM and the D genome contained 2,172 markers over 2,764.44 cM. The chromosomes with the highest and lowest marker density were D08 (0.69 cM) and A11 (2.75 cM), respectively. A total of 284 markers (7.258%) showed segregation distortion (Supplementary Table 5). As with the 40D population, the A genome had more distortion with 146 distorted loci over 12 SDRs. In contrast, the D genome had 138 distorted loci with 9 SDRs.

Both linkage maps showed the pollen color phenotype segregating with markers mapped to the A05 chromosome based on the marker's physical positions on the *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019). Using the reported anchor simple sequence repeat (SSR) marker JESPR-50, we located the *Gh-*$P_{A1}$ region to the short arm of the A05 chromosome (Shan *et al.*, 2016). Pairwise recombination fractions and LOD score plots for both linkage maps showed the groupings for the markers were correct with no signals representing relatedness of markers between the linkage groups (Supplementary figure 2 A & B). There was evidence of multiple linkage blocks denoted by the yellow squared in the linkage groups. The collinearity between the physical based sequence map with the linkage genetic maps also supported the correct ordering of the markers (Supplementary figure 2 C & B). There were no major chromosome rearrangements between the physical positions of the markers and either of the linkage maps generated. The 40D and 42D linkage maps covered 91.23% and 93.86% (2.08 Gb) of the genome on average (Supplementary Table 6; Supplementary Table 7). Most of the chromosomes had high coverage with most linkage groups covering between 90.06 to 99.60% of the chromosome in the 40D population. Chromosomes A01 (8.91%), D06 (82.60%), D07 (77.13%), and D11 (85.25%) all had relatively less coverage. The 42D population's linkage map also showed high coverage between 90.61 and 99.84% coverage. Chromosomes A01, A01, A04, A07, A09, A12, and D11 all had between 70.39 and 88.64% coverage. Single marker analysis (SMA) was performed on both linkage maps to identify any potential loci with a significant association with pollen color that was missed in GWAS and linkage mapping. SMA for both populations showed a single peak on the A05 chromosome matching the broad area previously identified (Supplementary figure 3). The absence of any additional peaks in any other regions of the genome shows a lack of any additional loci associated with pollen color.

The A05 short arm linkage information was taken and an iteration-based ordering was performed through RECORD for each linkage map. Redundant markers were removed and the reduced linkage maps were combined in a single consensus map (Figure 5). The resulting consensus map contains 60 markers and is 94.49 cM with a marker density of 1.58 cM (Supplementary Table 8). The consensus map corresponds to a 26.36 Mb genomic region between 82.49 and 108.84 Mb which accounts for 24.21% of the total length of the A05 chromosome. In the consensus map, *Gh-*$P_{A1}$ was mapped in a 5.36 cM region between the SNP markers i12529Gh and i12525Gh. *Gh-*$P_{A1}$ was mapped 3.32 cM from i12529Gh and 2.04 cM from i12525Gh on the

short arm of A05 (Figure 5). The i12525Gh SNP marker was also a flanking marker in both linkage maps. Population 40D mapped the *Gh-$P_{A1}$* locus between flanking markers i12525Gh (103.68 - 103.68 Mb proximal to the centromere) and i12524Gh (103.75 - 103.75 Mb proximal to the telomere) while population 42D mapped *Gh-$P_{A1}$* between i12537Gh (103.25 - 103.25 proximal to the centromere) and i12525Gh proximal to the telomere. The mapped flanking SNP markers in the consensus map correspond to a 112.83 Kb (5.36 cM) genomic region of between 103.57 and 103.68 Mb on chromosome A05. There are four candidate genes identified in this genomic region (Table 2).

**Figure 5:** Consensus mapping of the short arm of chromosome A05 made from combining two linkage maps developed from $F_5$ RIL populations (DES 56 × Sealand #2 and DES 56 × Southland M1). RIL populations genotyped using cottonSNP63K array (Hulse-Kemp *et al.*, 2015). Pollen color locus *Gh-*$P_{A1}$ was papped to a 5.36 cM region between flanking SNP markers i12529Gh and i12525Gh. Comparison of marker order of consensus genetic map and physical map of A05 short arm using the *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019).

**Figure 5** (continued).

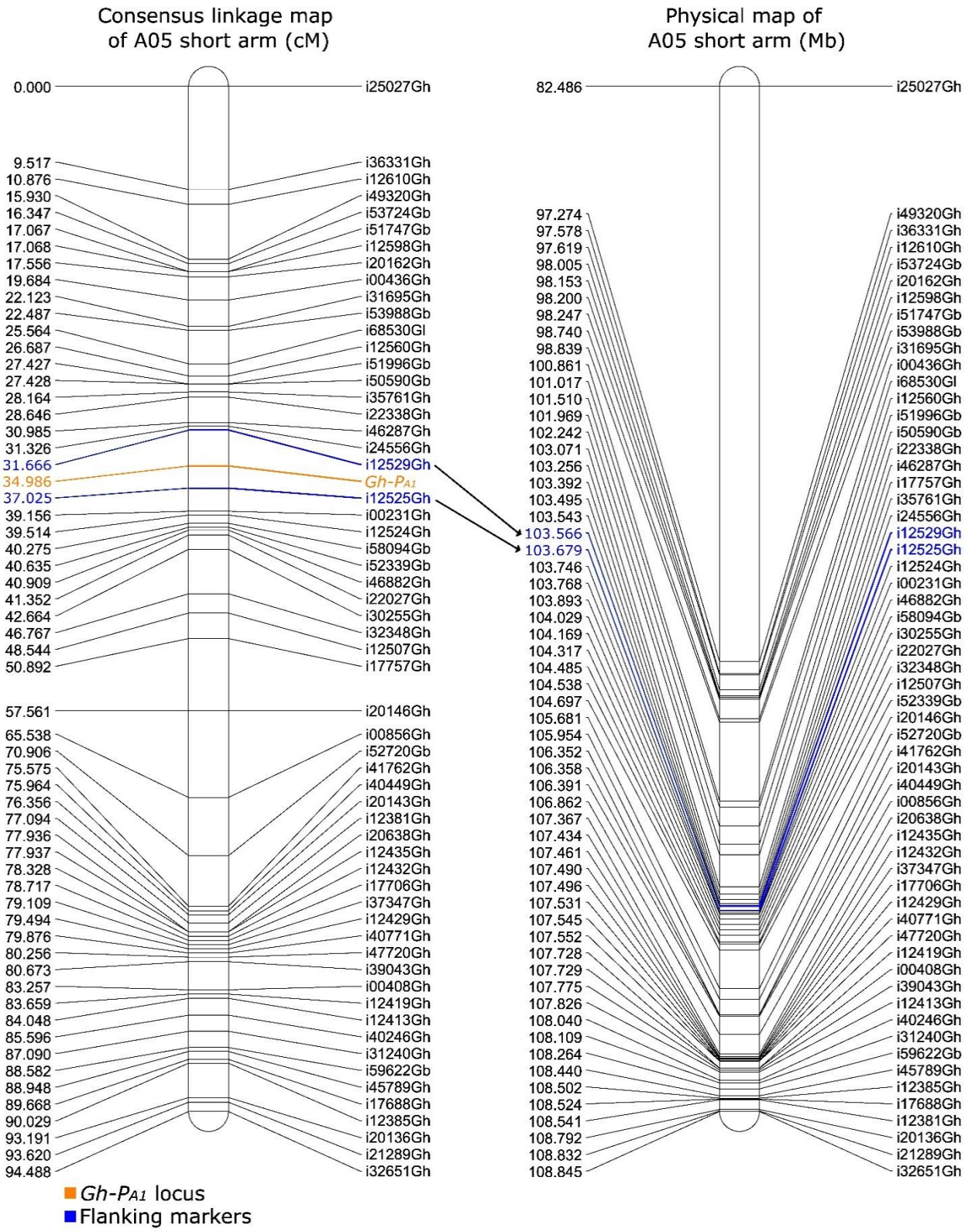


**Table 2:** Putative genes identified by combined GWAS and linkage analyses of *Gh-$P_{A1}$* in the Upland cotton. Biological process obtained from TAIR Arabidopsis database.

| Gene ID | Gene Name | Description | GO Biological process in Arabidopsis | Start (bp) | End (bp) | GWAS or Consensus mapping |
|---|---|---|---|---|---|---|
| Gh_A05G387800 | tmem56-b | Transmembrane protein 56-B | Involved in lipid homeostasis | 103,565,445 | 103,571,999 | GWAS |
| Gh_A05G387900 | PSY | Phytoene synthase, chloroplastic | Involved in the carotenoid biosynthetic process | 103,574,185 | 103,577,072 | GWAS & Consensus |
| Gh_A05G388000 | FER | Receptor-like protein kinase FERONIA | Involved in negative regulation of abscisic acid-activated signaling pathway, negative regulation of cell growth, protein autophosphorylation, response to brassinosteroid, response to ethylene, root development | 103,624,887 | 103,627,655 | GWAS & Consensus |
| Gh_A05G388100 | OsABCB25 | ABC transporter B family member 25 | Involved in transmembrane transport and acts upstream of or within the response to aluminum ion | 103,665,145 | 103,671,638 | GWAS & Consensus |
| Gh_A05G388200 | eif3m | Eukaryotic translation initiation factor 3 subunit M | Involved in cytoplasmic translational initiation and acts upstream of or within the RNA metabolic process | 103,675,744 | 103,678,848 | GWAS & Consensus |
| Gh_A05G388300 | RUB2 | Ubiquitin-NEDD8-like protein RUB2 | Involved in modification-dependent protein catabolic process, protein ubiquitination and acts upstream of or within embryo development ending in seed dormancy, the ethylene biosynthetic process, protein neddylation, the response to auxin | 103,684,425 | 103,686,752 | GWAS |
| Gh_A05G388400 | RUB2 | Ubiquitin-NEDD8-like protein RUB2 | Involved in modification-dependent protein catabolic process, protein ubiquitination and acts upstream of or within embryo development ending in seed dormancy, the ethylene biosynthetic process, protein neddylation, the response to auxin | 103,689,535 | 103,692,103 | GWAS |
| Gh_A05G388500 | CXE18 | Probable carboxylesterase 18 | Involved in pollen tube growth | 103,716,194 | 103,717,015 | GWAS |
| Gh_A05G388600 | CXE18 | Probable carboxylesterase 18 | Involved in pollen tube growth | 103,717,106 | 103,718,255 | GWAS |
| Gh_A05G388700 | SRT2 | NAD-dependent protein deacylase SRT2 | Encodes SRT2, a member of the SIR2 (sirtuin) family HDAC (histone deacetylase) (SRT1/AT5g55760, SRT2/AT5G09230).Homolog of mammalian SIRT7 homolog which was described as an rRNA deacetylase. | 103,746,019 | 103,752,276 | GWAS |

*2.4.5 Candidate gene identification and marker allele combinations*

The genomic region of *Gh-$P_{A1}$* was able to be identified through a combined GWAS and linkage mapping analysis in two RIL populations. GWAS analysis was able to identify a 226 Kb region (103.54-103.77 Mb) with 3 MTAs (i12529Gh, i12525Gh, and i12524Gh). That region was further delineated through linkage analysis using RILs. The consensus linkage map identified a 113 Kb overlapping target genomic region on the short arm of A05.

The three significant MTAs identified through GWAS (i12529Gh, i12525Gh, and i12524Gh) were used to look at marker allele combinations. In this targeted region, five allele combinations were identified (Figure 4 B). The majority of the accessions with yellow pollen were found to have the marker alleles TGC and TGT. Of the 20 lines with any of the "TXX" allele combinations, 16 of them are yellow (80%) and only 4 were cream (20%). The majority of cream pollen accessions were found to have the marker alleles CTT or CGT. Out of the 114 accessions with any of the "CXX" allele combinations, 110 had cream pollen (96.45%). As such, the alleles for the marker i12529Gh (C/T) had the best correlation with the pollen phenotype.

Scanning the *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019) between SNP marker i24556Gh (103.54 Mb) and i00231Gh (103.77 Mb) identified a 225.60 Kb candidate genomic region with ten putative genes (Table 2). Out of these ten genes, only one gene (id Gh_A05G387900) was reported to be involved in any pigmentation pathway. This gene, phytoene synthase, PSY, is known to be the rate limiting step in the carotenoid biosynthetic pathway (Cazzonelli & Pogson, 2010; Rodriguez-Concepcion, 2010). Further, the consensus linkage mapping using RILs showed that two of the three MTAs (i12529Gh and i12525Gh) identified as the two flanking markers to *Gh-$P_{A1}$* locus. Four of the ten putative genes identified in the GWAS delineated region were found to be between these two flanking markers (Table 2). These genes are PSY (103.57 - 103.58 Mb), FERONIA (FER) (103.62 - 103.63 Mb), OsABCB25 an ABC transporter (103.67 - 103.67 Mb), and eif3m a translation initiation factor 3 subunit (103.68 – 103.68 Mb).

## 2.5 Discussion

In this study, using a combination of linkage mapping and GWAS, we identified, mapped and genomically targeted a major locus on chromosome A05 controlling pollen color differences in Upland cotton. Our research also indicated the carotenoid biosynthesis gene phytoene synthase as

a potential candidate gene at the *Gh-$P_{A1}$* locus where the biochemical analyses confirmed that flavonoids are not the phytochemicals causing the pollen coloration in cotton.

*2.5.1 Linkage mapping*

In this study, two biparental $F_5$ RIL populations of 122 and 111 lines were genotyped with the cottonSNP63K array (Hulse-Kemp *et al.*, 2015). In these populations, 8,502 (13.48%) and 8,896 (14.11%) SNP markers were polymorphic with an average marker density of 1.67 and 1.52 respectively (Supplementary Table 4; Supplementary Table 5; Supplementary figure 1). The percentage of marker polymorphisms can vary due to how genetically diverse the parents of the crosses are. The percentage of polymorphic markers is comparable to the range seen in intraspecific *G. hirsutum* crosses in other studies (Gowda *et al.*, 2022; Shrestha *et al.*, 2023; Zhang *et al.*, 2019; Zhu *et al.*, 2021). The two linkage maps had a segregation distortion of 6.40% and 7.26% (Supplementary Table 4; Supplementary Table 5). Linkage distortion can alter the calculated mapping distance as the alleles of markers with linkage distortion are not segregating as expected (Xian *et al.*, 2006). The percentage of segregation distortion observed in this study is lower than what is found in other high-density linkage maps (Gowda *et al.*, 2022; Li *et al.*, 2016; Wang *et al.*, 2019). Segregation distortion may be affected by the parents used, cross type, species, and marker type (Xian *et al.*, 2006). To compare marker order a CIRCOS plot was made to demonstrate the collinearity between the linkage groups and their associated chromosomes from the *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019) (Supplementary figure 2 C & D). These analyses suggested that there were no major structural rearrangements that exist in the parents and mapping populations. Mapping also showed regions of low recombination within a chromosome (Supplementary Figure 2 C & D). Areas with low crossing over or low gene density especially around the centromeres can cause these regions with uneven distribution of recombination (Endo & Gill, 1996; Wang *et al.*, 2013).

Outside of our study, a previous mapping of *$P_1$* by Zhang *et al.* (2016) localized *$P_1$* to an 8.9 cM region of A05 between two simple sequence repeat (SSR) markers (NAU4106-390 and NAU792-400) through interspecific mapping population derived from *G. hirsutum* x *G. darwinii* cross. Yet, the physical positions of the SSR markers was unknown. However, in our current study we were able to accurately map the *Gh-$P_{A1}$* locus within the 5.36 cM region and target it to a small genomic region of 112.83 Kb in *G. hirsutum* in the short arm of A05 (Supplementary Table 8).

*2.5.2 Candidate gene analysis and PSY gene*

GWAS identified a 226 Kb region (A05: 103.54-103.77 Mb) for the pollen color locus *Gh-$P_{A1}$* flanked by the markers i24556Gh and i00231Gh. Using a consensus linkage map of the two RIL populations this region was further delineated to a 112.83 Kb genomic region (A05: 103.57 - 103.68 Mb) that overlaps with the GWAS results. This region is flanked by the SNP markers i12529Gh distal and i12525Gh proximal to the telomere. Scanning of the genomic region identified by GWAS identified 10 putative putative gene sequences. The reduced region identified in consensus mapping narrowed this down to four putative genes. Of the ten putative genes identified in the combined linkage and GWAS region, only one candidate gene, PSY (id Gh_A05G387900), was identified to be involved in pigmentation in functional genomic analysis. PSY is the first step in the carotenoid biosynthesis pathway which condenses two geranylgeranyl diphosphate (GGPP) into phytoene (Cazzonelli & Pogson, 2010; Rodriguez-Concepcion, 2010). This is known to be a rate limiting step in the carotenoid pathway making it of particular importance. As such, PSY is often used in plant studies to increase carotenoid or vitamin A content in crops including sweet potato (Park *et al.*, 2016), potato (Römer *et al.*, 2002), tomato (Fraser *et al.*, 2009) and more. The most well known example of PSY is from the transformation studies creating Golden Rice (Datta *et al.*, 2006; Schaub *et al.*, 2005; Ye *et al.*, 2000). Studies have found that one genome can have multiple PSY genes that may be tissue specific (Welsch *et al.*, 2008). Cotton itself has a number of PSY genes found on multiple chromosomes throughout its genome (Supplementary Table 9). The absence of any flavonoid genes is supportive of the flavonoid biochemical quantification included in this study. The flavonols chosen to be quantified in this study were based on previous studies into floral and anther biochemical analysis in cotton (Hanny, 1980; Ismailov *et al.*, 1994; Nix *et al.*, 2017). The absence of significant associations between any of the yellow and cream pollen supports previous work on flavonoid differentiation between yellow and cream anthers (Hanny, 1980). This suggests that the pollen coloration evident in cotton is not due to a difference in flavonoid profiles, and is likely due to carotenoid accumulation.

*2.5.3 Historic translocation event gave rise to Gh-$P_{A1}$*

Linkage analysis has identified *Gh-$P_{A1}$* to be the PSY gene on chromosome A05. However, the homeologous chromosome D05 does not have orthologous PSY gene sequence. Instead the PSY orthologue was found on chromosome D04 as identified by obtaining BLAST positions of SNP

markers as well as through GWAS. This area has markers on both A05 and D04 with equally high Bit scores. We were able to localize *Gh-$P_{A1}$* to A05 through the linkage grouping data. Through this we can conclude that this region was a part of the chromosome D04 to D05 translocation event in cotton's tetraploid evolution as described by Huang *et al.* (2020) and Chen *et al.* (2020). Huang *et al.* (2020) identified two large reciprocal translocations between chromosome D04 and D05 in *Gossypium* spp. after polyploidization. Chen *et al.* (2020) further identified structural variations between the A and D subgenomes in five tetraploid species of cotton. Further studies can be done on the PSY copy from the D04 chromosome to determine if its function has undergone subfunctionalization or neofunctionalization. Exploration into this homologous region may lead to insight into the other $P_2$ gene discussed in Pima that gives rise to orange pollen coloration.

Based on the complementary gene model proposed by Turcotte and Feaster (1966) in Pima cotton, if Upland cotton is also controlled by two complementary genes then *Gh-$P_{A1}$* functions as $P_1$. Under this model, the yellow pigmentation is attributed to the presence of both $P_1$ and $P_2$ through the allele combination of $P_1P_1P_2P_2$. As such, we can infer that the cream pollen observed in Upland cotton is normal "cream" ($p_1p_1P_2P_2$) and not "neocream" ($p_1p_1p_2p_2$). We observed yellow pollen in Upland cotton ($P_1P_1P_2P_2$) which can only be obtained if Upland cotton cream lines already have copies of $P_2$. The presence of a $P_2$ gene in Upland cotton is not confirmed as no orange pollen Upland cotton lines ($P_1P_1p_2p_2$) have been identified. As a result, further study into the $P_2$ gene may lead to a greater understanding of the molecular controls of pollen color in Upland cotton.

### *2.5.4 Origins of yellow pollen in Upland cotton*

Although the pollen color phenotypes of yellow and cream color are easily discernible using visual observations, a minor variation existed within the yellow color phenotypes where the yellow color in Sealand #2 is lighter than the phenotype in Southland M1 (Figure 1). However, the observed expected segregation ratios in both populations and the identification and mapping of *Gh-$P_{A1}$* locus in both RIL populations as well as a single peak in GWAS study controlling the color differences suggest variation within yellow color genotypes could be minor and be due to expressivity variation of the yellow color alleles. Previous studies in the plant *Campanula americana* also show an expression difference in pollen color (Koski & Galloway, 2018; Lau & Galloway, 2004). We

also screened 320 Pima lines (unpublished) for pollen coloration none of which displayed cream pollen.

Based on these results, *Gh-$P_{A1}$* was likely introduced into a number of Upland cotton through a history of Pima ancestry. In this study, the yellow accession Sealand #2 was derived from Sea Island cotton plants (Culp & Harrell, 1974; Kumar *et al.*, 2019; Smith & Cothren, 1999; Ulloa *et al.*, 2007). Sea Island plants are reported to be hybrid varieties of primitive forms of Pima and a wild West Indian form of Upland cotton grown in the tidewater and offshore islands on the east coast of the U.S. (Stephens, 1976). These Sea Islands were classified botanically as Pima lines and retain many of the superior fiber qualities of Pima (Stephens, 1976). These Sea Island lines were introgressed into Upland cultivars to create the Sealand accessions by the Pee Dee breeding program (Culp & Harrell, 1974; Kumar *et al.*, 2019; Smith & Cothren, 1999; Ulloa *et al.*, 2007). Therefore, it is likely that the yellow pollen observed in the Sealand Upland cotton accessions originated from Pima cotton. As such, it is possible to use pollen pigmentation to explore the relationship between Pima and Upland lines and may be incorporated into phylogenetic studies of *Gossypium* species. Additionally, the Pee Dee breeding program later utilized the Sealand accessions to develop "PD" lines including PD 62-164 which is the male parent of DES 56 (female parent Stoneville 213) (Bridge & Chism, 1978; Culp & Harrell, 1974). This suggests a possible distant relatedness between Sealand #2 and DES 56 which may explain the low recombination between markers and reduced marker coverage as compared to the 42D population.

While many studies looked at floral pigment profiles and dimorphisms, relatively few consider pollen color despite a number of plants displaying naturally occurring color polymorphisms (Maize, Coe *et al.*, 1981; *Campanula*, Lau & Galloway, 2004; *Nigella*, Andersson & Jorgensen, 2005; California poppy, Wakelin *et al.*, 2003; Sunflower, Fambrini *et al.*, 2010; Hazelnut, Mehlenbacher & Smith, 2002; Petunia, De Vlamine *et al.*, 1984). Some reports determined pollen color to be associated with floral traits like petal color (Wakelin *et al.*, 2003; Wolfe, 2001), while others found it to be unrelated to other floral traits (Koski *et al.*, 2020). This suggests that in some plants pollen pigmentation is either linked to other floral traits or they are controlled by the same gene(s) and in other plants pollen pigmentation can be independently controlled. While not many loci are mapped (exception of De Vlamine *et al.*, 1984 in petunia), segregation ratios of the pollen color dimorphism suggest simple genetic controls of one (Andersson & Jorgensen, 2005; De Vlamine *et al.*, 1984; Mehlenbacher & Smith, 2002; Rafinski,

1979; Wakelin *et al.*, 2003) or two genes (Coe *et al.*, 1981; Fambrini *et al.*, 2010). The pollen pigmentation in most of the studied dimorphic plants have attributed the color differentiation to be due to different expressions of flavonoids including maize (Coe *et al.*, 1981), the bristle-cone pine (Johnson & Critchfield, 1974), and flowering plants like *C. americana* (Koski *et al.*, 2020) and petunia (De Vlamine *et al.*, 1984; Napoli *et al.*, 1999). Similarly to cotton's phenotypes, the sunflower and california poppy have instances of cream and orange pollen both of which have been associated with an increase in carotenoid content in the yellow/ orange pollen through simple genetic control (Fambrini *et al.*, 2010; Wakelin *et al.*, 2003). Associations with pollen coloration have been found with geographic distribution (Jorgensen & Andersson, 2005; Jorgensen *et al.*, 2006; Rafinski, 1979; Wang *et al.*, 2018) suggesting a local adaptation in pollen color evolution. Additionally, differences in pollen coloration have been associated with light intensity (Wang *et al.*, 2018), drought tolerance (Jorgensen & Andersson, 2005), tolerance to nutrient deficiency (Jorgensen & Andersson, 2005), heat tolerance (Koski & Galloway, 2018; Koski *et al.*, 2020), and pollinator preference (Koski *et al.*, 2020). Considering the roles of carotenoids in plants, accumulation in pollen through *Gh-*$P_{A1}$ may provide insight into Pima's increased fiber quality through ABA synthesis and ROS maintenance (Baker, 2008; Latowski *et al.*, 2011). Yellow pollen may also explain Pima's increased pollen load and could potentially be more tolerant to abiotic stress due to the antioxidant properties of carotenoids (Baker, 2008; Chinnusamy *et al.*, 2008; Du *et al.*, 2010; Wang *et al.*, 2018). As an easily visible single major gene, *Gh-*$P_{A1}$ would be simple to select for to include in breeding programs to potentially improve Upland cotton's abiotic stress tolerance in one of the plant's most vulnerable and economically important structures. Genomic targeted mapping of the locus (*Gh-*$P_{A1}$) controlling major pollen color variation and identification of candidate gene in the current study is the first effort to systematically establish the molecular genetic architecture of *Gh-*$P_{A1}$ and the biochemical processes of carotenoids in pollen physiology and their functional role in reproductive success and plant adaptation all of which are critical for cotton productivity.

**Acknowledgements:** The authors acknowledge the funding support provided by North Carolina Cotton Producers' Association and Cotton Incorporated (CI) for the current research and an assistantship to Nicole Pettit. We also thank Kelli Kochan of Texas A&M University for her help with genotyping the lines using the CottonSNP63K array. Thanks to Mohammad Foteh Ali and

Colleen Roark from Dr. Gloria Muday's lab in Wake Forest University for their help with the biochemical LC-MS analysis and Adchayan Santhirasegaran, Tyler Reid, Christopher Lee, and Rakesh Gowda for NC State University for their help in pollen sample collections. We appreciate Josh Udall, James Frelichowski, and Janna Love of the Crop Germplasm Research, USDA-ARS, College Station, TX for providing the Upland and Pima cotton accessions used in this study.

**Conflict of interest:** Authors declare no conflict of interest

**Funding:** Funding for this research was supported by NC Cotton Growers Association and Cotton Incorporated through an assistantship to Nicole Pettit and salary support to Dr. Anajan Gowda.

**SUPPLEMENTAL MATERIAL**: Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Supplementary figure 1:** SNP marker density and distribution for 26 Upland cotton chromosomes in linkage mapping and the genome wide association study. **a.** Distribution of 3,264 polymorphic SNPs in 122 RILs from the 40D $F_5$ population (DES 56 × Sealand #2). SNP density was calculated by number of SNP markers per cM using the cottonSNP63K array (Hulse-Kemp *et al.*, 2015). **b.** Distribution of 3,913 polymorphic SNP markers in 111 RILs from the 42D $F_5$ population (DES 56 × Southland M1). Marker density calculated as described in part A. **C.** Distribution of 19,349 polymorphic SNP markers in 136 Upland accessions. Marker density calculated by a 1 Mb sliding interval window.

**Supplementary figure 2:** Heat maps comparing the precise grouping and ordering of SNP markers in the linkage group corresponding to the 26 Upland cotton chromosomes for the two $F_5$ RIL populations **a.** heat map for 40D (DES 56 × Sealand #2) population based map **b.** heat map for 42D (DES 56 × Southland M1) population based map . The upper left triangle plots recombination fractions of each marker in a linkage group by all the other linkage groups. The lower right triangle plots the LOD scores between each marker. **c-d.** Comparison of collinearity between the genetically mapped linkage groups and the sequence based physical chromosome map of *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019).

**Supplementary figure 3:** Single marker analysis of pollen color for all 26 Upland cotton chromosomes for **a.** 40D $F_5$ RIL population (DES 56 × Sealand #2) and **b.** 42D $F_5$ RIL

population (DES 56 × Southland M1). A significant peak above the LOD threshold level 2.5 was observed above the A05 chromosome.

**Supplementary Table 1:** List of the 99 accessions used from the elite diversity panel (Tyagi *et al.*, 2014) and 37 accessions from the tropical landrace panel (Kaur *et al.*, 2017) used in GWAS analysis of pollen color in *G. hirsutum*.

**Supplementary Table 2:** SNP density and coverage statistics of 19,349 polymorphic SNP identified in 136 *G. hirsutum* accessions used for GWAS.

**Supplementary Table 3:** Percentage of phenotypic variance explained by the SNP markers associated with *Gh-*$P_{A1}$ identified by GWAS analysis.

**Supplementary Table 4:** Chromosomal marker information of the linkage map of the $F_5$ RIL population of DES 56 × Sealand #2. Markers were taken from the Cotton SNP 63K genotyping array (Hulse-Kemp *et al.*, 2015).

**Supplementary Table 5:** Chromosomal marker information from the linkage map derived from the $F_5$ RIL population of DES 56 × Southland M1. Markers were taken from the Cotton SNP 63K genotyping array (Hulse-Kemp *et al.*, 2015).

**Supplementary Table 6:** Coverage statistics of the $F_5$ RIL population of DES 56 × Sealand #2 linkage map in the sequence-based physical map of the *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019).

**Supplementary Table 7:** Coverage statistics of the linkage map developed from an $F_5$ RIL population of DES 56 × Southland M1 in the sequence-based physical map of the *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019).

**Supplementary Table 8:** Consensus linkage map information of the *Gh-*$P_{A1}$ locus on the short arm of chromosome A05.

**Supplementary Table 9:** PSY and associated genes found in *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019).

## REFERENCES

Andersson, S., & Jorgensen, T. H. (2005). The genetic basis of naturally occurring pollen color dimorphisms in Nigella degenii (Ranunculaceae). *Journal of Heredity*, *96*(5), 550-556. https://doi.org/10.1093/jhered/esi098

Azhar, M. T., Wani, S. H., Chaudhary, M. T., Jameel, T., Kaur, P., & Du, X. (2020). Heat tolerance in cotton: morphological, physiological, and genetic perspectives. In: *Heat Stress Tolerance in Plants: Physiological, Molecular and Genetic Perspectives*, 1-22. https://doi.org/10.1002/9781119432401.ch1

Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annual Review of Plant Biology*, *59*, 89-113. https://doi.org/10.1146/annurev.arplant.59.032607.092759

Balls, W. L. (1912). *The cotton plant in Egypt: studies in physiology and genetics*. Macmillan and Company, limited.

Barrett, J. C., Fry, B., Maller, J. D. M. J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263-265. https://doi.org/10.1093/bioinformatics/bth457

Beasley, C. A., & Ting, I. P. (1973). The effects of plant growth substances on in vitro fiber development from fertilized cotton ovules. *American Journal of Botany*, *60*(2), 130-139. https://doi.org/10.1002/j.1537-2197.1973.tb10209.x

Bradshaw Jr, H. D., & Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, *426*(6963), 176-178. https://doi.org/10.1038/nature02106

Bridge, R. R., & Chism, J. F. (1978). Registration of DES 56 Cotton 1 (Reg. No. 70). *Crop Science*, *18*(3), 524-524. https://doi.org/10.2135/cropsci1978.0011183X001800030052x

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889-890. https://doi.org/10.1093/bioinformatics/btg112

Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, *108*(10), 1880-1890. https://doi.org/10.1016/j.ajhg.2021.08.005

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, *103*(3), 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Cazzonelli, C. I., & Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends in plant science*, *15*(5), 266-274. https://doi.org/10.1016/j.tplants.2010.02.003

Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., Ding, M., Ye, W., Kirkbride, R. C., Jenkins, J., Plott, C., Lovell, J., Lin, Y. M., Vaughn, R., Liu, B., Simpson, S., Scheffler, B. E., Li, W., Saski, C. A., Grover, C. E., Hu, G., Conover, J. L., Carlson, J. W., Shu, S., Boston, L. B., Williams, M., Peterson, D. G., McGee, K., Jones, D. C., Wendel, J. F., Stelly, D. M., Grimwood, J., & Schmutz, J. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nature genetics*, *52*(5), 525-533. https://doi.org/10.1038/s41588-020-0614-5

Cheong, W. H., Tan, Y. C., Yap, S. J., & Ng, K. P. (2015). ClicO FS: an interactive web-based service of Circos. *Bioinformatics*, *31*(22), 3685-3687. https://doi.org/10.1093/bioinformatics/btv433

Chinnusamy, V., Gong, Z., & Zhu, J. K. (2008). Abscisic acid-mediated epigenetic processes in plant development and stress responses. *Journal of integrative plant biology*, *50*(10), 1187-1195. https://doi.org/10.1111/j.1744-7909.2008.00727.x

Coe, E. H., McCormick, S. M., & Modena, S. A. (1981). White pollen in maize. *Journal of Heredity*, *72*(5), 318-320. https://doi.org/10.1093/oxfordjournals.jhered.a109514

Culp, T. W., & Harrell, D. C. (1974). Breeding quality cotton at the Pee Dee Experiment Station, Florence, South Carolina. *USDA-ARS Publication, ARS-S-30*.

Datta, K., Rai, M., Parkhi, V., Oliva, N., Tan, J., & Datta, S. K. (2006). Improved 'golden' indica rice and post-transgeneration enhancement of metabolic target products of carotenoids (β-carotene) in transgenic elite cultivars (IR64 and BR29). *Current Science*, 935-939. http://www.jstor.org/stable/24094294

De Vlaming, P., Gerats, A. G. M., Wiering, H., Wijsman, H. J. W., Cornu, A., Farcy, E., & Maizonnier, D. (1984). Petunia hybrida: a short description of the action of 91 genes, their origin and their map location. *Plant Molecular Biology Reporter*, *2*, 21-42. https://doi.org/10.1007/BF03015868

Du, H., Wang, N., Cui, F., Li, X., Xiao, J., & Xiong, L. (2010). Characterization of the β-carotene hydroxylase gene *DSM2* conferring drought and oxidative stress resistance by

increasing xanthophylls and abscisic acid synthesis in rice. *Plant physiology*, *154*(3), 1304-1318. https://doi.org/10.1104/pp.110.163741

Endelman, J. B., & Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics*, *30*(11), 1623-1624. https://doi.org/10.1093/bioinformatics/btu091

Endo, T. R., & Gill, B. S. (1996). The deletion stocks of common wheat. *Journal of Heredity*, *87*(4), 295-307. https://doi.org/10.1093/oxfordjournals.jhered.a023003

Fambrini, M., Michelotti, V., & Pugliesi, C. (2010). Orange, yellow and white-cream: inheritance of carotenoid-based colour in sunflower pollen. *Plant Biology*, *12*(1), 197-205. https://doi.org/10.1111/j.1438-8677.2009.00205.x

Feng, H., Tian, X., Liu, Y., Li, Y., Zhang, X., Jones, B. J., Sun, Y., & Sun, J. (2013). Analysis of flavonoids and the flavonoid structural genes in brown fiber of upland cotton. *PLoS One*, *8*(3). https://doi.org/10.1371/journal.pone.0058820

Fraser, P. D., Enfissi, E. M., & Bramley, P. M. (2009). Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. *Archives of Biochemistry and Biophysics*, *483*(2), 196-204. https://doi.org/10.1016/j.abb.2008.10.009

Gokani, S. J., Kumar, R., & Thaker, V. S. (1998). Potential role of abscisic acid in cotton fiber and ovule development. *Journal of plant growth regulation*, *17*, 1-5. https://doi.org/10.1007/PL00007005

Gowda, S. A., Shrestha, N., Harris, T. M., Phillips, A. Z., Fang, H., Sood, S., Zhang, K., Bourland, F., Bart, R., & Kuraparthy, V. (2022). Identification and genomic characterization of major effect bacterial blight resistance locus (*BB-13*) in Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*, *135*(12), 4421-4436. https://doi.org/10.1007/s00122-022-04229-2

Hanny, B. W. (1980). Gossypol, flavonoid, and condensed tannin content of cream and yellow anthers of five cotton (*Gossypium hirsutum* L.) cultivars. *Journal of Agricultural and Food Chemistry*, *28*(3), 504-506. https://doi.org/10.1021/jf60229a021

Hanny, B. W., Bailey, J. C., & Meredith Jr, W. R. (1979). Yellow cotton pollen suppresses growth of larvae of tobacco budworm. *Environmental Entomology*, *8*(4), 706-707. https://doi.org/10.1093/ee/8.4.706

Harland, S. C. (1929). The genetics of cotton. Part II. The inheritance of pollen colour in New World cottons. *Journal of Genetics*, *20*(3), 387-399.

Hashimoto, H., Uragami, C., & Cogdell, R. J. (2016). Carotenoids and photosynthesis. *Carotenoids in nature: Biosynthesis, regulation and function*, 111-139. https://doi.org/10.1007/978-3-319-39126-7_4

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800-802. https://doi.org/10.1093/biomet/75.4.800

Holladay, S. K., Bridges, W. C., Jones, M. A., & Campbell, B. T. (2021). Yield performance and fiber quality of Pima cotton grown in the southeast United States. *Crop Science*, *61*(4), 2423-2434. https://doi.org/10.1002/csc2.20505

Huang, G., Wu, Z., Percy, R. G., Bai, M., Li, Y., Frelichowski, J. E., Hu, J., Wang, K., Yu, J. Z., & Zhu, Y. (2020). Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nature genetics*, *52*(5), 516-524. https://doi.org/10.1038/s41588-020-0607-4

Hulse-Kemp, A. M., Lemm, J., Plieske, J., Ashrafi, H., Buyyarapu, R., Fang, D. D., Frelichowski, J., Giband, M., Hague, S., Hinze, L. L., Kochan, K. J., Riggs, P. K., Scheffler, J. A., Udal, J. A., Ulloa, M., Wang, S. S., Zhu, Q. H., Bag, S. K., Bhardwaj, A., . . . Stelly, D. M. (2015). Development of a 63K SNP array for cotton and high-density mapping of intraspecific and interspecific populations of *Gossypium* spp. *G3: Genes Genomes Genetics*, *5*(6), 1187–1209. https://doi.org/10.1534/g3.115.018416

Ismailov, A. I., Karimdzhanov, A. K., Islambekov, S. Y., & Rakhimkhanov, Z. B. (1994). Flavonoids of the cotton plant and plants close to it. *Chemistry of Natural Compounds*, *30*, 1-14. https://doi.org/10.1007/BF00638411

JOHNSON, L. C., & CRITCHFIELD, W. B. (1974). A white-pollen variant of bristle-cone pine. *Journal of Heredity*, *65*(2), 123-123. https://doi.org/10.1093/oxfordjournals.jhered.a108471

Jorgensen, T. H., & Andersson, S. (2005). Evolution and maintenance of pollen-colour dimorphisms in *Nigella degenii*: habitat-correlated variation and morph-by-environment interactions. *New Phytologist*, *168*(2), 487-498. https://doi.org/10.1111/j.1469-8137.2005.01525.x

Jorgensen, T. H., Richardson, D. S., & Andersson, S. (2006). Comparative analyses of population structure in two subspecies of *Nigella degenii*: evidence for diversifying selection on pollen-color dimorphisms. *Evolution*, *60*(3), 518-528. https://doi.org/10.1111/j.0014-3820.2006.tb01133.x

Kaur, B., Tyagi, P., & Kuraparthy, V. (2017). Genetic diversity and population structure in the landrace accessions of *Gossypium hirsutum*. *Crop Science*, *57*(5), 2457-2470. https://doi.org/10.2135/cropsci2016.12.0999

Khan, A. Q., Li, Z., Ahmed, M. M., Wang, P., Zhang, X., & Tu, L. (2019). Eriodictyol can modulate cellular auxin gradients to efficiently promote in vitro cotton fibre development. *BMC plant biology*, *19*(1), 1-15. https://doi.org/10.1186/s12870-019-2054-x

Koes, R. E., Quattrocchio, F., & Mol, J. N. (1994). The flavonoid biosynthetic pathway in plants: function and evolution. *BioEssays*, *16*(2), 123-132. https://doi.org/10.1002/bies.950160209

Koski, M. H., & Galloway, L. F. (2018). Geographic variation in pollen color is associated with temperature stress. *New Phytologist*, *218*(1), 370-379. https://doi.org/10.1111/nph.14961

Koski, M. H., Berardi, A. E., & Galloway, L. F. (2020). Pollen colour morphs take different paths to fitness. *Journal of evolutionary biology*, *33*(4), 388-400. https://doi.org/10.1111/jeb.13599

Kumar, P., Singh, R., Lubbers, E. L., Shen, X., Paterson, A. H., Campbell, B. T., Jones, D. C., & Chee, P. W. (2019). Genetic evaluation of exotic chromatins from two obsolete interspecific introgression lines of upland cotton for fiber quality improvement. *Crop Science*, *59*(3), 1073-1084. https://doi.org/10.2135/cropsci2018.12.0745

Latowski, D., Kuczyńska, P., & Strzałka, K. (2011). Xanthophyll cycle–a mechanism protecting plants against oxidative stress. *Redox Report*, *16*(2), 78-90. https://doi.org/10.1179/174329211X13020951739938

Lau, J. A., & Galloway, L. F. (2004). Effects of low-efficiency pollinators on plant fitness and floral trait evolution in *Campanula americana* (Campanulaceae). *Oecologia*, *141*, 577-583. https://doi.org/10.1007/s00442-004-1677-1

Li, C., Dong, Y., Zhao, T., Li, L., Li, C., Yu, E., Mei, L., Daud, M. K., He, Q., Chen, J., & Zhu, S. (2016). Genome-wide SNP linkage mapping and QTL analysis for fiber quality and

yield traits in the upland cotton recombinant inbred lines population. *Frontiers in plant science*, *7*, 1356. https://doi.org/10.3389/fpls.2016.01356

Li, H. B., Qin, Y. M., Pang, Y., Song, W. Q., Mei, W. Q., & Zhu, Y. X. (2007). A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. *New Phytologist*, *175*(3), 462-471. https://doi.org/10.1111/j.1469-8137.2007.02120.x

Li, X., Mitchell, M., Rolland, V., Allen, S., MacMillan, C., & Pettolino, F. (2023). 'Pink cotton candy'—A new dye-free cotton. *Plant Biotechnology Journal*, *21*(4), 677. https://doi.org/10.1111/pbi.13990

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., & Chang, W. (2004). Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature*, *428*(6980), 287-292. https://doi.org/10.1038/nature02373

Lu, N., Roldan, M., & Dixon, R. A. (2017). Characterization of two TT2-type MYB transcription factors regulating proanthocyanidin biosynthesis in tetraploid cotton, *Gossypium hirsutum*. *Planta*, *246*, 323-335. https://doi.org/10.1007/s00425-017-2682-z

Mehlenbacher, S. A., & Smith, D. C. (2002). Inheritance of pollen color in hazelnut. *Euphytica*, *127*, 303-307. https://doi.org/10.1023/A:1020345004297

Mei, W., Qin, Y., Song, W., Li, J., & Zhu, Y. (2009). Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. *Journal of Genetics and Genomics*, *36*(3), 141-150. https://doi.org/10.1016/S1673-8527(08)60101-0

Mierziak, J., Kostyn, K., & Kulma, A. (2014). Flavonoids as important molecules of plant interactions with the environment. *Molecules*, *19*(10), 16240-16265. https://doi.org/10.3390/molecules191016240

Muhlemann, J. K., Younts, T. L., & Muday, G. K. (2018). Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proceedings of the National Academy of Sciences*, *115*(47), E11188-E11197. https://doi.org/10.1073/pnas.1811492115

Napoli, C. A., Fahy, D., Wang, H. Y., & Taylor, L. P. (1999). White anther: a petunia mutant that abolishes pollen flavonol accumulation, induces male sterility, and is complemented

by a chalcone synthase transgene. *Plant Physiology*, *120*(2), 615-622. https://doi.org/10.1104/pp.120.2.615

Narbona, E., del Valle, J. C., & Whittall, J. B. (2021). Painting the green canvas: how pigments produce flower colours. *The Biochemist*, *43*(3), 6-12. https://doi.org/10.1042/bio_2021_137

Nix, A., Paull, C., & Colgrave, M. (2017). Flavonoid profile of the cotton plant, *Gossypium hirsutum*: A review. *Plants*, *6*(4), 43. https://doi.org/10.3390/plants6040043

Park, S., Kim, H. S., Jung, Y. J., Kim, S. H., Ji, C. Y., Wang, Z., Jeong, J. C., Lee, H-S., Lee, S. Y., & Kwak, S. S. (2016). Orange protein has a role in phytoene synthase stabilization in sweetpotato. *Scientific reports*, *6*(1), 33563. https://doi.org/10.1038/srep33563

Pierce, E. T., Graham, B. P., Stiff, M. R., Osborne, J. A., & Haigler, C. H. (2019). Cultures of *Gossypium barbadense* cotton ovules offer insights into the microtubule-mediated control of fiber cell expansion. *Planta*, *249*, 1551-1563. https://doi.org/10.1007/s00425-019-03106-5

Pogson, B. J., Rissler, H. M., & Frank, H. A. (2005). The roles of carotenoids in photosystem II of higher plants. In 'Photosystem II: the light-driven water: plastoquinone oxidoreductase,. (Eds T Wydrzynski, K Satoh) pp. 515-537.

Polívka, T., & Frank, H. A. (2010). Molecular factors controlling photosynthetic light harvesting by carotenoids. *Accounts of chemical research*, *43*(8), 1125-1134. https://doi.org/10.1021/ar100030m

Rafiński, J. N. (1979). Geographic variability of flower colour in *Crocus scepusiensis* (Iridaceae). *Plant Systematics and Evolution*, *131*, 107-125. https://doi.org/10.1007/BF00984125

Rezanejad, F. (2009). Air pollution effects on structure, proteins and flavonoids in pollen grains of *Thuja orientalis* L.(Cupressaceae). *Grana*, *48*(3), 205-213. https://doi.org/10.1080/00173130902949417

Rodríguez-Concepción, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. *Archives of Biochemistry and Biophysics*, *504*(1), 118-122. https://doi.org/10.1016/j.abb.2010.06.016

Römer, S., Lübeck, J., Kauder, F., Steiger, S., Adomat, C., & Sandmann, G. (2002). Genetic engineering of a zeaxanthin-rich potato by antisense inactivation and co-suppression of

carotenoid epoxidation. *Metabolic engineering*, *4*(4), 263-272. https://doi.org/10.1006/mben.2002.0234

Rutley, N., Miller, G., Wang, F., Harper, J. F., Miller, G., & Lieberman-Lazarovich, M. (2021). Enhanced reproductive thermotolerance of the tomato high pigment 2 mutant is associated with increased accumulation of flavonols in pollen. *Frontiers in Plant Science*, *12*, 672368. https://doi.org/10.3389/fpls.2021.672368

Schaub, P., Al-Babili, S., Drake, R., & Beyer, P. (2005). Why is golden rice golden (yellow) instead of red?. *Plant Physiology*, *138*(1), 441-450. https://doi.org/10.1104/pp.104.057927

Shan, W., Jiang, Y., Han, J., & Wang, K. (2016). Comprehensive cytological characterization of the *Gossypium hirsutum* genome based on the development of a set of chromosome cytological markers. *The Crop Journal*, *4*(4), 256-265. https://doi.org/10.1016/j.cj.2016.04.001

Shim, H., Chasman, D. I., Smith, J. D., Mora, S., Ridker, P. M., Nickerson, D. A., Krauss, R. M., & Stephens, M. (2015). A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PloS one*, *10*(4), e0120758. https://doi.org/10.1371/journal.pone.0120758

Shrestha, N., Zhang, K., Gowda, S. A., Abdelraheem, A., Jones, D. C., & Kuraparthy, V. (2023). Identification of quantitative trait loci for fiber quality, yield, and plant height traits in Upland cotton. *Crop Science*. https://doi.org/10.1002/csc2.20937

Smith, C. W., & Cothren, J. T. (Eds.). (1999). *Cotton: origin, history, technology, and production* (Vol. 4). John Wiley & Sons. https://doi.org/10.2135/cropsci2000.4051492ax

Stephens, S. G. (1954). Interspecific homologies between gene loci in *Gossypium*. I. Pollen color. *Genetics*, *39*(5), 701. https://doi.org/10.1093/genetics/39.5.701

Stephens, S. G. (1976). The origin of Sea Island cotton. *Agricultural History*, *50*(2), 391-399. http://www.jstor.org/stable/3741338

Tan, J., Tu, L., Deng, F., Hu, H., Nie, Y., & Zhang, X. (2013). A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant physiology*, *162*(1), 86-95. https://doi.org/10.1104/pp.112.212142

Turcotte, E. L., & Feaster, C. V. (1966). A Second Locus for Pollen Color in Pima Cotton, *Gossypium barbadense* L. 1. *Crop Science*, *6*(2), 117-119. https://doi.org/10.2135/cropsci1966.0011183X000600020003x

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. J. Open Source Softw. 3: 731.

Tuttle, J. R., Nah, G., Duke, M. V., Alexander, D. C., Guan, X., Song, Q., Chen. Z. J.m, Scheffler, B. E., & Haigler, C. H. (2015). Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. *BMC genomics*, *16*(1), 1-28. https://doi.org/10.1186/s12864-015-1708-9

Tyagi, P., Gore, M. A., Bowman, D. T., Campbell, B. T., Udall, J. A., & Kuraparthy, V. (2014). Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*, *127*, 283-295. https://doi.org/10.1007/s00122-013-2217-3

Ulloa, M., Brubaker, C., Chee, P. (2007). Cotton. In: Kole, C. (eds) Technical Crops. Genome Mapping and Molecular Breeding in Plants, vol 6. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-34538-1_1

USDA, Agricultural Research Service, National Plant Germplasm System. 2023. Germplasm Resources Information Network (GRIN Taxonomy). National Germplasm Resources Laboratory, Beltsville, Maryland. https://www.ars-grin.gov/. Accessed 17 October 2023.

USDA-ERS. 2022. Cotton Sector at a Glance. https://www.ers.usda.gov/topics/crops/cotton-and-wool/cotton-sector-at-a-glance/#. Accessed July 2023.

USDA-ERS. 2023. Oil Crops Sector at a Glance. https://www.ers.usda.gov/topics/crops/soybeans-and-oil-crops/oil-crops-sector-at-a-glance/. Accessed October 2023.

Van Ooijen JW (2018) JoinMap 5, software for the calculation of genetic linkage maps in experimental populations of diploid species. Kyazma. https://www.kyazma.nl/index.php/JoinMap/. Accessed September 2023.

Van Os, H., Stam, P., Visser, R. G., & Van Eck, H. J. (2005). RECORD: a novel method for ordering loci on a genetic linkage map. *Theoretical and Applied Genetics*, *112*, 30-40. https://doi.org/10.1007/s00122-005-0097-x

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity*, *93*(1), 77-78. https://doi.org/10.1093/jhered/93.1.77

Wakelin, A. M., Lister, C. E., & Conner, A. J. (2003). Inheritance and biochemistry of pollen pigmentation in California poppy (*Eschscholzia californica* Cham.). *International journal of plant sciences*, *164*(6), 867-875. https://doi.org/10.1086/378825

Walter, M. H., Floss, D. S., & Strack, D. (2010). Apocarotenoids: hormones, mycorrhizal metabolites and aroma volatiles. *Planta*, *232*, 1-17. https://doi.org/10.1007/s00425-010-1156-3

Wang S., C. J. Basten, and Z.-B. Zeng (2012). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm. Accessed September 2023.

Wang, F., Zhang, J., Chen, Y., Zhang, C., Gong, J., Song, Z., Zhou, J., Wang, J., Zhao, C., Jiao, M., Liu, A., Du, Z., Yuan, Y., Fan, S., & Zhang, J. (2020). Identification of candidate genes for key fibre-related QTL s and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G. barbadense* introgressions. *Plant Biotechnology Journal*, *18*(3), 707-720. https://doi.org/10.1111/pbi.13237

Wang, X. Y., Quan, Q. M., Wang, B., Li, Y. X., & Huang, S. Q. (2018). Discovery of androecium color polymorphism in *Epimedium pubescens* with habitat preference of anther/pollen color in the genus. *Journal of plant ecology*, *11*(4), 533-541. https://doi.org/10.1093/jpe/rtx034

Wang, Z., Zhang, D., Wang, X., Tan, X., Guo, H., & Paterson, A. H. (2013). A whole-genome DNA marker map for cotton based on the D-genome sequence of *Gossypium raimondii* L. *G3: Genes, Genomes, Genetics*, *3*(10), 1759-1767. https://doi.org/10.1534/g3.113.006890

Welsch, R., Wust, F., Bar, C., Al-Babili, S., & Beyer, P. (2008). A third phytoene synthase is devoted to abiotic stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. *Plant Physiology*, *147*(1), 367-380. https://doi.org/10.1104/pp.108.117028

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, TL., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D.,

Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software*, *4*(43), 1686. http://doi.org/10.21105/joss.01686

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, *64*(1), 3-19. https://doi.org/10.1016/S0031-9422(03)00300-5

Wink, M. (2016). Evolution of secondary plant metabolism. *eLS*, 1-11. https://doi.org/10.1002/9780470015902.a0001922.pub3

Wolfe, L. M. (2001). Associations among multiple floral polymorphisms in *Linum pubescens* (Linaceae), a heterostylous plant. *International Journal of Plant Sciences*, *162*(2), 335-342. https://doi.org/10.1086/319578

Xian-Liang, S., Xue-Zhen, S., & Tian-Zhen, Z. (2006). Segregation distortion and its effect on genetic mapping in plants. *Chinese Journal of Agricultural Biotechnology*, *3*(3), 163-169. https://doi.org/10.1079/CJB2006110

Xiao, Y. H., Yan, Q., Ding, H., Luo, M., Hou, L., Zhang, M., Yao, D., Liu, H. S., Li, X., Zhao, J., & Pei, Y. (2014). Transcriptome and biochemical analyses revealed a detailed proanthocyanidin biosynthesis pathway in brown cotton fiber. *PLoS One*, *9*(1), e86344. https://doi.org/10.1371/journal.pone.0086344

Yan, Q., Wang, Y., Li, Q., Zhang, Z., Ding, H., Zhang, Y., Liu, H., Lup, M., Liu, D., Song, W., Liu, H., Yao, D., Ouyand, X., Li, Y., Li, X., Pei, Y. & Xiao, Y. (2018). Up-regulation of *GhTT2-3A* in cotton fibres during secondary wall thickening results in brown fibres with improved quality. *Plant Biotechnology Journal*, *16*(10), 1735-1747. https://doi.org/10.1111/pbi.12910

Yang, Z., Ge, X., Yang, Z., Qin, W., Sun, G., Wang, Z., Li, Z., Liu, J., Wu, J., Wang, Y., Lu, L., Wang, P., Mo, H., Zhang, X., & Li, F. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nature communications*, *10*(1), 2989. https://doi.org/10.1038/s41467-019-10820-x

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P., & Potrykus, I. (2000). Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, *287*(5451), 303-305. http://doi.org/10.1126/science.287.5451.303

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, proteomics & bioinformatics*, *19*(4), 619-628. https://doi.org/10.1016/j.gpb.2020.10.007

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, *38*(2), 203-208. https://doi.org/10.1038/ng1702

Zahid, K. R., Ali, F., Shah, F., Younas, M., Shah, T., Shahwar, D., Hassan, W., Ahrmad, Z., Qi, C., Lu, Y., Iqbal, A., & Wu, W. (2016). Response and tolerance mechanism of cotton *Gossypium hirsutum* L. to elevated temperature stress: a review. *Frontiers in plant science*, *7*, 937. https://doi.org/10.3389/fpls.2016.00937

Zeng, J., Yao, D., Luo, M., Ding, L., Wang, Y., Yan, X., Ye, S., Wang, C., Wu, Y., Zhang, J., Li, Y., Ran, L., Dai, Y., Chen, Y., Wand, F., Lai, H., Liu, N., Fang, N., Pei, Y., & Xiao, Y. (2023). Fiber-specific increase of carotenoid content promotes cotton fiber elongation by increasing abscisic acid and ethylene biosynthesis. *The Crop Journal*. https://doi.org/10.1016/j.cj.2023.02.010

Zhang, D., Zhang, T., & Guo, W. (2010). Effect of H2O2 on fiber initiation using fiber retardation initiation mutants in cotton (*Gossypium hirsutum*). *Journal of plant physiology*, *167*(5), 393-399. https://doi.org/10.1016/j.jplph.2009.10.005

Zhang, K., Kuraparthy, V., Fang, H., Zhu, L., Sood, S., & Jones, D. C. (2019). High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (*Gossypium hirsutum* L.). *BMC genomics*, *20*(1), 889. https://doi.org/10.1186/s12864-019-6214-z

Zhang, Q., Liu, M., & Ruan, J. (2017). Metabolomics analysis reveals the metabolic and functional roles of flavonoids in light-sensitive tea leaves. *BMC plant biology*, *17*(1), 1-10. https://doi.org/10.1186/s12870-017-1012-8

Zhang, S., Lan, Q., Gao, X., Yang, B., Cai, C., Zhang, T., & Zhou, B. (2016). Mapping of genes for flower-related traits and QTLs for flowering time in an interspecific population of *Gossypium hirsutum*× *G. darwinii*. *Journal of genetics*, *95*, 197-201. https://doi.org/10.1007/s12041-016-0617-2

Zhu, L., Andres, R. J., Zhang, K., & Kuraparthy, V. (2021). High-density linkage map construction and QTL analysis of fiber quality and lint percentage in tetraploid cotton. *Crop Science, 61*(5), 3340-3360. https://doi.org/10.1002/csc2.20519

Zhu, T., Liang, C., Meng, Z., Sun, G., Meng, Z., Guo, S., & Zhang, R. (2017). CottonFGD: an integrated functional genomics database for cotton. *BMC plant biology, 17*(1), 1-9. https://doi.org/10.1186/s12870-017-1039-x

# CHAPTER 3: Development and validation of Kompetitive Allele-Specific PCR (KASP) markers for bacterial blight resistant locus *BB-13* in Upland cotton (*Gossypium hirsutum* L.)

As prepared for Publication in Crop Science

Nicole Pettit[1*], S. Anjan Gowda[1*], Navin Shrestha[1], Taylor Harris[2,3], Rebecca Bart[2], Fred Bourland[4], Gina Brown-Guedira[5], Don C. Jones[6], Vasu Kuraparthy[1§]

[1] Crop & Soil Sciences Department, North Carolina State University, Raleigh, NC 27695

[2] Danforth Plant Sciences Center, 975 N Warson Rd, St Louis, MO 63132

[3] Division of Biology & Biomedical Sciences, Washington University, St. Louis, MO 63110

[4] NE Research & Extension Center, Crop, Soil, and Environmental Sciences, University of Arkansas, Keiser, AR 72351

[5] U.S. Department of Agriculture's Agricultural Research Service, Plant Science Research, Raleigh, NC 27695

[6] Cotton Incorporated, 6399 Weston Parkway, Cary, NC 27513, USA

[*]Equal contribution

[§] Corresponding author (vasu_kuraparthy@ncsu.edu)

**Abbreviations**:

AFLP, amplified fragment length polymorphism; ASP, allele-specific forward primer; CBB, cotton bacterial blight; CRP, common reverse primer; DNA, deoxyribonucleic acid; KASP, kompetitive allele-specific PCR; MAS, marker-assisted selection; NTC, no template control; PACE, PCR allele competitive extension; PCR, polymerase chain reaction; QTL, quantitative trait locus; RAPD, randomly amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; STS, sequence-tagged site; Xcm, Xanthomonas citri subsp. malvacearum

## Abstract

Cotton bacterial blight (CBB), caused by the pathogen *Xanthomonas citri* subsp. *malvacearum* (*Xcm*), can inflict significant damage to cotton (*Gossypium*

*hirsutum* L.) production. Previously, we identified and mapped the broad-spectrum CBB-resistant locus *BB-13* on the long arm of chromosome D02 using array-based single nucleotide polymorphisms (SNPs). In the current study, linked SNPs were converted into easily assayable Kompetitive Allele-Specific PCR (KASP) markers to enable efficient detection and marker-assisted selection of alleles at the *BB-13* locus. The KASP marker's efficiency in detecting the *BB-13* resistant gene was validated using an Upland cotton diversity panel of 72 accessions phenotyped with *Xcm* race 18. The KASP marker NCBB-KASP4, derived from the CottonSNP63K array-based marker i25755Gh that is closely associated with *BB-13*, predicted the CBB response phenotypes with an error rate of 4.17% in the diversity panel. Additionally, two independent biparental recombinant inbred line populations segregating for resistance to *Xcm* race 18 were used for KASP marker validation and to test their utility in detecting the presence of the *BB-13* locus in the resistant accession CABD3CABCH-1-89. NCBB-KASP4, validated across breeding populations and broad germplasm, is a reliable KASP marker for detection and testing of *BB-13* locus in cotton. Further, diagnostic array-based SNP marker i25755Gh's allele pattern and the potential CBB response is described for 875 *Gossypium* accessions. These SNP-based phenotypic predictions for 875 accessions along with disease response phenotypes to *Xcm* race 18 for 253 accessions provide a reference for CBB resistance in diverse cotton germplasm in the United States.

**Key words**:

Cotton, bacterial leaf blight, KASP marker, SNPs, resistance, marker-assisted selection

**Core Ideas:**

- Linked single nucleotide polymorphisms (SNPs) converted into easily assayable Kompetitive allele-specific PCR (KASP) markers to enable detection and marker-assisted selection of alleles at the *BB-13* locus.
- Validated NCBB-KASP4 is a reliable KASP marker for detection and testing of *BB-13* locus in cotton.
- *BB-13* diagnostic KASP markers' utility was validated using a diversity panel and segregating inbred populations.

- SNP-based phenotypic predictions along with cotton bacterial blight response phenotypes for diverse cotton accessions were described.

## 3.1 INTRODUCTION

Cotton (*Gossypium* spp.) is an economically and commercially important source of natural fiber. In addition to its use in the textile industry, cotton seeds are used as a food source for humans and animals through the production of cotton seed oil and meal. Grown worldwide, *Gossypium* comprises over 50 recognized species (Wendel & Grover, 2015), four of which are commercially grown. Of these, *Gossypium hirsutum* L., or Upland cotton, is one of the most widely cultivated due to its productivity and adaptability across a wide range of environmental conditions. *Gossypium barbadense* L., or Pima cotton, has fibers that are longer and finer than other types of cotton, making it a valuable economic species. *G. hirsutum* and *G. barbadense* are allotetraploids ($2n = 4x = 52$) with A and D genomes derived from interspecific hybridization between two diploid species (Viot & Wendel, 2023; Wendel & Cronn, 2003).

Cotton bacterial blight (CBB) is a disease caused by the bacterial pathogen *Xanthomonas citri* subsp. *malvacearum* (*Xcm*). The disease was first reported in the United States in 1891 (Atkinson, 1892) and has historically been one of the most damaging diseases in cotton. There are over 22 races of *Xcm* identified around the world (Delannoy et al., 2005; Zhang et al., 2020). Different races are more common in different regions with race 18 being the most prevalent and virulent race found in the United States (El-Zik & Thaxton, 1994; Zhang et al., 2020). *Xcm* can infect cotton plants at any growth stage through openings on the plant such as stomata, lenticels, hydathodes, and wounds (Kemerait et al., 2017; Wallace & El-Zik, 1989). The infected areas first display symptoms of localized water-soaked lesions usually on the abaxial surface of leaves, on the stems, bracts, and bolls (Innes, 1983; Kemerait et al., 2017; Verma, 1986). The lesions darken as they grow and can appear angular accompanied by a colored halo in the surrounding tissue (Innes, 1983; Kemerait et al., 2017; Ridgway et al., 1984; Verma, 1986). As the disease progresses, the leaves can fray and defoliate prematurely and systemic infection can occur (Innes, 1983; Kemerait et al., 2017; Phillips et al., 2017; Ridgway et al., 1984; Verma, 1986). Infection in the bolls can damage the quality, yield, and color of the fiber and cause

contamination of the seeds (El-Zik & Thaxton, 1994; Kemerait et al., 2017; Phillips et al., 2017). The pathogen can overwinter on these seeds, residual lint, and crop residue from previous years (Yang, 2013). The disease is managed through practices such as the removal of the residual lint from the seeds through a process called acid delinting, treating seeds with a fungicide, sanitary practices including the clearing of crop residues after the growing season, crop rotation, and the use of resistant cultivars (Xiao et al., 2009; Zhang et al., 2020). Acid delinting and the cultivation of resistant cultivars effectively controlled CBB for over 50 years but in 2011 CBB reemerged in the United States in Mississippi, Missouri, and Arkansas (Allen, 2011). The disease quickly spread through the southern United States cotton belt due to large acreages of susceptible cultivars that were planted (Phillips et al., 2017). It is thought that the prevalence of resistant cultivars kept *Xcm* at minimal levels and after 50 years when susceptible cultivars became prevalent again, infection spread. To date, there are no effective chemical control measures available once the crop has been infected. The most effective strategy for limiting the disease severity and its damage in cotton is through the development and deployment of resistant cultivars.

The development of resistant cultivars began in the 1930s and continued through the 1980s. So far, over 20 major *B* genes and multiple quantitative trait loci (QTLs) regions across the genome have been identified for CBB resistance in different cotton varieties (Bird, 1982; Bird & Blank, 1951; Bourland, 2018; Green & Brinkerhof, 1956; Knight, 1963; Knight & Clouston, 1939; Zhang et al., 2020). Some resistance genes are specific to certain races of *Xcm* and may work in tandem for partial resistance (Elassbli, Abdelraheem et al., 2021; Wright et al., 1998). A number of CBB resistance genes (*B12* and *BB-13*) and QTLs have been reported on chromosome D02 (Elassbli et al., 2021; Gowda et al., 2022; Rungis et al., 2002; Wright et al., 1998; Xiao et al., 2009; Yang, 2013).

CBB resistance genes have been mapped using molecular markers (Zhang et al., 2020). However, DNA-based diagnostic markers for CBB resistance which can be easily assayable across diverse germplasm are not currently available in cotton. Kompetitive allele-specific PCR markers (KASP), are high-throughput molecular markers that utilize fluorescently labeled single nucleotide polymorphisms (SNPs) to determine the marker allele present at a specific locus in the genome (Semagn et al., 2014). KASP markers are proven to be a quick and cost-effective

method to analyze marker alleles on a large number of samples making them a useful tool for marker-assisted selection (MAS) in breeding (Makhoul et al., 2020; Rasheed et al., 2016; Semagn et al., 2014).

The *BB-13* locus, mapped to a 371 kb region on chromosome D02, showed broad-spectrum resistance to *Xcm* races, including race 18 (Gowda et al., 2022). We previously identified closely linked SNP markers (i25755Gh and i46775Gh) for *BB-13* (Gowda et al., 2022). However, the identified markers can be used only through CottonSNP63K array-based genotyping which is cost prohibitive for most breeding programs. A closely associated KASP marker that can diagnostically detect alleles of *BB-13* across diverse germplasm will enable quick and cost-effective detection of the *BB-13* locus for MAS in cotton breeding. This study aims to design an assemblage of high-throughput KASP markers to detect the presence of *BB-13* using the SNPs previously identified to be in close proximity to the locus.

## 3.2 METHODS AND MATERIALS

### 3.2.1 Plant material

Seventy-two lines from the elite diversity panel of *G. hirsutum* assembled by Tyagi et al. (2014) and phenotyped for CBB response to race 18 by Gowda et al. (2022) were used to validate KASP markers for CBB resistance. The panel represents a major portion of genetic diversity present in US Upland cotton (Table S1). One hundred and four recombinant inbred lines (RILs) derived from Acala Maxxa (PI 540885; DIV 017) ✗ Arkot 8102 (PI 595852; DIV042) (Bourland et al., 1997) that were previously phenotyped for CBB response to race 18 were used to validate the KASP markers (Gowda et al., 2022). The KASP markers most closely linked and diagnostic for the *BB-13* locus were further used to validate its applicability and to test for the presence and segregation of the *BB-13* locus in an independent RIL population. This RIL population was developed by crossing Acala Maxxa (PI 540885; DIV 017) as the susceptible female parent to the resistant male parent CABD3CABCH-1-89 (PI 603002; DIV063). The $F_1$ plants of each cross combination were manually selfed to obtain $F_2$ seeds. Individual plants from $F_2$ and subsequent generations were advanced to $F_6$ generation using the single-seed-descent method by manually selfing the flowers of the individual plants using glassine bags. Forty-five RILs and

their respective parental accessions of this RIL population were used for KASP marker genotyping.

Following marker validation, predictions for potential CBB resistance in 875 accessions of *Gossypium* spp. were made in order to provide an extensive resource of accessions estimated to be resistant and susceptible based on the linked marker to the *BB-13* locus. These 875 accessions include 387 *G. hirsutum* and 488 *G. barbadense* accessions. The 387 Upland accessions are sourced from an elite diversity panel (Tyagi et al., 2014) of 125 accessions that were not phenotyped, a tropical landrace panel with 181 accessions (Kaur et al., 2017), and 81 accessions from Dr. Fred Bourland's cotton breeding program (FB panel) of the University of Arkansas (Bourland & Jones, 2006, 2017; Bourland et al., 1997). The 488 Pima accessions are from a Pima photoperiod insensitive diversity panel of 357 accessions, a Pima photoperiod sensitive diversity panel of 128 accessions, and three *G. barbadense* lines from the elite diversity panel of Tyagi et al. (2014). Three accessions from the elite diversity panel did not have genotype information for the SNP marker and were not included in the prediction table. The SNP marker from the CottonSNP63K array associated with the best-performing KASP marker was used as a surrogate to predict the CBB response phenotypes of the 875 accessions. This SNP marker was also validated in 253 previously phenotyped accessions in the diversity panel.

### 3.2.2 Phenotyping of the diversity panel and RIL populations

Forty-five RILs along with two parental accessions were phenotyped for CBB response to race 18 of *Xcm* at the seedling stage in growth chambers as described by Gowda et al. (2022). The segregation ratios of phenotypes of the RIL population were calculated using a chi-square analysis under the assumption that a single dominant gene controls the CBB resistance.

### 3.2.3 DNA isolation

Leaf tissue from 3- to 4-week-old accessions of the diversity panel, the RIL populations, and their parental lines were collected in 2 mL centrifuge tubes. The samples were immediately frozen in liquid nitrogen and stored in a −80°C freezer. The samples were homogenized using a YG5 3422011 macerator (BTLab systems). The DNA was isolated from the ground leaf tissues using a DNeasy Plant Mini-Kit (Qiagen USA). The isolated DNA was evaluated for the concentration and

quality of DNA (260/230 > 1.5 and 260/280 = 1.8 to 2.0) using a Nanodrop spectrophotometer ND1000 (Thermo Fisher Scientific) or by picogreen assay.

### 3.2.4 Primer design

Ten SNPs closely linked with the CBB-resistant locus *BB-13* identified through combined genome-wide association mapping and linkage analysis from our previous study (Gowda et al., 2022) were selected for KASP marker development. These 10 markers, i04886Gh, i04940Gh, i15296Gh, i25755Gh, i46775Gh, i04879Gh, i04892Gh, i04941Gh, i04907Gh, and i04910Gh (henceforth referred to as NCBB-KASP1, NCBB-KASP2, NCBB-KASP3, NCBB-KASP4, NCBB-KASP5, NCBB-KASP6, NCBB-KASP7, NCBB-KASP8, NCBB-KASP9, and NCBB-KASP10, respectively), are from the CottonSNP63K array (Hulse-Kemp et al., 2015). The probe sequences of the 10 markers were obtained from the CottonGen database (Yu et al., 2021). The probe sequences were utilized for designing two allele-specific forward primers (ASP1 and ASP2) and a common reverse primer (CRP). The primers were designed using the PCR allele competitive extension (PACE) assay design service (3Cr Bioscience). The ASPs were tagged with HEX (GAAGGTCGGAGTCAACGGATT) and FAM (GAAGGTGACCAAGTTCATGCT) fluorophore tail sequences from the PACE assay designing service and primers were synthesized from Integrated DNA Technologies. The sequence information of all 10 primers is documented in Table S2.

### 3.2.5 KASP/PACE genotyping assay

KASP/PACE assays were performed with a final reaction volume containing 2.0 μL of 2xPACE mix, 0.5 μL of primer mix, and 2.5 μL of 15–25 ng DNA. The primer mix was composed of 46 μL of molecular grade water (Fisher Bioreagents), 12 μL each of ASP1 and ASP2, and 30 μL of the CRP (3CR Bioscience, 2022; Makhoul et al., 2020). Polymerase chain reaction (PCR) of the DNA samples of diversity panel accessions was performed on a 384-well PCR machine (Eppendorf) with fluorescent endpoints quantified on a PHERAstar Plus Plate Reader version 1.43 (BMG Labtech). RIL populations were amplified and analyzed using an Applied Biosystems Real-Time PCR QuantStudio 5 machine (Applied Biosystems). PCR cycling conditions on a 384-well PCR machine were performed over three steps following PACE guidelines (3CR Bioscience, 2022; Makhoul et al., 2020) consisting of (a) enzyme activation performed at 94°C for 15 min, (b) 10 touchdown cycles of denaturing and annealing/extension, and (c) final amplification with 24 cycles

of denaturation and annealing/extension. The 10 touchdown cycles were performed with a denaturation temperature of 94°C for 20 s and an annealing/extension temperature of 65°C for 1 min (decreasing 0.8°C per cycle). The 24 cycles of final amplification were performed at a denaturation temperature of 94°C for 20 s and an annealing/extension temperature of 57°C for 1 min. After the completion of PCR and quantifying reads, KlusterCaller software version 3.4.1.39 (LGC Genomics) was used to analyze the fluorescence reads and assign allelic calls. If fluorescent calls were not appropriately segregated due to lack of amplification, three additional PCR cycles were run and endpoints were quantified. Additional cycles were repeated until allelic calls were separated. The three additional PCR cycles include a denaturation step at 94°C for 20 s and an annealing/extension step at 57°C for 1 min. Cartesian plots displaying quantified fluorescence were made using the North Carolina State University Peanut Breeding and Genetics SNP caller (Andres & Dunne, 2022). The QuantStudio 5 Real time-PCR machine used similar cycling conditions with the exception of running 50 instead of 24 final amplification cycles. The fluorescent endpoints were read by the QuantStudio 5 machine and allelic calls were made using QuantStudio Design and Analysis version 1.5.2 software. The allelic calls for each sample cluster together in groups on a Cartesian plane based on the fluorescent endpoints that were detected. Alleles were scored based on the fluorescent endpoints as homozygous allele one or homozygous allele two, heterozygous calls as three, and zero for unknown calls in KlusterCaller as well as QuantStudio Design and Analysis version 1.5.2 software. Homozygous alleles were designated as resistant or susceptible in relation to the alleles of the accessions with known phenotypes.

### 3.2.6 Data analysis

In order to determine the ability of the marker to correctly distinguish the CBB response phenotype, the error rate of the markers was calculated (error rate = number of marker calls that did not co-segregated with the phenotype ÷ total number of marker calls excluding unknown and heterozygous alleles). The error rates were calculated for the KASP markers for 72 phenotyped diversity panel accessions, the RIL populations, and the 253 diversity panel accessions using SNP marker i25755Gh.

The i25755Gh, closely linked to *BB-13* locus, is the most reliable marker for CBB resistance validated through the KASP marker system in the current study. The i25755Gh SNP marker data

from CottonSNP63k genotyping array (unpublished) of 387 Upland and 488 Pima accessions were utilized as a surrogate for predicting the CBB response phenotypes among these 875 accessions.

## 3.3 RESULTS

### 3.3.1 Phenotyping of the diversity panel and RIL populations

A total of 253 diversity panel accessions and one of the RIL population (Acala Maxxa × Arkot 8102) were previously phenotyped (Figure 1) (Gowda et al., 2022). Out of the 72 diversity panel accessions that were used for KASP marker validation in the current study, 26 accessions were resistant (36.11%) and 46 were susceptible (63.89%) to *Xcm* race 18. In the biparental RIL population of cross Acala Maxxa × Arkot 8102, phenotypes segregated in a 1:1 ratio of resistant to susceptible (Gowda et al., 2022). Phenotypic evaluation of the other RIL population (Acala Maxxa × CABD3CABCH-1-89) in the present study also showed a 1:1 segregation ratio of resistant and susceptible phenotypes (Table S3). This suggests the CBB resistance to *Xcm* race 18 in accessions Arkot 8102 (DIV042) and CABD3CABCH-1-89 (DIV063) are controlled by major gene(s).

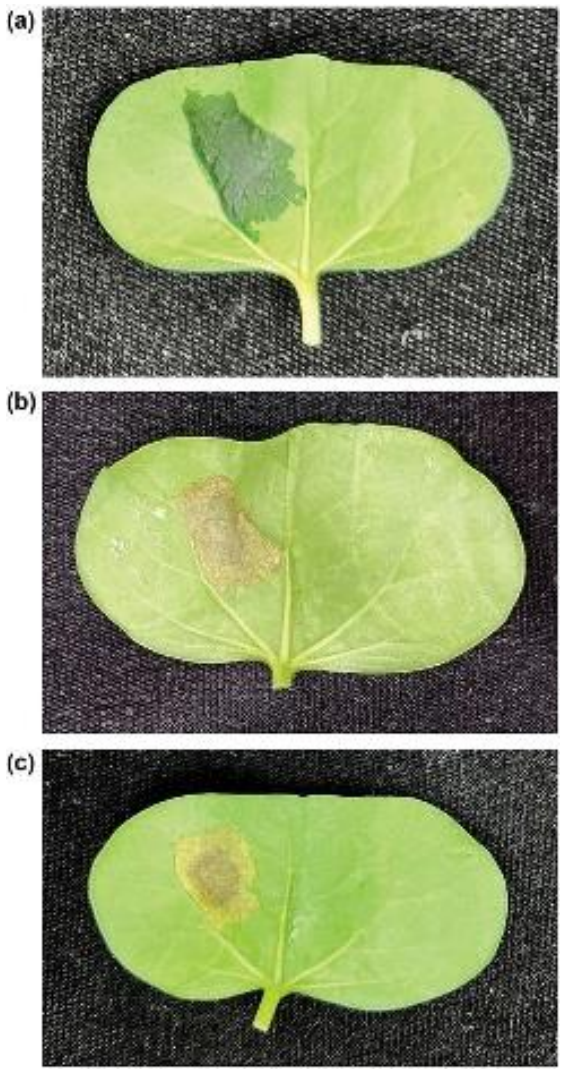



**FIGURE 1.** Phenotypic response of Upland cotton accessions to race 18 of *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) at seedling stage. (a) A susceptible response of water soaked lesions in Acala Maxxa. (b) A hypersensitive response of disease resistance in Arkot 8102. (c) A hypersensitive response of disease resistance in CABD3CABCH-1-89.

### 3.3.2 Validation of KASP markers in the diversity panel

Eight KASP markers (NCBB-KASP1 through NCBB-KASP8) were developed from array-based SNP markers linked to *BB-13* as identified in the previous genome-wide association studies (GWAS) analysis (Gowda et al., 2022). Testing these KASP markers on the 72 diversity panel accessions with known CBB response phenotypes to *Xcm* race 18 showed that all eight KASP

markers were polymorphic and heterozygous alleles were detectable for each marker (Figure 2a–c). The error rate for each marker was calculated (Table 1). NCBB-KASP4 had the lowest error rate of 4.17%. The markers NCBB-KASP1, NCBB-KASP3, NCBB-KASP5, and NCBB-KASP6 had few errors, with an error rate of each marker between 4.2% and 4.4%. One marker, NCBB-KASP7 performed poorly with 40 incorrect predictions and an error rate of 55.56%. With only three incorrect predictions out of 72 accessions with known CBB response phenotypes, NCBB-KASP4 had the fewest errors.

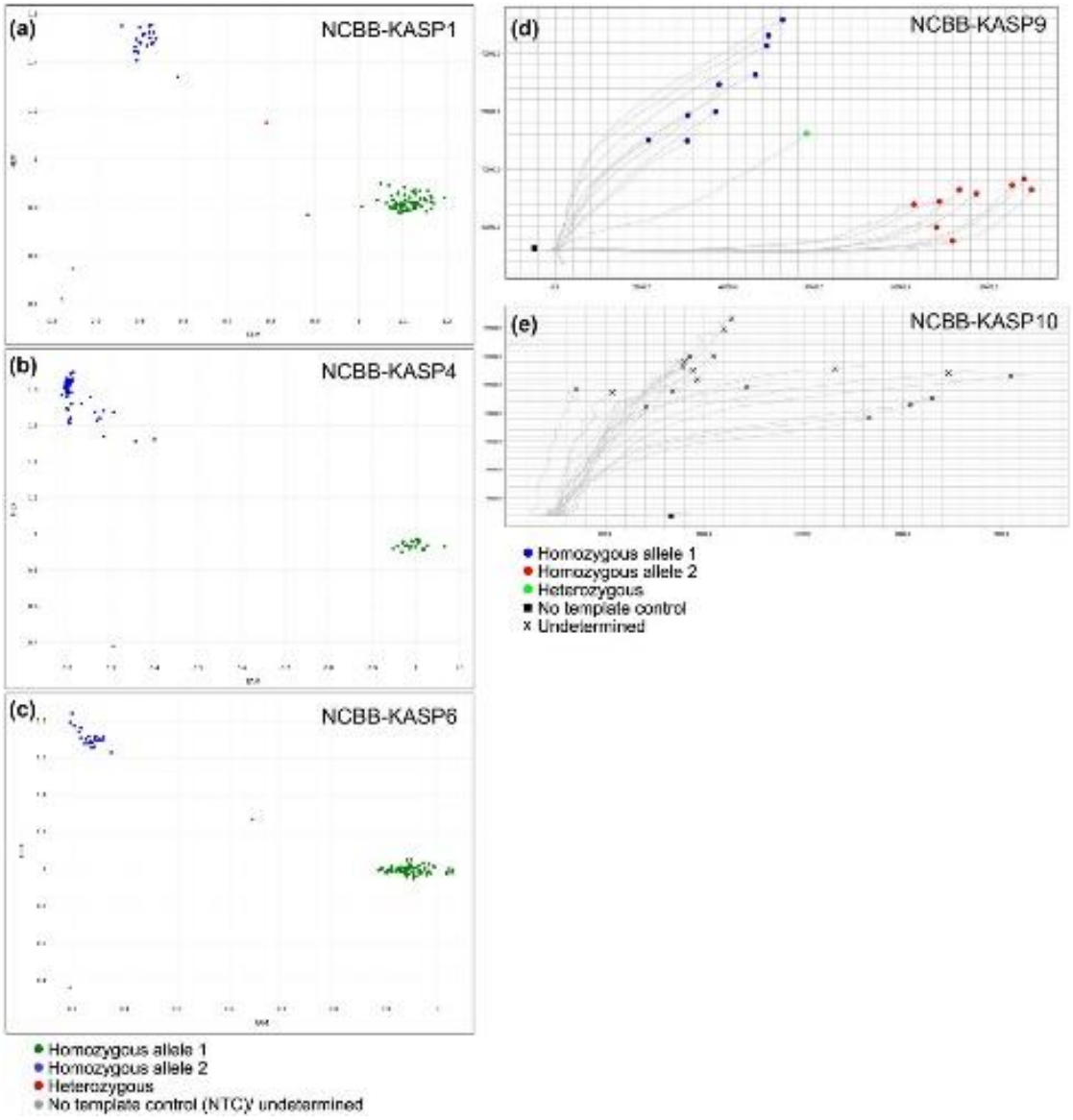


**FIGURE 2.** Allele discrimination Cartesian plots showing FAM and HEX fluorescence determining resistant and susceptible marker allele pattern in *Gossypium hirsutum* lines. (a–c) Kompetitive allele-specific PCR (KASP) marker allele calls in 72 elite diversity panel accessions quantified on a PHERAstar Plus Plate Reader. Fluorescent endpoints were read and allele calls assigned in KlusterCaller software and displayed using the North Carolina State University Peanut Breeding and Genetics single nucleotide polymorphism (SNP) caller (Andres & Dunne, 2022) (a, NCBB-KASP1; b, NCBB-KASP4; and c, NCBB-KASP6). (d–e) KASP marker allele calls in a recombinant inbred line (RIL) population from a cross of Acala Maxxa × Arkot 8102. Fluorescent endpoints were quantified on a QuantStudio 5 Real time-PCR machine. Allele calls were assigned using QuantStudio Design and Analysis software (d, NCBB-KASP9 and e, NCBB-KASP10).

**TABLE 1.** Allele pattern of eight Kompetitive allele-specific PCR (KASP) markers for cotton bacterial blight (CBB) resistance locus *BB-13* validated in 72 *Gossypium hirsutum* accessions of the elite diversity (DIV) panel of Tyagi et al. (2014).

| | | Allele pattern of the KASP marker | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Accession name | DIV number | NCBB-KASP1 | NCBB-KASP2 | NCBB-KASP3 | NCBB-KASP4 | NCBB-KASP5 | NCBB-KASP6 | NCBB-KASP7 | NCBB-KASP8 | Phenotypic response to Race-18 |
| Acala Maxxa | DIV017 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |
| Arkot 8102 | DIV042 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| BJA Glandless Nectariless | DIV046 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | Resistant |
| ACALA SJ 1 | DIV021 | 0 | 1 | 0 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| CAHUGARPIH 1 88 | DIV065 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | Resistant |
| DELFOS 9169 (ORIGINAL) | DIV104 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTAPINE SMOOTH LEAF | DIV118 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| EARLY FLOFF | DIV135 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| FTA | DIV152 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| PRONTO | DIV280 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| ACALA #111 ROGERS | DIV002 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| AK DJURA VIRESCENT | DIV033 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |
| DELTAPINE SR 383 | DIV119 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| EMPIRE GL2 GL2 | DIV137 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| GA 161 | DIV153 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | Susceptible |
| Pyramid | DIV283 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |

**Table 1** (continued).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ACALA 1064(New Mexico) | DIV003 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| ALLEN 33 | DIV034 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| CASCOT C 13 | DIV068 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| DELTATYPE WEBBER | DIV120 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| EMPIRE WR | DIV138 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| GEORGIA KING | DIV154 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| ROWDEN | DIV290 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| ACALA 1517 WILT | DIV004 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |
| ARIZONA RED | DIV037 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | Susceptible |
| CASCOT L 7 | DIV069 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| DELTAPINE 20 | DIV109 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| EWINGS LONG STAPLE | DIV139 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| GERMAIN'S ACALA GC 352 | DIV155 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |
| ROWDEN 41B TPSA | DIV292 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| CD3HCHULBH 1 88 | DIV072 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| DELTAPINE 41 | DIV110 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTATYPE WEBBER (253 1) T142 8 | DIV122 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| EWINGS LONG STAPLE x TIDEWATER | DIV140 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| GP 3755 | DIV160 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| TAM 86E 8 | DIV338 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 0 | Resistant |
| ACALA 1517D | DIV007 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |

**Table 1** (continued).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ARKANSAS 10 | DIV040 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| CD3HHARCIH 1 88 | DIV073 | 3 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | Resistant |
| DELTAPINE 45 | DIV111 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| H1220 | DIV176 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| ACALA 4 42 | DIV010 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTAPINE 50 | DIV112 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| LAMBRIGHT 2020A | DIV196 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| TAMCOT CAB CS | DIV342 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| COKER 201 | DIV079 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTAPINE 51 | DIV113 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| Fibermax 966 | DIV146 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| LANKART 511 | DIV200 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| TAMCOT CD3H | DIV344 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| ACALA 5 | DIV012 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| Bronze | DIV054 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Resistant |
| DEL CERRO | DIV100 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTAPINE 61 | DIV114 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DUNN 1047 | DIV130 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Resistant |
| LANKART 57 | DIV201 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| TAMCOT SP 37 | DIV352 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| DELCOT 277 | DIV101 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTAPINE 90 | DIV115 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |

**Table 1** (continued).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| DUNN 325 | DIV131 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| PD 4381 | DIV259 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| TAMCOT SP 37H | DIV353 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| ACALA 8 | DIV015 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |
| CA23 | DIV059 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Resistant |
| DELCOT 344 | DIV102 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | Resistant |
| DELTAPINE A | DIV116 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| PD 4461Q | DIV260 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| WILDS 34 4(411) T85 2 | DIV382 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| CABD3SHP3S 1 90 | DIV064 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | Resistant |
| DELFOS 6102 | DIV103 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| DELTAPINE PREMA | DIV117 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Susceptible |
| DURANGO | DIV133 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | Susceptible |
| Error rate | | 4.29% | 8.33% | 4.41% | 4.17% | 4.23% | 4.23% | 55.56% | 8.57% | |

### 3.3.3 Evaluation of KASP markers in the RIL populations

After evaluating the efficacy of the KASP markers in detecting the *BB-13* locus in the diversity panel, we then tested the KASP markers' utility in detecting the presence of the *BB-13* locus in the RIL population previously used to map the *BB-13* locus's linkage (Gowda et al., 2022). Three KASP markers (NCBB-KASP4, NCBB-KASP6, and NCBB-KASP7) were chosen based on the validation in the diversity panel along with two additional KASP markers (NCBB-KASP9 and NCBB-KASP10) from previous linkage mapping of the *BB-13* locus (Gowda et al., 2022). The alleles of NCBB-KASP10 did not segregate into two distinct clusters on Cartesian plots (Figure 2e), suggesting the marker is monomorphic. NCBB-KASP4, as with the other polymorphic KASP markers, was able to detect heterozygotic alleles (Figure 2d–e). Using all 104 tested lines for each marker and the two parental lines, error rates were calculated (Table S4). The SNPs associated with NCBB-KASP4 and NCBB-KASP9 are flanking markers for the *BB-13* locus with a linkage distance of 0.49 and 0.46 cM, respectively (Gowda et al., 2022), suggesting that their error rate should be the lowest due to a smaller frequency of recombination between the loci. NCBB-KASP4 and NCBB-KASP9 had the lowest error rate of 0.95% and 0.96%, respectively. NCBB-KASP7 had an error rate of 1.03% and NCBB-KASP6 had an error rate of 1.96% with two errors.

The four polymorphic KASP markers were run on 45 lines of an additional RIL population, Acala Maxxa × CABD3CABCH-1-89, and both parental accessions. Error rates for each marker were calculated with NCBB-KASP4 and NCBB-KASP9 having the lowest error rates of 2.17% (Table S4). The error rates were similar to the Acala Maxxa × Arkot 8102 population suggesting that the CBB resistance in CABD3CABCH-1-89 is likely due to the *BB-13* locus.

Out of the 10 KASP markers evaluated, NCBB-KASP1, NCBB-KASP3, NCBB-KASP4, NCBB-KASP5, NCBB-KASP6, and NCBB-KASP9 performed well based on low error rates in both the diversity panel and RIL populations, with NCBB-KASP4 performing the best (Table 2). NCBB-KASP7 had a very high error rate in the diversity panel and NCBB-KASP10 was monomorphic, suggesting both are poor markers for CBB resistance.

**TABLE 2.** Primer sequence information of the best performing Kompetitive allele-specific PCR (KASP) markers for cotton bacterial blight (CBB) resistance locus *BB-13* in Upland cotton.

| Marker | Associated SNP | KASP primer names | Primer sequence (5'-3') |
|---|---|---|---|
| NCBB-KASP1 | i04886Gh | P_i04886Gh ASP 1 | GAAGGTGACCAAGTTCATGCTATTTCAACAAATTTGATCGTGTACGGG |
| | | P_i04886Gh ASP 2 | GAAGGTCGGAGTCAACGGATTTCAACAAATTTGATCGTGTACGGA |
| | | P_i04886Gh CRP | ATACTGAAGCATTCCAGCATAACGGATTT |
| NCBB-KASP3 | i15296Gh | P_i15296Gh ASP 1 | GAAGGTGACCAAGTTCATGCTAACCAAGCACCTTCCTCTCTTAAG |
| | | P_i15296Gh ASP 2 | GAAGGTCGGAGTCAACGGATTAATAACCAAGCACCTTCCTCTCTTAAA |
| | | P_i15296Gh CRP | GACAATGACATCAGTCTATCACAGTACAA |
| NCBB-KASP4 | i25755Gh | P_i25755Gh ASP 1 | GAAGGTGACCAAGTTCATGCTCCTCATTGATTTCTCCATCCCGC |
| | | P_i25755Gh ASP 2 | GAAGGTCGGAGTCAACGGATTACCTCATTGATTTCTCCATCCCGT |
| | | P_i25755Gh CRP | GTACGAAGAACTGAATTCTGAAACCTGTA |
| NCBB-KASP5 | i46775Gh | P_i46775Gh ASP 1 | GAAGGTGACCAAGTTCATGCTATAAACCAGGCACTTGGTACAACG |
| | | P_i46775Gh ASP 2 | GAAGGTCGGAGTCAACGGATTATAAACCAGGCACTTGGTACAACA |
| | | P_i46775Gh CRP | ACATTCTTCCATCCACTGAC |
| NCBB-KASP6 | i04879Gh | P_i04879Gh ASP 1 | GAAGGTGACCAAGTTCATGCTGGCTCTGCAACTCTTTCGCACAT |
| | | P_i04879Gh ASP 2 | GAAGGTCGGAGTCAACGGATTGCTCTGCAACTCTTTCGCACAC |
| | | P_i04879Gh CRP | GGAGTCCAGCTAGTGTCCAGATATA |
| NCBB-KASP9 | i04907Gh | P_i04907Gh ASP 1 | GAAGGTGACCAAGTTCATGCTAATTGATACTTCGATCATCTCCTTCTTTAT |
| | | P_i04907Gh ASP 2 | GAAGGTCGGAGTCAACGGATTGATACTTCGATCATCTCCTTCTTTAC |
| | | P_i04907Gh CRP | GACGCCGTTTAGAGAATGGTATTATCAAA |

### 3.3.4 Phenotypic predictions based on allelic results

The CottonSNP63K array-based SNP marker i25755Gh that is tightly linked (0.49 cM) to the *BB-13* locus (Gowda et al., 2022) that was utilized to develop NCBB-KASP4 (error rate of 4.17% in the diversity panel), was used as a diagnostic SNP for predicting CBB resistance in a larger diversity panel of *Gossypium* germplasm collection. To evaluate the use of i25755Gh as a diagnostic tool, 253 of the diversity panel accessions were previously phenotyped for CBB resistance to *Xcm* race 18 (Gowda et al., 2022) and were compared to the alleles of i25755Gh (Table S5). With only nine errors, the SNP i25755Gh had an error rate of 3.56% showing it is effective as an SNP marker for CBB resistance. Hence, by utilizing the CottonSNP63K array

genotypic data of 875 accessions, we can predict their disease reaction solely based on i25755Gh SNP's allele patterns with a 96.44% confidence level (Table S6).

Out of 181 accessions of the tropical landrace panel, only nine accessions (4.97%) showed the resistant marker allele (CC) while 172 (95.03%) accessions carry the susceptible marker allele (TT) for SNP marker i25755Gh. Among 81 accessions of the FB panel, 72 (88.89%) accessions showed the resistant marker allele and only nine (11.11%) had the susceptible allele for i25755Gh. Of the 125 *G. hirsutum* accessions from the diversity panel that were not phenotyped, six (4.80%) had the resistant allele and 119 (95.20%) had the susceptible allele. Among 488 Pima accessions genotyped, only one (0.20%) accession (BAR 14/25-2 (PI_694252)) had the resistant allele while the remaining had the susceptible allele for SNP marker i25755Gh. In total, out of 875 accessions of *Gossypium* spp. analyzed for SNP marker i25755Gh, 88 (10.06%) are predicted to be resistant to *Xcm* race 18 due to the *BB-13* locus and 787 (89.94%) are predicted to be susceptible.

## 3.4 DISCUSSION

Molecular markers enable researchers to detect genes and predict traits without the need for phenotyping or complete genome sequencing and help improve efficiency in breeding. In the current study, we report the development and validation of KASP and SNP markers to diagnostically detect CBB-resistant locus *BB-13* in cotton and describe the marker-based prediction of CBB resistance in a diversity panel of *Gossypium* accessions. Restriction fragment length polymorphism, randomly amplified polymorphic DNA, simple sequence repeat (SSR), sequence-tagged sites (STS), amplified fragment length polymorphic markers, and SNPs have aided in the identification and mapping of over 20 CBB resistance genes including *B2*, *B3*, *B6*, *B12* and *BB-13* in the cotton genome (Amudha et al., 2003; Gowda et al., 2022; Rungis et al., 2002; Silva et al., 2014; Wright et al., 1998; Xiao et al., 2009; Yang, 2013; Zhang et al., 2020). Broad-spectrum CBB-resistant gene *B12* was first identified by Wallace and El-Zik (1989) in *G. hirsutum* L. germplasm S295. Molecular mapping localized the *B12* gene to the D02 chromosome (Rungis et al., 2002). Xiao et al. (2009) and Wright et al., 1998 further delineated the *B12* location on D02 and identified four SSR and four SNP markers associated with the gene. SSR markers CIR246, BNL3545, and BNL3644 were shown to be closely linked to the *B12* gene. Of those, CIR246 was the only one specific to the D subgenome. Faustine et al. (2015) further tested CIR246 and the SNP NG0207155 identified by

Xiao et al. (2009). Elassbli, Zhu et al. (2021) used two STS markers (NMBBD02-1 and NMBB-D02-2) to test for resistance gene *B12*. Both STS markers detected high heterozygosity of 21%–27.5% in a small population of released homozygous germplasm lines. These markers' (NMBBD02-1 and NMBB-D02-2) sequences and genomic sequence coordinates are not publicly available and their broad applicability across diverse germplasm has not been tested. Despite the heavy usage of markers for CBB resistance research, there is a lack of tested diagnostic markers for the CBB resistance genes in cotton. *BB-13* is a broad-spectrum CBB resistant locus recently identified in Upland cotton (Gowda et al., 2022). The KASP markers (NCBB-KASP1, NCBB-KASP3, NCBB-KASP4, NCBB-KASP5, NCBB-KASP6, and NCBB-KASP9) and array-based SNP marker (i25755Gh) developed and tested in the current study for *BB-13* locus are the first publicly available diagnostic markers for major CBB resistance gene in cotton.

Ideally, a perfect marker based on the causal gene at the *BB-13* locus would have no errors and would be the ideal for MAS. However, at present, the specific CBB resistance gene in the *BB-13* locus has not been identified. The *BB-13* locus on the long arm of the D02 chromosome encompassed 30 putative genes on the TM1 CRI v1 *G. hirsutum* reference genome (Yang et al., 2019), nine of which are reported to be involved in disease resistance (Gowda et al., 2022). Additionally, two NBS-LRR genes that are putatively involved in disease resistance in plants were found adjacent to the 371 Kb region of *BB-13* locus (Gowda et al., 2022). In the absence of the perfect marker, a KASP marker that is based on the closest linked SNP to *BB-13* would be useful to breeders for introgression of the *BB-13* locus into advanced breeding lines and enable its rapid deployment in adapted cultivars using MAS. As the SNPs used to develop KASP markers were not cosegregating with *BB-13* locus, error rates are expected for the KASP markers due to the recombinations between the marker and gene in both RIL populations and the diversity panel. This is evident from the KASP markers with a greater number of errors between assigned resistant and susceptible alleles and the true CBB phenotype showed a higher degree of recombination between it and the *BB-13* locus (Table 1 and Table S4). Nevertheless, the error rate of NCBB-KASP4 (4.17%) and four other KASP markers (4.2%–4.4%) were very low in the diversity panel, demonstrating the reliability of these markers. NCBB-KASP4 and NCBB-KASP9 are some of the closest linked flanking markers reported by Gowda et al. (2022) and are shown to overall have the lowest error rates in the two RIL populations suggesting that the RIL

populations segregate for the *BB-13* locus. The error rate of i25755Gh SNP was calculated using 253 diversity panel lines that have been manually phenotyped (Gowda et al., 2022). These include the 72 accessions used in validating NCBB-KASP4 in the diversity panel. Low error rates for not only the KASP marker (4.17%) but its parental SNP (3.56%) itself suggests that this marker locus is a useful tool in *BB-13* diagnostics. Of the nine errors in the SNP i25755Gh evaluation, eight resistant lines were incorrectly identified as susceptible, suggesting that the response of those accessions may be controlled by a resistance gene other than *BB-13* or are evidence of recombination events between i25755Gh and *BB-13*.

Our study showed that NCBB-KASP4 is a consistent and reliable high-throughput KASP marker for the *BB-13* CBB resistance gene in *G. hirsutum* as shown by evaluations in the diversity panel and biparental mapping populations. In the current study, 875 accessions were predicted for their potential CBB response phenotype (Table S6) and 253 accessions were previously phenotyped (Table S5) (Gowda et al., 2022). In total, we present CBB response phenotypes of 1128 accessions obtained either through manual phenotyping or prediction, which represent a major portion of germplasm in the US cotton breeding programs. The only *G. barbadense* line that is predicted to be resistant, BAR 14/25-2 (PI_694252), is a Pima accession developed by Knight (1963) to deliberately introduce CBB resistance into *G. barbadense* (Zhang et al., 2020). *G. barbadense* does not have much natural resistance to CBB and most plants are susceptible unless resistance was specifically introduced, further supporting the diagnostic potential of this marker in other *Gossypium* species (Jalloul et al., 2015; Zhang et al., 2020). The predicted CBB phenotypes can assist cotton breeders in choosing CBB-resistant parents for their crossing programs with 96.44% confidence by referencing these predictions. Tables S5 and S6 can serve as a reference to easily find resistant accessions. Overall, NCBB-KASP4 as a diagnostic marker for CBB resistance and the CBB response phenotypes of 1128 accessions in the study could be useful for breeders to access and rapidly deploy CBB resistance in cotton using MAS and germplasm resources.

**AUTHOR CONTRIBUTIONS: Nicole Pettit**: Data curation; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Satyanarayna Anjan Gowda**: Formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Navin Shrestha**: Resources. **Taylor Harris**: Investigation; methodology; resources. **Rebecca Bart**: Methodology;

resources; writing—review and editing. **Fred Bourland**: Resources; writing—review and editing. **Gina Brown-Guedira**: Resources. **Don C. Jones**: Resources; writing—review and editing. **Vasu Kuraparthy**: Conceptualization; funding acquisition; investigation; methodology; project administration; resources; supervision; visualization; writing—review and editing.

**ACKNOWLEDGMENTS:** The authors acknowledge the funding support provided by North Carolina Cotton Producers' Association and Cotton Incorporated (CI) for the current research and an assistantship to Nicole Pettit. We also thank Kelli Kochan of Texas A&M University for her help with genotyping the lines using the CottonSNP63K array. Thanks to Marisa Yoder and Alex Weil at the Danforth Plant Science Center for their help with bacterial infiltrations and cotton planting. We thank Jared Smith and Isabelle Gent of Small Grains Genotyping Laboratory, USDA-ARS for their technical help with the KASP assay. We appreciate Josh Udall, James Frelichowski, and Janna Love of the Crop Germplasm Research, USDA-ARS, College Station, TX for providing the Upland and Pima cotton accessions used in this study.

**CONFLICTS OF INTEREST**: The authors declare no conflicts of interest for this study.

**Funding information**: Funding support for this research provided by Cotton Incorporated (project 19-952) and North Carolina Cotton Producers' Association (project 19-796NC) to Vasu Kuraparthy, National Science Foundation 1928344 to Rebecca Bart. Taylor Harris is supported by an NIGMS-funded Initiative for Maximizing Student Development grant to Washington University (GM103757).

## REFERENCES


Allen, T. (2011, July 15). *ALERT Bacterial Blight of Cotton.* Mississippi Crops. https://www.mississippi-crops.com/2011/07/15/alert-bacterial-blight-of-cotton/

Amudha, J., Balasubramani, G., Meshram, M. K., Gotmare, V., & Mayee, C. D. (2003). Molecular characterization of gene introgression for bacterial blight resistance in cotton. *Plant Cell Biotechnology and Molecular Biology, 4*(3-4), 131–136. Retrieved from https://www.ikprress.org/index.php/PCBMB/article/view/1677

Atkinson, G. F. (1892). *Some Diseases of Cotton.* Alabama experiment station. http://hdl.handle.net/11200/2696

Bird, L. S. (1982). The MAR (multi-adversity resistance) system for genetic improvement of cotton. *Plant Disease, 66*, 172–176.

Bird, L. S., & Blank, L. M. (1951). *Breeding strains of cotton resistant to bacterial blight*. Texas Agricultural Experiment Station.

Bourland, F. M. & Jones, D. C. (2006). Registration of Arkot 9406ne, Arkot 9605ne, and Arkot 9631ne, three nectariless germplasm lines of cotton. *Crop Science, 46*(4), 1833-1834. https://doi.org/10.2135/cropsci2006.03-0184

Bourland, F. M. & Jones, D. C. (2007). Registration of Arkot 9304a, Arkot 9304b, Arkot 9308 and Arkot 9314 germplasm lines of cotton. *Journal of Plant Registrations, 1*(1), 56-57. https://doi.org/10.3198/jpr2006.08.0539crg

Bourland, F. M. & Jones, D. C. (2011). Registration of Arkot 0008, Arkot 0009, and Arkot 0012 germplasm lines of cotton. *Journal of Plant Registrations, 5*(3), 374-378. https://doi.org/10.3198/jpr2010.12.0703crg

Bourland, F. M. & Jones, D. C. (2014). Registration of Arkot 0111, Arkot 0113, and Arkot 0114 Germplasm Lines of Cotton. *Journal of Plant Registration, 8*(1), 68-72. https://doi.org/10.3198/jpr2012.09.0036crg

Bourland, F. M. & Jones, D. C. (2015). Registration of Arkot 0403ne, Arkot 0409, and Arkot 0410HG germplasm lines of cotton. *Journal of Plant Registration, 9*(3), 353-357. https://doi.org/10.3198/jpr2015.04.0028crg

Bourland, F. M. & Jones, D. C. (2017). Registration of Arkot 0502ne, Arkot 0504ne, Arkot 0506ne, and Arkot 0517HG Germplasm Lines of Cotton. *Journal of Plant Registration, 11*(1), 66-70. https://doi.org/10.3198/jpr2016.06.0031crg

Bourland, F. M. (2018). Breeding for bacterial blight resistance in cotton. In: *Proceedings of the beltwide cotton conference* (19–22). San Antonio, TX.

Bourland, F. M., McGowan, J. R. E., & Johnson, J. T. (1997). Registration of arkot 8102, arkot 8506, and arkot 8514 germplasm lines of cotton. *Crop Science, 37*, 1397. https://doi.org/10.2135/cropsci1997.0011183X003700040086x

Brown, J. K. (2001). Bacterial Blight. In T. L. Kirkpatrick & C. S. Rothrock (Eds.), *Compendium of cotton diseases* (2nd ed., pp. 34-35). APS Press.

Cunnac, S., Bolot, S., Forero, S. N., Ortiz, E., Szurek, B., Noel, L. D., Arlat, M., Jacques, M. A., Gagnevin, L., Carrere, S., Nicole, M., & Koebnik, R. (2013). High-Quality Draft Genome Sequences of Two *Xanthomonas citri* pv. malvacearum Strains. *Genome Announcements, 1*(4). https://doi.org/10.1128/genomeA.00674-13

Daniels, M. J., Barber, C. E., Turner, P. C., Cleary, W.G., Sawczyc, M. K. (1984). Isolation of mutants of *Xanthomonas campestris* pv. *campestris* showing altered pathogenicity. *Microbiology, 130*(9), 2447-2455. https://doi.org/10.1099/00221287-130-9-2447

Delannoy, E., Lyon, B. R., Marmey, P., Jalloul, A., Daniel, J. F., Montillet, J. L., Essenberg, M., & Nicole, M. (2005). Resistance of cotton towards *Xanthomonas campestris* pv. *malvacearum*. *Annual review of phytopathology, 43*, 63-82. https://doi.org/10.1146/annurev.phyto.43.040204.140251

Elassbli, H., Abdelraheem, A., Sanogo, S., Kuraparthy, V., Stelly, D. M., Hinze, L., Wheeler, T. A., Wedegaertner, T., & Zhang, J. (2019). A genome-wide association study of resistance to Fusarium wilt and bacterial blight in US Upland cotton germplasm. In: *Proceedings of beltwide cotton conference* (pp. 16–20).

Elassbli, H., Abdelraheem, A., Zhu, Y., Teng, Z., Wheeler, T. A., Kuraparthy, V., Hinze, L., Stelly, D. M., Wedegaertner, T., & Zhang, J. (2021). Evaluation and genome-wide association study of resistance to bacterial blight race 18 in US upland cotton germplasm. *Molecular Genetics and Genomics, 296*(3), 719–729. https://doi.org/10.1007/s00438-021-01779-w

Elassbli, H., Zhu, Y., Abdelraheem, A., Wheeler, T., Wedegaertner, T., & Zhang, J. (2021b). Genetic analysis of resistance to bacterial blight Race 18 and B*12*-linked marker analysis in U.S. upland cotton. *Crop Science*, *61*(5), 3458-3468. https://doi.org/10.1002/csc2.20567

El-Zik, K., & Thaxton, P. (1994). Breeding for resistance to bacterial blight of cotton in relation to races of the pathogen. In: *Challenging the future: World Cotton Research Conference*-1 (pp. 236- 241).

Faustine, C., Hoffmann, L. V., Lukonge, E., & Tibazarwa, F. I. (2015). Marker-assisted screening of cotton cultivars for bacterial blight resistance gene. *Huria Journal, 19*(1), 1–7.

Gowda, A. S., Shrestha, N., Harris, T. M. Phillips, A. Z., Fang, H., Sood, S., Zhang, K., Bourland, F., Bart, R., & Kuraparthy, V. (2022). Identifcation and genomic characterization of major effect bacterial blight resistance locus (*BB-13*) in Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics, 135*(12), 4421-4436. https://doi.org/10.1007/s00122-022-04229-2

Green, J. M., & Brinkerhof, L. A. (1956). Inheritance of three genes for bacterial blight resistance in upland Cotton. *Agronomy Journal, 48*(11), 481–485. https://doi.org/10.2134/agronj1956.00021962004800110001x

Hulse-Kemp, A. M., Lemm, J., Plieske, J., Ashrafi, H., Buyyarapu, R., Fang, D. D., Frelichowski, J., Giband, M., Hague, S., Hinze, L. L., Kochan, K. J., Riggs, P. K., Scheffler, J. A., Udal, J. A., Ulloa, M., Wang, S. S., Zhu, Q. H., Bag, S. K., Bhardwaj, A., … Stelly, D. M. (2015). Development of a 63K SNP Array for Cotton and High-Density Mapping of Intraspecific and Interspecific Populations of *Gossypium* spp. *G3-Genes Genomes Genetics, 5*(6), 1187-1209. https://doi.org/10.1534/g3.115.018416

Innes, N. L. (1983). Bacterial Blight of Cotton. *Biological reviews, 58*(1), 157-176. https://doi.org/10.1111/j.1469-185X.1983.tb00383.x

Jalloul, A., Sayegh, M., Champion, A., & Nicole, M. (2015). Bacterial blight of cotton. *Phytopathologia Mediterranea*, *54*(1), 3-20. https://doi.org/10.14601/Phytopathol_Mediterr-14690

Kaur, B., Tyagi, P., & Kuraparthy, V. (2017). Genetic Diversity and Population Structure in the Landrace Accessions of *Gossypium hirsutum*. *Crop Science, 57*(5), 2457-2470. https://doi.org/10.2135/cropsci2016.12.0999

Kemerait, B., Allen, T., Lu, S., Rothrock, C., Faske, T., Woorward, J., Wheeler, T., Isakeit, T., Bart, R., Phillips, A., Lawrence, K., Hagan, A., Price, P., Mehl, H., Dufault, N., Kelly, H., & Nichols, R. (2017). *Identification and Management of Bacterial Blight of Cotton.* Cotton Incorporated. https://www.cottoninc.com/cotton-production/ag-research/plant-pathology/management-bacterial-blight-cotton/

Knight, R. L., & Clouston, T. W. (1939). The genetics of black-arm resistance. Factors B2 and B3. *Journal of Genetics, 38*, 133–159

Knight, R. L., (1963). The genetics of blackarm resistance XII. Transference of resistance from *Gossypium herbaceum* to *G. barbadense*. *Journal of Genetics, 3*, 328–346.

Makhoul, M., Rambla, C., Voss-Fels, K. P., Hickey, L. T., Snowdon, R. J., & Obermeier, C. (2020). Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. *Theoretical and Applied Genetics, 133*(8), 2413-2430. https://doi.org/10.1007/s00122-020-03608-x

Phillips, A. Z., Berry, J. C., Wilson, M. C., Vihayaraghavan, A., Burke, J., Bunn, J. I., Allen, T. W., Wheeler, T., & Bart, R. S. (2017). Genomics-enabled analysis of the emergent disease cotton bacterial blight. *Public Library of Science Genetics, 13*(9), 1-23. https://doi.org/10.1371/journal.pgen.1007003

Rasheed, A., Wen, W., Gao, F. M., Zhai, S., Jin, H., Liu, J. D., Guo, Q., Zhang, Y., Dreisigacker, S., Xianchun, X., & Zhonghu, H. (2016). Development and validation of KASP assays for functional genes underpinning key economic traits in wheat. *Theoretical and Applied Genetics, 129*(10), 1843–1860. https://doi.org/10.1007/s00122-016-2743-x

Ridgway, R. L., Bell, A. A., Veech, J. A., & Chandler, J. M. (1984). Cotton protection practices in the USA and world. *Cotton* 289-295. Madison, WI. https://doi.org/10.2134/agronmonogr24.c9

Rungis, D., Llewellyn, D., Dennis, E. S., & Lyon, B. R. (2002). Investigation of the chromosomal location of the bacterial blight resistance gene present in an Australian cotton (*Gossypium*

*hirsutum* L.) cultivar. *Australian Journal of Agricultural Research, 53*, 551–560. https://doi.org/10.1071/AR01121

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding, 33*(1), 1–14. https://doi.org/10.1007/s11032-013-9917-x

Silva, R. A., Barroso, P. A. V., Hoffmann, L. V., Giband, M., & Coutinho, W. M. (2014). A SSR marker linked to the B12 gene that confers resistance to race 18 of *Xanthomonas axonopodis* pv. *malvacearum* in cotton is also associated with other bacterial blight resistance gene complexes. *Australasian Plant Pathology, 43*(1), 89–91. http://dx.doi.org/10.1007/s13313-013-0249-4

Thomson, M. J. (2014). High-throughput SNP genotyping to accelerate crop improvement. *Plant Breeding and Biotechnoloy, 2*(3), 195–212. https://doi.org/10.9787/PBB.2014.2.3.195

3CR Bioscience Ltd. (2022) *PACE® 2.0 master mix user guide.* Retrieved from https://3crbio.com/wp-content/uploads/2023/01/PACE-2.0-User-Guide-v2.pdf

Tyagi, P., Gore, M. A., Bowman, D. T., Campbell, B. T., Udall, J. A., & Kuraparthy, V. (2014). Genetic diversity and population structure in the US upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics, 127*(2), 283-295. https://doi.org/10.1007/s00122-013-2217-3

Van, H. O., Stam, P., Richard, G. F., Herman, V. J., Eck, V. (2005). Record: a novel method for ordering loci on a genetic linkage map. *Theoretical and Applied Genetics, 112*(1), 30-40. https://doi.org/10.1007/s00122-005-0097-x

Verna, J. P. (1986). *Bacterial Blight of Cotton*. CRC Press.

Viot, C. R., & Wendel, J. F. (2023). Evolution of the Cotton Genus, *Gossypium*, and Its Domestication in the Americas. *Critical Reviews in Plant Science*. https://doi.org/10.1080/07352689.2022.2156061

Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and qtls. *Journal of heredity, 93*(1), 77-78. https://doi.org/10.1093/jhered/93.1.77

Wallace, T. P., & El-Zik, K. M. (1989). Inheritance of Resistance in Three Cotton Cultivars to the HV1 Isolate of Bacterial Blight. *Crop Science, 29*(5), 1114-1119. https://doi.org/10.2135/cropsci1989.0011183X002900050003x

Wendel, J. F., & Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. *Advances in Agronomy, 78*, 139–186. https://doi.org/10.1016/S0065-2113(02)78004-8

Wendel, J. F., & Grover, C. E. (2015). Taxonomy and Evolution of the Cotton Genus, *Gossypium*. Cotton, 57, 2nd ed. *Agronomy Monographs*, 25-44. https://doi.org/10.2134/agronmonogr57.2013.0020

Wendel, J. F., Brubaker, C., & Percival, A. (1992). Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. *American Journal of Botany, 79*, 1291–1310. https://doi.org/10.1002/j.1537-2197.1992.tb13734.x

Wickens, G. M. (1956). Vascular infection of cotton by *Xanthomonas malvacearum* (E. F. Smith) Dowson. *Annals of Applied Biology, 44*(1), 129-137. https://doi.org/10.1111/j.1744-7348.1956.tb06851.x

Wright, R. J., Thaxton, P. M., El-Zik, K. M., & Paterson, A. H. (1998). D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. *Genetics, 149*(4), 1987–1996. https://doi.org/10.1093/genetics/149.4.1987

Xiao, J., Fang, D. D., Bhatti, M., Hendrix, B., & Cantrell, R. (2009). A SNP haplotype associated with a gene resistant to *Xanthomonas axonopodis* pv. *malvacearum* in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding, 25*, 593–602. https://doi.org/10.1007/s11032-009-9355-y

Yang, Z. (2013). *Fine mapping a bacterial blight resistance gene in the cotton cultivar S295 (Gossypium hirsutum L.)* [Master's Thesis, Texas Tech University]. TTU Digital Library. https://ttu-ir.tdl.org/bitstream/handle/2346/50304/YANG-THESIS-2013.pdf?sequence=1

Yang, Z., Ge, X., Qin, W., Sun, G., Wang, Z., Li, Z., Liu, J., Wu, J., & Wang, Y. (2019). Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nature Communications, 10*(1), 1–13. https://doi.org/10.1038/s41467-019-10820-x

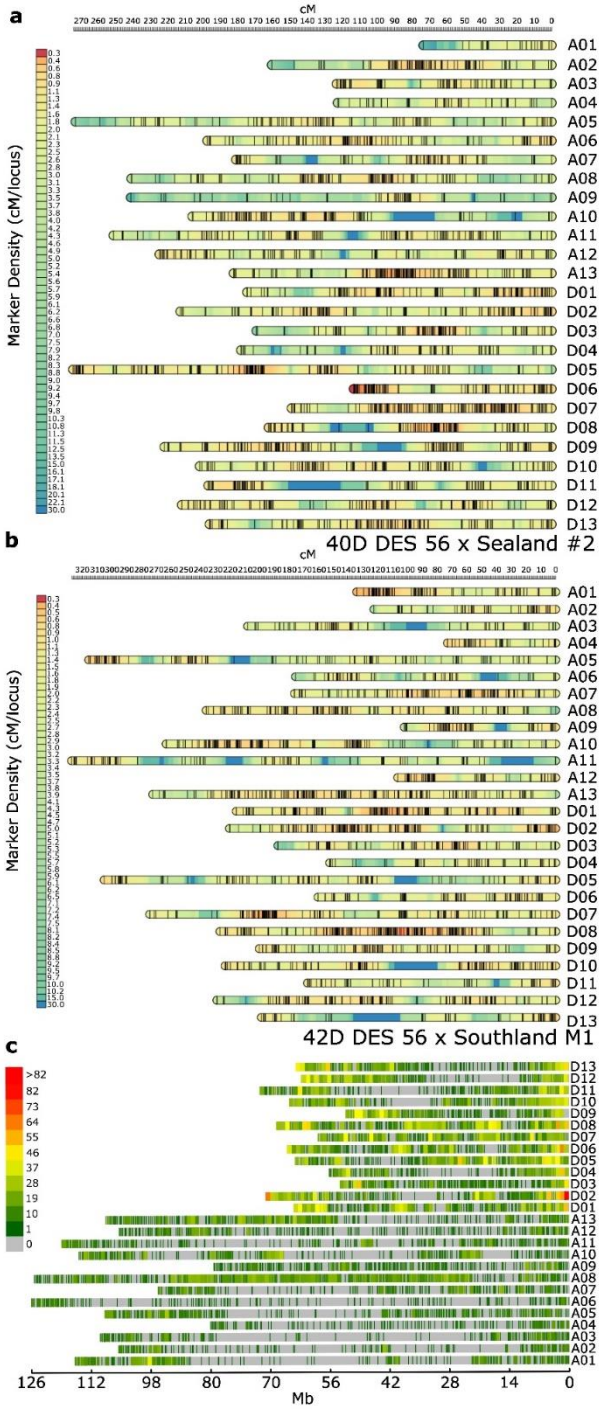
Yu, J., Jung, S., Cheng, C., Lee, T., Zheng, P., Buble, K., Crabb, J., Humann, J., Hough, H., Jones, D., Campbell, J. T., Udall, J., & Main, D. (2021). CottonGen: the community database for cotton genomics. *Plants (Basel), 10*(1), 2805. https://doi.org/10.3390/plants10122805

Zhang, J., Bourland, F., Wheeler, T., & Wallace, T. (2020) Bacterial blight resistance in cotton: genetic basis and molecular mapping. *Euphytica, 216*, 1–19. https://doi.org/10.1007/s10681-020-02630-w

# APPENDIX

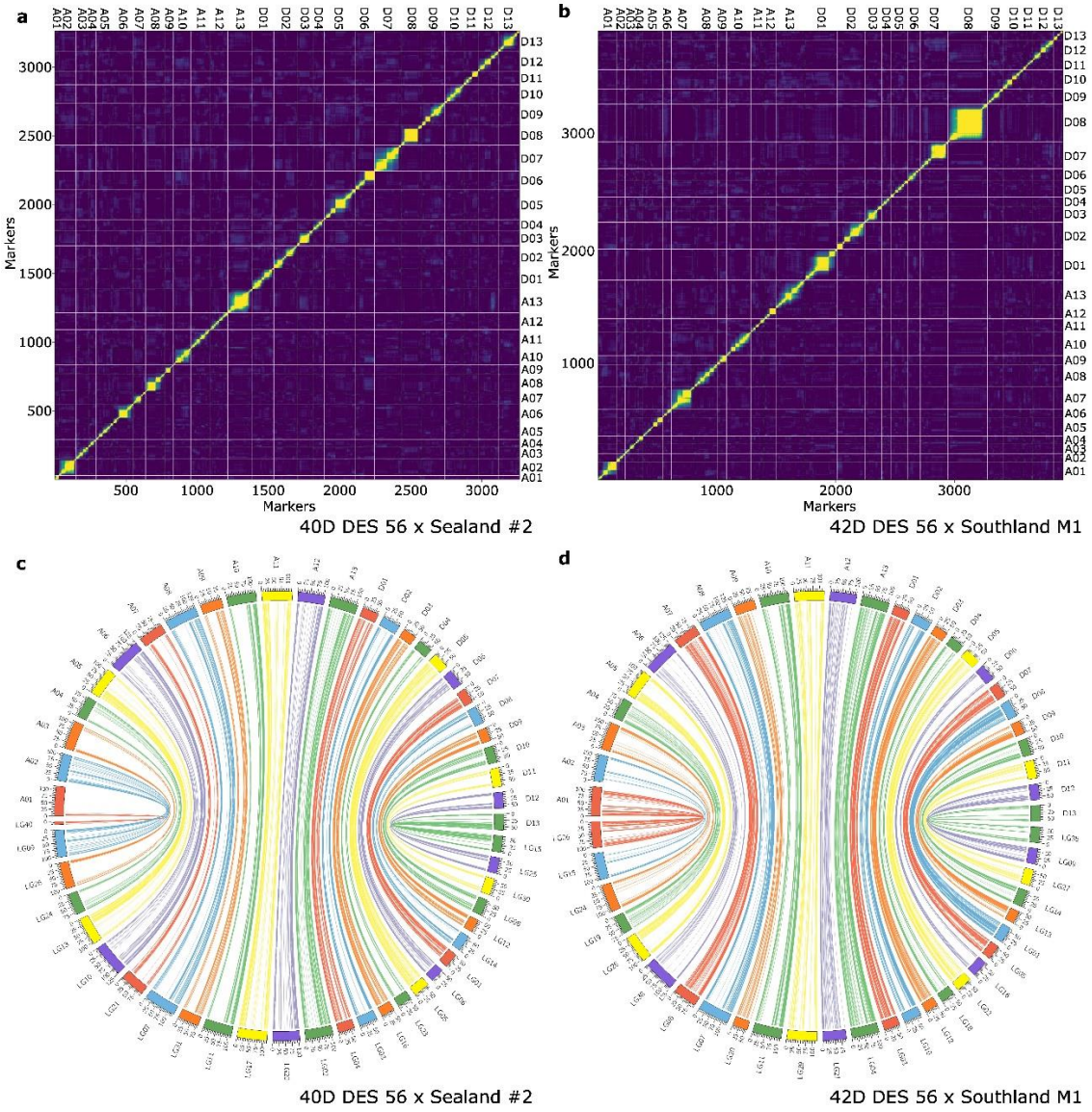
**Supplementary figure 1:** SNP marker density and distribution for 26 Upland cotton chromosomes in linkage mapping and the genome wide association study. **(a)** Distribution of 3,264 polymorphic SNPs in 122 RILs from the 40D $F_5$ population (DES 56 × Sealand #2). SNP density was calculated by number of SNP markers per cM using the cottonSNP63K array (Hulse-Kemp *et al.*, 2015). **(b)** Distribution of 3,913 polymorphic SNP markers in 111 RILs from the 42D $F_5$ population (DES 56 × Southland M1). Marker density calculated as described in part a. **(c)** Distribution of 19,349 polymorphic SNP markers in 136 Upland accessions. Marker density calculated by a 1 Mb sliding interval window.

**Supplementary Figure 1** (continued).



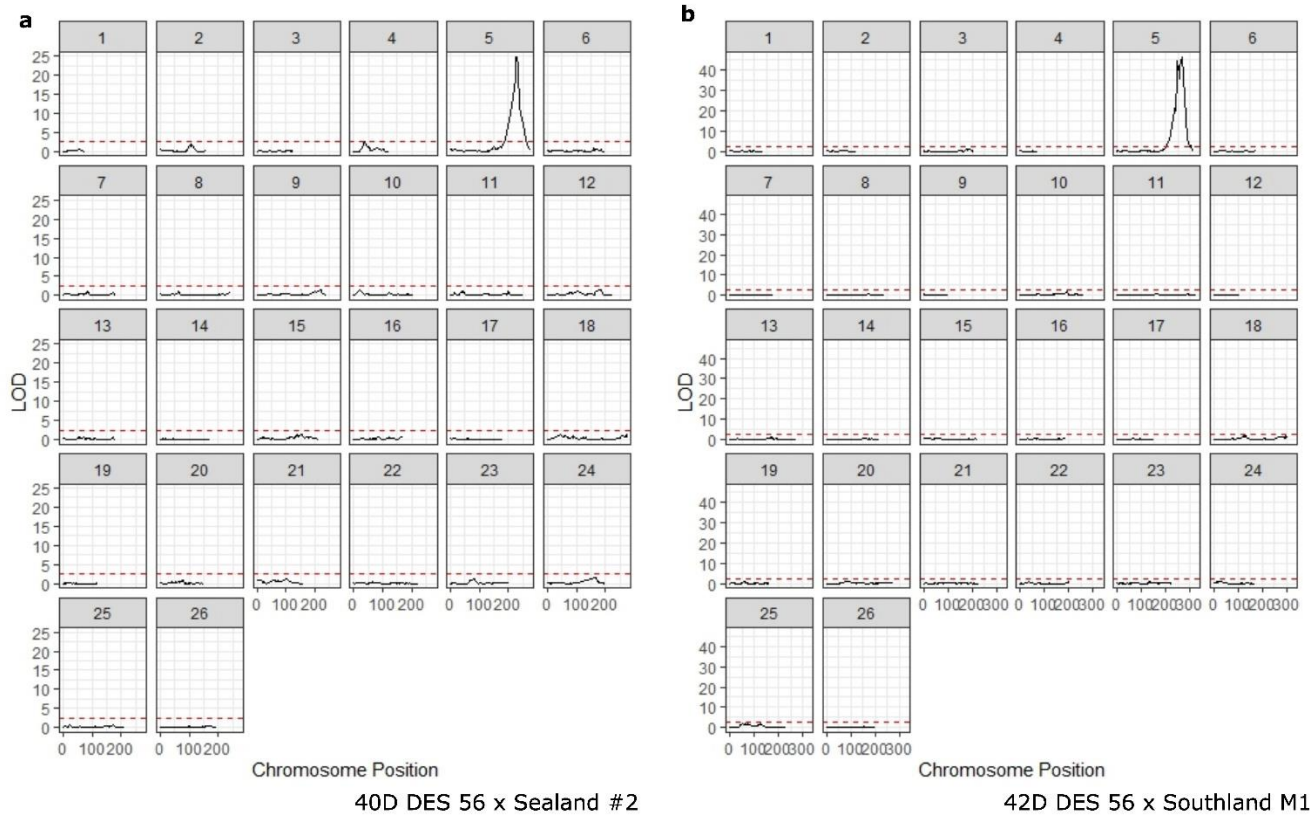
**Supplementary figure 2:** Heat maps comparing the precise grouping and ordering of SNP markers in the linkage group corresponding to the 26 Upland cotton chromosomes for the two $F_5$ RIL populations **(a)** heat map for 40D (DES 56 × Sealand #2) population based map **(b)** heat map for 42D (DES 56 × Southland M1) population based map. The upper left triangle plots recombination fractions of each marker in a linkage group by all the other linkage groups. The lower right triangle plots the LOD scores between each marker. **(c-d)** Comparison of collinearity between the genetically mapped linkage groups and the sequence based physical chromosome map of *G. hirsutum* cv. TM1 CRI v1 reference genome (Yang *et al.*, 2019).

**Supplementary Figure 2** (continued).



**Supplementary figure 3:** Single marker analysis of pollen color for all 26 Upland cotton chromosomes for **(a)** 40D $F_5$ RIL population (DES 56 × Sealand #2) and **(b)** 42D $F_5$ RIL population (DES 56 × Southland M1). A significant peak above the LOD threshold level 2.5 was observed above the A05 chromosome.

**Supplementary Figure 3** (continued).




**Supplementary Table 1:** List of the complete 382 *Gossypium* accessions in the elite diversity (DIV) panel. A subsection of 72 of the *G. hirsutum* accessions used for KASP marker validation for eight markers. Two hundred and fifty three of the *G. hirsutum* accessions were used as SNP i25755Gh marker evaluation. The remaining 129 accessions used SNP i25755Gh as a diagnostic marker for predictions for the *BB-13* cotton bacterial blight phenotypic response to *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) race 18.

**Supplementary Table 2:** Primer sequence information of the complete ten KASP markers utilized in the current study to identify a diagnostic marker for *BB-13* locus of cotton bacterial blight from *Xanthomonas citri* subsp. *malvacearum* (*Xcm*) race 18. Markers were chosen based SNPs from the CottonSNP63K array identified in previous GWAS mapping and linkage analysis (Gowda et al., 2022). SNP physical positions on the D02 chromosome are reported using the *Gossypium hirsutum* TM1 CRI v1 reference genome (Yang et al., 2019).

**Supplementary Table 3:** Phenotypic segregation ratios in the recombinant inbred line populations evaluated for resistance to race 18 of cotton bacterial blight pathogen *Xanthomonas*

*citri* subsp. *malvacearum* (*Xcm*). RIL populations were phenotyped in the growth chambers at the seedling stage. Both populations showed a 1:1 ratio of resistant and susceptible phenotypes.

**Supplementary Table 4:** Error rates for four polymorphic KASP markers in the biparental RIL populations according to true resistance phenotype against cotton bacterial blight pathogen *Xanthomonas citri* subsp. *malvacearum* (*Xcm*). Linkage distances between the markers and the *BB-13* locus are according to linkage mapping of Acala Maxxa × Arkot 8102 RIL population from Gowda et al. (2022).

**Supplementary Table 5:** Comparison of predicted cotton bacterial blight (CBB) phenotypes and actual CBB phenotypes of 253 *Gossypium hirsutum* accessions of elite diversity (DIV) panel. The CBB phenotype predictions were made solely based on SNP maker alleles of i25755Gh. The allele pattern of i25755Gh is sourced from CottonSNP63K array genotyping of the elite diversity panel. In the table, i25755Gh's allele “CC” is linked with CBB resistance at the *BB-13* locus and allele “TT” is linked with susceptibility.

**Supplementary Table 6:** Predictions of cotton bacterial blight (CBB) response phenotypes of 875 Gossypium germplasm lines based on i25755Gh SNP allele calls. The i25755Gh's allele calls were obtained through CottonSNP63K array genotyping of the 875 cotton accessions. In the table, i25755Gh's allele “CC” is linked with CBB resistance at the *BB-13* locus and susceptibility with allele “TT”.