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 resistant germplasm for over half a century. Recently, CBB re-emerged as an agronomic problem in the United States. Here, we report analysis of cotton variety planting statistics that indicate a steady increase in the percentage of susceptible cotton varieties grown each year since 2009. Phylogenetic analysis revealed that strains from the current outbreak cluster with race 18 *Xanthomonas citri* pv. *malvacearum* (*Xcm*) strains. Illumina based draft genomes were generated for thirteen *Xcm* isolates and analyzed along with 4 previously published *Xcm* genomes. These genomes encode 24 conserved and nine variable type three effectors. Strains in the race 18 clade contain 3 to 5 more effectors than other *Xcm* strains. SMRT sequencing of two geographically and temporally diverse strains of *Xcm* yielded circular chromosomes and accompanying plasmids. These genomes encode eight and thirteen distinct transcription activator-like effector genes. RNA-sequencing revealed 52 genes induced within two cotton cultivars by both tested *Xcm* strains. This gene list includes a homeologous pair of genes, with homology to the known susceptibility gene, MLO. In contrast, the two strains of *Xcm* induce different clade III SWEET sugar transporters. Subsequent genome wide analysis revealed patterns in the overall expression of homeologous gene pairs in cotton after inoculation by *Xcm*. These data reveal host-pathogen specificity at the genetic level 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 a 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 contributing factors for 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, including in Arkansas and Mississippi. Nonetheless,

CBB has received little research focus during the last several decades because, prior to 2011, losses from this disease were not substantial. Modern molecular and genomic technologies can now be employed expeditiously to deduce the underlying cause of the disease re-emergence and pinpoint optimized routes towards the development of durable resistance.

CBB is caused by *X. citri pv. malvacearum* (*Xcm*); however, the pathogen has previously been placed within other species groupings [6-9]. The *Xcm* pathovar can be further divided into at least 19 races according to virulence phenotypes on a panel of historical cotton cultivars: Acala-44, Stoneville 2B-S9, Stoneville 20, Mebane B-1, 1-10B, 20-3, and 101-102.B [10, 11]. Historically, the most common race observed in the U.S. has been race 18, which was first isolated in 1973 [12]. This race is highly virulent, causing disease on all cultivars in the panel except for 101-102.B. However, this diagnostic panel of cotton varieties used to race type strains is no longer available from the USDA/ARS, Germplasm Resources Information Network (GRIN).

CBB can occur at any stage in the plant's life cycle and on any aerial organ. Typical symptoms include seedling blight as either pre- or post-emergent damping-off, black arm on petioles and stems, water-soaked spots on leaves and bracts, and most importantly boll rot [10]. The most commonly observed symptoms are the angular-shaped lesions on leaves that can coalesce and result in a systemic infection. Disease at each of these stages can cause yield losses either by injury to the plant or direct damage to the boll. No effective chemical treatments for the disease have been released to date. Methods to reduce yield loss as a result of CBB include acid de-linting cotton seed prior to planting, field cultivation practices to reduce

sources of overwintering inoculum and planting cultivars with known sources of resistance [3, 4, 8, 13, 14].

Xanthomonads assemble the type three secretion system (T3SS), a needle-like structure, to inject diverse type three effectors (T3Es) into the plant cell to suppress immunity and promote disease [15-19]. For example, transcription activator-like (TAL) effectors influence the expression levels of host genes by binding directly to promoters in a sequence-specific way [20]. Up-regulated host genes that contribute to pathogen virulence are termed susceptibility genes and may be modified through genome editing for the development of resistant crop varieties [21].

Plants have specialized immune receptors, collectively known as nucleotide-binding leucine rich repeat receptors that recognize, either directly or indirectly, the pathogen effector molecules [22, 23]. Historically, this host-pathogen interaction has been termed the 'gene-for-gene' model of immunity, wherein a single gene from the host and a single gene from the pathogen are responsible for recognition [24]. Recognition triggers a strong immune response that often includes a localized hypersensitive response (HR) in which programmed cell death occurs around the infection site [25]. Nineteen CBB resistance loci have been reported in *Gossypium hirsutum* breeding programs; however, none have been molecularly identified [8, 13].

Here we combine comparative genomics of the pathogen *Xcm* with transcriptomics of the host to identify molecular determinants of Cotton Bacterial Blight. This will inform the development of durable resistance strategies.

Results

CBB Reemergence in the US

In 2011, farmers, extension specialists, and certified crop advisers in Missouri, Mississippi, and Arkansas observed cotton plants exhibiting symptoms of CBB. Widespread infected plant material was observed throughout much of the production area, but appeared to be centered around Clarksdale, Mississippi. In figure 1, we collate reports from this outbreak and overlay these data with US cotton planting statistics to reveal that this disease has spread through much of the cotton belt in the southern U.S. (Figs 1 and S1, Table S1). To date, CBB has been reported from at least eight out of the sixteen states that grow cotton (Fig 1). In 2014, we collected diseased cotton leaves from two sites across Mississippi and proved pathogen causality following Koch's postulates [26]. In addition, PCR amplification of the 16S rRNA gene confirmed that the causal agent was a member of the *Xanthomonas* genus. Multi locus sequence type (MLST) analysis and maximum-likelihood analysis were performed using concatenated sections of the *gltA*, *lepA*, *lacF*, *gyrB*, *fusA* and *gap-1* loci for increased phylogenetic resolution (Fig 2a). The newly sequenced strains were named MS14002 and MS14003 and were compared to four previously published *Xcm* genomes and thirty-six additional *Xanthomonas* genomes representing thirteen species (Tables 1, S2). MS14002 and MS14003 grouped with the previously published *Xcm* strains as a single unresolved clade, further confirming that the current disease outbreak is CBB and is caused by *Xcm*. The species designation reported here is consistent with previous reports [6, 7].

134 **Table 1:** Illumina and SMRT sequenced *Xcm* genomes described in this paper.

Strain Name	Identifier	Lat.	Long.	City/Country	Year	Provider	Collector	Platform	Contig #	Avg Contig Len	Total Bases	n50
MS14002		34.12	-90.52	Clarksdale, MS	2014	R.Bart	R.Bart	Illumina	545	9443.27	5146580	62542
MS14003		32.95	-90.51	Wilzone, MS	2014	R.Bart	R.Bart	Illumina	2577	1511.35	3894744	2209
Race1				US		S.lu	P.Thaxton	Illumina	523	10127.35	5296606	48599
Race2				US		S.lu	P.Thaxton	Illumina	387	13402.57	5186796	54804
Race3				US		S.lu	P.Thaxton	Illumina	725	7207.34	5225324	28344
Race12				US		S.lu	P.Thaxton	Illumina	632	8134.35	5140911	21428
Race18				US		S.lu	P.Thaxton	Illumina	369	13924.03	5137968	112543
AR81009	CFBP2035			Argentina	1981	CIRM-CFBP	Frossard P	Illumina	306	17182.59	5257872	86594
MA81010	CFBP2036			Mali	1981	CIRM-CFBP	Frossard P	Illumina	584	9033.09	5275326	23323
SU58011	CFBP2530			Sudan	1958	CIRM-CFBP	Last F.T. M2519	Illumina	1134	4563.33	5174819	9682
SU9012	CFBP5637			Sudan	1992	CIRM-CFBP	Schmit J. 11781	Illumina	377	13919.54	5247665	88522
Xcm013	MSCT4			US		S.lu		Illumina	2169	2151.5	4666607	3869
Xcm014	MSCT8			US		S.lu		Illumina	580	8929.58	5179156	88255
MS14003		32.95	-90.51	Wilzone, MS	2014	R.Bart	R.Bart	SMRT	4	1286176.5	5144706	5029617
AR81009	CFBP2035			Argentina	1981	CIRM-CFBP	Frossard P	SMRT	4	1352212	5408848	5267057

135

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

Race groups have been described for *Xcm* strains by analyzing compatible (susceptible) and incompatible (resistant) interactions on a panel of seven cotton cultivars. Different geographies often harbor different pathogen races [7]. Consequently, one possible explanation for the recent outbreak of CBB would be the introduction of a new race of *Xcm* capable of overcoming existing genetic resistance. Only 2 varieties of the original cotton panel plus three related cultivars, were available and these cultivars were not sufficient to determine whether a new race had established within the U.S. Thirteen *Xcm* strains were sequenced using Illumina technology to determine the phylogenetic relationship between recent isolates of *Xcm* and historical isolates. Isolates designated as race 1, race 2, race 3, race 12 and race 18 have been maintained at Mississippi State University with these designations. Additional isolates were obtained from the Collection Française de Bactéries associées aux Plantes (CFBP) culture collection. Together, these isolates include nine strains from the US, three from Africa, and one from South America and span collection dates ranging from 1958 through 2014 (Fig 1, Table 1). Illumina reads were mapped to the *Xanthomonas citri* subsp. *citri* strain Aw12879 (Genbank assembly accession: GCA_000349225.1) using Bowtie2 and single nucleotide polymorphisms (SNPs) were identified using Samtools [27, 28]. Only regions of the genome with at least 10x coverage for all genomes were considered. This approach identified 17,853 sites that were polymorphic in at least one genome. Nucleotides were concatenated and used to build a neighbor-joining tree (Fig 2b). This analysis revealed that recent U.S. *Xcm* isolates grouped with the race 18 clade. Notably, the race 18 clade is phylogenetically distant from the other *Xcm* isolates.

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

Xanthomonads deploy many classes of virulence factors to promote disease. Type three effectors (T3E) are of particular interest for their role in determining race designations. T3E profiles from sixteen *Xcm* isolates were compared to determine whether a change in the virulence protein arsenal of the newly isolated strains could explain the re-emergence of CBB. Genomes from 13 *Xcm* isolates were *de novo* assembled with SPAdes and annotated with Prokka based on annotations from the *X. euvesicatoria* (aka. *X. campestris* pv. *vesicatoria*) 85-10 genome (NCBI accession: NC_007508.1). T3Es pose a particular challenge for reference based annotation as no bacterial genome contains all effectors. Consequently, an additional protein file containing known T3Es from our previous work was included within the Prokka annotation pipeline [15, 29]. This analysis revealed 24 conserved and 9 variable *Xcm* T3Es (Fig 3a). Race 18 clade isolates contain more effectors than other isolates that were sequenced. The recent *Xcm* isolates (MS14002 and MS14003) were not distinguishable from the historical race 18 isolate, with the exception of XcmNI86 isolated from Nicaragua in 1986, which contains mutations in XopE2 and XopP.

Analysis of the genomic sequence of T3Es revealed presence/absence differences, frameshifts and premature stop codons. However, this analysis does not preclude potential allelic or expression differences among the virulence proteins that could be contributing factors to the re-emergence of CBB. Therefore, newly isolated strains may harbor subtle genomic changes that have allowed them to overcome existing resistance phenotypes. Many commercial cultivars of cotton are reported to be resistant to CBB [30-32]. Based on these

previous reports, we selected commercial cultivars resistant and susceptible (6 of each) to CBB. In addition, we included 5 available varieties that are related to the historical panel as well as 2 parents from a nested association mapping (NAM) population currently under development [33]. All varieties inoculated with the newly isolated *Xcm* strains exhibited inoculation phenotypes consistent with previous reports (Figs 3b,c). In these assays, bright field and near infrared (NIR) imaging were used to distinguish water-soaked disease symptoms from rapid cell death (HR) that is indicative of an immune response. These data confirm that existing resistance genes present within cotton germplasm are able to recognize the newly isolated *Xcm* strains and trigger a hypersensitive response. Together, the phylogenetic analysis, effector profile conservation and cotton inoculation phenotypes, confirm that the recent outbreak of *Xcm* in the US represents a re-emergence of a race 18 clade *Xcm* and is not the result of a dramatic shift in the pathogen.

The USDA Agricultural Marketing Service (AMS) releases reports on the percentage of upland cotton cultivars planted in the U.S. each year (www.ams.usda.gov/mnreports/cnavar.pdf). Most of these varieties are screened for resistance or susceptibility to multiple strains of *Xcm* by extension scientists and published in news bulletins [30, 31, 34-38]. These distinct datasets were cross referenced to reveal that only 25% of the total cotton acreage was planted with resistant cultivars in 2016 (Fig 3d, Table S3). This is part of a larger downward trend in which the acreage of resistant cultivars has fallen each year since at least 2009 when the percentage of acreage planted with resistant varieties was at 75%.

Comparative genome analysis for two *Xcm* strains

Differences in virulence were observed among *Xcm* strains at the molecular and phenotypic level. In order to gain insight into these differences, we selected two strains from our collection that differed in T3E content, virulence level, geography of origin and isolation date. AR81009 was isolated in Argentina in 1981 and is one of the most virulent strains investigated in this study; MS14003 was isolated in Mississippi in 2014 and is a representative strain of the race 18 clade (Fig S2). The latter strain causes comparatively slower and diminished leaf symptoms; however, both strains are able to multiply and cause disease on susceptible varieties of cotton (Fig S3). Full genome sequences were generated with Single Molecule Real-Time (SMRT) sequencing. Genomes were assembled using the PacBio Falcon assembler which yielded circular 5Mb genomes and associated plasmids. Genic synteny between the two strains was observed with the exception of two 1.05 Mb inversions (Fig 4). Regions of high and low GC content, indicative of horizontal gene transfer, were identified in both genomes. In particular, a 120kb insertion with low GC content was observed in AR81009. This region contains one T3E as well as two annotated type four secretion system-related genes, two conjugal transfer proteins, and two multi drug resistant genes (Fig 4 insert). MS14003 contains three plasmids (52.4, 47.4, and 15.3kb) while AR81009 contains two plasmids (92.6 and 22.9kb). Analysis of homologous regions among the plasmids was performed using progressiveMauve [39]. This identified four homologous regions greater than 1kb that were shared among multiple plasmids (Fig 4).

Both strains express TAL effector proteins as demonstrated through western blot analysis using a TAL effector specific polyclonal antibody (Fig 5) [40]. However, the complexity of TAL effector repertoires within these strains prevented complete resolution of each

223 individual TAL effector. The long reads obtained from SMRT sequencing are able to span whole
224 TAL effectors, allowing for full assemblies of the TAL effectors in each strain. The AR81009
225 genome encodes twelve TAL effectors that range in size from twelve to twenty three repeat
226 lengths, six of which reside on plasmids. The MS14003 genome encodes eight TAL effectors that
227 range in size from fourteen to twenty eight repeat lengths, seven of which reside on plasmids
228 (Fig 5). Three partial TAL effector-like coding sequences were also identified within these
229 genomes and are presumed to be non-functional. A 1-repeat gene with reduced 5' and 3'
230 regions was identified in both strains directly upstream of a complete TAL effector. In addition,
231 a large 4kb TAL effector was identified in AR81009 with a 1.5 kb insertion and 10 complete
232 repeat sequences. The tool AnnoTALE was used to annotate and group TAL effectors based on
233 the identities of the repeat variable diresidues (RVDs) in each gene [41]. Little homology was
234 identified among TAL RVD sequences within and between strains; only two TAL effectors were
235 determined to be within the same TAL class between strains (TAL19b of AR81009 and TAL19 of
236 MS14003) and two within strain MS14003 (TAL14b and TAL16).

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

238 An RNA-sequencing experiment was designed to determine whether AR81009 and
239 MS14003 incite different host responses during infection (Fig 6a, b). Isolates were inoculated
240 into the phylogenetically diverse *G. hirsutum* cultivars Acala Maxxa and DES 56 [33]. Infected
241 and mock-treated tissue were collected at 24 and 48 hours post inoculation. First, we
242 considered global transcriptome patterns of gene expression. Fifty-two genes were determined
243 to be induced in all *Xcm-G. hirsutum* interactions at 48 hours (Fig 6c, Table S4). Of note among
244 this list is a homeologous pair of genes with homology to the known susceptibility target MLO

[42-45]. Gene induction by a single strain was also observed; AR81009 and MS14003 uniquely induced 127 and 16 *G. hirsutum* genes, respectively (Fig 6c). In contrast, the average magnitude of gene induction between the two strains was not significantly different (Fig S4). Both *Xcm* strains caused more genes to be differentially expressed in DES 56 than in Acala Maxxa. Among the 52 genes significantly induced by both strains, sixteen conserved targets are homeologous pairs, whereas seventeen and fifteen genes are encoded by the A and D sub-genomes, respectively (Tables 2 and S4). It has been previously reported that homeologous genes encoded on the *G. hirsutum* A and D sub-genomes are differentially regulated during abiotic stress [46]. A set of approximately 10,000 homeologous gene pairs were selected and differential gene expression was assessed (Fig 7). For each pairwise comparison of *Xcm* strain and *G*. *hirsutum* cultivar, a similar number of genes were differentially expressed in each of the A and D subgenomes. However, some homeologous pairs were up- or down-regulated differentially in response to disease, indicating a level of sub-genome specific responses to disease. For example, SWEET sugar transporter gene Gh_D12G1898 in the D genome is induced over fourfold during infection with *Xcm* strain AR81009, while the homeolog Gh_A12G1747 in the A genome is induced to a much smaller extent.

Table 2: Eight homeologous pairs of *Gossypium hirsutum* genes are upregulated in both Acala Maxxa and DES 56 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

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

SWEET sugar transporter genes have been reported to be targets of and upregulated by *Xanthomonas* TAL effectors in *Manihot esculenta*, *Oryza sativa*, and *Citrus sinensis* [21, 40, 47, 48]. In rice and cassava, the SWEET genes are confirmed susceptibility genes that contribute to disease symptoms. The previously reported susceptibility genes and the SWEETs identified here, are clade III sugar transporters (Fig S5). The NBI *Gossypium hirsutum* genome encodes 54 putative SWEET sugar transporter genes. Of these 54 genes, three were upregulated greater than fourfold in response to inoculation by one of the two *Xcm* strains (Fig 8). Predicted TAL effector binding sites were identified using the program TALEnt [49]. MS14003 significantly induces the homeologs Gh_A04G0861 and Gh_D04G1360 and contains the TAL effectors M14b, M28a, and M28b, which are predicted to bind within the 300bp promoter sequences of at least one of these genes. Of note is TAL M28a, which is predicted to bind both homeologs (Fig S6a). In contrast, AR81009 induces Gh_D12G1898 to a greater extent than its homeolog Gh_A12G1747. TAL effectors A14c and A16b from AR81009 are predicted to bind to the Gh_D12G1898 and Gh_A12G1747 promoters; however, TAL A14a is predicted to bind only the Gh_D12G1898 promoter (Fig S6b). We note that while Gh_A12G1747 did not pass the fourfold cut off for gene induction, this gene is slightly induced compared to mock inoculation.

Discussion

Cotton Bacterial Blight was considered controlled in the U.S. until an outbreak was observed during the 2011 growing season in Missouri, Mississippi and Arkansas [50]. Until 2011,

284 seed sterilization, breeding for resistant varieties, and farming techniques such as crop rotation
285 and sterilizing equipment prevented the disease from becoming an economic concern [51]. The
286 number of counties reporting incidence of CBB has increased from 17 counties in 2011 to 77
287 counties in 2015 [38, 52, 53]. This paper investigates the root of the re-emergence and
288 identifies several routes towards control of the disease.

289 When CBB was first recognized as re-emerging, several possible explanations were
290 proposed including: (1) A highly virulent race of the pathogen that had been introduced to the
291 U.S.; (2) Historical strains of *Xcm* that had evolved to overcome existing resistance (e.g. an
292 effector gene change or host shift); and (3) Environmental conditions over the last several years
293 that had been particularly conducive to the disease. Here, we present evidence that the re-
294 emergence of CBB is not due to a large genetic change or race shift in the pathogen. Rather, the
295 re-emergence of the disease is likely due to agricultural factors such as large areas of
296 susceptible cultivars being planted. The presented data do not rule out potential environmental
297 conditions that may also have contributed to the re-emergence. In this context, environmental
298 conditions include disease conducive temperature and humidity as well as potentially
299 contaminated seed or other agronomic practices that may have perpetuated spread of the
300 disease outbreaks. Importantly, the presented data confirm that the presence of resistance loci
301 could be deployed to prevent further spread of this disease. However, since many of the most
302 popular farmer preferred varieties lack these resistance traits, additional breeding or
303 biotechnology strategies will be needed to maximize utility. Notably, the current *Xcm* isolates
304 characterized in this study all originate from Mississippi cotton fields in 2014. During the 2015
305 and 2016 growing seasons, resistant cotton cultivars were observed in Texas with symptoms

306 indicative of bacterial infection distinct from CBB. Additional work is underway to identify and
307 characterize the causal agent(s) of these disease symptoms.

308 Recent work on CBB in the US has focused on the most prevalent US *Xcm* race: race 18.
309 However, races are not necessarily phylogenetically distinct clades. Race 18 isolates have been
310 reported overseas, indicating that there may be independent origins of the race or cross-
311 continent movement of this pathogen. Phenotypic race delineations were created before
312 modern genetic and phylogenetic techniques were developed. However, modern genetics
313 presents the opportunity to begin classifying strains based upon phylogenetic and effector
314 profiles rather than phenotypes on a limited range of host varieties. Here, we identify all known
315 and putative race 18 isolates as phylogenetically grouped into a single clade and distinct from
316 other *Xcm* isolates. Future efforts can further explore phylogenetic relatedness among diverse
317 isolates.

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

326 Over the past decade, susceptibility genes have become targets for developing disease
327 tolerant plants [54, 55]. These genes are typically highly induced during infection [56].

328 Therefore, RNA-Seq of infected plants has become a preferred way to identify candidate
329 susceptibility genes. Once identified, genome editing can be used to block induction of these
330 genes [57]. We report a homeologous pair of genes that are homologs of the MLO gene as
331 targeted by both *Xcm* strains in both cotton cultivars. These genes are excellent candidates for
332 future biotechnology efforts. Because the potential importance of these genes in cotton biology
333 is unknown, their role in cotton physiology must first be explored. Knock-out mutations of MLO
334 genes in other systems has led to durable resistance against powdery mildew as well as
335 oomycetes and bacteria such as *Xanthomonas* [42, 45]. The dual purpose of host susceptibility
336 genes has been observed previously. For example, the rice *Xa13* (aka. *Os8N3* and *OsSWEET11*)
337 gene is required for pollen development but also targeted by a rice pathogen during infection
338 [58]. Xa13 is a member of the clade III SWEET sugar transporters implicated in many
339 pathosystems. In this case, the induction of *Xa13* for pathogen susceptibility is mediated by a
340 TAL effector. Of the 54 SWEET genes in the *G. hirsutum* genome, at least three are significantly
341 upregulated during *Xcm* infection. In contrast to MLO, no single SWEET gene was induced by
342 both pathogen strains in both hosts.

343 Analysis of SWEET gene expression after inoculation revealed a context for polyploidy in
344 the *G. hirsutum-Xcm* pathosystem. This relatively unexplored area of plant-microbe interactions
345 arose from our observation of a potential difference in induction magnitude between the
346 homeologous Gh_A12G1747 and Gh_D12G1898 SWEET genes. Further analysis revealed many
347 examples of preferentially induced or down-regulated homeologs in response to *Xcm* infection.
348 Characterization of sub-genome specialization may lead to new insights regarding durability of
349 resistance and susceptibility loci in polyploid crops. Future research may investigate the diploid

ancestors of tetraploid cotton to further explore the evolution of host and pathogen in the context of ploidy events [59]

Multiple putative TAL effector binding sites were identified within each up-regulated SWEET promoter. These observations suggest that TAL M28a from MS14003 may induce the homeologs Gh_A04G0861 and Gh_D04G1360. Further, TAL effector A14a from AR81009 is likely responsible for the upregulation of Gh_D12G1898. Whether additional TAL effectors are involved in these responses is not clear. Genome organization in the host, such as histone modifications or other epigenetic regulations may also be affecting these interactions. Future research will investigate these mechanisms further.

Collectively, the data presented here suggest that the wide-spread planting of CBB-susceptible cultivars has contributed to the re-emergence of CBB in the southern U.S. It is possible that a reservoir of race 18 *Xcm* was maintained in cotton fields below the level of detection due to resistant cultivars planted in the 1990s and early 2000s. Alternatively, the pathogen may have persisted on an alternate host or was re-introduced by contaminated seed [9, 10]. Regardless of the cause of the re-emergence, the genomic comparisons among pathogen races and host cultivars has identified several possible routes towards resistance. These include the use of existing effective resistance loci as well as the potential disruption of the induction of susceptibility genes through genome editing. The latter is an attractive strategy in part because of recent progress in genome editing [60, 61]. In summary, within a relatively short time frame, through the deployment of modern molecular and genomic techniques, we were able to identify factors that likely contribute to the re-emergence of cotton bacterial

371 blight and generate data that can now be rapidly translated to effective disease control
372 strategies.

373 ## Materials & Methods

374 **_Xcm_ strain isolation and manipulation:**

375 New *Xcm* strains were isolated from infected cotton leaves by grinding tissue in 10mM
376 $MgCl_2$ and culturing bacteria on NYGA media. The most abundant colony type was selected,
377 single colony purified and then 16S sequencing was used to confirm the bacterial genus as
378 previously described [62]. In addition, single colony purified strains were re-inoculated into
379 cotton leaves and the appearance of water soaked symptoms indicative of CBB infection was
380 confirmed. Both newly isolated strains as well as strains received from collaborators were used
381 to generate a rifampicin resistance version of each strain. Wildtype strains were grown on
382 NYGA, then transferred to NYGA containing 100µg/ml rifampicin. After approximately 4-5 days,
383 single colonies emerged. These were single colony purified and stored at -80C. The rifampicin
384 resistant version of each *Xcm* strain was used in all subsequent experiments reported in this
385 manuscript unless otherwise noted.

386 **Plant inoculations**

387 Cotton varieties from the original cotton panel for determining *Xcm* race designations
388 were obtained from the USDA/ARS, Germplasm Resources Information Network (GRIN).
389 Varieties included in the *G. hirsutum* NAM population were provided by Vasu Kuraparthy [33].
390 Other commercial varieties were obtained from Terry Wheeler and Tom Allen. Disease assays
391 were conducted in a growth chamber set at 30°C and 80% humidity. *Xcm* strains were grown on
392 NYGA plates containing 100µg/ml rifampicin at 30°C for two days before inoculations were

393 performed. Inoculations were conducted by infiltrating a fully expanded leaf with a bacterial
394 solution in 10mM $MgCl_2$ (OD_{600} specified within each assay).

395 The field tests were conducted as follows: Cotton cultivars are planted in two row plots
396 (10 – 11 m in length, 1 m row spacing), in a randomized complete block design with four
397 replications. Approximately 60 to 80 days after planting, *Xcm* was applied to the test area
398 similar to that described in Wheeler et al. (2007) [37]. Briefly, Xcm is grown in trypticase soy
399 broth (30 g/L) for 1 ½ days and then 19 L of the concentrated bacterial solution (10^8 cfu/ml) are
400 diluted into 189 L of water (resulting in 10^6 cfu/ml) . The surfactant Silwet L-77
401 (polyalkyleneoxide modified heptamethyltrisiloxane, Loveland Industries, Greely, CO) is added
402 at 0.2% v/v. The suspension of bacteria are sprayed over the top of the cotton at a pressure of
403 83 kpa and rate of 470 L/ha. The nozzles used were TeeJet 8008. Symptoms were typically
404 visible 14 days after application and plots were rated for incidence of symptoms 17-21 days
405 after application [34-37].

406 **Cotton Cultivar Statistics**

407 Area of cotton planted per county in the United States in 2015 was obtained from the
408 USDA National Agricultural Statistics Service:
409 www.nass.usda.gov/Statistics_by_Subject/result.php?7061F36A-A4C6-3C65-BD7F-
410 129B702CFBA2&sector=CROPS&group=FIELD%20CROPS&comm=COTTONUSDA. Estimated
411 percentage of upland cotton planted for each variety was obtained from the Agricultural
412 Marketing Service (AMS): www.ams.usda.gov/mnreports/canvar.pdf.

413 **Bacterial Sequencing and Phylogenetics**

Illumina based genomic datasets were generated as previously described [29]. Paired-end Illumina reads were trimmed using Trimmomatic v0.32 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) [63]. Genome assemblies were generated using the SPAdes *de novo* genome assembler [64]. Strain information is reported in Supplemental Table 1. Similar to our previously published methods [29], the program Prokka was used in conjunction with a T3E database to identify type three effector repertoires for each of the 12 *Xcm* isolates as well as four *Xcm* genomes previously deposited on NCBI (S2Table) [65].

Multi-locus sequence analysis was conducted by concatenating sequences of the *gltA*, *lepA*, *lacF*, *gyrB*, *fusA* and *gap-1* loci obtained from the Plant-Associated Microbes Database (PAMDB) for each strain as previously described [66]. A maximum-likelihood tree using these concatenated sequences was generated using CLC Genomics 7.5.

Variant Based Phylogeny

A variant based dendrogram was created by comparing 12 Illumina sequenced *Xcm* genomes to the complete *Xanthomonas citri* subsp. *citri* strain Aw12879 reference genome (Genbank assembly accession: GCA_000349225.1) on NCBI. Read pairs were aligned to the reference genome using Bowtie2 v2.2.9 with default alignment parameters [27]. From these alignments, single nucleotide polymorphisms (SNPs) were identified using samtools mpileup v1.3 and the bcftools call v1.3.1 multi-allelic caller [28]. Using Python v2.7, the output from samtools mpileup was used to identify loci in the *X. citri* subsp. *citri* reference genome with a minimum coverage of 10 reads in each *Xcm* genome used Python version 2.7 available at http://www.python.org. Vcftools v0.1.14 and bedtools v2.25.0 were used in combination to

remove sites marked as insertions or deletions, low quality, or heterozygous in any of the genomes [67, 68]. Remaining loci were concatenated to create a FASTA alignment of confident loci. Reference loci were used where SNP's were not detected in a genome. The resulting FASTA alignment contained 17853 loci per strain. This alignment was loaded into the online Simple Phylogeny Tool from the ClustalW2 package to create a neighbor joining tree of the assessed strains [69, 70]. Trees were visualized using FigTree v1.4.2.

Genome Assembly

Single Molecule, Real Time (SMRT) sequencing of *Xcm* strains MS14003 and AR81009 was obtained from DNA prepped using a standard CTAB DNA preparation. Blue Pippin size selection and library preparation was done at the University of Deleware Sequencin Facility. The genomes were assembled using FALCON-Integrate (https://github.com/PacificBiosciences/FALCON-integrate/commit/cd9e93) [71]. The following parameters were used: Assembly parameters for MS14003: length_cutoff = 7000; length_cutoff_pr = 7000; pa_HPCdaligner_option = -v -dal8 -t16 -e.70 -l2000 -s240 -M10; ovlp_HPCdaligner_option = -v -dal8 -t32 -h60 -e.96 -l2000 -s240 -M10; falcon_sense_option = --output_multi --min_idt 0.70 --min_cov 5 --local_match_count_threshold 2 --max_n_read 300 --n_core 6; overlap_filtering_setting = --max_diff 80 --max_cov 160 --min_cov 5 --bestn 10; Assembly parameters for AR81009: length_cutoff = 8000; length_cutoff_pr = 8000; pa_HPCdaligner_option = -v -dal8 -t16 -e.72 -l2000 -s240 -M10; ovlp_HPCdaligner_option = -v -dal8 -t32 -h60 -e.96 -l2000 -s240 -M10; falcon_sense_option = --output_multi --min_idt 0.72 --min_cov 4 --local_match_count_threshold 2 --max_n_read 320 --n_core 6; overlap_filtering_setting = --max_diff 90 --max_cov 300 --min_cov 10 --bestn 10. Assemblies

458 were polished using iterations of pbalign and quiver, which can be found at
459 https://github.com/PacificBiosciences/pbalign/commit/cda7abb and
460 https://github.com/PacificBiosciences/GenomicConsensus/commit/43775fa. Two iterations
461 were run for *Xcm* strain MS14003 and 3 iterations for AR81009. Chromosomes were then
462 reoriented to the DnaA gene and plasmids were reoriented to ParA. The assemblies were
463 checked for overlap using BLAST, and trimmed to circularize the sequences [72]. TAL effectors
464 were annotated and grouped by RVD sequences using AnnoTALE [41]. Homologous regions
465 among plasmids that are greater than 1 kb were determined using progressiveMauve [39].
466 Genomic comparisons between the MS14003 and AR81009 chromosomes were visualized using
467 Circos [73]. Single-copy genes on each of the chromosomes were identified and joined using
468 their annotated gene IDs. Lines connecting the two chromosomes represent these common
469 genes and their respective positions in each genome. A sliding window of 1KB was used to
470 determine the average GC content. Methylation was determined using the Base Modification
471 and Motif Analysis workflow from pbsmrtpipe v0.42.0 at
472 https://github.com/PacificBiosciences/pbsmrtpipe.

473 **Western Blot Analysis**

474 Western Blot analysis of Transcription Activator-Like (TAL) effectors was performed
475 using a polyclonal TAL specific antibody [40]. Briefly, bacteria were suspended in 5.4 pH
476 minimal media for 4.5 hours to induce effector production and secretion. Bacteria were
477 pelleted and then suspended in laemmli buffer and incubated at 95 degrees Celsius for three
478 minutes to lyse the cells. Freshly boiled samples were loaded onto a 4-6% gradient gel and run
479 for several hours to ensure sufficient separation of the different sized TAL effectors.

480 **Gene Expression Analysis**

481 Susceptible cotton were inoculated with *Xcm* using a needleless syringe at an OD_{600} of
482 0.5. Infected and mock-treated tissue were collected and flash frozen at 24 and 48 hours post
483 inoculation. RNA was extracted using the Sigma tRNA kit. RNA-sequencing libraries were
484 generated as previously described [74].

485 Raw reads were trimmed using Trimmomatic [63]. The Tuxedo Suite was used for
486 mapping reads to the TM-1 NBI *Gossypium hirsutum* genome [75], assembling transcripts, and
487 quantifying differential expression [27].

488 Read mapping identified several mis-annotated SWEET genes that skewed differential
489 expression results. The annotations of SWEET genes Gh_A12G1747, Gh_D07G0487, and
490 Gh_D12G1898 were shortened to exclude 20-30kb introns. Two exons were added to
491 Gh_D05G1488. The 2.7kb scaffold named Scaffold013374 was also removed from analysis
492 because its gene Gh_Sca013374G01 has exact sequence homology to Gh_A12G1747 and
493 created multi-mapped reads that interfered with expression analysis.

494 Homeologous pairs were identified based on syntenic regions with MCScan [76]. A
495 syntenic region was defined as a region with a minimum of five genes with an average
496 intergenic distance of two and within extended distance of 40. All other values were set to the
497 default. Comparisons between homeologs was performed by examining cuffdiff differential
498 expression and classifying them according to the sub-genome expression pattern. Genes
499 considered up or down regulated meet both differential expression from mock significance of
500 q-value < 0.05 and the absolute value of the log2 fold change is greater than 2.

501 **TAL Binding Sites**

Bioinformatic prediction of TAL effector binding sites on the *G. hirsutum* promoterome was performed using the TAL Effector-Nucleotide Targeter (TALEnt) [50]. In short, the regions of the genome that were within 300 basepairs of annotated genes were queried with the RVD's of MS14003 and AR81009 using a cutoff score of 4. Promiscuously binding TALs 16 from MS14003 and 16a from AR81009 were removed from analysis.

Acknowledgements

The authors would like to acknowledge Dr. Robert Nichols for useful discussions throughout the presented research and preparation of this manuscript.

References

1. Cotton: World Markets and Trade [Internet]. 2016 [cited August 2016]. Available from: http://apps.fas.usda.gov/psdonline/circulars/cotton.pdf.
2. World Agricultural Supply and Demand Estimates [Internet]. 2016 [cited August 12, 2016]. Available from: http://usda.mannlib.cornell.edu/usda/current/wasde/wasde-08-12-2016.pdf.
3. Blank LM, Bird LS. Breeding Strains of Cotton Resistant to Bacterial Blight. Texas: Texas Agricultural Experiment Station; 1951. p. 1-25.
4. Green JM, Brinkerhoff LA. Inheritance of Three Genes for Bacterial Blight Resistance in Upland Cotton1. Agronomy Journal. 1956;48(11):481-5. doi: 10.2134/agronj1956.00021962004800110001x.
5. Simpson D, Weindling R. Bacterial blight resistance in a strain of Stoneville cotton. Journal of the American Society of Agronomy. 1946.
6. Huang X, Zhai J, Luo Y, Rudolph K. Identification of a highly virulent strain of Xanthomonas axonopodis pv. malvacearum. European Journal of Plant Pathology. 2008;122(4):461-9. doi: 10.1007/s10658-008-9312-3.
7. Cunnac S, Bolot S, Forero Serna N, Ortiz E, Szurek B, Noel LD, et al. High-Quality Draft Genome Sequences of Two Xanthomonas citri pv. malvacearum Strains. Genome Announc. 2013;1(4). doi: 10.1128/genomeA.00674-13. PubMed PMID: 23990578; PubMed Central PMCID: PMCPMC3757453.
8. Delannoy E, Lyon BR, Marmey P, Jalloul A, Daniel JF, Montillet JL, et al. Resistance of cotton towards Xanthomonas campestris pv. malvacearum. Annu Rev Phytopathol. 2005;43:63-82. doi: 10.1146/annurev.phyto.43.040204.140251. PubMed PMID: 16078877.
9. Brinkerhoff LA. Variation in Xanthomonas malvacearum and its relation to control. Annu Rev Phytopathol. 1970;8:85-110.
10. Verma JP. Bacterial Blight of Cotton. CRC Press1986.
11. Hunter RE, Brinkerhoff LA, Bird LS. The Development of a Set of Upland Cotton Lines for Differentiating Races of Xanthomonas malvacearum. Phytopathology. 1968;58:830-2.
12. Baldwin Jr C, editor Report of the Bacterial Blight Committee--1975 [Cotton, USA]. Proceedings Beltwide Cotton Production Research Conferences; 1976.

13. Essenberg M, Bayles MB, Pierce ML, Verhalen LM. Pyramiding B genes in cotton achieves broader but not always higher resistance to bacterial blight. Phytopathology. 2014;104(10):1088-97. doi: 10.1094/PHYTO-06-13-0167-R. PubMed PMID: 24655289.
14. Gabriel DW, Burges A, Lazo GR. Gene-for-gene interactions of five cloned avirulence genes from Xanthomonas campestris pv. malvacearum with specific resistance genes in cotton. Proceedings of the National Academy of Sciences. 1986;83:6415-9.
15. Schwartz AR, Potnis N, Timilsina S, Wilson M, Patané J, Martins J, et al. Phylogenomics of Xanthomonas field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. Frontiers in Microbiology. 2015;6. doi: 10.3389/fmicb.2015.00535.
16. Jacques MA, Arlat M, Boulanger A, Boureau T, Carrere S, Cesbron S, et al. Using Ecology, Physiology, and Genomics to Understand Host Specificity in Xanthomonas. Annu Rev Phytopathol. 2016;54:163-87. doi: 10.1146/annurev-phyto-080615-100147. PubMed PMID: 27296145.
17. Macho AP, Zipfel C. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. Current Opinion in Microbiology. 2015;23:14-22. doi: http://dx.doi.org/10.1016/j.mib.2014.10.009.
18. Kim J-G, Stork W, Mudgett Mary B. Xanthomonas Type III Effector XopD Desumoylates Tomato Transcription Factor SlERF4 to Suppress Ethylene Responses and Promote Pathogen Growth. Cell Host & Microbe. 2013;13(2):143-54. doi: http://dx.doi.org/10.1016/j.chom.2013.01.006.
19. Wang S, Sun J, Fan F, Tan Z, Zou Y, Lu D. A Xanthomonas oryzae pv. oryzae effector, XopR, associates with receptor-like cytoplasmic kinases and suppresses PAMP-triggered stomatal closure. Science China Life Sciences. 2016:1-8. doi: 10.1007/s11427-016-5106-6.
20. Doyle EL, Stoddard BL, Voytas DF, Bogdanove AJ. TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. Trends Cell Biol. 2013;23(8):390-8. doi: 10.1016/j.tcb.2013.04.003. PubMed PMID: 23707478; PubMed Central PMCID: PMCPMC3729746.
21. Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B. Five phylogenetically close rice SWEET genes confer TAL effector-mediated susceptibility to Xanthomonas oryzae pv. oryzae. New Phytol. 2013;200(3):808-19. doi: 10.1111/nph.12411. PubMed PMID: 23879865.
22. Schreiber KJ, Baudin M, Hassan JA, Lewis JD. Die another day: Molecular mechanisms of effector-triggered immunity elicited by type III secreted effector proteins. Seminars in Cell & Developmental Biology. 2016;56:124-33. doi: http://dx.doi.org/10.1016/j.semcdb.2016.05.001.
23. Lee HA, Yeom SI. Plant NB-LRR proteins: tightly regulated sensors in a complex manner. Brief Funct Genomics. 2015;14(4):233-42. doi: 10.1093/bfgp/elv012. PubMed PMID: 25825425.
24. Flor HH. Current Status of the Gene-for-Gene Concept. Annu Rev Phytopathol. 1971;9:275-96.
25. Cui H, Tsuda K, Parker JE. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. Annual Review of Plant Biology. 2015;66(1):487-511. doi: 10.1146/annurev-arplant-050213-040012.
26. Koch R. Untersuchungen ueber Bakterien V. Die Aetiologie der Milzbrand-Krankheit, begruendent auf die Entwicklungsgeschichte des Bacillus Anthracis. Beitr. z. Biol. D. Pflanzen 2: 277-310. Milestones in Microbiology. 1876;1556.
27. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562-78. doi: 10.1038/nprot.2012.016. PubMed PMID: 22383036; PubMed Central PMCID: PMCPMC3334321.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352.
29. Bart R, Cohn M, Kassen A, McCallum EJ, Shybut M, Petriello A, et al. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable

586 resistance. Proceedings of the National Academy of Sciences. 2012;109(32):13130-. doi:
587 10.1073/pnas.1211014109.
588 30. Dever JK, Wheeler TA, Boman RK, Kerns D, Foster M, Nesmith D, et al. Cotton Performance Tests
589 in the Texas High Plains and Trans Pecos Areas of Texas --2009. Texas AgriLife Research and Extension
590 Service Center. 2010;10(2):1-74.
591 31. Dever JK, Morgan V, Kelly CM, Wheeler TA, Elkins H, Mendoza V, et al. Cotton Performance Tests
592 in the Texas High Plains, 2014. Texas AgriLife Research and Extension Service Center. 2015;15(1).
593 32. Organization CSaIR, inventorCotton Variety Fibermax 9892004 March 3, 2004.
594 33. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthy V. Genetic diversity and
595 population structure in the US Upland cotton (Gossypium hirsutum L.). Theor Appl Genet.
596 2014;127(2):283-95. doi: 10.1007/s00122-013-2217-3. PubMed PMID: 24170350.
597 34. Wheeler TA, Woodward JE. Response of cotton varieties to bacterial blight race 18 ain 2016.
598 Texas AgriLife Research and Extension Service Center. 2016. doi:
599 Lubbock.tamu.edu/files/2016/12/Bacterial-blight-trials-2016-2.pdf.
600 35. Wheeler TA, Woodward JE. Response of cotton varieties to diseases on the Texas High Plains.
601 Texas AgriLife Research and Extension Service Center. 2012. doi: Lubbock.tamu.edu/files/2013/02/2012-
602 cotton-disease-ratings.pdf.
603 36. Wheeler TA, Woodward JE. Response of cotton varieties to diseases on the Southern High Plains
604 of Texas, 2010. Texas AgriLife Research and Extension Service Center. 2010. doi:
605 Lubbock.tamu.edu/files/2011/11/DiseaseRecommendations.pdf.
606 37. Wheeler TA, Sagram US, Schuster GL, Gannaway JR. Identification of Factors that Influence
607 Screening for Bacterial Blight Resistance. The Journal of Cotton Science. 2007;11(2):91-7. doi:
608 http://www.cotton.org/journal/2007-11/2/upload/jcs11-91.pdf
609 38. Woodward JE, Wheeler TA. Bacterial Blight Update for Cotton on the High Plains 2015. Available
610 from: http://plantclinic.tamu.edu/files/2015/08/WoodwardWheeler2015.pdf.
611 39. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain,
612 Loss and Rearrangement. PLOS ONE. 2010;5(6):e11147. doi: 10.1371/journal.pone.0011147.
613 40. Cohn M, Bart RS, Shybut M, Dahlbeck D, Gomez M, Morbitzer R, et al. Xanthomonas axonopodis
614 virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar
615 transporter in cassava. Mol Plant Microbe Interact. 2014;27(11):1186-98. doi: 10.1094/MPMI-06-14-
616 0161-R. PubMed PMID: 25083909.
617 41. Grau J, Reschke M, Erkes A, Streubel J, Morgan RD, Wilson GG, et al. AnnoTALE: bioinformatics
618 tools for identification, annotation, and nomenclature of TALEs from Xanthomonas genomic sequences.
619 Sci Rep. 2016;6:21077. doi: 10.1038/srep21077. PubMed PMID: 26876161; PubMed Central PMCID:
620 PMCPMC4753510.
621 42. Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, et al. The Barley Mlo
622 Gene: A Novel Control Element of Plant Pathogen Resistance. Cell. 1997;88(5):695-705. doi:
623 http://dx.doi.org/10.1016/S0092-8674(00)81912-1.
624 43. Jarosch B, Kogel K-H, Schaffrath U. The Ambivalence of the Barley Mlo Locus: Mutations
625 Conferring Resistance Against Powdery Mildew (Blumeria graminis f. sp. hordei) Enhance Susceptibility
626 to the Rice Blast Fungus Magnaporthe grisea. Molecular Plant-Microbe Interactions. 1999;12(6):508-14.
627 doi: 10.1094/MPMI.1999.12.6.508.
628 44. Jørgensen IH. Discovery, characterization and exploitation of Mlo powdery mildew resistance in
629 barley. Euphytica. 1992;63(1):141-52. doi: 10.1007/BF00023919.
630 45. Kim DS, Hwang BK. The pepper MLO gene, CaMLO2, is involved in the susceptibility cell-death
631 response and bacterial and oomycete proliferation. The Plant Journal. 2012;72(5):843-55. doi:
632 10.1111/tpj.12003.

633 46. Dong S, Adams KL. Differential contributions to the transcriptome of duplicated genes in
634 response to abiotic stresses in natural and synthetic polyploids. New Phytol. 2011;190(4):1045-57. doi:
635 10.1111/j.1469-8137.2011.03650.x. PubMed PMID: 21361962.
636 47. Chen LQ. SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytol.
637 2014;201(4):1150-5. PubMed PMID: 24649486.
638 48. Verdier V, Triplett LR, Hummel AW, Corral R, Cernadas RA, Schmidt CL, et al. Transcription
639 activator-like (TAL) effectors targeting OsSWEET genes enhance virulence on diverse rice (Oryza sativa)
640 varieties when expressed individually in a TAL effector-deficient strain of Xanthomonas oryzae. The New
641 phytologist. 2012;196(4):1197-207. doi: 10.1111/j.1469-8137.2012.04367.x. PubMed PMID: 23078195.
642 49. Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, VanDyk JK, et al. TAL Effector-
643 Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids
644 Research. 2012;40(W1):W117-W22.
645 50. Allen T. ALERT Bacterial Blight of Cotton 2011 [updated July 15, 2011]. Available from:
646 http://www.mississippi-crops.com/2011/07/15/alert-bacterial-blight-of-cotton/.
647 51. Aaron S. Alexander JEW, Randal K. Boman, Terry A. Wheeler, Norman W. Hopper. Effect of the
648 Easiflo Cottonseed Processing Method on Recovery of Xanthomonas axonopodis pv. malvacearum. The
649 Texas Journal of Agriculture and Natural Resource. 2012;25:13-23.
650 52. Allen TW. Epidemiology of Bacterial Blight of Cotton: 2011 to 2015. 2016 Cotton Incorporated -
651 Bacterial Blight Review. 2016.
652 53. Kemerait B. Cotton bacterial blight is back and this is what you need to know 2016 [updated
653 August 10, 2016]. Available from: http://southeastfarmpress.com/cotton/cotton-bacterial-blight-back-
654 and-what-you-need-know.
655 54. Li T, Liu B, Spalding MH, Weeks DP, Yang B. High-efficiency TALEN-based gene editing produces
656 disease-resistant rice. Nature Biotechnology. 2012;30(5):390-2. doi: 10.1038/nbt.2199. PubMed PMID:
657 22565958.
658 55. Jia H, Orbovic V, Jones JB, Wang N. Modification of the PthA4 effector binding elements in Type I
659 CsLOB1 promoter using Cas9/sgRNA to produce transgenic Duncan grapefruit alleviating
660 XccDeltapthA4:dCsLOB1.3 infection. Plant Biotechnol J. 2016;14(5):1291-301. doi: 10.1111/pbi.12495.
661 PubMed PMID: 27071672.
662 56. Strauß T, van Poecke RMP, Strauß A, Römer P, Minsavage GV, Singh S, et al. RNA-seq pinpoints a
663 Xanthomonas TAL-effector activated resistance gene in a large-crop genome. Proceedings of the
664 National Academy of Sciences. 2012;109(47):19480-5.
665 57. Yang B, Sugio A, White FF. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice.
666 Proceedings of the National Academy of Sciences. 2006;103(27):10503-8.
667 58. Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, et al. Promoter mutations of an essential gene for
668 pollen development result in disease resistance in rice. Genes & Development. 2006;20(10):1250-5. doi:
669 10.1101/gad.1416306.
670 59. Yoo M-J, Wendel JF. Comparative Evolutionary and Developmental Dynamics of the Cotton
671 (*Gossypium hirsutum*) Fiber Transcriptome. PLoS Genet. 2014;10(1):e1004073. doi:
672 10.1371/journal.pgen.1004073.
673 60. Janga MR, Campbell LM, Rathore KS. CRISPR/Cas9-mediated targeted mutagenesis in upland
674 cotton (Gossypium hirsutum L.). Plant Molecular Biology. 2017:1-12. doi: 10.1007/s11103-017-0599-3.
675 61. Li C, Unver T, Zhang B. A high-efficiency CRISPR/Cas9 system for targeted mutagenesis in Cotton
676 (Gossypium hirsutum L.). Sci Rep. 2017;7:43902. doi: 10.1038/srep43902. PubMed PMID: 28256588;
677 PubMed Central PMCID: PMCPMC5335549.
678 62. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for
679 phylogenetic study. Journal of bacteriology. 1991;173(2):697-703. PubMed PMID: 1987160.

63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404; PubMed Central PMCID: PMCPMC4103590.
64. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 2012;19(5):455-77. doi: 10.1089/cmb.2012.0021.
65. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
66. Almeida NF, Yan S, Cai R, Clarke CR, Morris CE, Schaad NW, et al. PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes. Phytopathology. 2010;100(3):208-15. doi: 10.1094/PHYTO-100-3-0208. PubMed PMID: 20128693.
67. 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.
68. 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.
69. 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.
70. 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.
71. 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
72. 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.
73. 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.
74. 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.
75. 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.
76. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and Collinearity in Plant Genomes. Science. 2008;320(5875):486.

Figure Legends

Fig 1: Cotton Bacterial Blight (CBB) symptoms and reemergence across the southern United States. (Left) Typical CBB symptoms present in cotton fields near Lubbock, TX during the 2015 growing season include angular leaf spots, boll rot, and black arm rot. Yellow shading within world map (top) indicates origin of strains included in this study. Acres of cotton planted per county in the United States in 2015 (blue) and counties with confirmed CBB in 2015 (red outline). Statistics on the area of cotton planted in the U.S. were acquired from the USDA. CBB

722 was reported by extension agents, extension specialists, and certified crop advisers in their
723 respective states.

724 **Table 1: Illumina and SMRT sequenced *Xcm* genomes described in this paper.**

725 **Fig 2: Phylogenetic analysis of *Xcm* isolates and 13 species of *Xanthomonas*** A) MLST (Multi
726 Locus Sequence Typing) and maximum likelihood analysis of 13 Illumina sequenced *Xcm*
727 isolates (this paper) and 40 other Xanthomonads using concatenated sections of the gltA, lepA,
728 lacF, gyrB, fusA and gap-1 loci. B) SNP based neighbor-joining tree generated from 17,853
729 variable loci between 13 *Xcm* isolates and the reference genome *Xanthomonas citri* subsp. *citri*
730 strain Aw12879. The tree was made using the Simple Phylogeny tool from ClustalW2.

731 **Fig 3: Molecular and phenotypic analysis of *Xcm* and *G. hirsutum* interactions.** A) Type three
732 effector profiles of *Xcm* isolates were deduced from *de novo*, Illumina based genome
733 assemblies. Effector presence or absence was determined based on homology to known type
734 three effectors using the program Prokka. B) Commercial and public *G. hirsutum* cultivars were
735 inoculated with 13 *Xcm* isolates. Susceptible (S) indicates water soaking symptoms. Resistant
736 (R) indicates a visible hypersensitive response. Plants were screened with a range of inoculum
737 concentration from OD_{600} = 0.001-0.5. C) Disease symptoms on *G. hirsutum* cultivars Stoneville
738 5288 B2F and DES 56 after inoculation with *Xcm* strain AR81009 (OD_{600} = 0.05). Symptoms are
739 visualized under visible (VIS) and near infrared (NIR) light. D) The proportion of US fields
740 planted with susceptible and resistant cultivars of *G. hirsutum* was determined using planting
741 acreage statistics from the USDA-AMA and disease phenotypes based on previous reports for
742 common cultivars [34-36].

743 **Fig 4: SMRT sequencing of two phenotypically and geographically diverse *Xcm* isolates:**
744 **MS14003 and AR81009.** Circos plot visualization of two circular *Xcm* genomes. Tracks are as
745 follows from inside to outside: synteny of gene models; GC Content; DNA Methylation on + and
746 – strands; location of type three effectors (teal) and TAL effectors (red), and position. On each
747 side, accompanying plasmids are cartooned. Type three effector repertoires and the type IV
748 secretion systems were annotated using Prokka. Homologous regions greater than 1kb were
749 identified using MAUVE, and TAL effectors were annotated using AnnoTALE.

Fig 5: SMRT sequencing and western blot reveal diverse TAL effector repertoires between *Xcm* strains MS14003 and AR81009. Western Blot of TAL effectors using polyclonal TAL-specific antibody and gene models of TAL effectors identified by AnnoTALE. Blue and Green highlighted gene models represent TALs grouped in the same clade by repeat variable di-residue (RVD) sequence using AnnoTALE.

Fig 6: RNA-Sequencing analysis of infected *G. hirsutum* tissue demonstrates transcriptional changes during CBB. A) Disease phenotypes of *Xcm* strains MS14003 and AR81009 on *G. hirsutum* cultivars Acala Maxxa and DES 56, 7 days post inoculation. B) Acala Maxxa and DES 56 were inoculated with *Xcm* strains MS14003 and AR81009 at an OD of 0.5 and a mock treatment of 10mM $MgCl_2$. Inoculated leaf tissue was collected at 24 and 48 hpi (before disease symptoms emerged). C) Venn diagram of upregulated *G. hirsutum* genes (Log2(fold change in FPKM) ≥ 2 and p value ≤ 0.05) in response to *Xcm* inoculation. Venn diagram was created using the VennDiagram package in R.

Table 2: Homeologous pairs of *Gossypium hirsutum* genes upregulated in Acala Maxxa and DES 56 after inoculation with *Xcm* strains MS14003 and AR81009.

Fig 7: Expression of homeologous pairs across the A and D *G. hirsutum* genomes in response to *Xcm* inoculation. Genes are considered up or down regulated if the absolute value of gene expression change after inoculation as compared to mock treatment was Log2(fold change in FPKM) ≥ 2 and p value ≤ 0.05. By these criteria, pink shading indicates no significant gene expression change. A-D-: both members of the homeologous gene pair are down regulated; A-D0: only the 'A' sub-genome homeolog is down regulated; A-D+: 'A' sub-genome homeolog is down regulated, 'D' sub-genome homeolog is upregulated; etc. Number of gene pairs (n) meeting each expression pattern is indicated within the grey bar. For all genes meeting each expression pattern, the distribution of expression patterns is displayed as a box plot. Rectangles indicate the interquartile range and the whiskers show 1.5 times the interquartile range. A) Acala Maxxa inoculated with MS14003 B) DES 56 inoculated with MS14003 C) Acala Maxxa inoculated with AR 81009 D) DES 56 inoculated with AR81009.

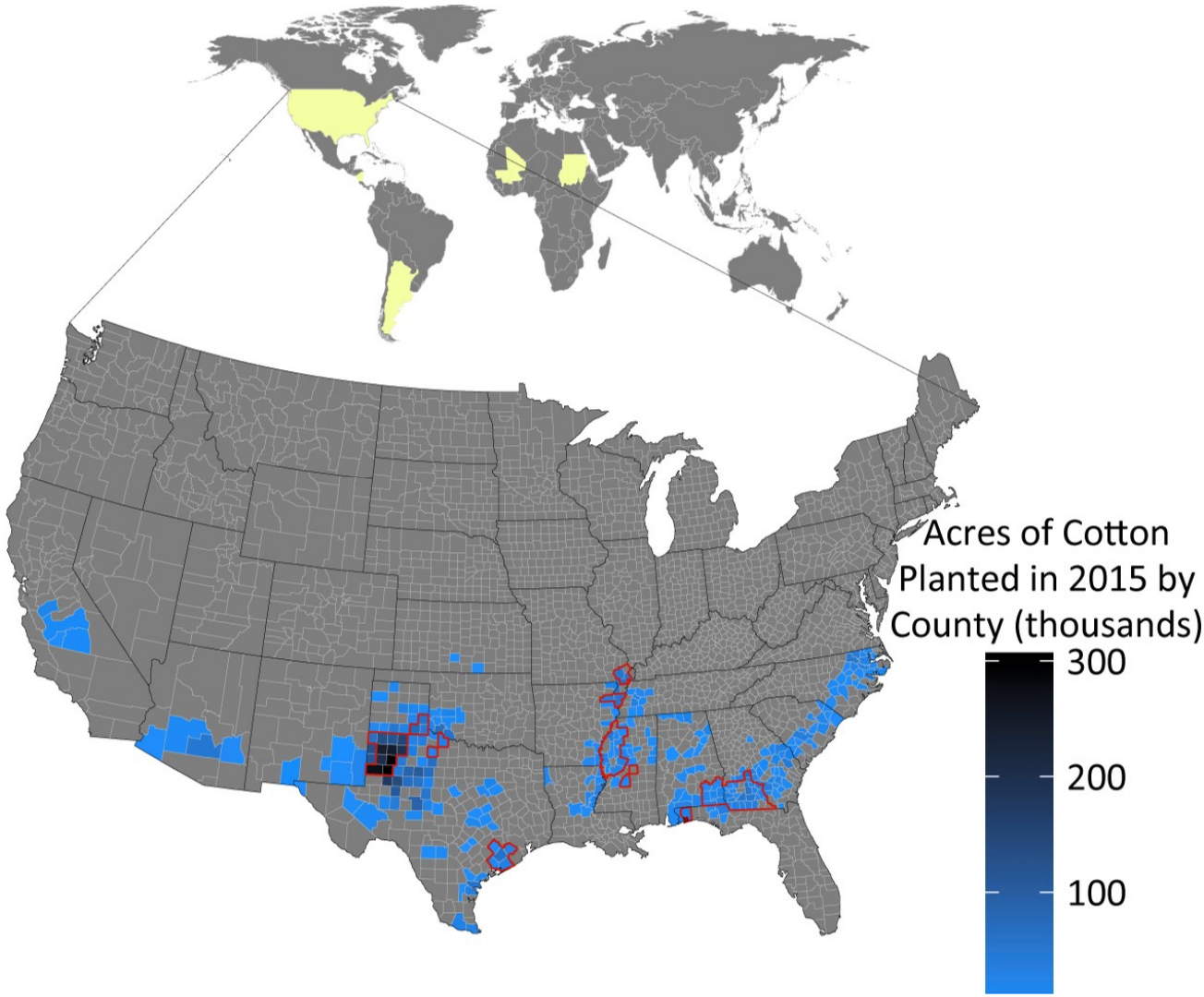
Fig 8: Three candidate *G. hirsutum* susceptibility genes are targeted by two different *Xcm* strains. A) The homeologous pair of SWEET genes A04_G0861 and D04_G1360 are upregulated

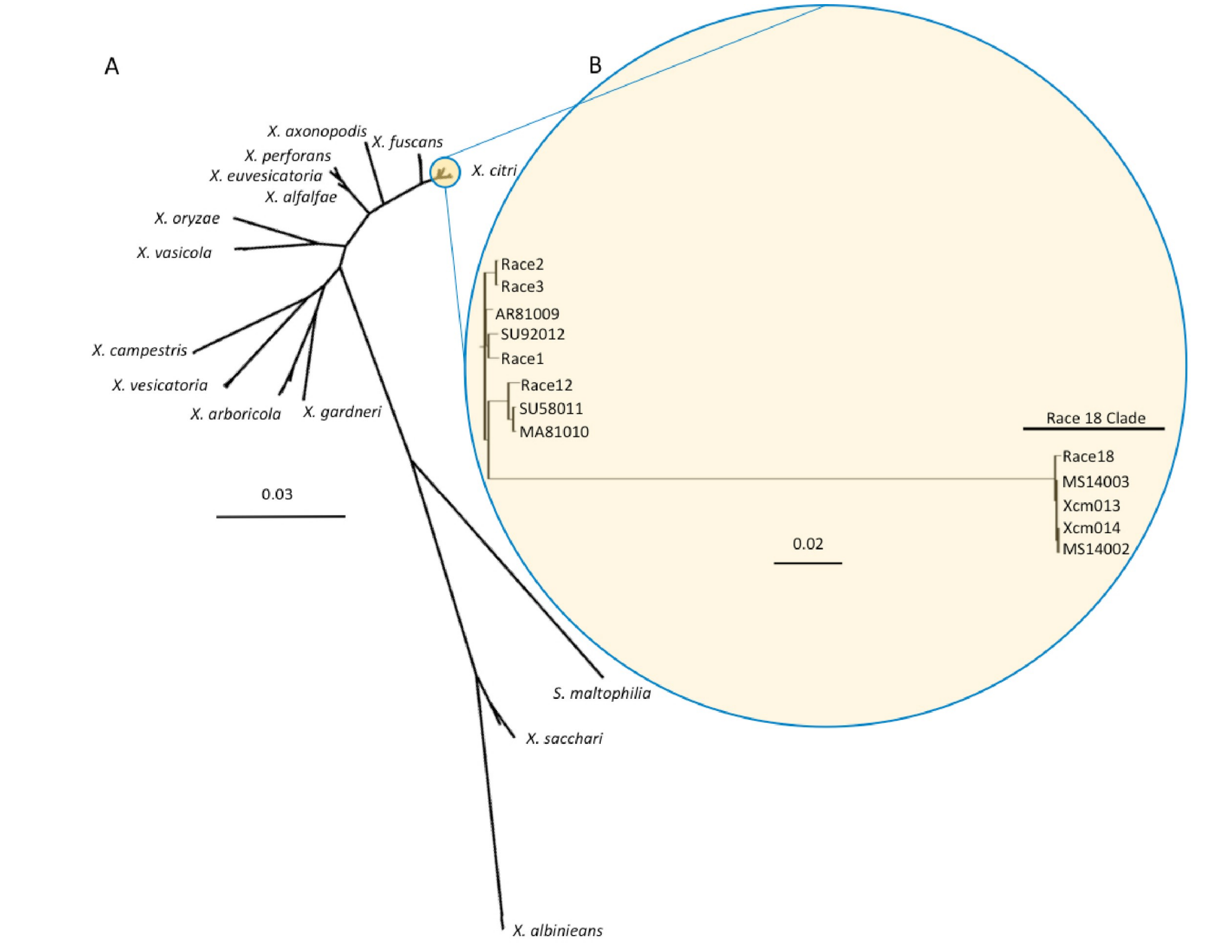
779 in the presence of *Xcm* strain MS14003. (top) Cartoon summary of 300bp promoters of
780 A04_G0861 and D04_G1360. (bottom) Heat-map of the expressions of A04_G0861 and
781 D04_G1360 48 hours after mock or *Xcm* inoculation. B) The SWEET gene D12_G1898 is
782 upregulated in the presence of *Xcm* strain AR81009. (top) Cartoon summary of 300bp
783 promoters of D12_G1898 and A12_G1747. (bottom) Heat-map of the expressions of
784 A12_G1747 and D12_G1898 48 hours after mock or *Xcm* inoculation. TAL effector binding sites
785 were predicted with TALEsf using a quality score cutoff of 4. Gene promoter cartoon legend:
786 Arrow: TAL effector binding site; Black dot: Deletion; Black bar: SNP; Pink bar: TATA box; Teal
787 section: 5'UTR.
788
789 ## Supporting Information Legends
790
791 **S1 Table: US Counties with reported CBB incidence from 2009 to 2016.**
792 **S2 Table: *Xanthomonas* genomes previously deposited on NCBI that are referenced in this**
793 **paper.**
794 **S3 Table: Disease phenotypes and percent acreage of commercial *G. hirsutum* varieties**
795 **planted in the US from 2009-2016.**
796 **S4 Table: RNA-Seq analysis reveals that 52 genes are induced in all *Xcm-G. hirsutum***
797 **interactions at 48 hours ($p \leq 0.05$) with a Log2 (fold change in FPKM) ≥ 2).**
798 **S1 Fig: Maps of CBB incidence in the US from 2011-2012 and 2014-2016.** CBB incidence was
799 reported by extension agents, extension specialists and certified crop advisers in their
800 respective states for the years 2011-2012 and 2014-2016, and compiled by Tom Allen. CBB
801 reports for 2013 were infrequent.
802 **S2 Fig: Disease phenotypes of historical Race18 strain and MS14003 strain.** *Xcm* strains Race18
803 and MS14003 were inoculated into *G. hirsutum* variety PHY499 WRF at an OD600 of 0.01 and
804 imaged at 8 dpi.
805 **S3 Fig: Growth assay of MS14003 and AR81009 on cotton varieties Acala Maxxa and DES 56.**
806 *G. hirsutum* varieties were inoculated with *Xcm* at an OD600: 0.05. Tissue was collected at day 0
807 and day 3 and processed as described in materials and methods.

S4 Fig: Expression levels of significantly upregulated genes with a Log2 fold change of 2 in *G. hirsutum* A) All significantly upregulated genes with a Log2 fold change of 2 B) All significantly upregulated genes ($p \leq 0.05$) with a Log2 (fold change in FPKM) ≥ 2 that are unique to each cultivar/*Xcm* disease interaction in *G. hirsutum.* Numbers in grey bar indicate the total number of genes for each condition.

S5 Fig: Phylogeny of SWEET genes from *Gossypium hirsutum*, *Manihot esculenta,* and *Arabidopsis thaliana.* Four predicted *G. hirsutum* SWEET genes are compared to classified *A. thaliana* SWEET genes and the MeSWEET10a *M. esculenta* susceptibility gene. A protein alignment and phylogenetic tree was generated by Clustal Omega, and the tree was visualized using Figtree v1.4.2.

S6 Fig: Alignment of predicted TAL effector binding sites on induced *G. hirsutum* SWEET genes. A) TAL M28a is predicted to bind to and up-regulate the homeologous pair of SWEET genes: A04_G0861 and D04_G1360 in *G. hirsutum* varieties Acala Maxxa and DES56 after inoculation with *Xcm* strain MS14003. B) TAL A14a is predicted to bind to and up-regulate the SWEET gene D12_G1898 G1360 in *G. hirsutum* varieties Acala Maxxa and DES56 after inoculation with *Xcm* strain AR81009.


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
MS14003
Xcm013
Xcm014
MS14002
0.02

A

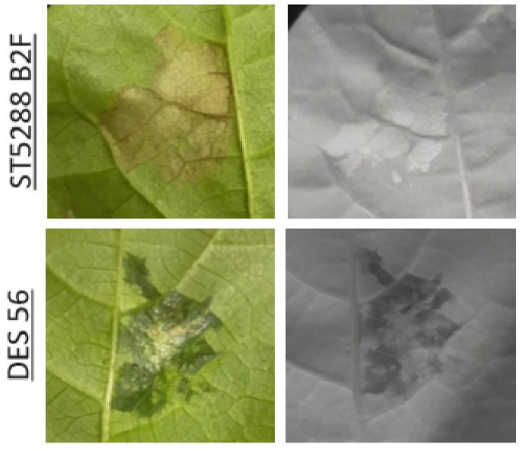
Type Three Effectors

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

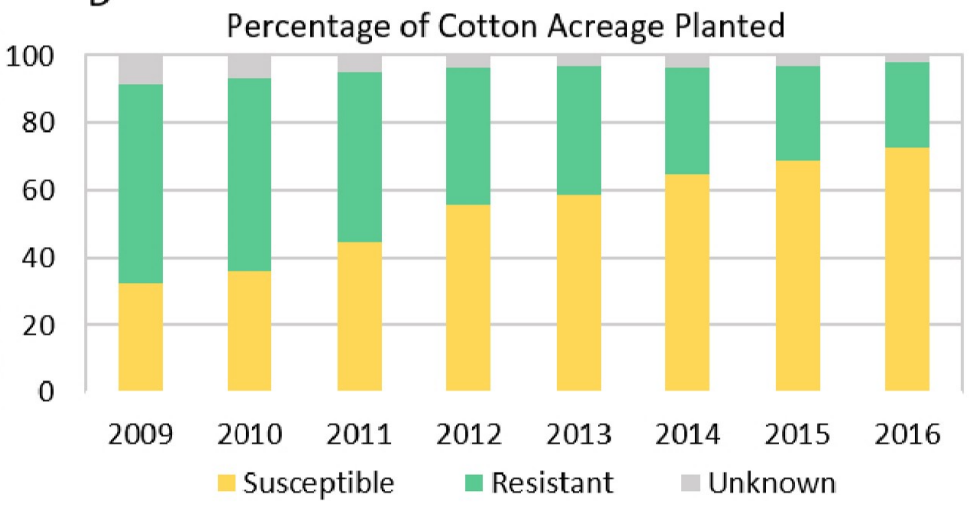
B

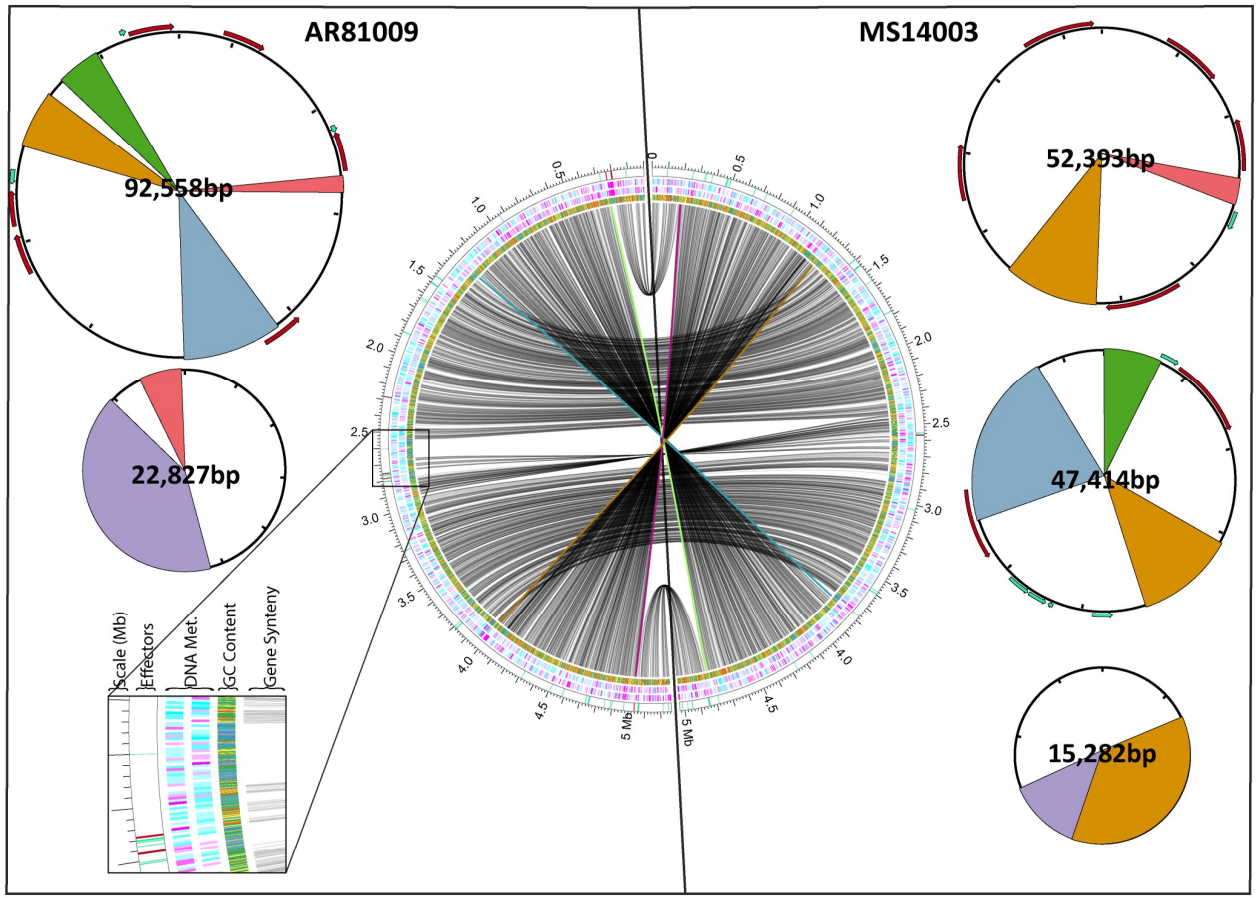
	Variety	R/S
Commercial Varieties	Fibermax 989	R
	Fibermax 1830 GLT	R
	Fibermax 2334 GLT	R
	Fibermax 2484 B2F	R
	Deltapine 1133 B2RF	R
	Stoneville 5288 B2F	R
	Deltapine 0912 B2RF	S
	Deltapine 1028 B2RF	S
	Deltapine 1034 B2RF	S
	Deltapine 1048 B2RF	S
	PhytoGen 499 WRF	S
	Stoneville 4946 GLB2	S
CBB Panel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

C



D





Methylation

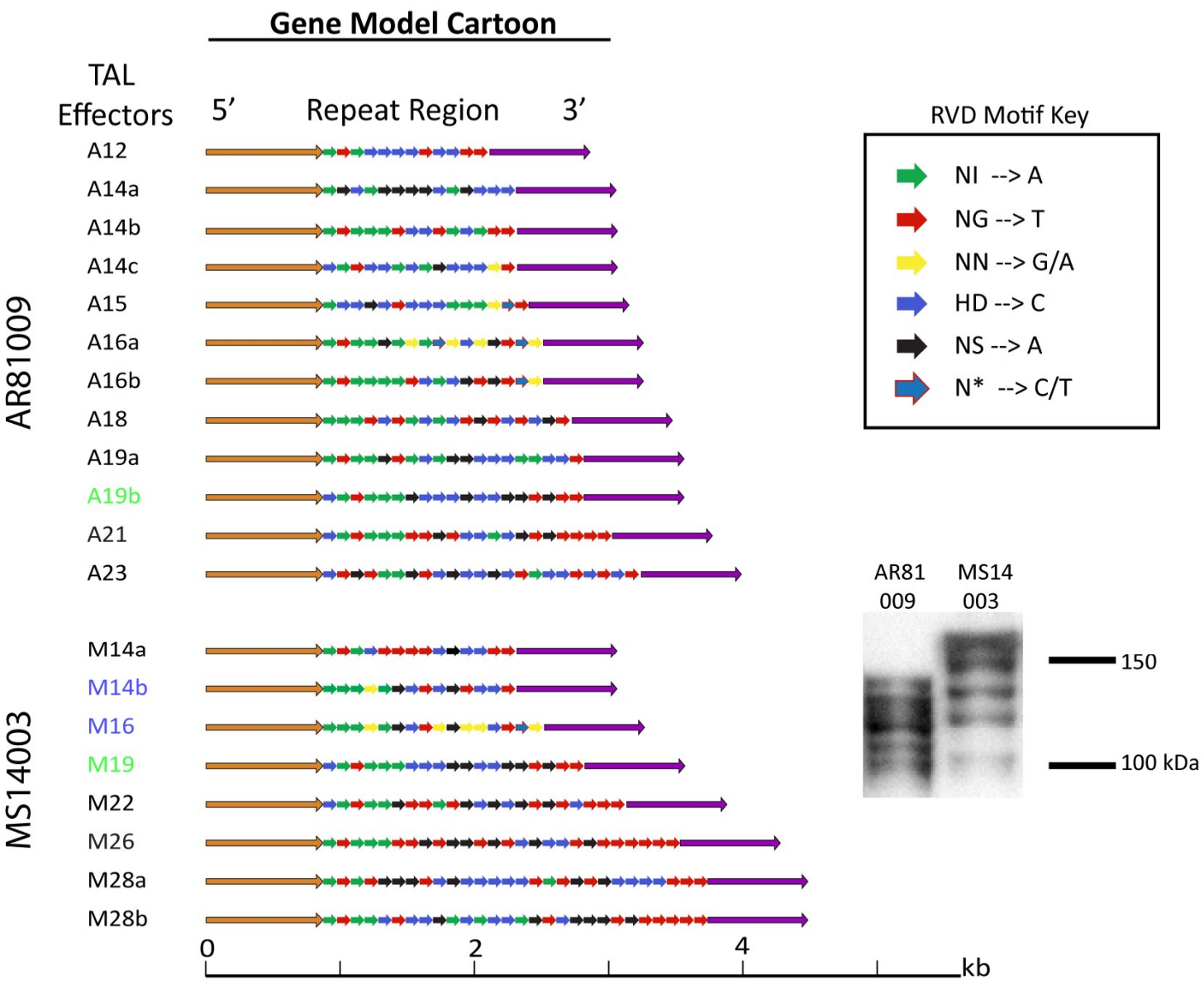
>80
- 70
- 60
- 50
- 40
<30

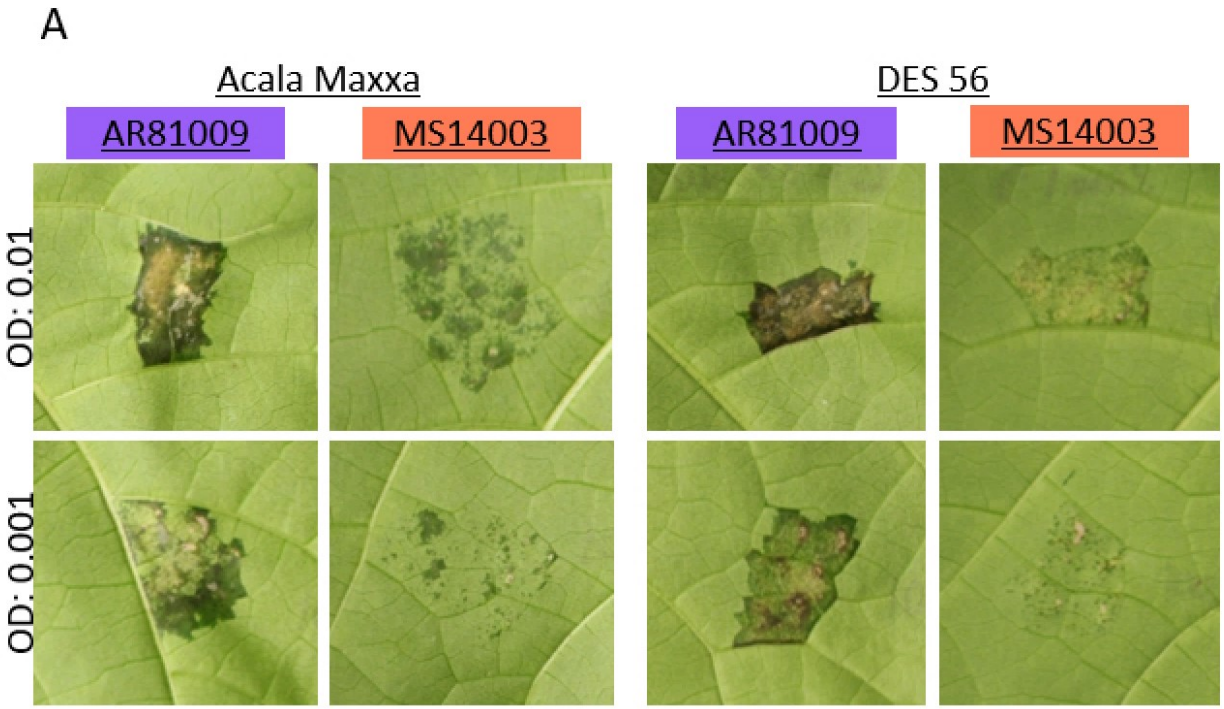
GC Content

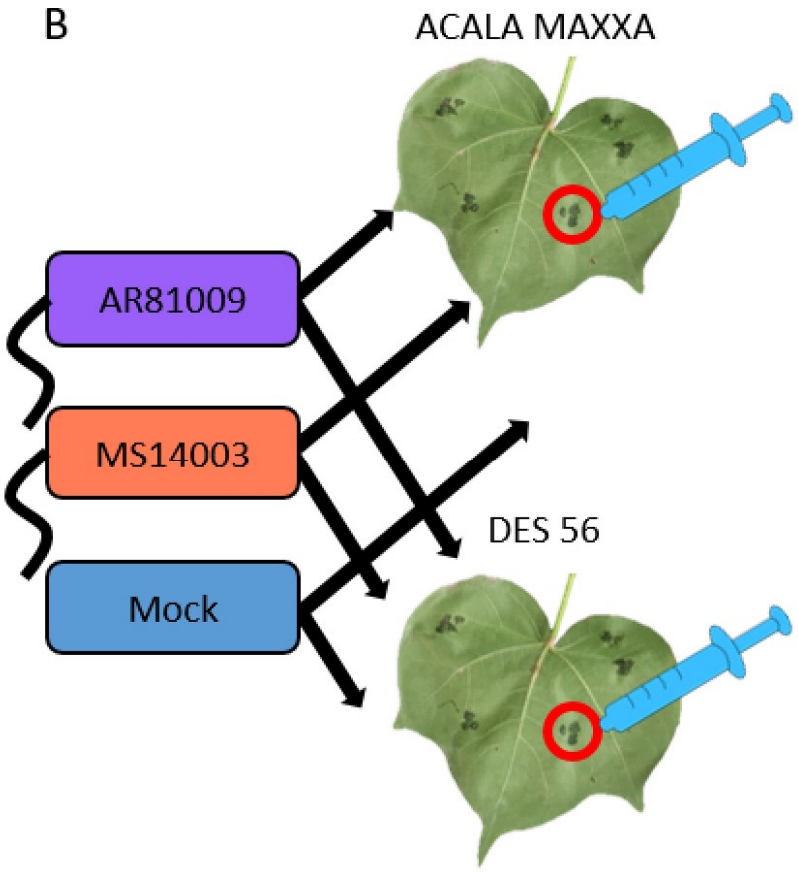
< 0.45
0.45
0.56
0.61
0.65
0.67
0.71
> 0.71

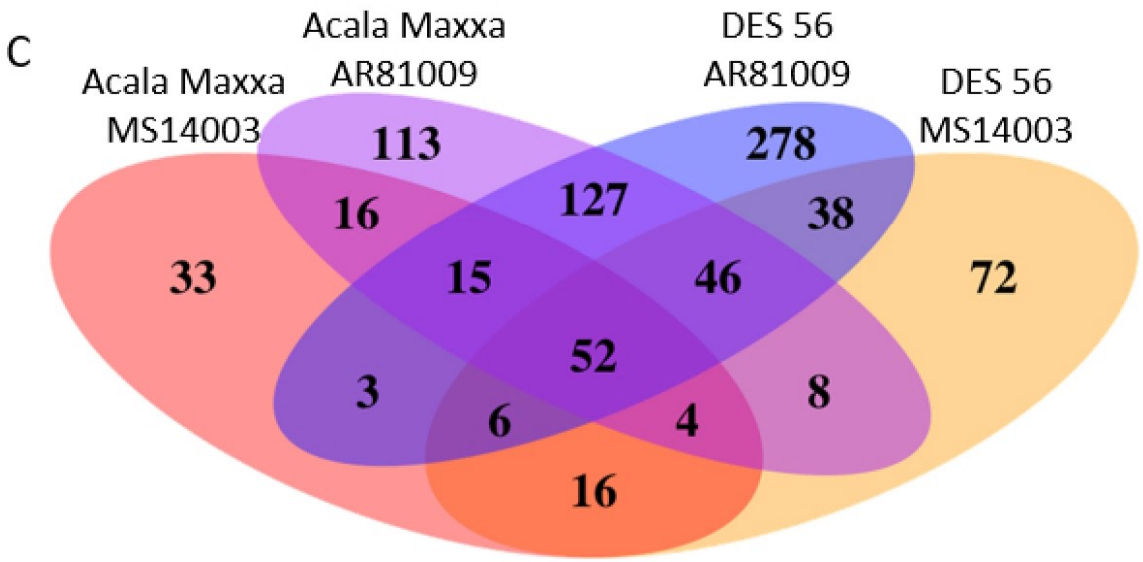
Genes and Homologous Regions

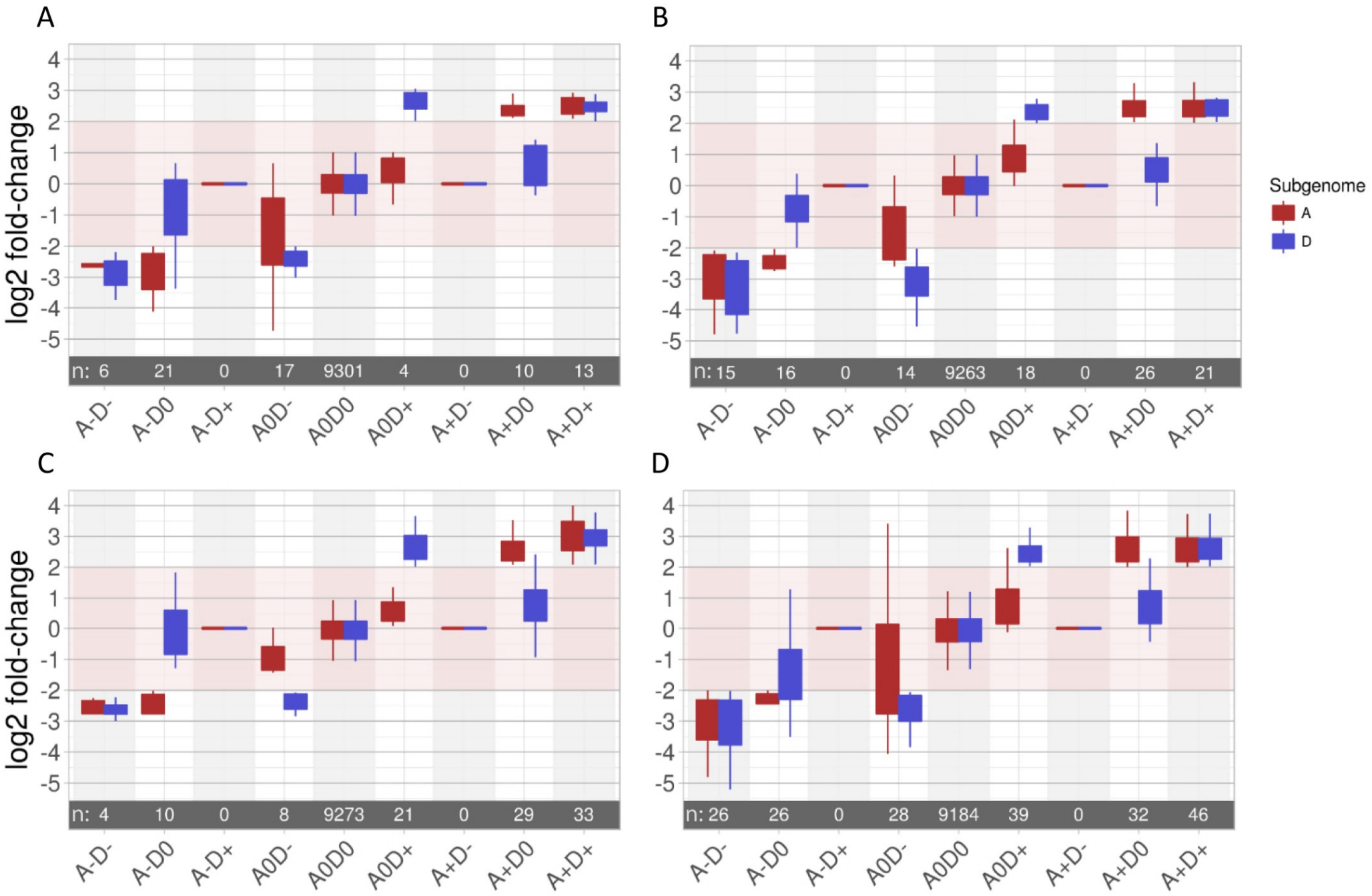
- TAL Effector
- T3E
- Type IV Secretion
- Homologous Region I
- Homologous Region II
- Homologous Region III
- Homologous Region IV

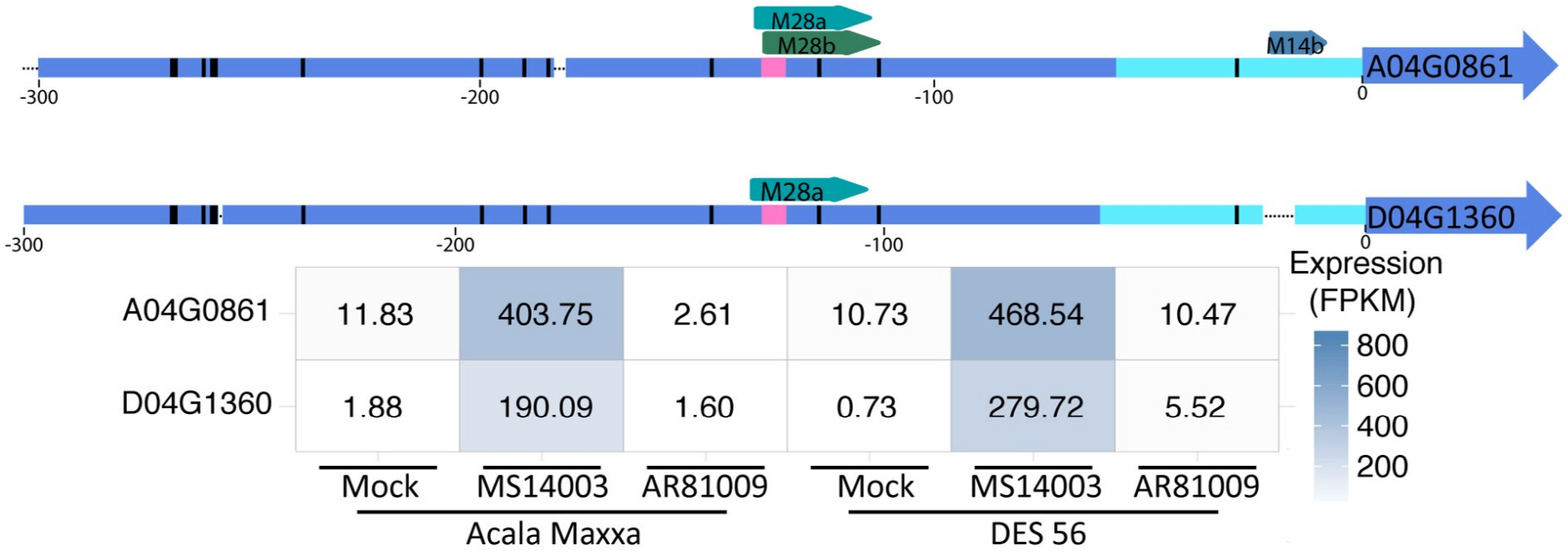

Gene Model Cartoon
TAL Effectors
5’
Repeat Region
3’
AR81009
A12
A14a
A14b
A14c
A15
A16a
A16b
A18
A19a
A19b
A21
A23
MS14003
M14a
M14b
M16
M19
M22
M26
M28a
M28b
0
2
4
kb
RVD Motif Key
NI --> A
NG --> T
NN --> G/A
HD --> C
NS --> A
N* --> C/T
AR81 009
MS14 003
150
100 kDa

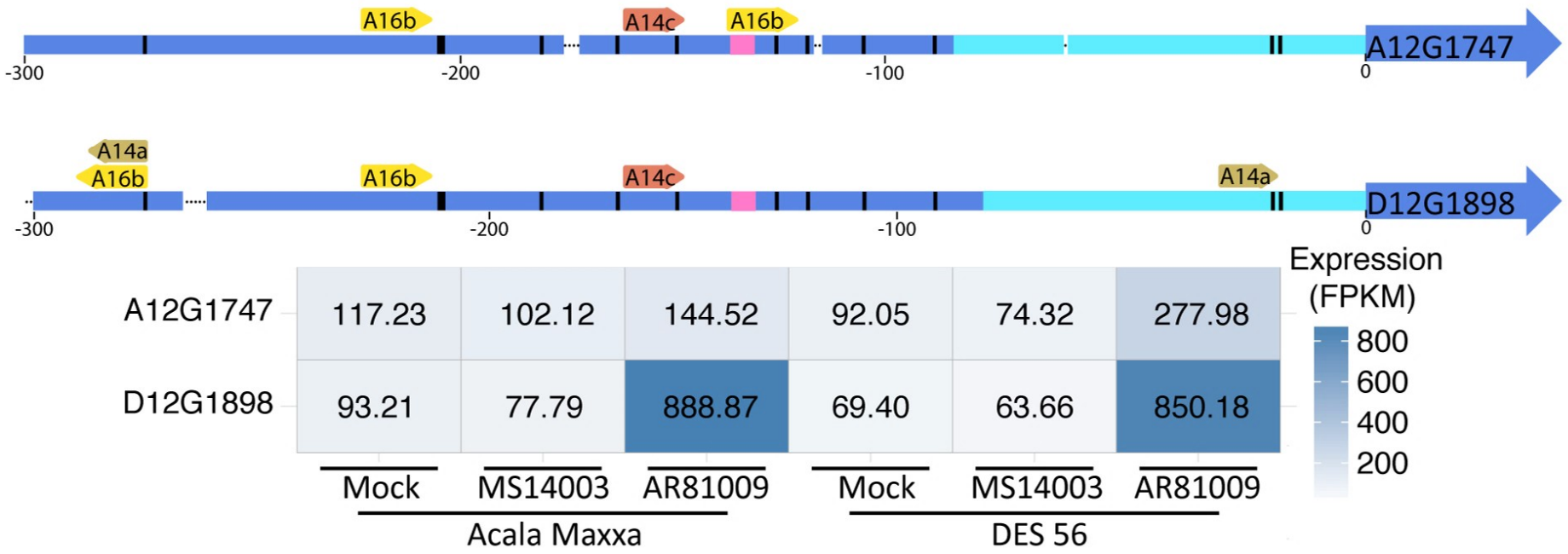

A
Acala Maxxa
DES 56
AR81009
MS14003
AR81009
MS14003
OD: 0.01
OD: 0.001

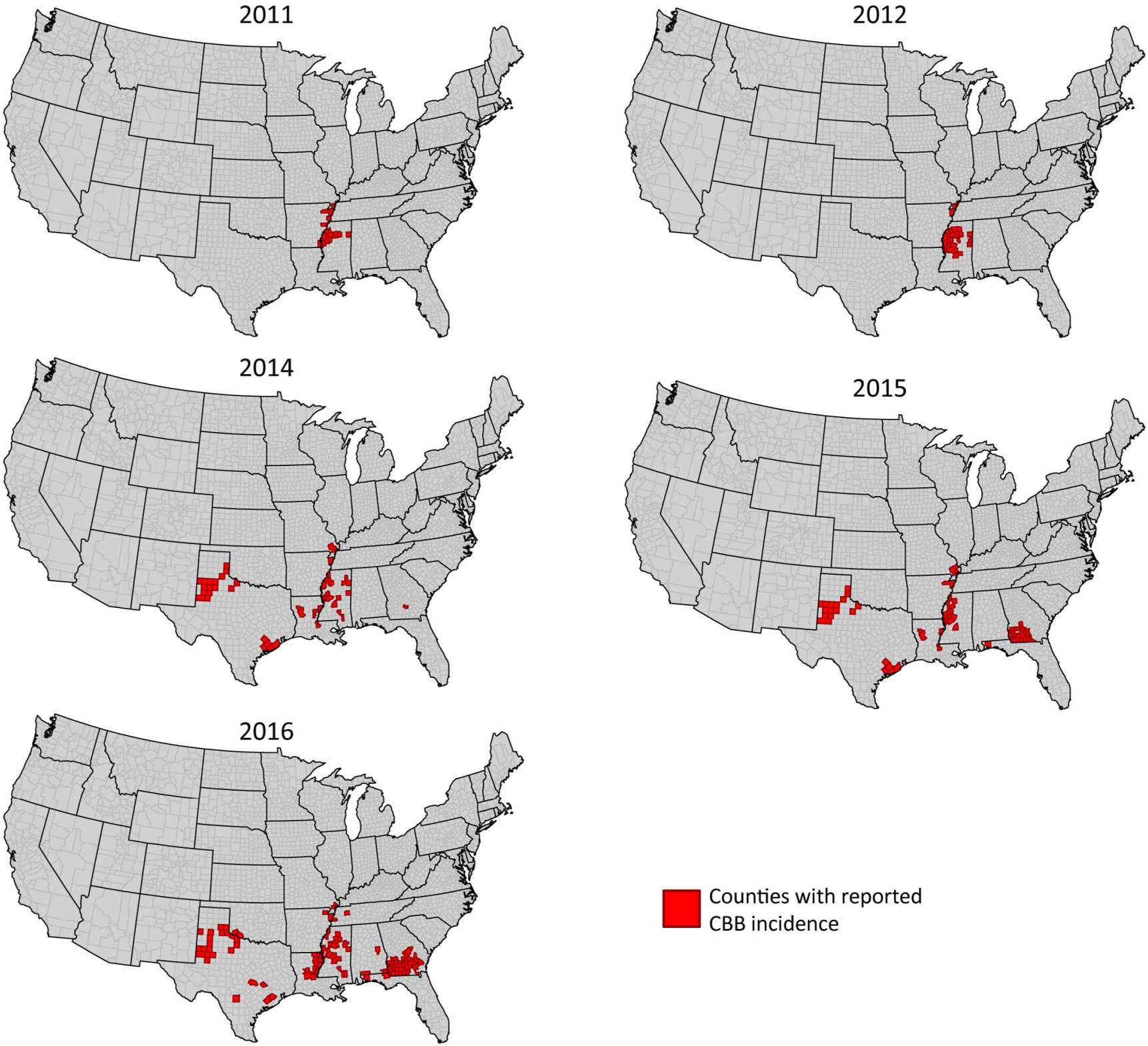

B
ACALA MAXXA
AR81009
MS14003
Mock
DES 56

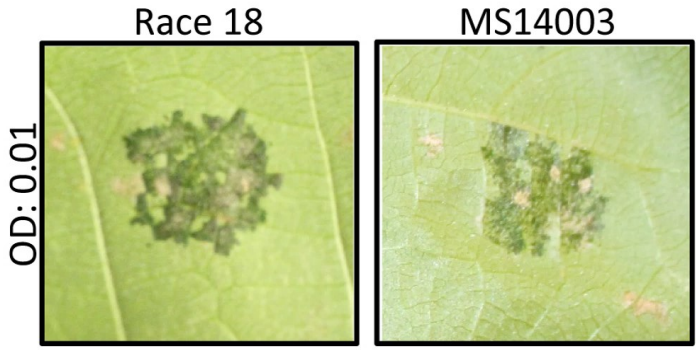

C
Acala Maxxa
MS14003
Acala Maxxa
AR81009
DES 56
AR81009
DES 56
MS14003
113
278
16
127
38
33
15
46
72
52
3
8
6
4
16

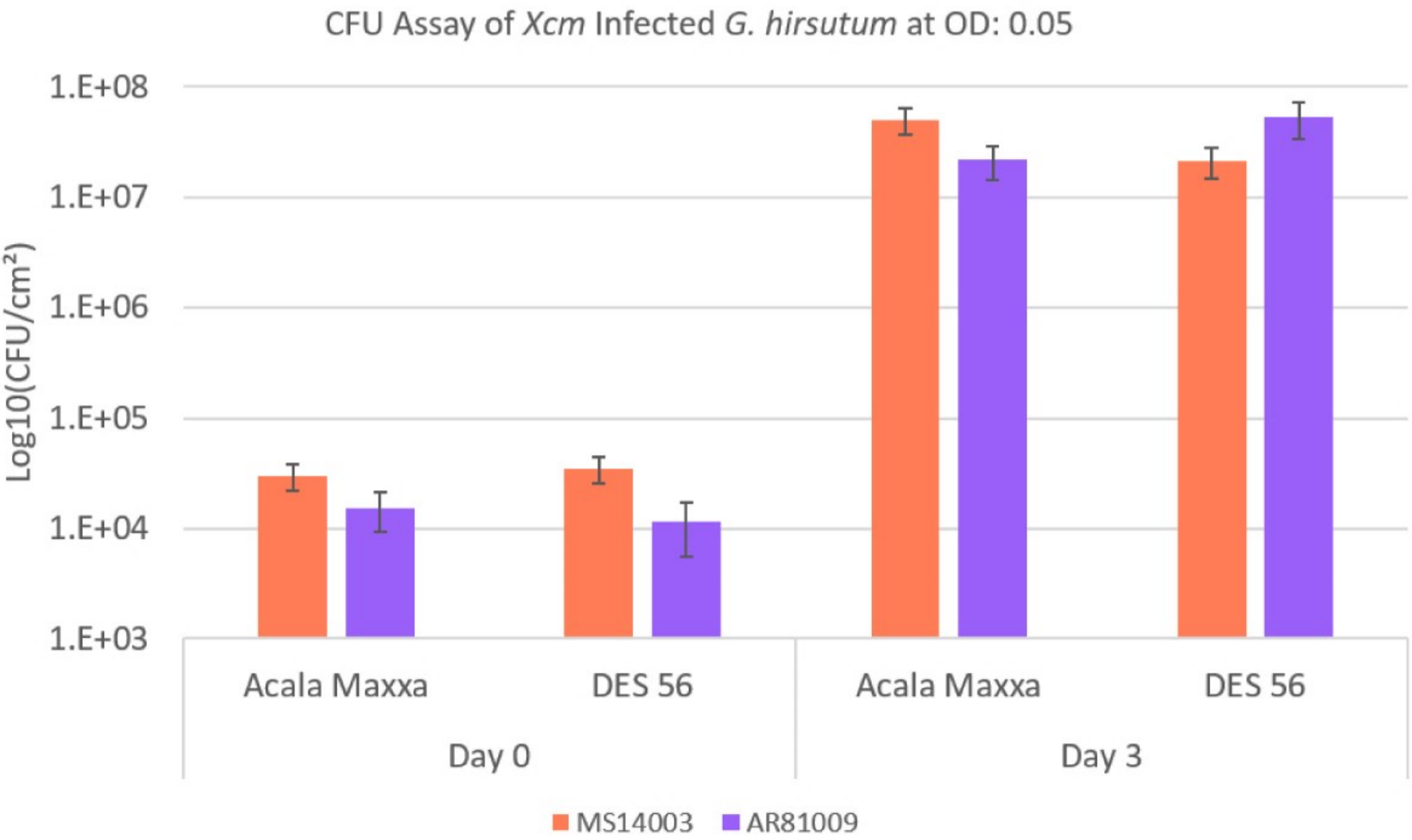

A
B
C
D
log2 fold-change
Subgenome
A
D
n: 6 21 0 17 9301 4 0 10 13
n: 15 16 0 14 9263 18 0 26 21
n: 4 10 0 8 9273 21 0 29 33
n: 26 26 0 28 9184 39 0 32 46
A-D-
A-D0
A-D+
A0D-
A0D0
A0D+
A+D-
A+D0
A+D+

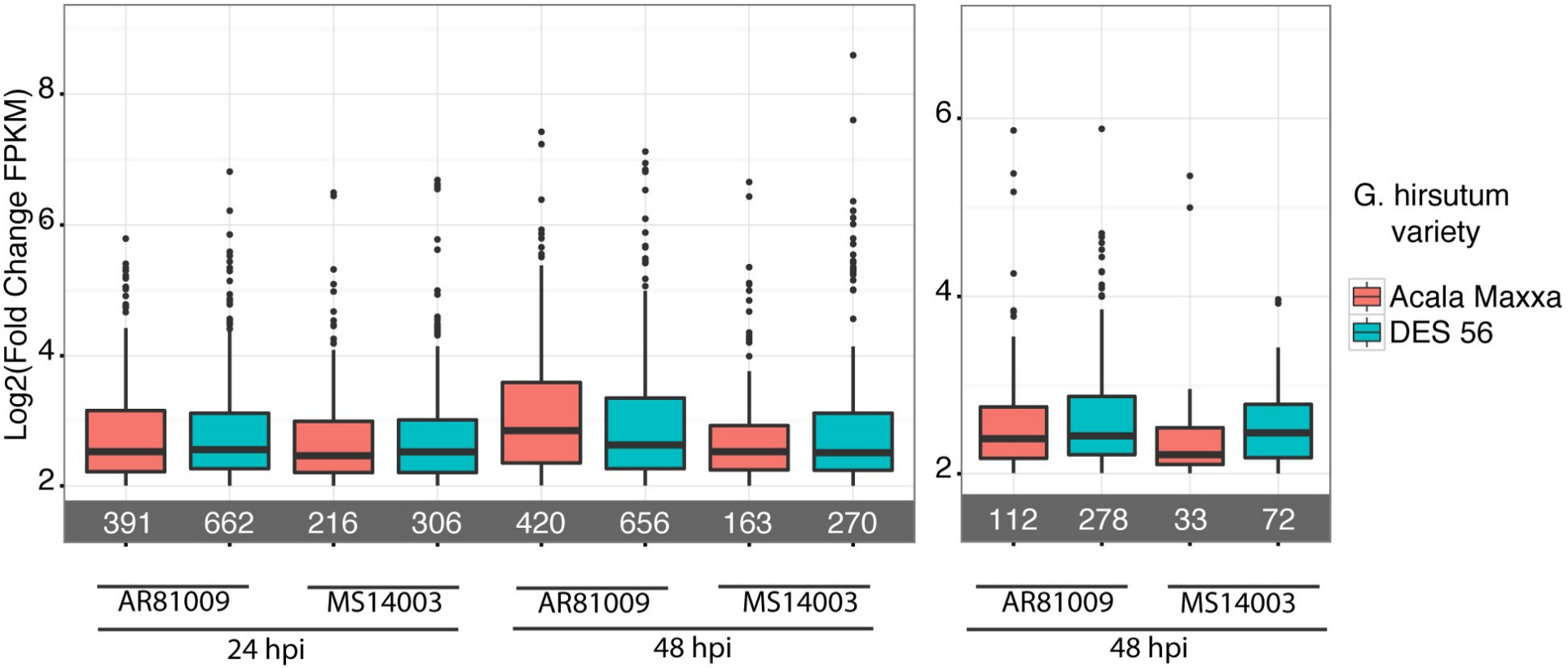

M28a
M28b
M14b
A04G0861
-300
-200
-100
0
M28a
D04G1360
-300
-200
-100
0
Expression (FPKM)
A04G0861
11.83
403.75
2.61
10.73
468.54
10.47
D04G1360
1.88
190.09
1.60
0.73
279.72
5.52
800
600
400
200
Mock
MS14003
AR81009
Mock
MS14003
AR81009
Acala Maxxa
DES 56

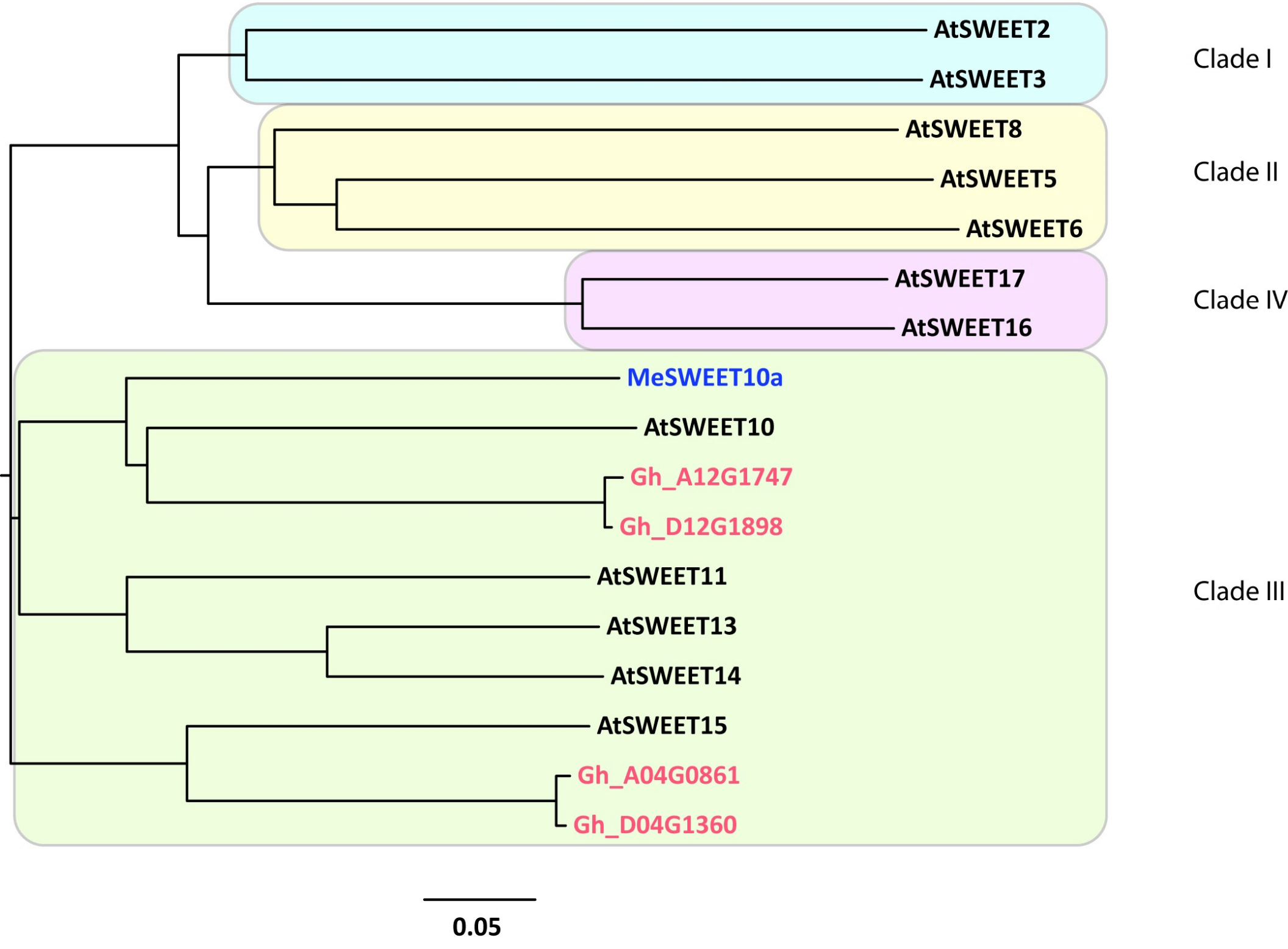

A16b
A14c
A16b
A12G1747
-300
-200
-100
0
A14a
A16b
A16b
A14c
A14a
D12G1898
-300
-200
-100
0
Expression (FPKM)
A12G1747
117.23
102.12
144.52
92.05
74.32
277.98
D12G1898
93.21
77.79
888.87
69.40
63.66
850.18
800
600
400
200
Mock
MS14003
AR81009
Mock
MS14003
AR81009
Acala Maxxa
DES 56

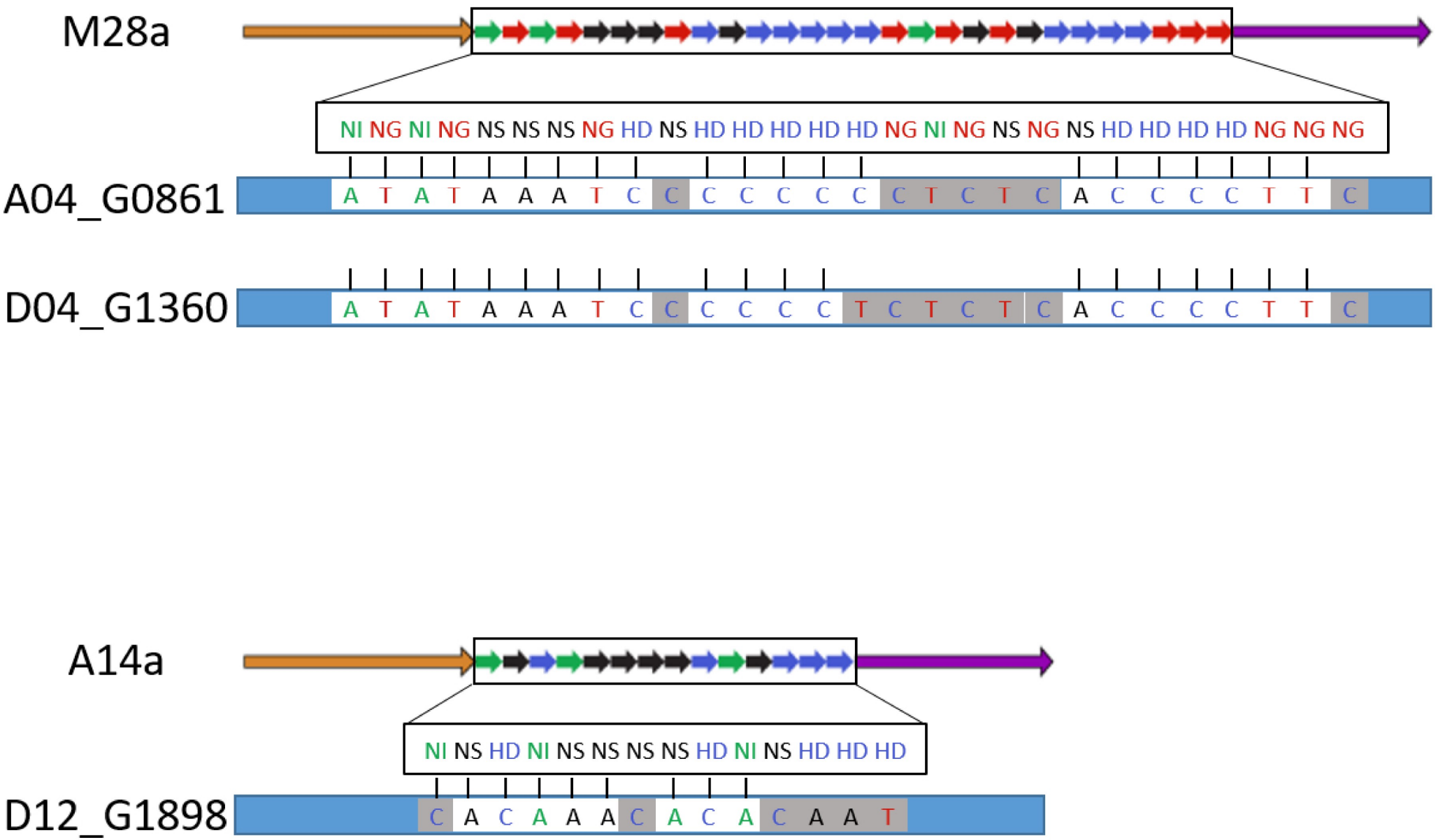

2011
2012
2014
2015
2016
Counties with reported
CBB incidence


Race 18
MS14003
OD: 0.01


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
6
4
2
112
278
33
72
AR81009
MS14003
AR81009
MS14003
24 hpi
48 hpi
AR81009
MS14003
48 hpi
G. hirsutum variety
Acala Maxxa
DES 56


AtSWEET2
AtSWEET3
Clade I
AtSWEET8
AtSWEET5
AtSWEET6
Clade II
AtSWEET17
AtSWEET16
Clade IV
MeSWEET10a
AtSWEET10
Gh_A12G1747
Gh_D12G1898
AtSWEET11
AtSWEET13
AtSWEET14
AtSWEET15
Gh_A04G0861
Gh_D04G1360
Clade III
0.05


M28a
NI NG NI NG NS NS NS NG HD NS HD HD HD HD HD NG NI NG NS NG NS HD HD HD HD NG NG NG
A04_G0861
A T A T A A A T C C C C C C C T C T C A C C C C T T C
D04_G1360
A T A T A A A T C C C C C C T C T C T C A C C C C T T C
A14a
NI NS HD NI NS NS NS NS HD NI NS HD HD HD
D12_G1898
C A C A A A C A C A C A A T