Genomics-enabled analysis of the emergent disease cotton bacterial blight

Anne Z. Phillips[1,2], Jeffrey C. Berry[1¶], Mark C. Wilson[1¶], Anupama Vijayaraghavan[1], Jillian Burke[1], J. Imani Bunn[1], Tom W. Allen[3], Terry Wheeler[4], Rebecca Bart[1*]

[1]Donald Danforth Plant Science Center, St. Louis MO, United States of America

[2]Department of Biology, Washington University in Saint Louis, St. Louis, MO, United States of America

[3]Delta Research and Extension Center, Mississippi State University, Stoneville, MS, United States of America

[4]Texas AgriLife Research, Texas AgriLife Extension Service, Lubbock, TX, United States of America

*Corresponding Author

Email: rbart@danforthcenter.org

[¶]these authors contributed equally to this work.

Abstract

Cotton bacterial blight (CBB), an important disease of (*Gossypium hirsutum*) in the early 20th century, had been controlled by resistance genes for over half a century. Recently, CBB re-emerged as an agronomic problem in the United States. Here, a comparative genomics analysis of host and pathogen was conducted. Phylogenetic analysis revealed that strains from the current outbreak cluster with race 18 *Xanthomonas citri* pv. *malvacearum* (*Xcm*) strains. Type three effector repertoires of 16 *Xcm* isolates reveal 24 conserved effectors as well as nine variable effectors. In addition, virulent race 18 strains contain 3 to 5 more effectors than non-race 18 strains. Genome assemblies for two geographically and temporally divergent strains of *Xcm*, yielded circular chromosomes and accompanying plasmids. These genomes encode eight and thirteen distinct transcription activator-like effector genes. RNA-sequencing revealed that both strains induced 52 conserved gene targets in diverse cotton cultivars, including a homeologous pair of genes, with homology to the known susceptibility gene, MLO. In contrast, the two strains of *Xcm* induced different SWEET sugar transporters and in one case, only one homeolog was significantly induced. Subsequent genome wide analysis revealed the overall expression patterns of the homeologous gene pairs in cotton after inoculation by *Xcm*. These data reveal host-pathogen specificity in the *Xcm-G. hirsutum* pathosystem, give explanations for the CBB reemergence, and strategies for future development of resistant cultivars.

Author Summary

Cotton bacterial blight (CBB), caused by *Xanthomonas citri* pv. *malvacearum* (*Xcm*), significantly limited cotton yields in the early 20th century but has been controlled by classical resistance genes for more than 50 years. In 2011, the pathogen re-emerged with vengeance. In this study, we compare diverse pathogen isolates and cotton varieties to further understand the virulence mechanisms employed by *Xcm* and to identify promising resistance strategies. We generate fully contiguous genome assemblies for two diverse *Xcm* strains and identify pathogen proteins used to modulate host transcription and promote susceptibility. RNA-Sequencing of infected cotton reveals novel putative gene targets for the development of durable *Xcm* resistance. Together, the data presented reveal the underlying cause of CBB re-emergence in the U.S. and highlight several promising routes towards the development of durable resistance including classical resistance genes and potential manipulation of susceptibility targets.

Introduction

Upland cotton (*Gossypium hirsutum* L.) is the world's leading natural fiber crop. Cotton is commercially grown in over 84 countries and in the United States is responsible for $74 billion annually [1, 2]. Numerous foliar diseases affect cotton throughout the world's cotton growing regions. Historically, one of the most significant foliar diseases has been bacterial blight, caused by *Xanthomonas citri* pv. *malvacearum*. Cotton bacterial blight significantly limited cotton yield in the late 20th century. In the 1940's and 1950's, breeders identified and introgressed multiple resistance loci into elite germplasm [3-5].This strategy proved durable for over half a century. In 2011, cotton bacterial blight (CBB) returned and caused significant losses to farmers in the southern United States, more specifically in Arkansas and Mississippi.

66 Nonetheless, CBB has received little research focus during the last several decades because this
67 disease had been considered "tamed". Modern molecular and genomic technologies can now
68 be employed expeditiously to deduce the underlying cause of the disease re-emergence and
69 pinpoint optimized routes towards the development of durable resistance.

70 CBB is caused by *X. citri pv. malvacearum* (*Xcm*); however, the pathogen has previously
71 been placed within other species groupings [6-9]. The *Xcm* pathovar can be further divided into
72 at least 19 races according to virulence phenotypes on a panel of historical cotton cultivars:
73 Acala-44, Stoneville 2B-S9, Stoneville 20, Mebane B-1, 1-10B, 20-3, and 101-102.B [10, 11].
74 Historically, the most common race observed in the U.S. has been race 18, which was first
75 isolated in 1973 [12]. This race is highly virulent, causing disease on all cultivars in the panel
76 except for 101-102.B. CBB can occur at any stage in the plant's life cycle and on any aerial
77 organ. Typical symptoms include seedling blight as either pre or post-emergent damping-off,
78 black arm on petioles and stems, water-soaked spots on leaves and bracts, and most
79 importantly boll rot. The most commonly observed symptoms are the angular-shaped lesions
80 on leaves that, in some cases, can coalesce and result in a systemic vein infection where leaf
81 lesions coalesce on major leaf veins. Disease at each of these stages can cause yield losses
82 either by injury to the plant or direct damage to the boll. No effective chemical treatments for
83 the disease have been released to date. Therefore, the most important methods to reduce loss
84 as a result of CBB include field methods that rely on cultivation to reduce potential sources of
85 overwintering inoculum and planting cultivars with known sources of resistance.

86 Most pathogenic bacteria assemble the type three secretion system (T3SS), a needle-
87 like structure, to inject diverse type three effectors (T3Es) into the plant cell to suppress

88 immunity and promote disease [13-17]. For example, transcription activator-like (TAL) effectors
89 influence the expression levels of host genes by binding directly to host gene promoters in a
90 sequence-specific way [18]. Up-regulated host genes that contribute to pathogen virulence are
91 termed susceptibility genes and may be modified through genome editing for the development
92 of resistant crop varieties [19].

93 Plants have specialized immune receptors, collectively known as nucleotide-binding
94 leucine rich repeat receptors (NLRs), that recognize, either directly or indirectly, the pathogen
95 effector molecules [20, 21]. Historically, this host-pathogen interaction has been termed the
96 'gene-for-gene' model of immunity, wherein a single gene from the host and a single gene from
97 the pathogen are responsible for recognition [22]. Recognition triggers a strong immune
98 response that often includes a localized hypersensitive response (HR) in which programmed cell
99 death occurs around the infection site [23]. Nineteen CBB resistance loci have been reported in
100 *Gossypium hirsutum* breeding programs; however, none have been molecularly identified [8,
101 24].

102 Here we combine comparative genomics of the pathogen *Xcm* with transcriptomics of
103 the host to identify the molecular interactions underlying this re-emergent disease. This will
104 inform the development of durable resistance strategies.

105 Results

106 CBB Reemergence in the US

107 In 2011, farmers, Extension specialists, and Certified Crop Advisers in Missouri,
108 Mississippi, and Arkansas observed cotton plants exhibiting symptoms of CBB. While a major
109 limiting factor for cotton production through the 1950s, this disease had been controlled by

110 agricultural practices such as acid-delinting seed as well as planting resistant cultivars. Prior to
111 the widespread observation of CBB in the mid-southern U.S., isolated, sporadic instances of the
112 disease were generally detected on an annual basis. Reemergence of the disease occurred
113 rapidly during 2011. Widespread infected plant material was observed throughout much of the
114 production area, but appeared to be centered around Clarksdale, Mississippi. Much of the
115 infestation in the Arkansas production system was reported to have originated from several
116 infested seed lots [25]. The disease has since spread through much of the cotton belt in the
117 southern U.S. (Figs 1 and S1).

118 In 2014, diseased cotton leaves were collected from three sites across Mississippi and
119 Koch's postulates were conducted to prove causality [26]. PCR amplification of the 16S rRNA
120 gene confirmed that the causal agent was a member of the *Xanthomonas* genus. Multi locus
121 sequence type (MLST) analysis and maximum-likelihood analysis were performed using
122 concatenated sections of the gltA, lepA, lacF, gyrB, fusA and gap-1 loci (Fig 2a) for increased
123 phylogenetic resolution. The newly sequenced strains were named MS14001, MS14002 and
124 MS14003 and were compared to four previously published *Xcm* genomes and thirty-six
125 additional *Xanthomonas* genomes representing thirteen species (Tables S1, S2). MS14001,
126 MS14002 and MS14003 grouped with the previously published *Xcm* strains as a single
127 polytomy, further confirming that the current disease outbreak is CBB and is caused by *Xcm*.
128 The species designation reported here is consistent with previous reports [6, 7]. To date, CBB
129 has been reported from at least eight out of the sixteen states that grow cotton (Fig 1).

130 **Contemporary U.S. *Xcm* strains cluster phylogenetically with historical race 18 strains.**

131 Race groups have been described for *Xcm* strains by analyzing compatible (susceptible)
132 and incompatible (resistant) interactions on a panel of seven cotton cultivars. In general, race
133 groups tend to be geographically distinct. For example, as mentioned previously, race 18 is
134 prevalent in the U.S. while race 20 is a highly virulent strain reported from several African
135 countries [7]. Consequently, one possible explanation for the recent outbreak of CBB would be
136 the introduction of a new race of *Xcm* capable of overcoming existing genetic resistance.
137 Unfortunately, only 2 cultivars of the original cotton panel plus three related cultivars, were
138 available and these cultivars were not sufficient to determine whether a new race had
139 established within the U.S. Consequently, twelve *Xcm* strains were sequenced using Illumina
140 technology to determine the phylogenetic relationship between recent isolates of *Xcm* and
141 historical isolates. Isolates designated as race 1, race 2, race 3, race 12 and race 18 have been
142 maintained at Mississippi State University. Additional isolates were obtained from the
143 Collection Française de Bactéries associées aux Plantes (CFBP) culture collection. Together,
144 these isolates include eight strains from the US, three from Africa, and one from South America
145 and span collection dates ranging from 1958 through 2014 (Fig 1). Illumina reads were mapped
146 to the *Xanthomonas citri* subsp. *citri* strain Aw12879 (565918 [RefSeq]) using Bowtie and single
147 nucleotide polymorphisms (SNPs) were identified using Samtools [27, 28]. Only regions of the
148 genome with at least 10x coverage for all genomes were considered. This approach identified
149 17,853 sites that were polymorphic in at least one genome. Nucleotides were concatenated
150 and used to build a neighbor-joining tree (Fig 2b). This analysis revealed that recent *Xcm*
151 isolates grouped with the race 18 clade. Notably, the race 18 clade is phylogenetically distant
152 from the other *Xcm* isolates.

153 **Contemporary US *Xcm* strains have conserved virulence protein arsenals and disease**
154 **phenotypes with historical race 18 strains.**

155 Type three effector (T3E) profiles from sixteen *Xcm* isolates were compared to
156 determine whether a change in the virulence protein arsenal of the newly isolated strains could
157 explain the re-emergence of CBB. Genomes from 12 *Xcm* isolates were de novo assembled with
158 SPAdes and annotated with Prokka based on annotations from the *X. euvesicatoria* (aka. *X.*
159 *campestris* pv. *vesicatoria*) 85-10 genome (NCBI accession: NC_007508.1). T3Es pose a
160 particular challenge for reference based annotation as no bacterial genome contains all
161 effectors. Consequently, an additional protein file containing known T3Es from our previous
162 work was included within the Prokka annotation pipeline [13, 29]. This analysis revealed 24
163 conserved and 9 variable *Xcm* T3Es (Fig 3a). Most race 18 isolates contain more effectors than
164 other isolates that were sequenced. The recent *Xcm* isolates (MS14002 and MS14003) were not
165 distinguishable from historical race 18 isolates, with the exception of XcmNI86 isolated from
166 Nigeria in 1986, which contains mutations in XopE2 and XopP.

167 Analysis of the genomic sequence of T3E revealed presence/absence differences,
168 frameshifts and premature stop codons. However, this analysis does not preclude potential
169 allelic or expression differences among the virulence proteins that could be contributing factors
170 to the re-emergence of CBB. Therefore, newly isolated strains may harbor subtle genomic
171 changes that have allowed them to overcome existing resistance phenotypes. Many
172 commercial cultivars of cotton are reported to be resistant to CBB [30-32]. Based on these
173 previous reports, we selected commercial cultivars resistant and susceptible (6 of each) to CBB.
174 In addition, we included 5 available varieties that are related to the historical panel as well as 2

175 parents from a nested association mapping (NAM) population currently under development
176 [33]. All varieties inoculated with the newly isolated *Xcm* strains exhibited inoculation
177 phenotypes consistent with previous reports for these varieties (Figs 3b,c). In these assays,
178 brightfield and near infrared (NIR) imaging were used to distinguish water-soaked disease
179 symptoms from rapid cell death (hypersensitive response) that is indicative of an immune
180 response. These data confirm that existing resistance genes present within cotton germplasm
181 are able to recognize the newly isolated *Xcm* strains and trigger a hypersensitive response.
182 Together, the phylogenetic analysis, effector profile conservation and cotton inoculation
183 phenotypes, confirm that the recent outbreak of *Xcm* in the US represents a re-emergence of
184 race 18 *Xcm* and is not the result of a dramatic shift in the pathogen.
185 The USDA Agricultural Marketing Service (AMS) publishes the percentage of upland
186 cotton cultivars planted in the U.S. each year (www.ams.usda.gov/mnreports/cnavar.pdf). In 2016,
187 only 25% of the total cotton acreage was planted with resistant cultivars (Fig 3d), based on
188 previously published CBB phenotypes for these cultivars. This is part of a larger downward
189 trend in which the acreage of resistant cultivars has fallen each year since at least 2009.
190 **Comparative genome analysis for two *Xcm* strains**
191 Differences in virulence were observed among *Xcm* strains at the molecular and
192 phenotypic level. In order to gain insight into these differences, we selected two strains from
193 our collection that differed in T3E content, virulence level, geography of origin and isolation
194 date. AR81009 was isolated in Argentina in 1981 and is one of the most virulent strains
195 investigated in this study; MS14003 was isolated in Mississippi in 2014 and causes
196 comparatively slower and diminished leaf symptoms. However, both strains are able to multiply

197 and cause disease on susceptible varieties of cotton (S2 Fig). Full genome sequences were
198 generated with Single Molecule Real-Time (SMRT) sequencing. Genomes were assembled using
199 the PacBio Falcon assembler which yielded circular 5Mb genomes and associated plasmids.
200 Genic synteny between the two strains was observed with the exception of two 1.05 Mb
201 inversions (Fig 4). Regions of high and low GC content, indicative of horizontal gene transfer,
202 were identified in both genomes. In particular, a 120kb insertion with low GC content was
203 observed in AR81009. This region contains one T3E as well as two annotated type four
204 secretion system related genes, two conjugal transfer proteins, and two multi drug resistant
205 genes. MS14003 contained three plasmids of the sizes 52.4, 47.4, and 15.3kb while AR81009
206 contained two plasmids of the sizes 92.6 and 22.9kb. Analysis of homologous regions among
207 the plasmids was performed using progressiveMauve [34]. This identified four homologous
208 regions greater than 1kb that were shared among multiple plasmids (Fig 4).
209 The AR81009 genome encodes twelve TAL effectors that range in size from twelve to
210 twenty three repeat lengths, six of which reside on plasmids. The MS14003 genome encodes
211 eight TAL effectors that range in size from fourteen to twenty eight repeat lengths, seven of
212 which reside on plasmids (Fig 5a). Three incomplete TAL effectors were also identified within
213 these genomes. A 1-repeat gene with reduced 5’ and 3’ regions was identified in both strains
214 directly upstream of a complete TAL effector. In addition, a large 4kb TAL effector was
215 identified in AR81009 with a 1.5 kb insertion and 10 complete repeat sequences. The tool
216 AnnoTALE was used to annotate and group TAL effectors based on the identities of the repeat
217 variable diresidues (RVDs) in each gene [35]. Little homology was identified among TAL
218 effectors within and between strains; only two TAL effectors were determined to be within the

219 same TAL class between strains (TAL19b of AR81009 and TAL19 of MS14003) and two within
220 strain MS14003 (TAL14b and TAL16). Both strains express TAL effector proteins as
221 demonstrated through western blot analysis using a TAL effector specific antibody (Fig 5b).
222 However, the complexity of TAL effector repertoires within these strains prevented complete
223 resolution of each individual TAL effector.

224 **Transcriptome changes induced by *Xcm* in *G. hirsutum*.**

225 An RNA-sequencing experiment was designed to determine whether AR81009 and
226 MS14003 incite different host responses during infection (Fig 6a). Isolates were inoculated into
227 the phylogenetically diverse *G. hirsutum* cultivars Acala Maxxa and DES 56 (Fig 6b) [33].
228 Infected and mock-treated tissue were collected at 24 and 48 hours post inoculation. First, we
229 considered global transcriptome patterns of gene expression. Fifty-two genes were determined
230 to be induced in all *Xcm-G. hirsutum* interactions at 48 hours (S3 Table). Of note among this list
231 of genes is a homeologous pair of genes with homology to the known susceptibility target, MLO
232 [36-39]. Gene induction by a single strain was also observed; AR81009 and MS14003 uniquely
233 induced 127 and 16, *G. hirsutum* genes, respectively (Fig 6c). The increased number of genes
234 induced by AR81009 correlates with the observed severe leaf symptoms caused by this strain.
235 In contrast, the average magnitude of gene induction between the two strains was not
236 significantly different (S3 Fig). Both *Xcm* strains caused more genes to be differentially
237 expressed in DES 56 than in Acala Maxxa. Among the 52 genes significantly induced by both
238 strains, sixteen conserved targets are homeologous pairs, whereas seventeen and fifteen genes
239 are encoded by the A and D sub-genomes, respectively (Tables 1 and S3). It has been previously
240 reported that homeologous genes encoded on the *G. hirsutum* A and D sub-genomes are

241 differentially regulated during abiotic stress [40]. A set of approximately 10,000 homeologous
242 gene pairs were selected and differential gene expression was assessed (Fig 7). For each
243 pairwise comparison of *Xcm* strain and *G*. *hirsutum* cultivar, a similar number of genes were
244 differentially expressed in each of the A and D subgenomes. However, some homeologous pairs
245 were up or down regulated differentially in response to disease, indicating a level of sub-
246 genome specific responses to disease. For example, SWEET sugar transporter gene
247 Gh_D12G1898 in the D genome is induced over four fold during infection with *Xcm* strain
248 AR81009, but the homeolog Gh_A12G1747 in the A genome is not. This relationship is further
249 explored below.

250 **Table 1:** Eight homeologous pairs of *Gossypium hirsutum* genes are upregulated in both Acala Maxxa and DES 56
251 varieties 48 hours post inoculation with *Xanthomonas citri* pv. *malvacearum* strains MS14003 and AR81009.

A Genome	D Genome	Gene Annotation
Gh_A02G0615	Gh_D02G0670	Seven transmembrane MLO family protein
Gh_A03G0560	Gh_D03G0971	Pectate lyase family protein
Gh_A05G2012	Gh_D05G2256	Protein of unknown function DUF688
Gh_A06G0439	Gh_D06G0479	basic chitinase
Gh_A07G1129	Gh_D07G1229	Protein of unknown function (DUF1278)
Gh_A10G0257	Gh_D10G0257	Protein E6
Gh_A10G1075	Gh_D10G1437	Pectin lyase-like superfamily protein
Gh_A13G1467	Gh_D13G1816	pathogenesis-related 4

252

253 **Different strains of *Xcm* target distinct SWEET transporters in *G. hirsutum*.**

254 SWEET sugar transporter genes are commonly targeted and upregulated by TAL
255 effectors in *Xanthomonas* plant interactions [19, 41-43]. Surprisingly, no SWEET genes were
256 detected in the above list of conserved targets. However, of the 54 SWEET sugar transporter
257 genes encoded by the *G. hirsutum* genome, three were upregulated greater than 4 fold in
258 response to inoculation by one of the two *Xcm* strains (Fig 8). Potential TAL effector binding

259 sites were identified using the program TALEnt [44]. MS14003 significantly induces the
260 homeologs Gh_A04G0861 and Gh_D04G1360 and contains three TAL effectors predicted to
261 bind within the 300bp promoter sequences of at least one of these genes (Fig 8a). In contrast,
262 AR81009 significantly induces Gh_D12G1898 but not its homeolog Gh_A12G1747 (Fig 8b).
263 TAL14a, TAL14c, and TAL16b from AR81009 are all predicted to bind to the Gh_D12G1898
264 promoter however the latter two are also predicted to bind to the homeolog Gh_A12G1747.
265 We note that while Gh_A12G1747 did not pass the four fold cut off for gene induction, this
266 gene is slightly induced in DES 56 compared to mock inoculation.

267 **Discussion**

268 Cotton Bacterial Blight was considered controlled in the U.S. until an outbreak was
269 observed during the 2011 growing season in Missouri, Mississippi and Arkansas [45]. Until 2011,
270 seed sterilization, breeding for resistant varieties, and farming techniques such as crop rotation
271 and sterilizing equipment prevented the disease from becoming an economic concern [46]. The
272 number of counties reporting incidence of CBB has increased from 17 counties in 2011 to 77
273 counties in 2015 [47-49]. This paper investigates the root of the re-emergence and identifies
274 several routes towards control of the disease.

275 When the disease was first recognized as re-emerging, several possible explanations
276 were proposed including: (1) A highly virulent race of the pathogen that had been introduced to
277 the U.S.; (2) Historical strains of *Xcm* that had evolved to overcome existing resistance; and (3)
278 Environmental conditions over the last several years that had been particularly conducive to
279 the disease. Here, we present evidence that the re-emergence of CBB is not due to a large
280 genetic change or race shift in the pathogen as has been previously suggested. Rather, the re-

281 emergence of the disease is likely due to large areas of susceptible cultivars being planted. The
282 presented data do not rule out potential environmental conditions that may also have
283 contributed to the re-emergence. In this context, environmental conditions includes disease
284 conducive temperature and humidity as well as potentially contaminated seed or other
285 agronomic practices that may have perpetuated spread of the disease outbreaks. Importantly,
286 the presented data confirm that the presence of resistance loci could be deployed to prevent
287 further spread of this disease. However, since many of the most popular farmer preferred
288 varieties lack these resistance traits, additional breeding or biotechnology strategies will be
289 needed to maximize utility. Notably, the current *Xcm* isolates characterized in this study all
290 originate from Mississippi cotton fields in 2014. During the 2015 and 2016 growing seasons,
291 resistant cotton cultivars were observed in Texas with symptoms indicative of bacterial
292 infection yet distinct from CBB. Additional work is underway to identify and characterize the
293 causal agent(s) of these disease symptoms.

294 While resistant cotton cultivars were identified for all strains in this study, variability in
295 symptom severity was observed for different strains when inoculated into susceptible cultivars.
296 Two strains in particular, MS14003 and AR81009, have different effector profiles as well as
297 different disease phenotypes. Comparative genomic analysis of the two pathogens revealed
298 many differences that may contribute to the relative disease severity phenotypes. Similarly,
299 transcriptomic analysis of two cultivars of *G. hirsutum* inoculated with these strains confirm
300 that the genomic differences between the two strains result in a divergence in their molecular
301 targets in the host.

302 Over the past decade, susceptibility genes have become targets for developing disease
303 tolerant plants [50, 51]. These genes are typically highly induced during infection [52].
304 Therefore, RNA-Seq of infected plants has become a preferred way to identify candidate
305 susceptibility genes. Once identified, genome editing can be used to block induction of these
306 genes [53]. We report a homeologous pair of genes that are homologs of the MLO gene as
307 targeted by both *Xcm* strains in both cotton cultivars. This conservation makes it an excellent
308 candidate for future biotechnology efforts. Because the potential importance of this gene in
309 cotton biology is unknown, the effect of disrupting this gene in cotton physiology must first be
310 explored. However, knock-outs of MLO genes in other systems has led to durable resistance
311 against powdery mildew but also oomycetes and bacteria such as *Xanthomonas* [36, 39]. The
312 dual purpose of host susceptibility genes has been observed previously. For example, the rice
313 Xa13 (aka. Os8N3 and OsSWEET11) gene is required for pollen development but also targeted
314 by a rice pathogen during infection [54]. Xa13 is a member of the SWEET sugar transporters
315 implicated in many pathosystems. In this case, the induction of Xa13 for pathogen susceptibility
316 is mediated by a TAL effector. Of the 54 SWEET genes in the *G. hirsutum* genome, three are
317 significantly upregulated during *Xcm* infection. In contrast to MLO, no single SWEET gene was
318 induced by both pathogen strains in both hosts. Analysis of SWEET gene expression after
319 inoculation revealed a context for polyploidy in the *G. hirsutum-Xcm* pathosystem. We
320 observed a difference in induction between the Gh_A12G1747 and Gh_D12G1898 SWEET
321 genes. Future research may investigate the diploid ancestors of tetraploid cotton to further
322 explore the evolution of host and pathogen in the context of ploidy events [55].

323 Multiple putative TAL effector binding sites were identified within each up-regulated
324 SWEET promoter. These observations suggest that TAL14b, TAL28a and TAL28b from MS14003
325 may work independently or in concert to induce the homeologs Gh_A04G0861 and
326 Gh_D04G1360. Further, TAL14a from AR81009 is likely responsible for the upregulation of
327 Gh_D12G1898. Whether additional TAL effectors are involved in these responses is not clear. It
328 is possible that not all the TAL effectors are expressed. Similarly, genome organization in the
329 host, such as histone modifications or other epigenetic regulation may be affecting these
330 interactions. Future research will investigate these mechanisms further. However, these
331 experiments will be difficult as most *Xcm* strains are not amenable to conjugation nor
332 electroporation.
333 Collectively, the data presented here suggest that the wide-spread planting of CBB-
334 susceptible cultivars has contributed to the re-emergence of CBB in the southern U.S. It is
335 possible that a reservoir of race 18 *Xcm* was maintained in cotton fields below the level of
336 detection due to resistant cultivars planted in the 1990s and early 2000s. Alternatively, the
337 pathogen may have persisted on an alternate host or was brought in on contaminated seed as
338 has previously been suggested [9, 10]. Regardless of the cause of the re-emergence, the
339 genomic comparisons among pathogen races and host cultivars has identified several possible
340 routes towards resistance. These include the use of existing effective resistance loci as well as
341 the potential disruption of the induction of susceptibility genes through genome editing. The
342 latter is an attractive strategy in part because of recent progress in genome editing [56, 57]. In
343 summary, within a relatively short time frame, through the deployment of modern molecular
344 and genomic techniques, we were able to identify the cause of the re-emergence of cotton

345 bacterial blight and generate data that can now be rapidly translated to effective disease
346 control strategies.
347

348 ## Methods

349 ***Xcm* strain isolation and manipulation:**

350 New strains were isolated from infected cotton leaves by grinding tissue in 10mM $MgCl_2$
351 and culturing bacteria on NYGA media. The most abundant colony type was selected, single
352 colony purified and then 16S sequencing was used to confirm the bacterial genus as previously
353 described [58]. In addition, single colony purified strains were re-inoculated into cotton leaves
354 and the appearance of water soaked symptoms indicative of CBB infection was confirmed. Both
355 newly isolated strains as well as strains received from collaborators were used to generate a
356 rifampicin resistance version of each strain. Wildtype strains were grown on NYGA, then
357 transferred to NYGA containing 100µg/ml rifampicin. After approximately 4-5 days, single
358 colonies emerged. These were single colony purified and stored at -80C. The rifampicin
359 resistant version of each *Xcm* strain was used in all subsequent experiments reported in this
360 manuscript unless otherwise noted.

361 **Plant inoculations**

362 *Xcm* strains were grown on NYGA plates containing 100µg/ml rifampicin at 30°C for two
363 days before inoculations were performed. Disease assays were conducted in a growth chamber
364 set at 30°C and 80% humidity. Inoculations were conducted by infiltrating a fully expanded leaf
365 with a bacterial solution in 10mM $MgCl_2$ (OD_{600} specified within each assay).

366 **Cotton Cultivar Statistics**

367 Area of cotton planted per county in the United States in 2015 was obtained from the
368 USDA National Agricultural Statistics Service:
369 www.nass.usda.gov/Statistics_by_Subject/result.php?7061F36A-A4C6-3C65-BD7F-
370 129B702CFBA2&sector=CROPS&group=FIELD%20CROPS&comm=COTTONUSDA. Estimated
371 percentage of upland cotton planted for each variety was obtained from the Agricultural
372 Marketing Service (AMS): www.ams.usda.gov/mnreports/canvar.pdf. CBB disease phenotyping
373 data from 2009-2016 was determined via cotyledon scratch assays and/or field trials sprayed
374 with virulent *Xcm* isolates and has previously been described [59-61].

375 **Bacterial Sequencing and Phylogenetics**

376 Illumina based genomic datasets were generated as previously described [29]. Paired-
377 end Illumina reads were trimmed using Trimmomatic v0.32 (ILLUMINACLIP:TruSeq3-
378 PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) [62]. Genome
379 assemblies were generated using the SPAdes *de novo* genome assembler [63]. Strain
380 information is reported in Supplemental Table 1. Similar to our previously published methods
381 [29], the program Prokka was used in conjunction with a T3E database to identify type three
382 effector repertoires for each of the 12 *Xcm* isolates as well as four *Xcm* genomes previously
383 deposited on NCBI (S2Table) [64].

384 Multi-locus sequence analysis was conducted by concatenating sequences of the gltA,
385 lepA, lacF, gyrB, fusA and gap-1 loci obtained from the Plant-Associated Microbes Database
386 (PAMDB) for each strain as previously described [65]. A maximum-likelihood tree using these
387 concatenated sequences was generated using CLC Genomics 7.5.

388 **Variant Based Phylogeny**

389 A variant based dendrogram was created by comparing 12 Illumina sequenced *Xcm*
390 genomes to the complete *Xanthomonas citri* subsp. *citri* strain Aw12879 reference genome
391 (565918 [RefSeq]) on NCBI. Read pairs were aligned to the reference genome using Bowtie2
392 v2.2.9 with default alignment parameters [27]. From these alignments, single nucleotide
393 polymorphisms (SNPs) were identified using samtools mpileup v1.3 and the bcftools call v1.3.1
394 multi-allelic caller [28]. Using Python v2.7, the output from samtools mpileup was used to
395 identify loci in the *X. citri* subsp. *citri* reference genome with a minimum coverage of 10 reads in
396 each *Xcm* genome used Python version 2.7 available at http://www.python.org. Vcftools
397 v0.1.14 and bedtools v2.25.0 were used in combination to remove sites marked as indel, low
398 quality, or heterozygous in any of the genomes [66, 67]. Remaining loci were concatenated to
399 create a FASTA alignment of confident loci. Reference loci were used where SNP's were not
400 detected in a genome. The resulting FASTA alignment contained 17853 loci per strain. This
401 alignment was loaded into the online Simple Phylogeny Tool from the ClustalW2 package to
402 create a neighbor joining tree of the assessed strains [68, 69]. Trees were visualized using
403 FigTree v1.4.2.

404 **Genome Assembly**

405 Single Molecule, Real Time (SMRT) sequencing of *Xcm* strains MS14003 and AR81009
406 was obtained from DNA prepped using a standard CTAB DNA preparation. Blue Pippin size
407 selection and library preparation was done at the University of Deleware Sequencin Facility.
408 The genomes were assembled using FALCON-Integrate
409 (https://github.com/PacificBiosciences/FALCON-integrate/commit/cd9e93) [70]. The following
410 parameters were used: Assembly parameters for MS14003: length_cutoff = 7000;

411 length_cutoff_pr = 7000; pa_HPCdaligner_option = -v -dal8 -t16 -e.70 -l2000 -s240 -M10;
412 ovlp_HPCdaligner_option = -v -dal8 -t32 -h60 -e.96 -l2000 -s240 -M10; falcon_sense_option = --
413 output_multi --min_idt 0.70 --min_cov 5 --local_match_count_threshold 2 --max_n_read 300 --
414 n_core 6; overlap_filtering_setting = --max_diff 80 --max_cov 160 --min_cov 5 --bestn 10;
415 Assembly parameters for AR81009: length_cutoff = 8000; length_cutoff_pr = 8000;
416 pa_HPCdaligner_option = -v -dal8 -t16 -e.72 -l2000 -s240 -M10; ovlp_HPCdaligner_option = -v -
417 dal8 -t32 -h60 -e.96 -l2000 -s240 -M10; falcon_sense_option = --output_multi --min_idt 0.72 --
418 min_cov 4 --local_match_count_threshold 2 --max_n_read 320 --n_core 6;
419 overlap_filtering_setting = --max_diff 90 --max_cov 300 --min_cov 10 --bestn 10. Assemblies
420 were polished using iterations of pbalign and quiver, which can be found at
421 https://github.com/PacificBiosciences/pbalign/commit/cda7abb and
422 https://github.com/PacificBiosciences/GenomicConsensus/commit/43775fa. Two iterations
423 were run for *Xcm* strain MS14003 and 3 iterations for AR81009. Chromosomes were then
424 reoriented to the DnaA gene and plasmids were reoriented to ParA. The assemblies were
425 checked for overlap using BLAST, and trimmed to circularize the sequences [71]. TAL effectors
426 were annotated and grouped by RVD sequences using AnnoTALE [35]. Homologous regions
427 among plasmids that are greater than 1 kb were determined using progressiveMauve [34].
428 Genomic comparisons between the MS14003 and AR81009 chromosomes were visualized using
429 Circos [72]. Single-copy genes on each of the chromosomes were identified and joined using
430 their annotated id's. Lines connecting the two chromosomes represent these common genes
431 and their respective positions in each genome. A sliding window of 1KB was used to determine
432 the average GC content. Methylation was determined using the Base Modification and Motif

433 Analysis workflow from pbsmrtpipe v0.42.0 at

434 https://github.com/PacificBiosciences/pbsmrtpipe.

435 **Western Blot Analysis**

436 Western Blot analysis of Transcription Activator-Like (TAL) effectors was performed

437 using a TAL specific antibody [43]. Briefly, bacteria were suspended in 5.4 pH minimal media for

438 4.5 hours to induce effector production and secretion. Pellet was then suspended in laemmli

439 buffer at 95 degrees Celsius for three minutes to lyse the cells. Freshly boiled samples were

440 then loaded onto a 4-6% gradient gel and run for several hours to ensure sufficient separation

441 of the different sized TAL effectors. Polyclonal rbTal10 antibody was used to visualize all TALs.

442 **Gene Expression Analysis**

443 Susceptible cotton were inoculated with *Xcm* using a needleless syringe at an OD_{600} of

444 0.5. Infected and mock-treated tissue were collected and flash frozen at 24 and 48 hours post

445 inoculation. RNA was extracted using the Sigma tRNA kit. RNA-sequencing libraries were

446 generated as previously described [73].

447 Raw reads were trimmed using Trimmomatic [62]. The Tuxedo Suite was used for

448 mapping reads to the TM-1 NBI *Gossypium hirsutum* genome [74], assembling transcripts, and

449 quantifying differential expression [27, 75].

450 Homeologous pairs were identified based on syntenic regions with MCScan [76]. A

451 syntenic region is defined as a region with a minimum of five genes with an average intergenic

452 distance of 2 and within extended distance of 40. All other values are set to the default.

453 Bioinformatic prediction of TAL effector binding sites on the *G. hirsutum* promoterome

454 was performed using the TAL Effector-Nucleotide Targeter (TALEnt) [44]. In short, the regions of

455 the genome that were within 300 basepairs of annotated genes were queried with the RVD's of
456 MS14003 and AR81009 using a cutoff score of 4. Promiscuously binding TALs 16 from MS14003
457 and 16a from AR81009 were removed from analysis.

458 Fig Legends

459 **Fig 1: Cotton Bacterial Blight (CBB) symptoms and reemergence across the southern United**
460 **States.** (Left) Typical CBB symptoms present in cotton fields near Lubbock, TX during the 2015
461 growing season include angular leaf spots, boll rot, and black arm rot. Acres of cotton planted
462 per county in the United States in 2015 (blue) and counties with confirmed CBB in 2015 (red
463 outline). Statistics on cotton planted in the U.S. were acquired from the USDA. CBB was
464 reported by Extension agents, Extension specialists, and Certified Crop Advisers in their
465 respective states, and compiled by Tom Allen.

466 **Fig 2: Phylogenetic analysis of *Xcm* isolates and 13 species of *Xanthomonas*** A) MLST (Multi
467 Locus Sequence Typing) analysis of 12 Illumina sequenced *Xcm* isolates (this paper) and 40
468 other Xanthomonads using concatenated sections of the gltA, lepA, lacF, gyrB, fusA and gap-1
469 loci. B) SNP based Neighbor-Joining Tree generated from 17853 variable loci between 14 *Xcm*
470 isolates and the reference genome *Xanthomonas citri* subsp. *citri* strain Aw12879. The tree was
471 made using the Simple Phylogeny tool from ClustalW2.

472 **Fig 3: Molecular and phenotypic analysis of *Xcm* and *G. hirsutum* interactions.** A) Type three
473 effector profiles of *Xcm* isolates were deduced from de novo, Illumina based genome
474 assemblies. Effector presence absence was determined based on homology to known type
475 three effectors using the program Prokka. B) Commercial and public *G. hirsutum* cultivars were
476 inoculated with 14 *Xcm* isolates. Susceptible (S) indicates water soaking symptoms. Resistant
477 (R) indicates a visible hypersensitive response. Plants were screened with a range of inoculum
478 concentration from OD_{600} = 0.001-0.5. C) Disease symptoms on *G. hirsutum* cultivars Stoneville
479 5288 B2F and DES 56 after inoculation with *Xcm* strain AR81009 (OD_{600} = 0.05). Symptoms are
480 visualized under visible (VIS) and near infrared (NIR) light. D) The proportion of US fields
481 planted with susceptible and resistant cultivars of *G. hirsutum* was determined based on

482 planting acreage statistics from the USDA-AMA and disease phenotypes based on previous
483 reports for common cultivars [59-61].
484 **Fig 4: SMRT sequencing of two phenotypically and geographically diverse *Xcm* isolates:**
485 **MS14003 and AR81009.** Circos plot comparing the circular genomes. Tracks are as follows from
486 inside to outside: synteny of gene models; GC Content; Methylation on + and – strands;
487 location of type three effectors (teal) and TAL effectors (red). On each side, accompanying
488 plasmids are cartooned. Type three effector repertoires and the type IV secretion systems were
489 annotated using Prokka, homologous regions greater than 1kb were identified using MAUVE,
490 and TAL effectors were annotated using AnnoTALE.
491 **Fig 5: SMRT sequencing and western blot reveal diverse TAL effector repertoires between**
492 ***Xcm* strains MS14003 and AR81009.** A) Gene models of TAL-effectors identified by AnnoTALE.
493 Blue and Green highlighted gene models represent TALs grouped in the same clade by RVD
494 sequence using AnnoTALE. B) Western Blot of TAL effectors using polyclonal TAL-specific
495 antibody.
496 **Fig 6: RNA-Sequencing analysis of infected *G. hirsutum* tissue demonstrates transcriptional**
497 **changes during CBB.** A) Disease phenotypes of *Xcm* strains MS14003 and AR81009 on *G.*
498 *hirsutum* cultivars Acala Maxxa and DES 56, 7dpi. B) RNA-Seq Experimental Design: Acala Maxxa
499 and DES 56 were inoculated with *Xcm* strains MS14003 and AR81009 at an OD of 0.5 and a
500 mock treatment of 10mM $MgCl_2$. Inoculated leaf tissue was collected at 24 and 48 hpi (before
501 disease symptoms emerged). C) Venn diagram of upregulated *G. hirsutum* genes (Log2(fold
502 change in FPKM) ≥ 2 and p value ≤ 0.05) in response to *Xcm* inoculation. Cuffdiff output was
503 parsed using a custom script and visualized with the VennDiagram package in R.
504 **Table 1: RNA-Seq analysis reveals that 8 homeologous pairs of *G. hirsutum* genes are**
505 **upregulated in both Acala Maxxa and DES 56 cultivars 48 hours post inoculation with *Xcm***
506 **strains MS14003 and AR81009** at Log2(fold change in FPKM) ≥ 2 and p value ≤ 0.05).
507 Homeologous pairs were identified using genic synteny.
508 **Fig 7: Expression of homeologous pairs across the A and D *G. hirsutum* genomes in response**
509 **to *Xcm* inoculation.** Genes considered up or down regulated meet both differential expression
510 from mock significance of q-value < 0.05 and the absolute value of the log2 fold change is

511 greater than 2. A) Acala Maxxa inoculated with MS14003 B) DES 56 inoculated with MS14003 C)
512 Acala Maxxa inoculated with AR 81009 D) DES 56 inoculated with AR81009.
513 **Fig 8: Three candidate *G. hirsutum* susceptibility genes are targeted by two different *Xcm***
514 **strains.** (left) Bioinformatically predicted *Xcm* TAL Effector binding sites on the 300bp promoter
515 region of four SWEET genes. These were predicted with TALEsf using a quality score cutoff of 4.
516 (right) Heat-map of Cuffdiff results of significantly upregulated *G. hirsutum* SWEET genes ($p \leq$
517 0.05) with a Log2 (fold change in FPKM) $\geq$ 2, 48 hours after inoculation with *Xcm.*
518

519 ## References

520 1. Cotton: World Markets and Trade [Internet]. 2016 [cited August 2016]. Available from:
521 http://apps.fas.usda.gov/psdonline/circulars/cotton.pdf.
522 2. World Agricultural Supply and Demand Estimates [Internet]. 2016 [cited August 12, 2016].
523 Available from: http://usda.mannlib.cornell.edu/usda/current/wasde/wasde-08-12-2016.pdf.
524 3. Blank LM, Bird LS. Breeding Strains of Cotton Resistant to Bacterial Blight. Texas: Texas
525 Agricultural Experiment Station; 1951. p. 1-25.
526 4. Green JM, Brinkerhoff LA. Inheritance of Three Genes for Bacterial Blight Resistance in Upland
527 Cotton1. Agronomy Journal. 1956;48(11):481-5. doi: 10.2134/agronj1956.00021962004800110001x.
528 5. Simpson D, Weindling R. Bacterial blight resistance in a strain of Stoneville cotton. Journal of the
529 American Society of Agronomy. 1946.
530 6. Huang X, Zhai J, Luo Y, Rudolph K. Identification of a highly virulent strain of Xanthomonas
531 axonopodis pv. malvacearum. European Journal of Plant Pathology. 2008;122(4):461-9. doi:
532 10.1007/s10658-008-9312-3.
533 7. Cunnac S, Bolot S, Forero Serna N, Ortiz E, Szurek B, Noel LD, et al. High-Quality Draft Genome
534 Sequences of Two Xanthomonas citri pv. malvacearum Strains. Genome Announc. 2013;1(4). doi:
535 10.1128/genomeA.00674-13. PubMed PMID: 23990578; PubMed Central PMCID: PMCPMC3757453.
536 8. Delannoy E, Lyon BR, Marmey P, Jalloul A, Daniel JF, Montillet JL, et al. Resistance of cotton
537 towards Xanthomonas campestris pv. malvacearum. Annu Rev Phytopathol. 2005;43:63-82. doi:
538 10.1146/annurev.phyto.43.040204.140251. PubMed PMID: 16078877.
539 9. Brinkerhoff LA. Variation in Xanthomonas malvacearum and its relation to control. Annu Rev
540 Phytopathol. 1970;8:85-110.
541 10. Verma JP. Bacterial Blight of Cotton. CRC Press1986.
542 11. R.E. Hunter LAB, and L.S. Bird. The Development of a Set of Upland Cotton Lines for
543 Differentiating Races of Xanthomonas malvacearum. Phytopathology. 1968;58:830-2.
544 12. Baldwin Jr C, editor Report of the Bacterial Blight Committee--1975 [Cotton, USA]. Proceedings
545 Beltwide Cotton Production Research Conferences; 1976.
546 13. Schwartz AR, Potnis N, Timilsina S, Wilson M, PatanÃ© J, Martins J, et al. Phylogenomics of
547 Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and
548 identifies determinants of host specificity. Frontiers in Microbiology. 2015;6. doi:
549 10.3389/fmicb.2015.00535.
550 14. Jacques MA, Arlat M, Boulanger A, Boureau T, Carrere S, Cesbron S, et al. Using Ecology,
551 Physiology, and Genomics to Understand Host Specificity in Xanthomonas. Annu Rev Phytopathol.
552 2016;54:163-87. doi: 10.1146/annurev-phyto-080615-100147. PubMed PMID: 27296145.

553 15. Macho AP, Zipfel C. Targeting of plant pattern recognition receptor-triggered immunity by
554 bacterial type-III secretion system effectors. Current Opinion in Microbiology. 2015;23:14-22. doi:
555 http://dx.doi.org/10.1016/j.mib.2014.10.009.
556 16. Kim J-G, Stork W, Mudgett Mary B. Xanthomonas Type III Effector XopD Desumoylates Tomato
557 Transcription Factor SlERF4 to Suppress Ethylene Responses and Promote Pathogen Growth. Cell Host &
558 Microbe. 2013;13(2):143-54. doi: http://dx.doi.org/10.1016/j.chom.2013.01.006.
559 17. Wang S, Sun J, Fan F, Tan Z, Zou Y, Lu D. A Xanthomonas oryzae pv. oryzae effector, XopR,
560 associates with receptor-like cytoplasmic kinases and suppresses PAMP-triggered stomatal closure.
561 Science China Life Sciences. 2016:1-8. doi: 10.1007/s11427-016-5106-6.
562 18. Doyle EL, Stoddard BL, Voytas DF, Bogdanove AJ. TAL effectors: highly adaptable phytobacterial
563 virulence factors and readily engineered DNA-targeting proteins. Trends Cell Biol. 2013;23(8):390-8. doi:
564 10.1016/j.tcb.2013.04.003. PubMed PMID: 23707478; PubMed Central PMCID: PMCPMC3729746.
565 19. Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B. Five phylogenetically close rice SWEET
566 genes confer TAL effector-mediated susceptibility to Xanthomonas oryzae pv. oryzae. New Phytol.
567 2013;200(3):808-19. doi: 10.1111/nph.12411. PubMed PMID: 23879865.
568 20. Schreiber KJ, Baudin M, Hassan JA, Lewis JD. Die another day: Molecular mechanisms of
569 effector-triggered immunity elicited by type III secreted effector proteins. Seminars in Cell &
570 Developmental Biology. 2016;56:124-33. doi: http://dx.doi.org/10.1016/j.semcdb.2016.05.001.
571 21. Lee HA, Yeom SI. Plant NB-LRR proteins: tightly regulated sensors in a complex manner. Brief
572 Funct Genomics. 2015;14(4):233-42. doi: 10.1093/bfgp/elv012. PubMed PMID: 25825425.
573 22. Flor HH. Current Status of the Gene-for-Gene Concept. Annu Rev Phytopathol. 1971;9:275-96.
574 23. Cui H, Tsuda K, Parker JE. Effector-Triggered Immunity: From Pathogen Perception to Robust
575 Defense. Annual Review of Plant Biology. 2015;66(1):487-511. doi: 10.1146/annurev-arplant-050213-
576 040012.
577 24. Essenberg M, Bayles MB, Pierce ML, Verhalen LM. Pyramiding B genes in cotton achieves
578 broader but not always higher resistance to bacterial blight. Phytopathology. 2014;104(10):1088-97. doi:
579 10.1094/PHYTO-06-13-0167-R. PubMed PMID: 24655289.
580 25. Rothrock CS, Kirkpatrick TL, Barber T, Coker CM, Smith SE, editors. The resurgence of bacterial
581 blight of cotton in Arkansas. Beltwide Cotton Conferences; 2012 January 3-6; Orlando, Florida.
582 26. Koch R. Untersuchungen ueber Bakterien V. Die Aetiologie der Milzbrand-Krankheit,
583 begruendent auf die Entwicklungsgeschichte des Bacillus Anthracis. Beitr. z. Biol. D. Pflanzen 2: 277-310.
584 Milestones in Microbiology. 1876;1556.
585 27. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript
586 expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-78.
587 doi: 10.1038/nprot.2012.016. PubMed PMID: 22383036; PubMed Central PMCID: PMCPMC3334321.
588 28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
589 format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.
590 29. Bart R, Cohn M, Kassen A, McCallum EJ, Shybut M, Petriello A, et al. High-throughput genomic
591 sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable
592 resistance. Proceedings of the National Academy of Sciences. 2012;109(32):13130-. doi:
593 10.1073/pnas.1211014109.
594 30. Dever JK, Wheeler TA, Boman RK, Kerns D, Foster M, Nesmith D, et al. Cotton Performance Tests
595 in the Texas High Plains and Trans Pecos Areas of Texas --2009. Texas AgriLife Research and Extension
596 Service Center. 2010;10(2):1-74.
597 31. Dever JK, Morgan V, Kelly CM, Wheeler TA, Elkins H, Mendoza V, et al. Cotton Performance Tests
598 in the Texas High Plains, 2014. Texas AgriLife Research and Extension Service Center. 2015;15(1).
599 32. Organization CSaIR, inventorCotton Variety Fibermax 9892004 March 3, 2004.

600 33. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and
601 population structure in the US Upland cotton (Gossypium hirsutum L.). Theor Appl Genet.
602 2014;127(2):283-95. doi: 10.1007/s00122-013-2217-3. PubMed PMID: 24170350.
603 34. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain,
604 Loss and Rearrangement. PLOS ONE. 2010;5(6):e11147. doi: 10.1371/journal.pone.0011147.
605 35. Grau J, Reschke M, Erkes A, Streubel J, Morgan RD, Wilson GG, et al. AnnoTALE: bioinformatics
606 tools for identification, annotation, and nomenclature of TALEs from Xanthomonas genomic sequences.
607 Sci Rep. 2016;6:21077. doi: 10.1038/srep21077. PubMed PMID: 26876161; PubMed Central PMCID:
608 PMCPMC4753510.
609 36. Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, et al. The Barley Mlo
610 Gene: A Novel Control Element of Plant Pathogen Resistance. Cell. 1997;88(5):695-705. doi:
611 http://dx.doi.org/10.1016/S0092-8674(00)81912-1.
612 37. Jarosch B, Kogel K-H, Schaffrath U. The Ambivalence of the Barley Mlo Locus: Mutations
613 Conferring Resistance Against Powdery Mildew (Blumeria graminis f. sp. hordei) Enhance Susceptibility
614 to the Rice Blast Fungus Magnaporthe grisea. Molecular Plant-Microbe Interactions. 1999;12(6):508-14.
615 doi: 10.1094/MPMI.1999.12.6.508.
616 38. Jørgensen IH. Discovery, characterization and exploitation of Mlo powdery mildew resistance in
617 barley. Euphytica. 1992;63(1):141-52. doi: 10.1007/BF00023919.
618 39. Kim DS, Hwang BK. The pepper MLO gene, CaMLO2, is involved in the susceptibility cell-death
619 response and bacterial and oomycete proliferation. The Plant Journal. 2012;72(5):843-55. doi:
620 10.1111/tpj.12003.
621 40. Dong S, Adams KL. Differential contributions to the transcriptome of duplicated genes in
622 response to abiotic stresses in natural and synthetic polyploids. New Phytol. 2011;190(4):1045-57. doi:
623 10.1111/j.1469-8137.2011.03650.x. PubMed PMID: 21361962.
624 41. Chen LQ. SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytol.
625 2014;201(4):1150-5. PubMed PMID: 24649486.
626 42. Verdier V, Triplett LR, Hummel AW, Corral R, Cernadas RA, Schmidt CL, et al. Transcription
627 activator-like (TAL) effectors targeting OsSWEET genes enhance virulence on diverse rice (Oryza sativa)
628 varieties when expressed individually in a TAL effector-deficient strain of Xanthomonas oryzae. The New
629 phytologist. 2012;196(4):1197-207. doi: 10.1111/j.1469-8137.2012.04367.x. PubMed PMID: 23078195.
630 43. Cohn M, Bart RS, Shybut M, Dahlbeck D, Gomez M, Morbitzer R, et al. Xanthomonas axonopodis
631 virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar
632 transporter in cassava. Mol Plant Microbe Interact. 2014;27(11):1186-98. doi: 10.1094/MPMI-06-14-
633 0161-R. PubMed PMID: 25083909.
634 44. Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, VanDyk JK, et al. TAL Effector-
635 Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids
636 Research. 2012;40(W1):W117-W22.
637 45. Allen T. ALERT Bacterial Blight of Cotton 2011 [updated July 15, 2011]. Available from:
638 http://www.mississippi-crops.com/2011/07/15/alert-bacterial-blight-of-cotton/.
639 46. Aaron S. Alexander JEW, Randal K. Boman, Terry A. Wheeler, Norman W. Hopper. Effect of the
640 Easiflo Cottonseed Processing Method on Recovery of Xanthomonas axonopodis pv. malvacearum. The
641 Texas Journal of Agriculture and Natural Resource. 2012;25:13-23.
642 47. Allen TW. Epidemiology of Bacterial Blight of Cotton: 2011 to 2015. 2016 Cotton Incorporated -
643 Bacterial Blight Review. 2016.
644 48. Wheeler JWaT. Bacterial Blight Update for Cotton on the High Plains 2015. Available from:
645 http://plantclinic.tamu.edu/files/2015/08/WoodwardWheeler2015.pdf.

646 49. Kemerait B. Cotton bacterial blight is back and this is what you need to know 2016 [updated
647 August 10, 2016]. Available from: http://southeastfarmpress.com/cotton/cotton-bacterial-blight-back-
648 and-what-you-need-know.
649 50. Li T, Liu B, Spalding MH, Weeks DP, Yang B. High-efficiency TALEN-based gene editing produces
650 disease-resistant rice. Nature Biotechnology. 2012;30(5):390-2. doi: 10.1038/nbt.2199. PubMed PMID:
651 22565958.
652 51. Jia H, Orbovic V, Jones JB, Wang N. Modification of the PthA4 effector binding elements in Type I
653 CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating
654 XccDeltapthA4:dCsLOB1.3 infection. Plant Biotechnol J. 2016;14(5):1291-301. doi: 10.1111/pbi.12495.
655 PubMed PMID: 27071672.
656 52. Strauß T, van Poecke RMP, Strauß A, Römer P, Minsavage GV, Singh S, et al. RNA-seq pinpoints a
657 Xanthomonas TAL-effector activated resistance gene in a large-crop genome. Proceedings of the
658 National Academy of Sciences. 2012;109(47):19480-5.
659 53. Yang B, Sugio A, White FF. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice.
660 Proceedings of the National Academy of Sciences. 2006;103(27):10503-8.
661 54. Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, et al. Promoter mutations of an essential gene for
662 pollen development result in disease resistance in rice. Genes & Development. 2006;20(10):1250-5. doi:
663 10.1101/gad.1416306.
664 55. Yoo M-J, Wendel JF. Comparative Evolutionary and Developmental Dynamics of the Cotton
665 (*Gossypium hirsutum*) Fiber Transcriptome. PLoS Genet. 2014;10(1):e1004073. doi:
666 10.1371/journal.pgen.1004073.
667 56. Janga MR, Campbell LM, Rathore KS. CRISPR/Cas9-mediated targeted mutagenesis in upland
668 cotton (Gossypium hirsutum L.). Plant Molecular Biology. 2017:1-12. doi: 10.1007/s11103-017-0599-3.
669 57. Li C, Unver T, Zhang B. A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton
670 (Gossypium hirsutum L.). Sci Rep. 2017;7:43902. doi: 10.1038/srep43902. PubMed PMID: 28256588;
671 PubMed Central PMCID: PMCPMC5335549.
672 58. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for
673 phylogenetic study. Journal of bacteriology. 1991;173(2):697-703. PubMed PMID: 1987160.
674 59. Wheeler TA, Woodward JE. Response of cotton varieties to bacterial blight race 18 ain 2016.
675 Texas AgriLife Research and Extension Service Center. 2016. doi:
676 Lubbock.tamu.edu/files/2016/12/Bacterial-blight-trials-2016-2.pdf.
677 60. Wheeler TA, Woodward JE. Response of cotton varieties to dieases on the Texas High Plains.
678 Texas AgriLife Research and Extension Service Center. 2012. doi: Lubbock.tamu.edu/files/2013/02/2012-
679 cotton-disease-ratings.pdf.
680 61. Wheeler TA, Woodward JE. Response of cotton varieties to diseases on the Southern High Plains
681 of Texas, 2010. Texas AgriLife Research and Extension Service Center. 2010. doi:
682 Lubbock.tamu.edu/files/2011/11/DiseaseRecommendations.pdf.
683 62. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
684 Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404;
685 PubMed Central PMCID: PMCPMC4103590.
686 63. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New
687 Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational
688 Biology. 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021.
689 64. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
690 65. Almeida NF, Yan S, Cai R, Clarke CR, Morris CE, Schaad NW, et al. PAMDB, a multilocus sequence
691 typing and analysis database and website for plant-associated microbes. Phytopathology.
692 2010;100(3):208-15. doi: 10.1094/PHYTO-100-3-0208. PubMed PMID: 20128693.

66. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8. doi: 10.1093/bioinformatics/btr330.
67. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. doi: 10.1093/bioinformatics/btq033.
68. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8. doi: 10.1093/bioinformatics/btm404.
69. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;4(4):406-25. doi: 10.1093/oxfordjournals.molbev.a040454.
70. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Meth. 2016;13(12):1050-4. doi: 10.1038/nmeth.4035

http://www.nature.com/nmeth/journal/v13/n12/abs/nmeth.4035.html#supplementary-information.
71. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. doi: 10.1186/1471-2105-10-421. PubMed PMID: 20003500; PubMed Central PMCID: PMCPMC2803857.
72. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45. doi: 10.1101/gr.092759.109. PubMed PMID: 19541911; PubMed Central PMCID: PMCPMC2752132.
73. Wilson MC, Mutka AM, Hummel AW, Berry J, Chauhan RD, Vijayaraghavan A, et al. Gene expression atlas for the food security crop cassava. New Phytol. 2017;213(4). doi: 10.1111/nph.14443. PubMed PMID: 28116755.
74. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531-7. doi: 10.1038/nbt.3207. PubMed PMID: 25893781.
75. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524-30. doi: 10.1038/nbt.3208. PubMed PMID: 25893780.
76. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and Collinearity in Plant Genomes. Science. 2008;320(5875):486.

Supporting Information

S1 Table: Illumina and SMRT sequenced *Xcm* genomes described in this paper.

S2 Table: *Xanthomonas* genomes previously deposited on NCBI that are referenced in this paper.

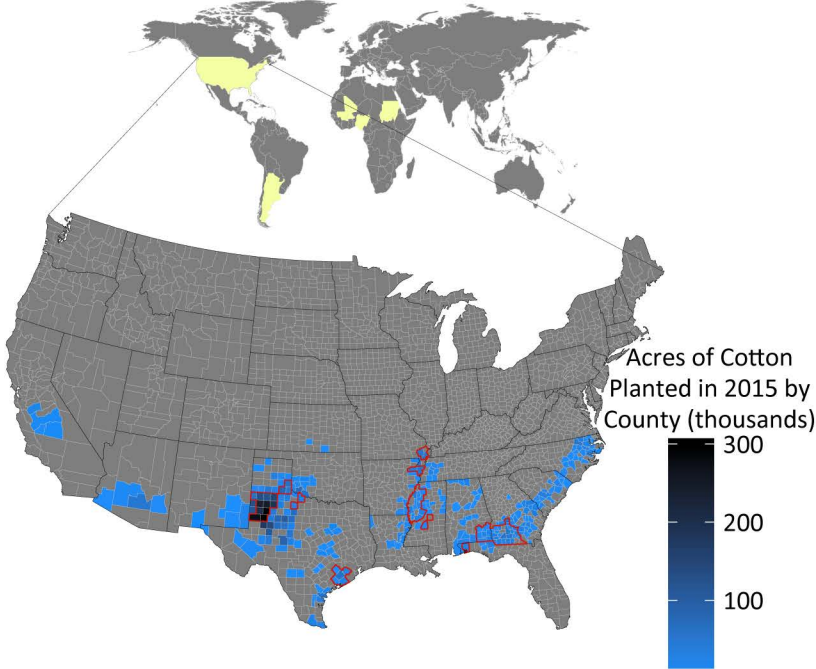
S3 Table: RNA-Seq analysis reveals that 52 genes are induced in all *Xcm-G. hirsutum* interactions at 48 hours (($p \leq 0.05$) with a Log2 (fold change in FPKM) $\geq$ 2).

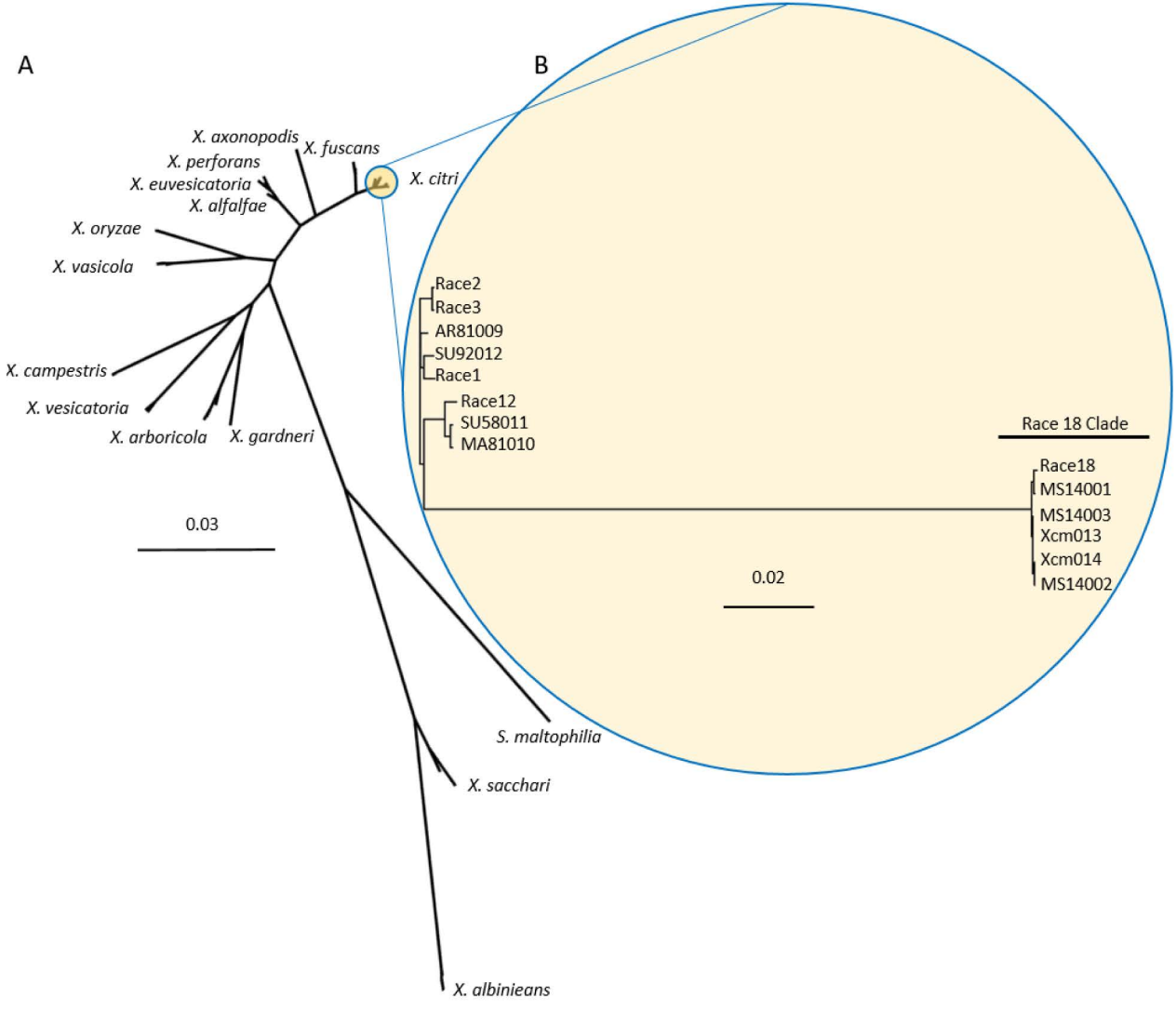
S1 Fig: Maps of CBB incidence in the US from 2011-2012 and 2014-2016. CBB incidence was reported by farmers, Extension specialists and Certified Crop Advisers in their respective states for the years 2011-2012 and 2014-2016, and compiled by Tom Allen. CBB reports for 2013 were infrequent.

734 **S2 Fig: Growth assay of MS14003 and AR81009 on cotton cultivars Acala Maxxa and DES 56.**

735 **S3 Fig: Expression levels of significantly upregulated genes with a Log2 fold change of 2 in *G.***
736 ***hirsutum*** A) All significantly upregulated genes with a Log2 fold change of 2 B) All significantly
737 upregulated genes ($p \leq 0.05$) with a Log2 (fold change in FPKM) ≥ 2 that are unique to each
738 cultivar/pathovar disease interaction in *G. hirsutum.*

739


Acres of Cotton
Planted in 2015 by
County (thousands)
300
200
100


A
B
X. axonopodis
X. fuscans
X. perforans
X. euvesicatoria
X. alfalfae
X. citri
X. oryzae
X. vasicola
X. campestris
X. vesicatoria
X. arboricola
X. gardneri
0.03
S. maltophilia
X. sacchari
X. albinieans
Race2
Race3
AR81009
SU92012
Race1
Race12
SU58011
MA81010
Race 18 Clade
Race18
MS14001
MS14003
Xcm013
Xcm014
MS14002
0.02

A

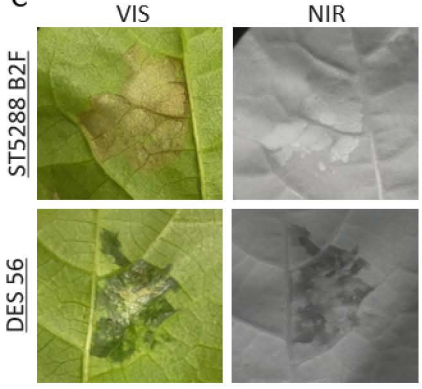
Type Three Effectors

+ Effector Present; • Pseudogene; - Effector Absent	24 Conserved	XopAK	XopC	HopAJ2	XopE2	XopJ	XopAI	HopAJ1	XopP	XopAO	TAL Effectors
SU58011	+	+	+	+	•	-	-	-	•	-	+
SU92012	+	+	+	+	•	-	-	•	•	-	+
Race 3	+	+	+	+	•	-	-	•	•	-	+
Race 12	+	+	+	+	•	-	-	•	-	-	+
MA81010	+	+	+	+	•	-	-	•	+	-	+
Xcm004	+	+	+	•	•	-	-	-	+	-	+
AR81009	+	•	+	+	+	-	-	•	+	-	+
SU44	+	+	•	+	•	-	+	•	+	+	+
Race 2	+	+	+	+	+	-	-	•	+	-	+
BF1	+	+	+	+	+	-	-	•	+	-	+
Race 18: NI86	+	+	+	+	•	+	+	+	•	-	+
Race 18: MS14002	+	+	+	+	+	+	+	+	+	-	+
Race 18: Race 18	+	+	+	+	+	+	+	+	+	-	+
Race 18: Xcm014	+	+	+	+	+	+	+	+	+	-	+
Race 18: MS14003	+	+	+	+	+	+	+	+	+	-	+
Race 18: BF2	+	+	+	+	+	+	+	+	+	-	+

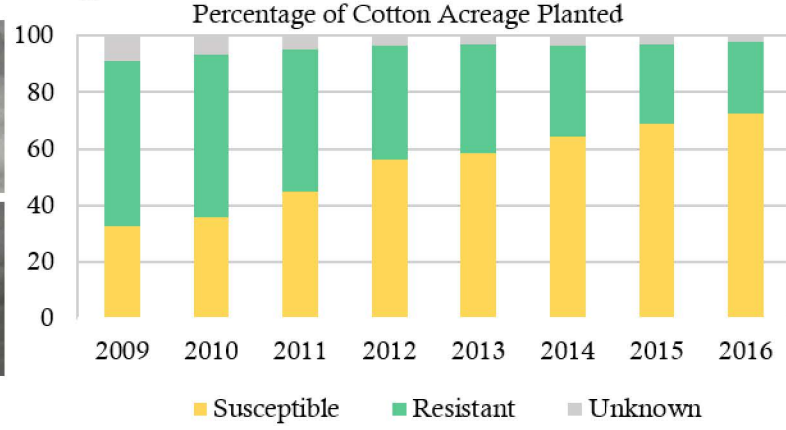
B

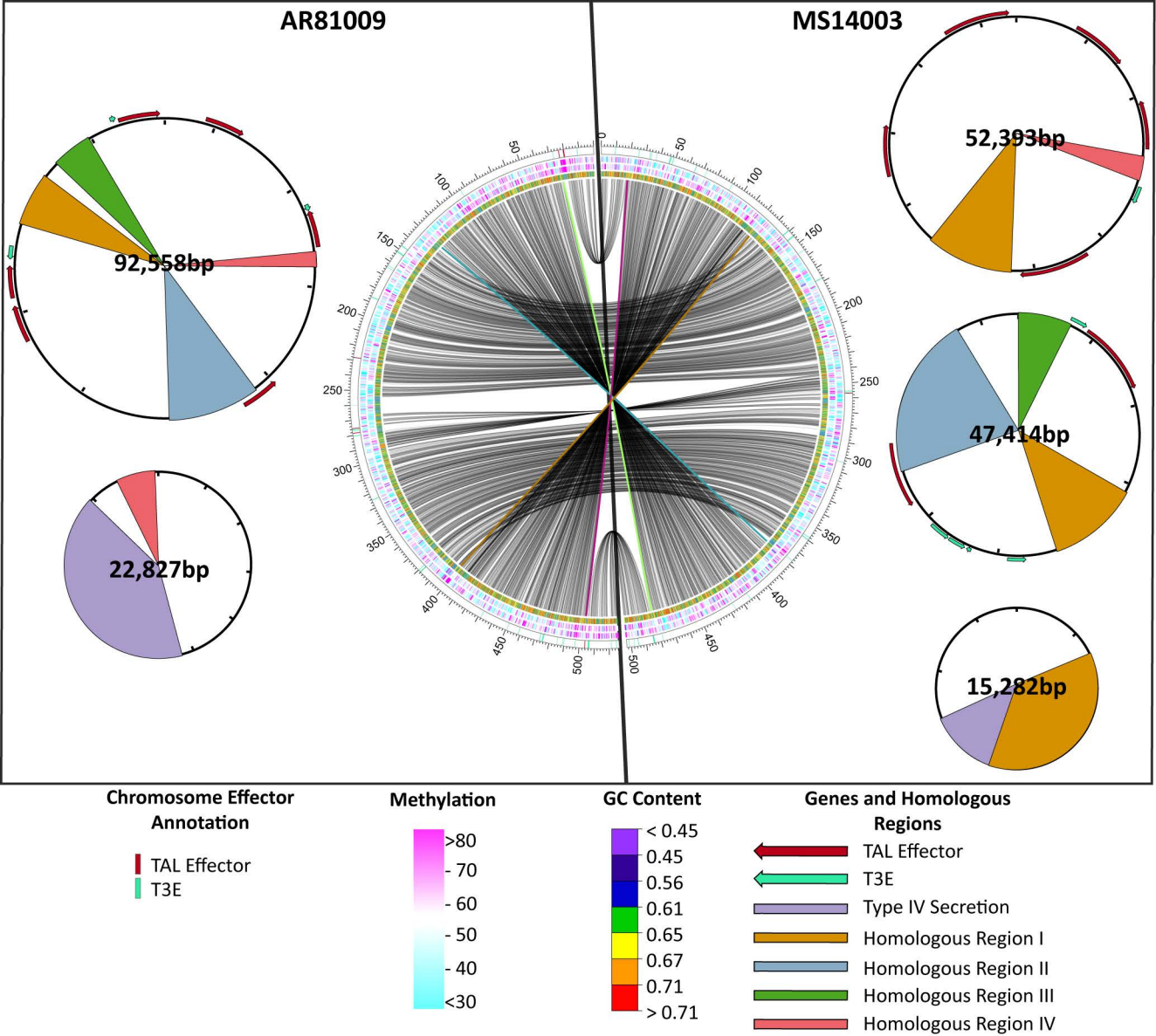
	Variety	R/S
Commercial Varieties	Fibermax 989	R
Commercial Varieties	Fibermax 1830 GLT	R
Commercial Varieties	Fibermax 2334 GLT	R
Commercial Varieties	Fibermax 2484 B2F	R
Commercial Varieties	Deltapine 1133 B2RF	R
Commercial Varieties	Stoneville 5288 B2F	R
Commercial Varieties	Deltapine 0912 B2RF	S
Commercial Varieties	Deltapine 1028 B2RF	S
Commercial Varieties	Deltapine 1034 B2RF	S
Commercial Varieties	Deltapine 1048 B2RF	S
Commercial Varieties	PhytoGen 499 WRF	S
Commercial Varieties	Stoneville 4946 GLB2	S
CBB Panel	Acala-44	S
CBB Panel	Gregg	S
CBB Panel	Mebane	S
CBB Panel	Stoneville 2B	S
CBB Panel	Stoneville 20	S
NAM	Acala-Maxxa	S
NAM	DES 56	S

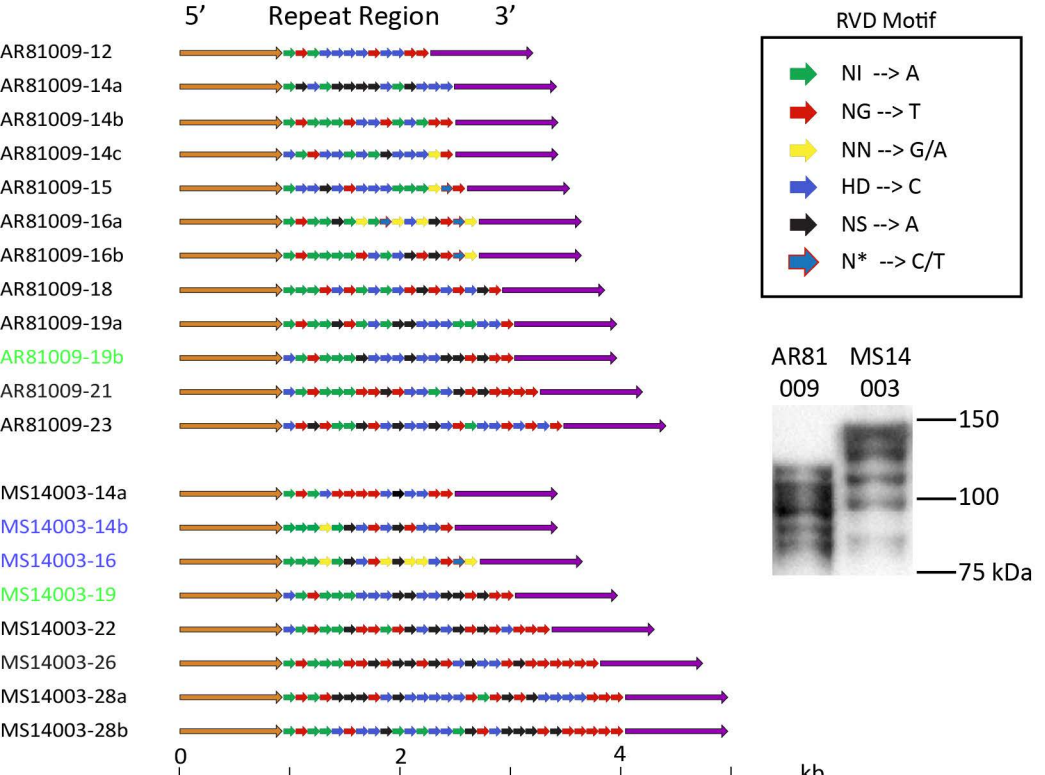
C

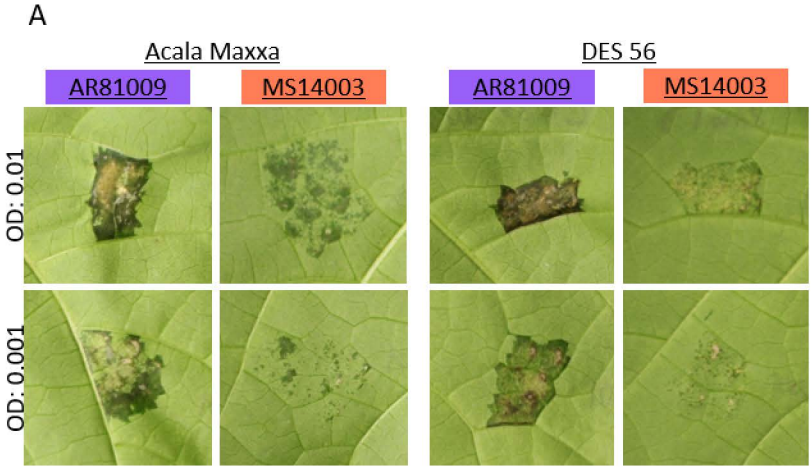


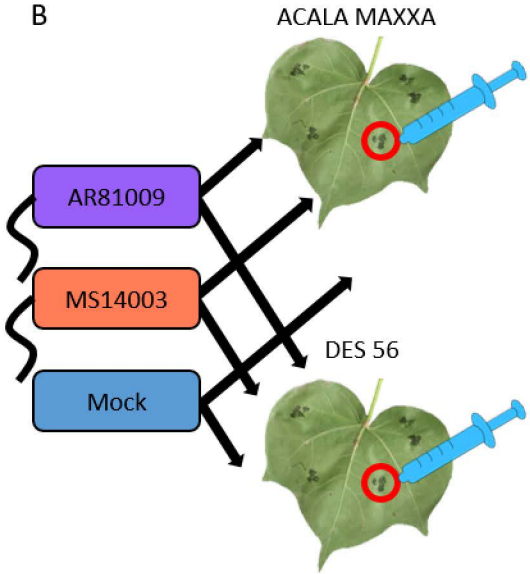
D

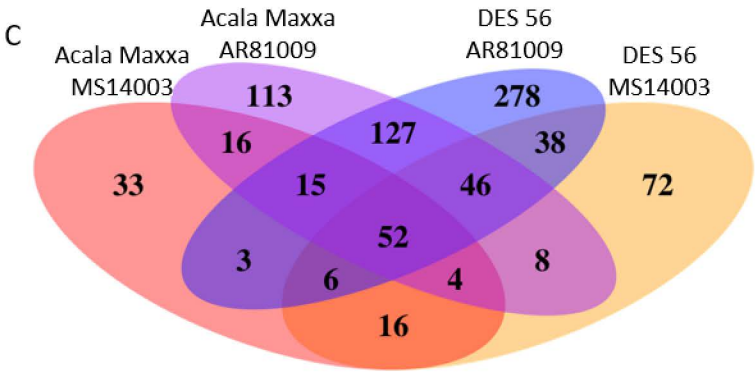


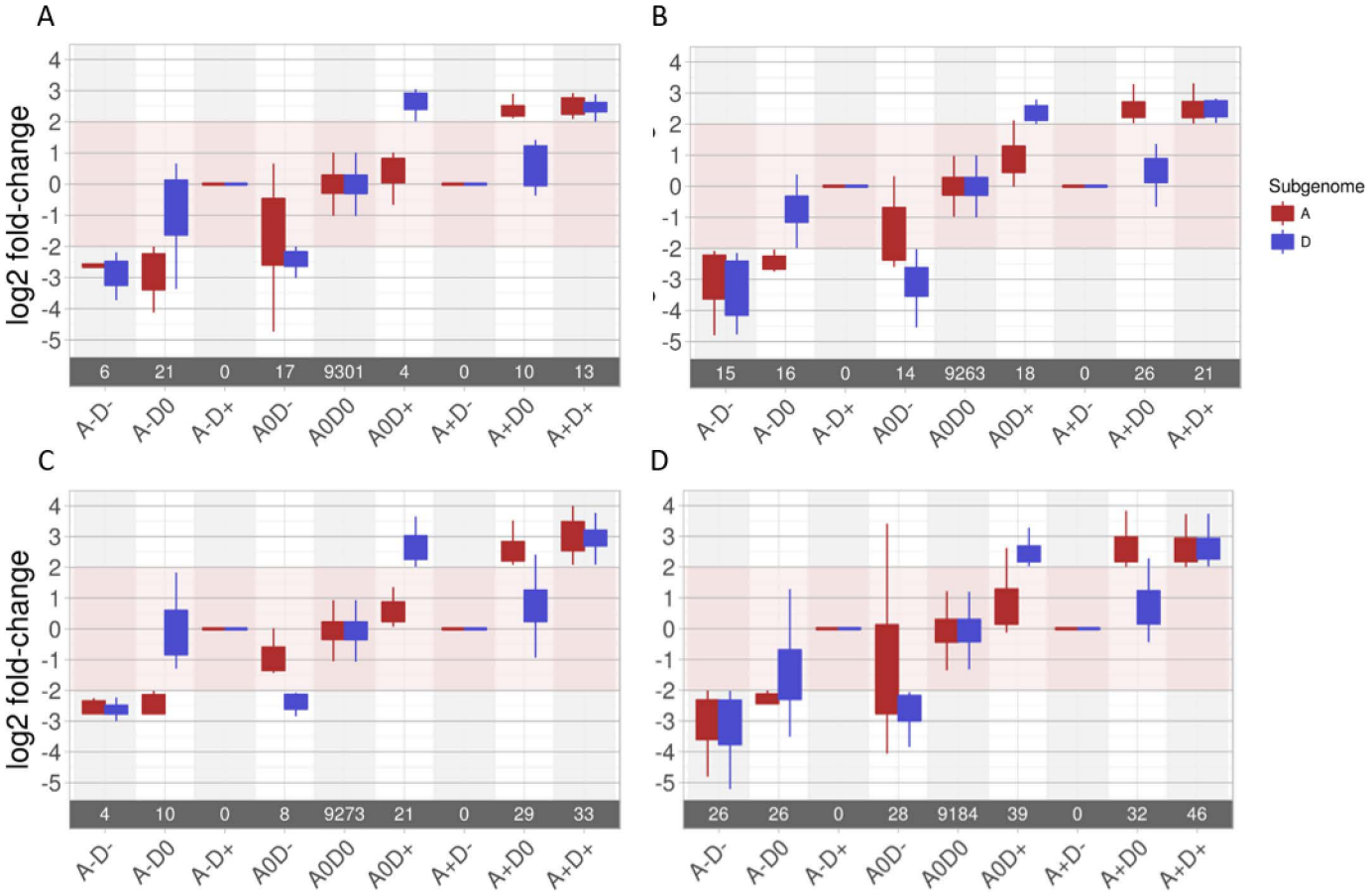

AR81009
MS14003
92,558bp
22,827bp
52,303bp
47,414bp
15,282bp
Chromosome Effector Annotation
TAL Effector
T3E
Methylation
>80
- 70
- 60
- 50
- 40
<30
GC Content
< 0.45
0.45
0.56
0.61
0.65
0.67
0.71
> 0.71
Genes and Homologous Regions
TAL Effector
T3E
Type IV Secretion
Homologous Region I
Homologous Region II
Homologous Region III
Homologous Region IV

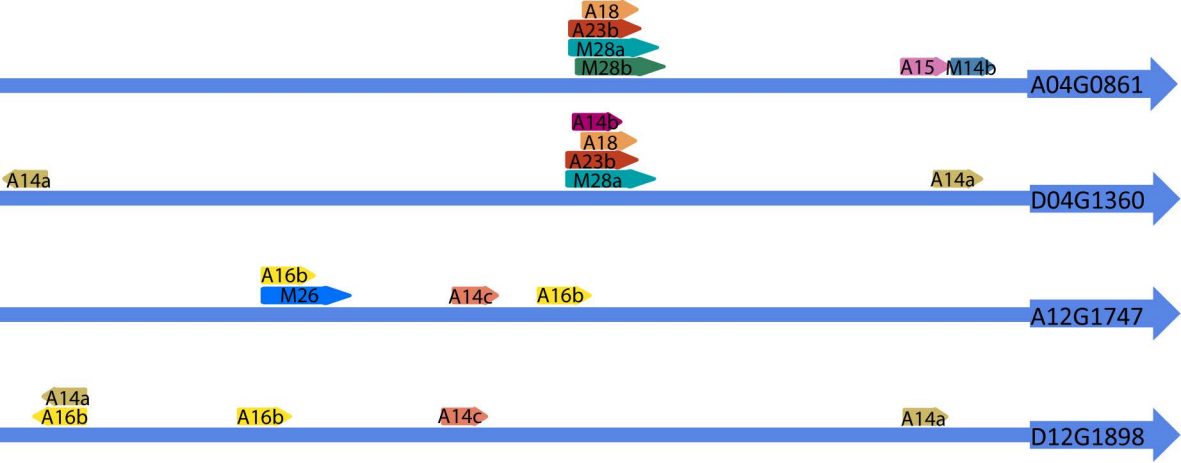

5'
Repeat Region
3'
AR81009-12
AR81009-14a
AR81009-14b
AR81009-14c
AR81009-15
AR81009-16a
AR81009-16b
AR81009-18
AR81009-19a
AR81009-19b
AR81009-21
AR81009-23
MS14003-14a
MS14003-14b
MS14003-16
MS14003-19
MS14003-22
MS14003-26
MS14003-28a
MS14003-28b
0
2
4
kb
RVD Motif
NI --> A
NG --> T
NN --> G/A
HD --> C
NS --> A
N* --> C/T
AR81 009
MS14 003
150
100
75 kDa


A
Acala Maxxa
DES 56
AR81009
MS14003
AR81009
MS14003
OD: 0.01
OD: 0.001


B
ACALA MAXXA
AR81009
MS14003
Mock
DES 56


C
Acala Maxxa MS14003
Acala Maxxa AR81009
DES 56 AR81009
DES 56 MS14003
113
278
16
127
38
33
15
46
72
52
3
6
4
8
16


A
B
C
D
log2 fold-change
Subgenome
A
D
6 21 0 17 9301 4 0 10 13
15 16 0 14 9263 18 0 26 21
4 10 0 8 9273 21 0 29 33
26 26 0 28 9184 39 0 32 46
A-D- A-D0 A-D+ A0D- A0D0 A0D+ A+D- A+D0 A+D+

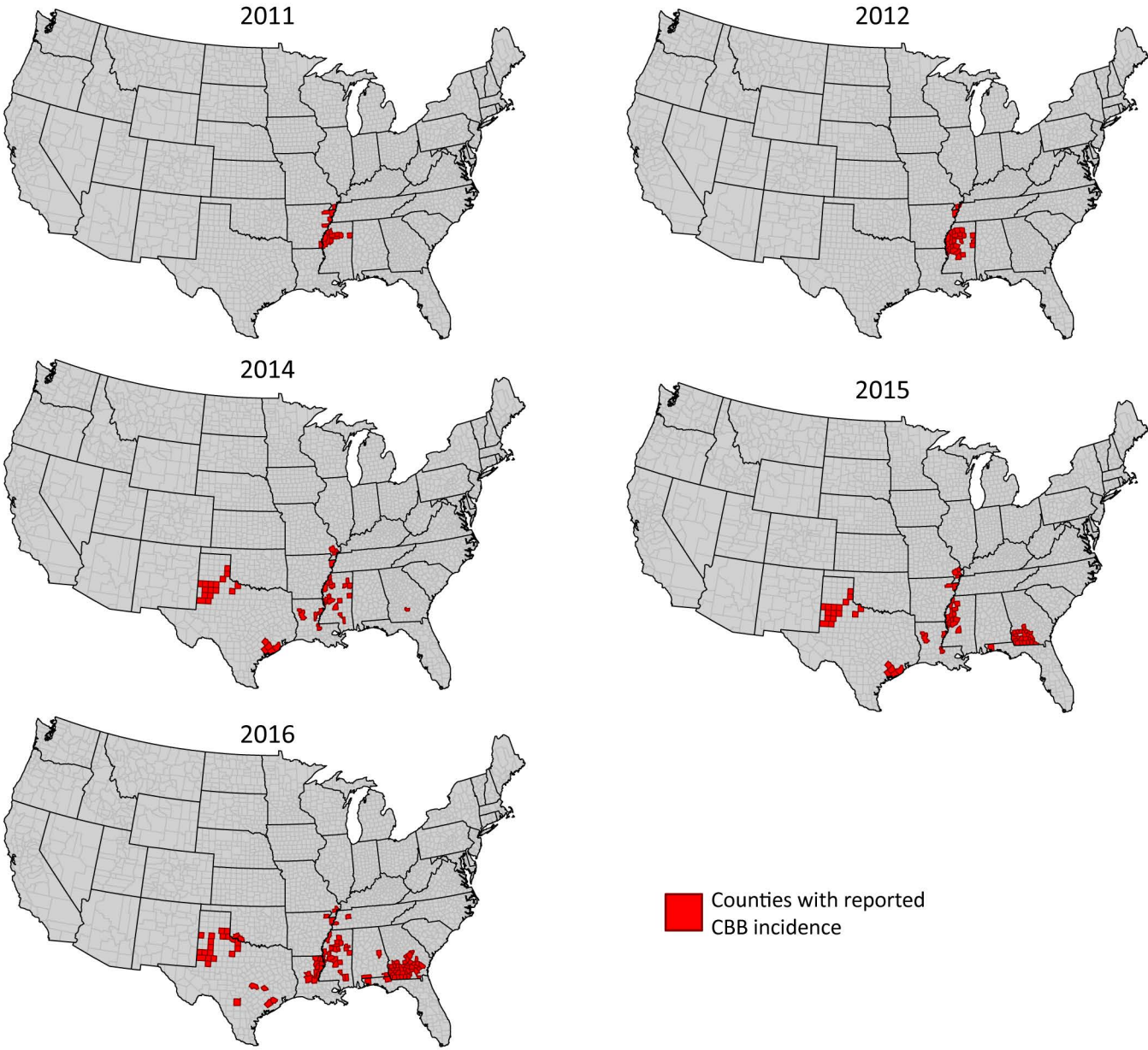


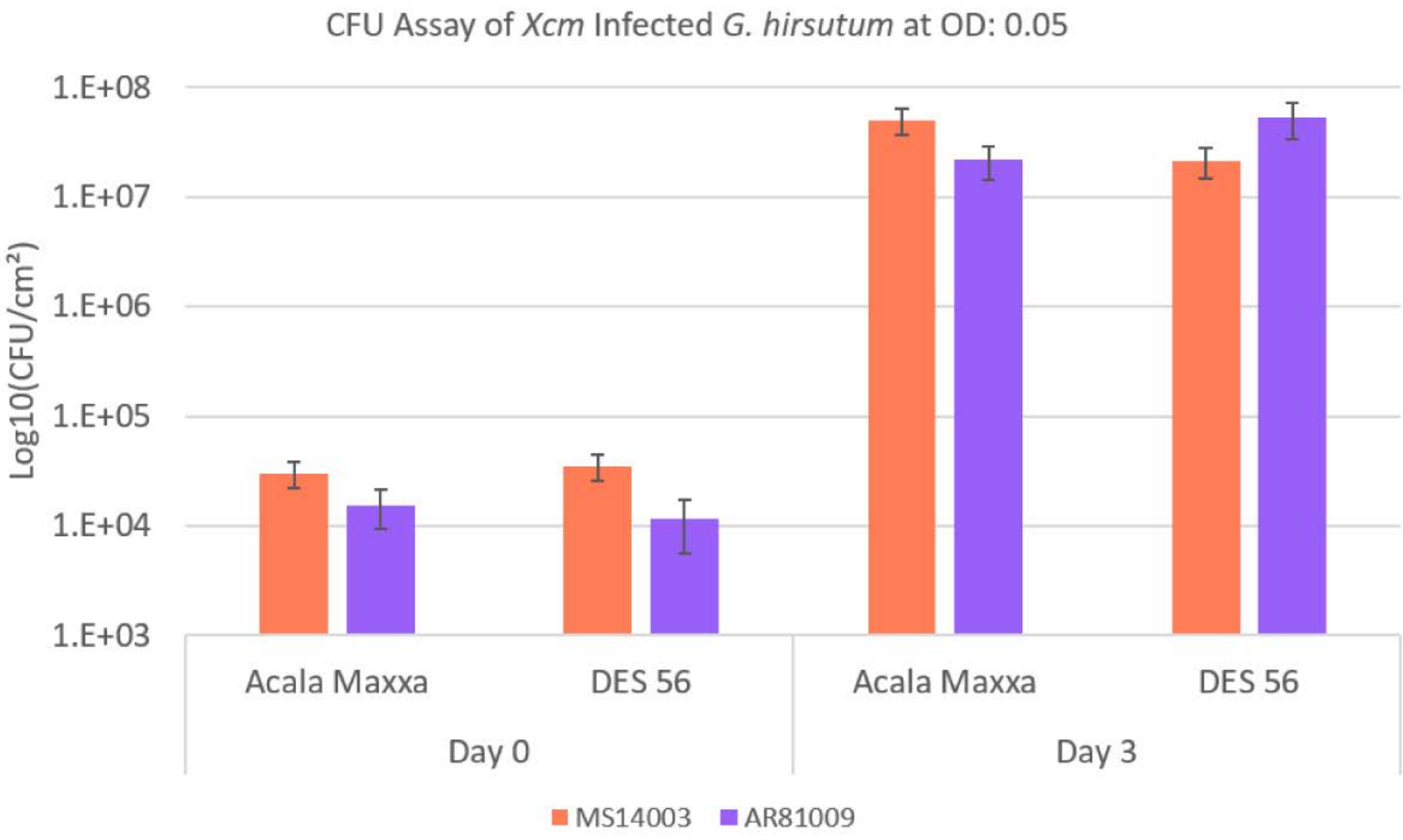
Expression of Selected SWEET Genes in Mock and Inoculated Cotton, 48 Hpi

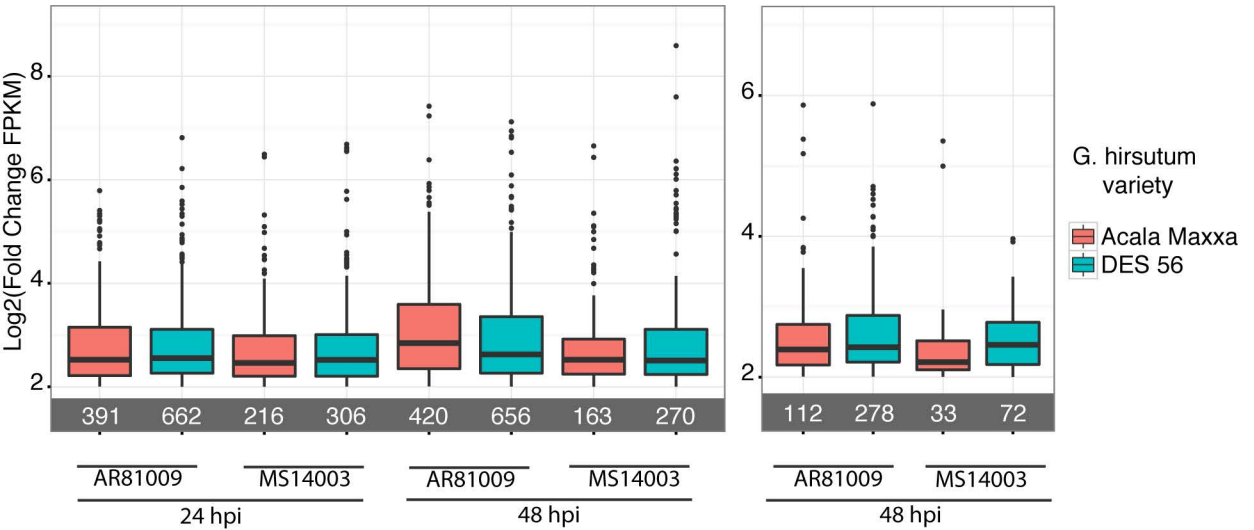
	Mock	MS14003	AR81009	Mock	MS14003	AR81009
	Acala Maxxa			DES 56		
A04G0861	11.83	403.75	2.61	10.73	468.54	10.47
D04G1360	1.88	190.09	1.60	0.73	279.72	5.52
A12G1747	117.23	102.12	144.52	92.05	74.32	277.98
D12G1898	93.21	77.79	888.87	69.40	63.66	850.18

Expression (FPKM)

800
600
400
200


2011
2012
2014
2015
2016
Counties with reported CBB incidence


CFU Assay of *Xcm* Infected *G. hirsutum* at OD: 0.05
1.E+08
1.E+07
1.E+06
1.E+05
1.E+04
1.E+03
Log10(CFU/cm²)
Acala Maxxa
DES 56
Acala Maxxa
DES 56
Day 0
Day 3
MS14003
AR81009


Log2(Fold Change FPKM)
8
6
4
2
391
662
216
306
420
656
163
270
AR81009
MS14003
AR81009
MS14003
24 hpi
48 hpi
6
4
2
112
278
33
72
AR81009
MS14003
48 hpi
G. hirsutum variety
Acala Maxxa
DES 56