Functional analysis of African *Xanthomonas oryzae* pv. *oryzae* TALomes reveals a new susceptibility gene in bacterial leaf light of rice

Tuan Tu Tran[1,#a¶], Alvaro L Pérez-Quintero[1¶], Issa Wonni[1,#c], Sara C. D. Carpenter[2], Yanhua Yu[1,#d], Li Wang[2], Jan E. Leach[3], Valérie Verdier[1,3], Sébastien Cunnac[1], Adam J. Bogdanove[2], Ralf Koebnik[1], Mathilde Hutin[2,*] and Boris Szurek[1,*]

[1]UMR IPME, IRD-CIRAD-Université Montpellier, Montpellier, France

[2]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, New York, United States of America

[3]Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, United States of America

[#a]Agricultural Genetics Institute, Hanoi, Vietnam

[#b]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, New York, United States of America

[#c]Institut de l'Environnement et de Recherches Agricoles, Laboratoire Mixte International, Observatoire des agents phytopathogènes en Afrique de l'Ouest, Bobo Dioulasso, Burkina Faso

[#d]State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, The Key Laboratory of Ministry of Education for Microbial and Plant Genetic Engineering, and College of Life Science and Technology, Guangxi University, Nanning, Guangxi, China.

*Corresponding authors : boris.szurek@ird.fr; mathildehutin1@gmail.com

[¶]These authors contributed equally to this work.

Abstract (354 words)

Most Xanthomonas species translocate Transcription Activator-Like (TAL) effectors into plant cells where they function like plant transcription factors via a programmable DNA-binding domain. Characterized strains of rice pathogenic *X. oryzae* pv. *oryzae* harbor 9-16 different *tal* effector genes, but the function of only a few of them has been decoded. Using sequencing of entire genomes, we first performed comparative analyses of the complete repertoires of TAL effectors, herein referred to as TALomes, in three *Xoo* strains forming an African genetic lineage different from Asian *Xoo*. A phylogenetic analysis of the three TALomes combined with in silico predictions of TAL effector targets showed that African *Xoo* TALomes are highly conserved, genetically distant from Asian ones, and closely related to TAL effectors from the bacterial leaf streak pathogen *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*). Nine clusters of TAL effectors could be identified among the three TALomes, including three showing higher levels of variation in their repeat variable diresidues (RVDs). Detailed analyses of these groups revealed recombination events as a possible source of variation among TAL effector genes. Next, to address contribution to virulence, nine TAL effector genes from the Malian *Xoo* strain MAI1 and four allelic variants from the Burkinabe *Xoo* strain BAI3, thus representing most of the TAL effector diversity in African *Xoo* strains, were expressed in the TAL effector-deficient *X. oryzae* strain X11-5A for gain-of-function assays. Inoculation of the susceptible rice variety Azucena lead to the discovery of three TAL effectors promoting virulence, including two TAL effectors previously reported to target the susceptibility (*S*) gene *OsSWEET14* and a novel major virulence contributor, TalB. RNA profiling experiments in rice and *in silico* prediction of EBEs were carried out to identify candidate targets of TalB, revealing *OsTFX1*, a bZIP transcription factor previously identified as a bacterial blight *S* gene, and *OsERF#123*, which encodes a subgroup IXc AP2/ERF transcription factor. Use of designer TAL effectors demonstrated that induction of either gene resulted in greater susceptibility to strain X11-5A. The induction of *OsERF#123* by BAI3Δ1, a *talB* knockout derivative of BAI3, carrying these designer TAL effectors increased virulence of BAI3Δ1, validating *OsERF#123* as a new, bacterial blight *S* gene.

Author Summary

The ability of most *Xanthomonas* plant pathogenic bacteria to infect their hosts relies on the action of a specific family of proteins called TAL effectors, which are transcriptional activators injected into the plant by the bacteria. TAL effectors enter the plant cell nucleus and bind to the promoters of specific plant genes. Genes that when induced can benefit pathogen multiplication or disease development are called susceptibility (*S*) genes. Here, we perform a comparative analysis of the TAL effector repertoires of three strains of *X. oryzae* pv. *oryzae*, which causes bacterial leaf blight of rice, a major yield constraint in this staple crop. Using sequencing of entire genomes, we compared the large repertoires of TAL effectors in three African *Xoo* strains which form a genetic lineage distinct from Asian strains. We assessed the individual contribution to pathogen virulence of 13 TAL effector variants represented in the three strains, and identified one that makes a major contribution. By combining host transcriptome profiling and TAL effector binding sites prediction, we identified two targets of this TAL effector that function as *S* genes, one previously identified, and one, new *S* gene. We validated the new *S* gene by functional characterization using designer TAL effectors. Both *S* genes encode transcription factors and can therefore be considered as susceptibility hubs for pathogen manipulation of the host transcriptome. Our results provide new insights into the diversified strategies underlying the roles of TAL effectors in promoting plant disease.

Introduction

X. oryzae pv. *oryzae* (*Xoo*) and *X. oryzae* pv. *oryzicola* (*Xoc*) respectively cause bacterial leaf blight (BLB) and bacterial leaf streak (BLS) of rice, two important constraints for production worldwide, causing yield losses ranging from 0-30% for BLS and up to 50% for BLB [1]. Although originally discovered in Asia, both diseases were reported in West Africa in the 1980's. Since then, reports of both BLB and BLS in different countries across the continent have increased, illustrating their potential effect on African rice production [2]. Although closely related genetically, *Xoo* and *Xoc* cause different symptoms in early stages of infection and colonize the host in different ways. While *Xoo* is a vascular pathogen, entering rice leaves via hydathodes or wounds and colonizing the xylem parenchyma [3], *Xoc* is an intercellular pathogen infecting leaves through stomata and multiplying in the mesophyll apoplast. Genetic diversity analysis revealed that *Xoo* can be classified into distinct genetic lineages according to geographical distribution where Asian and African strains define two separate clades. Intriguingly, African *Xoo* strains seem most closely related to and share several features with *Xoc* [4-6]. Pathogenicity assays grouped African *Xoo* into three novel races [5], and other studies indicated that they trigger specific resistance responses [7].

To successfully colonize their host, both *Xoc* and *Xoo* rely to some extent on host transcriptional changes caused by Transcription Activator-Like (TAL) effectors. TAL effectors form a unique class of type III effectors conserved in many *Xanthomonas spp.* [8]. Upon injection via the type III secretion system, TAL effectors localize to the host nucleus where they directly bind to effector-specific DNA sequences and activate transcription at those locations [9]. When induction of targeted genes promotes bacterial colonization *in planta* and/or disease symptom development, the activated gene is known as a susceptibility (*S*) gene [10]. Alternatively, in some incompatible interactions, TAL effectors may activate an “executor” resistance gene leading to successful host defense [11].

Among the few *S* genes that have been discovered to date, a well-studied case is the *SWEET* gene family of sugar transporters. With one recently described exception [12], induction of any of several clade-3 *SWEET* genes is essential for BLB [13, 14]. It has been suggested that

induction of these genes drives accumulation of sugars in the xylem thus providing nutrients for the bacteria, though other possibilities, including osmotic protection for the bacteria or signaling might be relevant outcome. At least three members of the *SWEET* family in rice are direct targets for several *Xoo* TAL effectors, including *OsSWEET11* induced by PthXo1, *OsSWEET13* induced by PthXo2, and *OsSWEET14*, which is induced by multiple TAL effectors from strains of various geographic origins and diverse genetic lineages [10]. Other known virulence targets in rice, although contributing to a lesser extent to host susceptibility, are *OsTFIIAγ1* and *OsTFX1*, which respectively encode a general transcription factor subunit induced by PthXo7 and a bZIP transcription factor induced by PthXo6 [15-17]. Interestingly, *OsTFX1* is induced by all *Xoo* strains assessed to date, but its role in promoting disease is not yet understood. *S* genes targeted by TAL effectors from other *Xanthomonas* include *UPA20*, which is responsible for mesophyll cell hypertrophy in bacterial spot of pepper [18], *CsLOB1*, which is involved in pustule formation during citrus bacterial canker [19], *OsSULTR3;6*, which encodes a sulfate transporter important in BLS of rice [20], and a gene encoding a bHLH transcription factor that upregulates a pectate lyase in bacterial spot of tomato [21].

Some TAL effectors act as avirulence factors. Most of them trigger defense responses through the induction of so-called executor (*E*) resistance genes like *Xa27*, *Xa23* and *Xa10* in rice [22-24] or *Bs3* [25] and *Bs4c* in pepper [26]. In *Xoo*, all *avr* genes known so far encode TAL effectors, further underscoring the importance of these proteins in BLB. Some TAL effectors act either as virulence or as avirulence factors, depending on the host genotype. An example is AvrXa7, which induces the *S* gene *OsSWEET14* [27] but also triggers resistance in rice cultivars harboring the executor *R* gene *Xa7* [28]. Another type of resistance triggered by TAL effectors was reported recently involving the recognition of *Xanthomonas oryzae* TAL effectors largely non-specifically by the nucleotide-binding leucine-rich repeat rice Xa1 protein [29] or by a yet unidentified protein encoded at the *Xo1* locus [30].

TAL effectors bind DNA in a specific way owing to their central repeat region, which is composed of almost identical tandem repeats (typically ranging from 33 to 35 amino acids), in which residues at positions 12 and 13 are hypervariable (referred to as repeat variable di-residues, i.e. RVDs) and together dictate a specific interaction with a single nucleotide of the

target DNA sequence. Hence, successive RVDs in a TAL effector protein are involved in specific attachment to a sequence of contiguous nucleotides (called the 'effector binding element,' EBE) located in the promoter of the gene to be activated [31, 32]. This code-like mechanism underlying TAL effector-DNA specificity enables prediction of TAL effector binding sites in plant genomes [9, 33].

Characterized African strains of *Xoo* have a reduced number of TAL effectors (usually nine) [5], when compared to strains of *Xoc* (up to 28) [34, 35] and Asian *Xoo* (up to 19) [36, 37]. By contrast, weakly virulent US *X. oryzae* strains are devoid of TAL effector genes [38], a characteristic that can be exploited as a genetic tool to study the function of TAL effectors individually [39].

In previous work, we identified two TAL effectors from African *Xoo* strains that play a role in pathogen virulence by targeting the promoter of *SWEET14* at two different EBEs: TalC from the Burkina Faso strain BAI3 [40] and Tal5 from the Malian strain MAI1 [14]. *SWEET14* is the only *S* gene so far known to be targeted by TAL effectors of African strains, as other African *Xoo* TAL effectors have not yet been characterized.

In this study, we employed long-read, single molecule real-time (SMRT) sequencing technology to define the complete TAL effector repertoires ('TALomes') of the two African *Xoo* strains MAI1, and BAI3, as has been done recently for a number of *Xo* strains [34, 36, 41, 42]. We also sequenced the genome of strain CFBP1947 [5], from Cameroon, which has been sequenced also by another group [43]. These sequences were used to gain insight into the phylogeny of the strains and the evolution of their TALomes. Further, to evaluate the relative virulence function of each TAL effector within a TALome, we assayed the ability of each to increase the virulence of the TAL effector-less *X. oryzae* strain X11-5A. This strain is less virulent than *Xoo* strains, and the introduction of *Xoo* TAL effectors that induce major *S* genes increases symptom development and pathogen growth *in planta* [39]. X11-5A is therefore a useful tool for dissecting the relative virulence contributions of different TAL effectors. This analysis highlighted TalB as a new, major virulence factor. A search for host genes induced *in planta* by TalB identified the transcription factor genes *OsTFX1* and *OsERF#123* as dual virulence targets

of TalB. Activation by designer TAL effectors (dTALEs) delivered by X11-5A or a *talB* knockout derivative of African *Xoo* strain BAI3 revealed *OsERF#123* to be a new class of *S* gene. Our results provide new insights into the diversified strategies underlying the roles of TAL effectors in promoting plant disease.

Results

Whole-genome sequence analysis of three African *X. oryzae* pv. *oryzae* strains

SMRT whole genome sequencing was performed for the African *Xoo* strains MAI1, BAI3 and CFBP1947 isolated from Mali, Burkina Faso and Cameroon, respectively [5]. The genomes were assembled using HGAP.3, resulting in one chromosome per strain. The CFBP1947 sequence agreed with the previously published one [accession number CP013666, 43], and whole genome alignments of all three strains revealed only a small (434 kb) inversion in the BAI3 and MAI1 genomes relative to the CFBP1947 genome, yet several rearrangements relative to representative Asian *Xoc* and *Xoo* whole genome sequences, with overall greater structural similarity to the *Xoc* genomes (Fig. 1B, Fig S1A, and Fig S1B). These alignments also revealed the genomes of Asian *Xoo* strains to contain more repetitive sequences overall than *Xoc* or African *Xoo* (Fig S1A).

Phylogenetic trees based on multi-locus sequence alignment (MLSA) using 33 housekeeping genes were constructed using available fully sequenced or draft *X. oryzae* genomes, revealing that the three African *Xoo* strains form a closely related group along with the *Xoo* strain NAI8 (draft genome, accession number AYSX00000000.1) collected in Niger [5] (Fig 1A). This analysis also showed that, as previously reported [5, 6], African *Xoo* strains are genetically distant from Asian *Xoo* and appear more closely related to *Xoc*. Also, the *X. campestris* pv. *leersiae* strain NCPPB4346, a pathogen of southern cutgrass shown to cluster with *X. oryzae* strains [38], was revealed to be closely related to both *Xoc* and African *Xoo*. Analyses of Average Nucleotide Identity (Fig S1C) [44] and in-silico DNA-DNA Hybridization [45] support the groupings found using MLSA (Fig 1A). African *Xoo* strains were also found to have, like *Xoc*, an overall reduced number of insertion sequences (IS) elements *(*Fig S2), and to contain very few (3 to 5) Tn*Xax1*-like short inverted repeats (Fig 1B), when compared to Asian *Xoo*. This particular type of IS which has been found in association with TAL effectors in *Xoo* as well as in other *Xanthomonas*, is hypothesized to be involved in the horizontal transfer of TAL effector genes and their transposition within a genome [46].

Comparative analysis of African *X. oryzae* pv. *oryzae* TALomes and candidate targets

tal gene sequences were extracted from the genomes using in-house perl scripts. This identified nine TAL effectors in each strain, including the previously characterized TAL effectors TalC from BAI3 [40] and Tal5 from MAI1 [14]. We used the program DisTAL [47] to assess similarities between the African *Xoo* TAL effectors and TAL effectors from all other fully sequenced *Xoo* and *Xoc* strains by alignment of the RVD sequences. A tree based on these alignments shows that the African *Xoo* TAL effectors group together with TAL effectors from Asian and African *Xoc* strains, suggesting a common ancestral origin (Fig 1C).

Consistent with the fact that the few observed Tn*Xax1*-like inverted repeat sequences discussed above do not flank any TAL effector genes on both sides, across strains the genomic positions of the loci corresponding to the nine TAL effector genes are conserved when accounting for the inversion around the replication origin (Fig 1B). And, the TAL effector gene(s) at each locus across strains are clearly orthologous to one another, falling into nine groups based on DisTAL alignments (Fig 1C). Each locus or TAL effector group is thus identified in this paper with a unique letter, assigned based on the number of repeats, from longest (A) to shortest (I). Specific alleles for each *tal* group are indicated with the strain name in subscript. Tal5 from MAI1 which contains 17.5 repeats was consequently renamed $TalF_{MAI1}$. Alternative names for these effectors following the nomenclatures proposed by others [36, 37] are included in Table S1.

The repeat and RVD sequences are highly conserved across half of the African *Xoo tal* loci, with up to 100% identity at the nucleotide level (e.g., group TalC) (Fig 2). However some loci display variation at the RVD level, particularly the groups TalA, TalB, TalF and TalI, which contain respectively two, six, six, and two polymorphic RVD positions, and the TalH group, where a four-repeat deletion is observed in strain BAI3 (Fig 2). Some of the groups display evidence of recombination. For example, the DNA sequences encoding repeats 3 to 8 of $TalF_{MAI1}$ are different from those in $TalF_{BAI3}$ and $TalF_{CFBP1947}$ but identical to repeats 12 to 17 of $TalG_{MAI1}$ (Fig 2, highlighted in thick frames). Similarly, repeats 1 to 5 of $TalF_{BAI3}$ are identical to those of $TalH_{BAI3}$ (Fig 2, dashed frames), thus suggesting recombination between the TalF and TalG loci in MAI1, and between TalF and TalH in BAI3.

For each of the TAL effectors we predicted EBEs in the rice (cv. Nipponbare) promoterome and looked at the overlap between the predictions for each TAL effector (Fig S3). Based on this analysis, the RVD differences for the polymorphic TAL groups TalA, TalB, TalF, TalH and TaIl can be expected to lead to differences in the activation of some targets. For example, only 60% of the predicted targets are shared between $TalA_{MAI1}$ and $TalA_{BAI3}$ or $TalA_{CFBP1947}$ due to the two RVD polymorphisms found (Fig 2). The differences were more dramatic in the highly variable TalF group where less than 10% of predicted targets for $TalF_{MAI1}$ overlap with those of $TalF_{BAI3}$ or $TalF_{CFBP1947}$ (Fig S3).

We next compared the prediction results with expression data. For this, RNA-seq libraries were generated from rice leaves harvested 24 hours after inoculation with the strain MAI1 or a water control. The reads were mapped to the rice genome, and differential expression was determined using the Tuxedo suite [48-50]. 1366 genes were found to be significantly induced in response to strain MAI1. Of these, 152 genes had predicted EBEs in their promoters for MAI1 TAL effectors and were considered candidate TAL effector targets (Table S2). In the case of the strain BAI3, micro-array data were already available for rice plants inoculated with this strain compared to a water control, collected at 24 hours post inoculation (GEO accession GSE19844) [40]. Out of 377 genes induced in response to BAI3, 45 contained predicted EBEs for BAI3 TAL effectors and were therefore considered candidate TAL effector targets (Table S3). We then compared the lists of TAL effector targets for both strains and found that, as shown in Table 1, 20 pairs of TAL effectors and targets were shared between strains BAI3 and MAI1, including *OsSWEET14* being targeted by both $TalC_{MAI1}$ and $TalC_{BAI3}$.

Systematic functional analysis of African *Xoo* TALomes

We reasoned that the list of shared targets described above may include novel *S* genes that contribute to the severity of disease caused by each strain. As a first step to identify such genes, we subcloned the TALomes of strains BAI3 and MAI1 to test for virulence contributions by gain of function assays using X11-5A, the *Xanthomonas oryzae* strain that is naturally devoid of TAL effector genes [38]. A strategy based on the enrichment of *Bam*HI-TAL effector sequences from genomic DNA libraries [51], led to the isolation of all nine *tal* genes from strain MAI1 and eight

from strain BAI3 (we were not able to isolate *talB*$_{BAI3}$ despite several attempts). The sequences of the cloned genes matched those from the genome assembly.

For functional assays, we aimed at testing TAL effectors representing most of the TALome diversity found in African *Xoo* so far. To that end we selected the nine TAL effectors from MAI1, four from BAI3 that displayed variations at the RVD level with respect to their counterpart in MAI1 (namely TalA$_{BAI3}$, TalF$_{BAI3}$, TalH$_{BAI3}$ and TalI$_{BAI3}$, Fig 2), and, for reference and as a positive control the previously characterized major virulence factor TalC$_{BAI3}$ [40]. Each *tal* gene was conjugated individually into X11-5A and the resulting strains were assayed by leaf clip inoculation to leaves of the susceptible rice variety Azucena for increases in virulence (lesion length after 15 days) relative to an empty vector conjugant. Three of the nine MAI1 TAL effectors, namely TalB$_{MAI1}$, TalC$_{MAI1}$ and TalF$_{MAI1}$, enhanced X11-5A virulence (Fig 3). TalC$_{MAI1}$ and TalF$_{MAI1}$ are discussed below, and TalB$_{MAI1}$ is discussed in the following section.

The RVD array of TalC$_{MAI1}$ is 100% identical to that of TalC$_{BAI3}$ (Fig 2), which activates the expression of the *S* gene *OsSWEET14* [40], thus explaining its ability to increase the virulence of X11-5A. TalF$_{MAI1}$ was shown previously also to induce *OsSWEET14* by binding at a non-overlapping EBE [14], likewise explaining its ability to enhance X11-5A virulence. To our knowledge, this is the first example of a strain carrying two TAL effectors that target the same *S* gene promoter at unrelated EBEs, and may reflect a pathogen adaptation to overcome host loss-of-susceptibility alleles such as *xa41*(t) that disrupt an EBE [52]. In contrast to TalF$_{MAI1}$, TalF$_{BAI3}$ did not increase the virulence of X11-5A (Fig 3). Indeed no EBE is predicted in the *SWEET14* promoter for TalF$_{BAI3}$, which as noted above differs at 6 out of 18 RVDs from TalF$_{MAI1}$.

TalB$_{MAI1}$ contributes to virulence by targeting both a known *S* gene and a new one

To understand how TalB$_{MAI1}$ promotes disease, we sought to identify its relevant targets. Among the nine genes induced by BAI3 that each has a predicted TalB EBE in its promoter (Table S3), we focused on two that were also induced by MAI1 (Table S2). *Os09g39810*, which was the most highly induced by MAI1 and had the highest TalB$_{MAI1}$ EBE prediction score (Table 1,

Table S2), and *Os09g29820* (*OsTFX1*), which was previously reported as a BLB *S* gene [15, 16]. *Os09g39810* is annotated as an expressed protein containing an AP2 domain and is from here on referred to as *OsERF#123* [annotated as an expressed protein containing an AP2 domain, 53]. Both $TalB_{MAI1}$ and $TalB_{BAI3}$, despite their six polymorphic RVDs, are predicted with high scores to bind to the same EBE in its promoter. This EBE ranks number 30 in the genome for $TalB_{MAI1}$ and number 6 for $TalB_{BAI3}$ (Table 1, Table S2 and Table S3). *OsTFX1* encodes a bZIP transcription factor, it is highly induced upon infection of rice with MAI1, and its predicted EBE ranks 570 in the genome for $TalB_{MAI1}$ and 135 for $TalB_{BAI3}$.

To ascertain whether *OsERF#123* and *OsTFX1* are bona fide targets of $TalB_{MAI1}$, rice leaves were infiltrated with MAI1 and X11-5A(p*talB*$_{MAI1}$), respectively carrying *talB*$_{MAI1}$ on the chromosome or on a plasmid, and X11-5A carrying an empty vector. As expected, RT-PCR amplification showed that both *OsERF#123* and *OsTFX1* were highly and specifically induced upon infection of rice with the strains expressing $TalB_{MAI1}$ (Fig S4). With the aim of deciphering which of the two targets is biologically relevant, i.e. an *S* gene, we engineered dTALEs to specifically induce each of them individually. Two dTALEs per candidate gene were devised, targeting sites located in the vicinity of the native EBE (Fig 4A). Leaf-clipping inoculation assays using X11-5A derivatives showed that dTALE-mediated induction of *OsERF#123* enhanced virulence similarly to $TalB_{MAI1}$ (Fig 4B). dTALEs for *OsTFX1* also conferred higher virulence to X11-5A albeit to a lower extent than the *OsERF#123* dTALEs. The virulence contribution was similar to that conferred on X11-5A by TAL effector PthXo6 from the Asian *Xoo* strain $PXO99^{A}$ [15, 16], which was included as a positive control. Importantly, X11-5A strains expressing the *OsERF#123*-targeted dTALEs, $dTALE_{ERF-1}$ or $dTALE_{ERF-2,}$ induced *OsERF#123* and not *OsTFX1*, as shown by qRT-PCR analysis (Fig 4C). Last, we evaluated patterns of *OsERF#123* transcript initiation using 5'-RACE experiments, and discovered that TalB initiates transcription primarily at sites residing 41-69 bp downstream of the 3' end of its EBE (Fig S5), as is the case for many TAL effectors and their targets. Altogether, our results reveal that $TalB_{MAI1}$ is a virulence factor that activates two *S* genes, *OsTFX1* and *OsERF#123*, each of which functions independently.

To confirm the function of TalB as a virulence factor for African *Xoo*, we generated a library of BAI3 TAL effector gene mutants by transformation with the suicide plasmid pSM7 [20] and identified a *talB* mutant by screening for strains unable to induce *OsERF#123*, assessed by qRT-PCR 24h after infiltration (data not shown). We identified such a strain and designated it BAI3Δ1 (Fig 5A). Assayed by leaf clip inoculation, BAI3Δ1 showed reduced virulence relative to the wildtype strain (Fig 5B). Induction of *OsERF#123* and virulence were fully restored by plasmid-borne $talB_{\mathrm{MAI1}}$, while induction of OsERF#123 was fully restored and virulence partially restored by either of the two *OsERF#123*-targeted dTALEs (Fig 5A and Fig 5B), thus establishing that TalB contributes to virulence in its native context by activating the newly identified *S* gene *OsERF#123*.

Discussion

In this study we sequenced the genomes of three African strains of *Xoo* and characterized their TALomes systematically using a gain-of-virulence assay, providing a first glimpse into the function and evolution of TAL effectors in this particular group of *X. oryzae* strains. Based on RNA profiling and *in silico* prediction, and both gain and loss of virulence assays and targeted dTALEs, we identified TalB as a virulence factor and TalB target *OsERF#123* as a novel *S* gene for bacterial blight of rice.

Our analysis revealed each African *Xoo* strain to contain nine TAL effectors that can be classified according to their repeat and RVD sequences [36, 47] in nine groups, corresponding to nine syntenic loci. Five of these loci showed some degree of variation in RVD sequence between strains, variation that can translate into different effects on the plant transcriptome. When comparing these sequences with the results of inoculation assays (see below), no clear relation was seen between contribution to virulence and evolutionary conservation. Among the three TAL effectors from MAI1 that contributed to virulence, $TalC_{MAI1}$ is highly conserved (i.e. identical at the nucleotide level) in all three strains and is expected to contribute to virulence in the same way in all strains harboring it, by inducing *OsSWEET14*, as previously reported [40]. Meanwhile, $TalB_{MAI1}$ shows some RVD differences from $TalB_{BAI3}$ and $TalB_{CFBP1947}$; nonetheless, the two latter alleles are expected also to induce the relevant targets *OsTFX1* and *OsERF#123* and thus to contribute to virulence similarly. In the case of $TalB_{BAI3}$ there is transcriptomic (Table 1) and functional (Fig 5) evidence that this is the case. Finally, $TalF_{MAI1}$, which also induces *OsSWEET14*, at an EBE distinct from that of TalC, belongs to a highly variable group in which the alleles in strains BAI3 and CFBP1947 may be inducing completely different targets. And at least in the case of BAI3, no significant contribution to virulence by its allele $talF_{BAI3}$ was observed in inoculated Azucena plants (Fig 3). The degree of variation in each of these groups may be dictated by interaction with the host, specifically the degree of variation at the target promoters across different host genotypes.

The evolution of the *talF* locus is of particular interest. The sequence analysis presented in this work hints at possible recombination events leading to variants of this TAL effector in the strains BAI3 and MAI1, where in each case a set of contiguous repeats was found to match exactly

those of another TAL effector in the genome (different in each case). The overall high similarity of TAL effector repeats likely favors recombination, and recombination of repeats may indeed be responsible for fast evolution of TAL effector variants in *Xanthomonas* [46, 47, 54]. Recombination in this case may have been further favored by the genomic context given that the loci involved are found close to the replication origin (Fig 1B), a region known to be a hotspot for recombination [55].

The fact that $talF_{\mathrm{MAI1}}$ targets the *S* gene *OsSWEET14* convergently with TalC in MAI1 raises questions about the selective forces acting upon this locus. It is possible that the MAI1 variant $talF_{\mathrm{MAI1}}$ was positively selected after a recombination event allowing for its induction of *OsSWEET14;* indeed, the recombined repeats match the corresponding nucleotides in the EBE almost perfectly according to the TAL code (Table 1, Fig 6). However, how could this selection have occurred in the presence of the *OsSWEET14* inducer $\mathrm{TalC}_{\mathrm{MAI1}}$ in the same strain? One possibility is that these two TAL effectors have been maintained in the *Xoo* strain MAI1 in response to genomic variants in the host population. Indeed, at least one type of variation (named *xa41*(*t*)) has been described in African rice varieties that impairs binding of $\mathrm{TalF}_{\mathrm{MAI1}}$ to its EBE in the promoter of *OsSWEET14* [52]. It is possible that a similar loss-of-susceptibility allele that impairs binding of $\mathrm{TalC}_{\mathrm{MAI1}}$ also exists in rice and is driving the selection of $\mathrm{TalF}_{\mathrm{MAI1}}$. Alternatively, $\mathrm{TalF}_{\mathrm{MAI1}}$ may be an ancient TAL effector that was lost in strains BAI3 and CFBP1947 after escaping selection due to its redundancy with TalC. Studying the TALomes of more African *Xoo* strains and genetic variation at *OsSWEET14* will likely clarify the evolutionary history of these loci.

With the aim of evaluating the function of the TAL effector repertoires of two African *Xoo* strains, we used the TAL effector-less strain of *X. oryzae* X11-5A [38]. Previous studies had shown that three TAL effectors that contribute to virulence of their strains of origin by targeting clade-III members of the SWEET family of sugar transporters conferred higher virulence to X11-5A [39]. Here, we took advantage of this gain-of-function system to systematically decipher the individual function of each TAL effector within the TALomes of the three African *Xoo* strains. This approach is easier than loss-of-function mutagenesis approaches, and unlike such approaches, it avoids the pitfall of redundancy among effectors. Indeed, the role of $talF_{\mathrm{MAI1}}$

could not have been identified by loss-of-function screening since *talC*$_{MAI1}$ activates the same *S* gene it does.

Our gain-of-function approach led also to the discovery that TalB$_{MAI1}$ is a novel major virulence factor. Combining *in silico* prediction of EBEs and transcriptomic data, we pinpointed two candidate targets, *OsERF#123* and *OsTFX1*. By specifically inducing the expression of each using dTALEs, we found that each target could explain some of the disease promoting effect of TalB$_{MAI1}$. Interestingly *OsTFX1* had been reported as a *S* gene for Asian *Xoo* strains and is known to be targeted by the TAL effector PthXo6 of *Xoo* strain PXO99^{A} [15, 16]. Our data show that African *Xoo* strains also have acquired the capacity to activate *OsTFX1* albeit with a TAL effector harboring a totally different RVD array (Fig 6). In addition to the well-known example of *OsSWEET14* whose promoter is targeted by several unrelated TAL effectors [56], *OsTFX1* represents a new case of functional convergence wherein two completely unrelated TAL effectors in strains of different geographic origin and distant genetic lineages have evolved to bind to overlapping EBEs. In line with this conclusion, *OsTFX1* is induced by several strains of Asian *Xoo* [15], suggesting that its induction is as essential as the induction of clade-III sugar transporters of the SWEET family. Yet, the regulatory targets of *OsTFX1* and the physiological consequences of its induction remain to be characterized.

The other gene targeted by TalB$_{MAI1}$, *OsERF#123*, encodes a putative protein of 131 amino acids (BAT09530) characterized by an AP2/ERF domain (spanning from residues 16 to 75). BLAST analysis showed that *OsERF#123* is single-copy in the rice cv. Nipponbare genome and broadly conserved across the *Oryza* genus (not shown). It attracted our attention because it was the most highly induced of the candidate TalB$_{MAI1}$ targets, and it had never been described before. In a genome-wide analysis of the ERF family in rice and Arabidopsis, Nakano *et al*. [53] showed that rice ERF proteins sort into 15 groups which can be further divided into subgroups, based on sequence polymorphisms. Interestingly, group IX consists of proteins specialized in mediating disease resistance. OsERF#123 was assigned to the subgroup IXc which includes a few proteins whose biological function is related to the induction of defense responses such as SlPti5, described as a pathogenesis-related (PR) protein [57, 58]. Nevertheless, it cannot be excluded that OsERF#123 acts as a negative regulator of defense responses as is the case for some WRKY

transcription factors that tightly regulate other, positive regulatory WRKY factors [59]. In such a scenario, TAL effector-mediated induction of *OsERF#123* could result in increased susceptibility. Strikingly, surveying publicly available expression data, we found evidence of *OsERF#123* expression only in response to African *Xoo* [40], *Xoc* [34], or chilling stress [60]. Semi-quantitative RT-PCR confirmed that a panel of 15 African *Xoo* strains originating from Burkina Faso and Mali induced *OsERF#123*, while the Asian *Xoo* strain PXO99^{A} did not (not shown). Also, examination of RNAseq data generated from rice leaf tissue 48h after inoculation with ten geographically diverse *Xoc* strains [34], revealed that *OsERF#123* was highly induced by nine out of the ten strains. This result suggests that *OsERF#123* might be targeted by strains of *Xoc* to promote bacterial leaf streak and is in agreement with our current understanding that African *Xoo* are closer to *Xoc* than to Asian *Xoo* strains. In separate work, Wang *et al*. [61] have recently shown that Tal3c of the *Xoc* strain BLS256, which is entirely distinct from $TalB_{MAI1}$ in RVD sequence, is responsible for *OsERF#123* induction, highlighting yet another striking case of evolutionary convergence.

Altogether, our results show that despite having a reduced TALome relative to their Asian counterparts or *Xoc*, African *Xoo* strains have successfully adapted that TALome to target physiologically relevant genes. As evidence, we show two features of African *Xoo* TALomes that have not been reported previously. First, we show redundant and convergent targeting of the same *S* gene by two TAL effectors from one strain (MAI1) with different RVD sequences: $TalC_{MAI1}$ and $TalF_{MAI1}$ (Fig 6), a feature that may have arisen as a response to loss-of-susceptibility alleles in the host population [52]. In the case of $TalF_{MAI1}$ we also hint for the first time at a recombination event leading to the evolution of virulence activity.

Second, we describe for the first time a TAL effector, $TalB_{MAI1}$, with dual targeting activity, binding to two different EBEs in unrelated genes, where both targets contribute to susceptibility. While TAL effectors often have been shown to induce multiple genes, so far at most only one of the induced genes has been shown to be required for virulence [19, 20]. Interestingly, in at least one case, the TAL effector may have higher affinity for the physiologically relevant target, based on EBE prediction scores [20]. It is then curious that the BAI3 and CFBP1947 variants of TalB are predicted to bind with even higher affinity to the EBEs of both *OsTFX1* and *OsERF#123*

than $TalB_{MAI1}$ is (Table 1), suggesting that the *talB* locus is undergoing some evolutionary fine-tuning of its binding specificity to target both genes simultaneously. Dual targeting of two susceptibility genes has recently been hypothesized for another African *Xoo* TAL effector, TalC, although its secondary target is yet to be discovered [12] .

Finally, both TalB targets, *OsTFX1* and *OsERF#123*, similarly to *SWEET* genes may constitute widely exploited hubs for pathogenicity, targeted convergently by distinct TAL effectors of different strains. Unlike *SWEET* genes, as transcription factor genes they may also serve uniquely as regulatory hubs for broad manipulation of the host transcriptome by TAL effectors, although their targets remain to be elucidated. Since *OsERF#123* is induced by African *Xoo* strains and diverse *Xoc* strains, it may be of particular interest as an *S* gene for both bacterial blight and bacterial leaf streak. However its role in bacterial leaf streak has yet to be determined [61]. Along with *OsHen1* (LOC_Os07g06970), *OsERF#123* is so far one of only a few genes targeted both by *Xoc* and by *Xoo* TAL effectors [32, 62, 63]. If it contributes to leaf streak susceptibility it would be the first example of a susceptibility gene for both diseases.

Deciphering the virulence function of individual TAL effectors and identifying the host target(s) is key to discovering or engineering novel sources of resistance to combat bacterial blight. This is particularly true in the case of African *Xoo* strains, which differ in many aspects from their Asian counterparts, and for which sources of resistance effective against Asian strains may not be effective. In this study, we have laid a foundation for such inquiry and discovered both convergent targeting (of *S* genes *OsTFX1* and *OsSWEET14*) by Asian and African *Xoo* strains, as well as a novel susceptibility gene (*OsERF#123*) for bacterial blight that is targeted by African *Xoo* strains and diverse *Xoc* strains, but so far not by Asian *Xoo* strains. Further experiments are in progress to investigate how this novel class of virulence target confers susceptibility to bacterial blight, and whether it plays a role in leaf streak.

Methods

Bacterial strains, growth conditions and plasmids

The bacterial strains used in this study were *Escherichia coli* DH5α (Stratagene, La Jolla, CA, USA), *X. oryzae* pv. *oryzae* strains BAI3, MAI1 and CFBP1947 [5], and *X. oryzae* strain X11-5A [38]. *E. coli* cells were cultivated at 37° C in Luria-Bertani broth (LB) medium, and *X. oryzae* strains at 28°C in PSA medium (10 g l^{-1} peptone, 10 g l^{-1} sucrose, 1 g l^{-1} glutamic acid, 16 g l^{-1} agar). Plasmids were introduced into *E. coli* by electroporation and into *Xo* by biparental conjugation using *E. coli* strain S17-1 as previously described [40]. The plasmid used to subclone individual *tal* genes is a derivative of pSKX1-*talC* obtained upon digestion with *Bam*HI and direct self-religation, leading to a construct harboring the ATG start codon of *talC* and 151 3' bp [14]; expression is driven by the *lac* promoter, which is constitutive in *Xanthomonas*. BAI3Δ1 was generated by transformation of BAI3 with the suicide plasmid pSM7 and mutant strains were selected with kanamycin at 25µg/mL [20].

Plant material and plant inoculations

Experiments were performed under glasshouse conditions under cycles of 12 hours of light at 28-30°C and 80% relative humidity (RH), and 12 hours of dark at 25°C and 70% RH. The rice varieties Azucena and IR24 were inoculated by leaf-clipping of four-week-old plants and evaluated as reported previously [40]. In each experiment, at least eight leaves from eight individuals were inoculated and experiments were replicated three times independently. Growth curves experiments were achieved as reported previously [40].

DNA preparation and purification

Total genomic DNA of the three African *Xoo* strains was prepared using the QIAGEN midi-prep genomic DNA preparation protocol for bacteria and using the Genomic-tip 100/G (Qiagen, Valencia, CA, USA). DNA was re-suspended in 500 µl of ddH_2O and aliquoted into several tubes and stored at -20°C for SMRT sequencing and sub-cloning of *tal* genes. All the plasmid preps were carried out by using the QIAprep Spin Miniprep Kit (Qiagen). DNA purification was performed by using the QIAquick PCR Purification Kit (Qiagen).

RNA isolation and qRT-PCR

Leaves of 3-week-old rice plants were infiltrated with water or *Xo* strains using an OD_{600} of 0.5. At 1 dpi, leaf segments from three plants were ground into a fine powder using the Qiagen Tissue-Lyser system (30 rps for 30 s). Total RNA was isolated from rice leaves using the RNeasy kit (Qiagen). cDNA was generated as reported previously [14]. Real-time PCR was performed on the Mx3005p multiplex quantitative PCR system (Stratagene, La Jolla, CA, USA) using SYBR Green (Eurogentec, Liège, Belgium). For each gene, three independent biological replicates were analyzed. DNA amounts were normalized using actin as a reference gene. Primer sequences are provided in Table S4.

RNAseq analysis

RNAseq libraries (paired-end strand specific TRUEseq Illumina) were constructed from RNA extracted from leaf tissues of Nipponbare plants collected 24 hours post inoculation with the *Xoo* strain MAI1 and a water control. Libraries were sequenced in a HiSeq2000 machine (Illumina) at the INRA Unité de Recherche en Génomique Végétale (Evry, France). Library quality was verified using FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped against the rice genome (Nipponbare v. MSU7) using the Tuxedo suite v2.1.1 [48] with the following parameters: Tophat (--max-multihits 20--b2-very-fast); Cufflinks (--frag-bias-correct --library-type fr-secondstrand --multi-read-correct); Cuffdiff (--upper-quartile-norm --multi-read-correc --FDR 0.05). Differentially expressed genes were determined using the R packages cummerbund [48-50]. Genes identified as differentially expressed (pvalue >0.01) by at least one of these approaches were kept for analyses. Libraries are available under accession GSE108504 at the NCBI-GEO.

Construction of designer TAL effectors

The *OsERF#123* and *OsTFX1* promoter sequences from the rice variety Nipponbare were analyzed in order to design the appropriate designer TAL effectors (dTALE) binding sites as described previously [14]. The specificity of the selected EBEs were controlled using Talvez v.3.1 (http://bioinfo.mpl.ird.fr/cgi-bin/talvez/talvez.cgi; [62]). dTALEs were generated using the Golden TAL technology and expressed as FLAG fusions under the control of a *lac* promoter in

the Golden Gate-compatible broad host range vector pSKX1 [14]. The RVD sequences of the dTALEs used in this study are provided in Table S5.

5'-RACE PCR analysis

Total RNA was isolated from leaves of Azucena 24 h after inoculation with *Xoo* strain MAI1 or *Xo* X11-5A (p*talB*) using the RNeasy kit (Qiagen), and subjected to 5'-RACE RT-PCR analysis using primer OsERF#123-Rev2 and the 5' RACE System for Rapid Amplification of cDNA Ends (Invitrogen). Individual cDNAs were cloned in pGEM-T easy vector (Promega) and sequenced.

Bioinformatics

Genomes of strain *Xoo* BAI3 and MAI1 were sequenced using SMRT whole genome sequencing (Icahn School of Medicine at Mount Sinai, New York, USA) and assembled using HGAP 3.0 as described [41]. Genomes were circularized using the AMOS package, specifically minimus2 under default parameters [64]. Genome sequences and raw sequences were deposited at the NCBI SRA (ncbi.nlm.nih.gov) under Bioproject PRJNA427174. TAL effector sequences were extracted from genomes using in-house perl scripts, and verified using the program PBX [43].

Genomes of other *Xanthomonas strains* were obtained from the NCBI under accession numbers: *Xoc* B8-12: CP011955.1, *Xoc* BLS279: CP011956.1, *Xoc* BLS256: CP003057.1, *Xoc* BXOR1: CP011957.1, *Xoc* CFBP2286: CP011962.1, *Xoc* CFBP7331: CP011958.1, *Xoc* CFBP7341: CP011959.1, *Xoc* CFBP7342: CP007221.1, *Xoc* L8: CP011960.1, *Xoc* MAI10: AYSY01, *Xoc* RS105: CP011961.1, *Xoo* CFBP1947: CP013666.1, *Xoo* KACC10331: AE013598.1, *Xoo* MAFF31101: AP008229.1, *Xoo* NAI8: AYSX01, *Xoo* PXO83 CP012947, *Xoo* PXO99^{A}: CP000967.1, *Xoo* PXO86: CP007166.1, *Xoo* X11-5A: LHUJ01, *Xoo* X8-1A: AFHL01, *Xtu* CFBP2541: NZ_CM003052.1, NZ_CM003053.1.

Amino-acid sequences of housekeeping genes were obtained from each genome using AMPHORA [65], and then aligned using MUSCLE with default parameters v3.8.31 [66]. Poorly aligned positions were removed using Gblocks v. 0.91b [67]. Phylogenetic trees were then generated with PhyML v. 3.1 [68]. Trees were then visualized using the package ape from R

[69]. ANI blast values were calculated for all pairs of genomes analyzed using the ANI calculator form the enveomics suite [44]. *In silico* DNA-DNA hybridization values were calculated using the Genome-to-Genome Distance Calculator version 1 [45]. Dot-plots of paired whole genome alignments were generated using Gepard v. 1.4 [70], with word-size = 30 and window= 30. Multiple genome alignments were made using progressive Mauve v. 2.4.0 [71]. Alignments were visualized using GenomeRing [72]. Circular genome representations were also made with BRIG v.0.95 [73]. Insertion Sequences families in *Xo* genomes were identified using IS-Saga [74]. In the case of Tn*Xax*1-like sequences, nucleotide sequences were obtained from Ferreira *et al*. [46] and were aligned against *Xo* genomes using BLAST 2.2.28+ (blastn -evalue 0.0001 -word_size 7).

Sequences of non-TAL effectors were extracted from The Xanthomonas resource (http://xanthomonas.org) and aligned against *Xo* genomes using BLAST 2.2.28+ (blastn BLAST 2.2.28+). TAL effector sequences were extracted from genomes using in-house perl scripts. Repeat-based trees were generated using DisTAL with default parameters [47] and trees based on predicted DNA-binding specificity were created using FuncTAL with default parameters [47]. TAL effectors were also classified in families using the suite AnnoTALE with default parameters [36]. N- and C-terminal sequences of TAL effectors were aligned using MUSCLE with default parameters v3.8.31 [66] and phylogenetic trees were then generated with PhyML v. 3.1 [68] as above. Unique TAL effector central repeats were also identified using DisTAL [47], and repeat sequences were visualized in R using the heatmap.2 function. A vector of colors was created to represent each repeat using the rainbow function of R, based on distances between repeats calculated from Needleman-Wunsch nucleotide alignments obtained from DisTAL [47].

Target predictions were made for RVD sequences of African *Xoo* TAL effectors using Talvez v.3.1 [47] against the promoter region of annotated rice genes (Nipponbare v. MSU7), 1000bp before the translation start site. The top 600 genes with the highest EBE prediction score were kept for intersection with expression data to identify candidate targets for each TAL effector.

Microarray data used correspond to the Gene Expression Omnibus (GEO) dataset GSE19844. The Affymetrix probe-level expression data in CEL files was processed using the Bioconductor

affy package [75] with default parameters of the rma function that computes the RMA (Robust Multichip Average) expression measure after background correction, normalization and probe summarization. The Bioconductor limma package was used to identify differentially expressed genes as described previously in [76]. The probesets that displayed a value of the log2-transformed fold change (logFC) equal or above 1 and an associated adjusted p-value (Benjamini and Hochberg's adjustment method) equal or below 0.1 were considered as differentially expressed. The mapping of Affymetrix probe sets on MSU7 Gene Model IDs was obtained from the www.ricechip.org website.

Other routine computational tasks were performed in R with utilities provided by packages from the Bioconductor project [77]. All the expression data, RVD sequences for TAL effectors, and TAL binding sites predictions here included can be explored in our recently released tool daTALbase (http://bioinfo-web.mpl.ird.fr/cgi-bin2/datalbase/home.cgi) [78].

Acknowledgments

This project was supported by grants from the Agence Nationale de la Recherche (http://www.agence-nationale-recherche.fr; ANR-14-CE19-443-0002 to BS), the CGIAR Research Program on rice agri-food systems (http://www.grisp.net/main/summary ; MENERGEP NewFrontier program to BS), the US National Science Foundation (https://www.nsf.gov ; IOS-1238189 to AB). TTT and AP-Q were also supported by doctoral fellowships awarded by the Erasmus Mundus Action 2 PANACEA and PRECIOSA program of the European Community, respectively (http://eacea.ec.europa.eu/erasmus_mundus/index_en.php). We are grateful to J. M. Jacobs for valuable comments.

Author Contribution

T.T.T., A.P.Q., M.H., S.C., I.W. and Y.Y. performed the experiments. T.T.T., A.P.Q., M.H, S.C., L.W., S.C., A.B., R.K. and B.S. analyzed the data. J.L. and V.V. contributed reagents/materials/analysis tools. T.T.T., A.P.Q., A.B., R.K., M.H. and B.S. planned and designed the research and wrote the manuscript.

References

1. Nino-Liu DO, Ronald PC, Bogdanove AJ. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. Molecular Plant Pathology. 2006;7(5):303-24.
2. Verdier V, Vera Cruz C, Leach JE. Controlling rice bacterial blight in Africa: needs and prospects. J Biotechnol. 2012;159(4):320-8.
3. Hilaire E, Young SA, Willard LH, McGee JD, Sweat T, Chittoor JM, et al. Vascular defense responses in rice: peroxidase accumulation in xylem parenchyma cells and xylem wall thickening. Mol Plant Microbe Interact. 2001;14(12):1411-9.
4. Poulin L, Grygiel P, Magne M, Gagnevin L, Rodriguez RL, Forero Serna N, et al. New multilocus variable-number tandem-repeat analysis tool for surveillance and local epidemiology of bacterial leaf blight and bacterial leaf streak of rice caused by *Xanthomonas oryzae*. Appl Environ Microbiol. 2015;81(2):688-98. PubMed PMID: 25398857.
5. Gonzalez C, Szurek B, Manceau C, Mathieu T, Séré Y, Verdier V. Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. Molecular Plant-Microbe Interactions. 2007;20(5):534-46.
6. Hajri A, Brin C, Zhao S, David P, Feng JX, Koebnik R, et al. Multilocus sequence analysis and type III effector repertoire mining provide new insights into the evolutionary history and virulence of *Xanthomonas oryzae*. Mol Plant Pathol. 2012;13(3):288-302.
7. Djedatin G, Ndjiondjop MN, Sanni A, Lorieux M, Verdier V, Ghesquiere A. Identification of novel major and minor QTLs associated with *Xanthomonas oryzae* pv. *oryzae* (African strains) resistance in rice (*Oryza sativa* L.). Rice 2016;9(1):18. doi: 10.1186/s12284-016-0090-9. PubMed PMID: 27107993; PubMed Central PMCID: PMCPMC4842194.
8. 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.
9. Bogdanove AJ, Schornack S, Lahaye T. TAL effectors: finding plant genes for disease and defense. Curr Opin Plant Biol. 2010;13(4):394-401. doi: 10.1016/j.pbi.2010.04.010. PubMed PMID: 20570209.
10. Hutin M, Pérez-Quintero AL, Lopez C, Szurek B. MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. Frontiers in plant science. 2015;6(July):535-. doi: 10.3389/fpls.2015.00535.
11. Boch J, Bonas U, Lahaye T. TAL effectors-pathogen strategies and plant resistance engineering. New Phytol. 2014;204(4):823-32. PubMed PMID: 25539004.
12. Blanvillain-Baufumé S, Reschke M, Solé M, Auguy F, Doucoure H, Szurek B, et al. Targeted promoter editing for rice resistance to Xanthomonas oryzae pv. oryzae reveals differential activities for SWEET14-inducing TAL effectors. Plant Biotechnology Journal. 2017;15(3):306-17. doi: 10.1111/pbi.12613.
13. Yuan M, Wang S. Rice *MtN3/saliva/SWEET* family genes and their homologs in cellular organisms. Mol Plant. 2013;6(3):665-74. doi: 10.1093/mp/sst035. PubMed PMID: 23430047.
14. 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.
15. Sugio A, Yang B, Zhu T, White FF. Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes *OsTFIIAg1* and *OsTFX1* during bacterial blight of rice. Proceedings of the National Academy of Sciences. 2007;104(25):10720-5. doi: 10.1073/pnas.0701742104.
16. Römer P, Recht S, Strauß T, Elsaesser J, Schornack S, Boch J, et al. Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. oryzae. New Phytologist. 2010;187:1048–57.
17. Yuan M, Ke Y, Huang R, Ma L, Yang Z, Chu Z, et al. A host basal transcription factor is a key component for infection of rice by TALE-carrying bacteria. Elife. 2016;5. doi: 10.7554/eLife.19605. PubMed PMID: 27472897; PubMed Central PMCID: PMCPMC4993585.
18. Kay S, Hahn S, Marois E, Hause G, Bonas U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science. 2007;318(5850):648-51. doi: 10.1126/science.1144956.
19. Hu Y, Zhang J, Jia H, Sosso D, Li T, Frommer WB, et al. *Lateral organ boundaries 1* is a disease susceptibility gene for citrus bacterial canker disease. Proc Natl Acad Sci U S A. 2014;111(4):E521-9. doi: 10.1073/pnas.1313271111. PubMed PMID: 24474801; PubMed Central PMCID: PMC3910620.

20. Cernadas RA, Doyle EL, Nino-Liu DO, Wilkins KE, Bancroft T, Wang L, et al. Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. PLoS Pathog. 2014;10(2):e1003972. doi: 10.1371/journal.ppat.1003972. PubMed PMID: 24586171; PubMed Central PMCID: PMC3937315.
21. Schwartz AR, Morbitzer R, Lahaye T, Staskawicz BJ. TALE-induced bHLH transcription factors that activate a pectate lyase contribute to water soaking in bacterial spot of tomato. Proc Natl Acad Sci U S A. 2017;114(5):E897-E903. doi: 10.1073/pnas.1620407114. PubMed PMID: 28100489; PubMed Central PMCID: PMCPMC5293091.
22. Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, et al. *R* gene expression induced by a type-III effector triggers disease resistance in rice. Nature. 2005;435(7045):1122-5. doi: 10.1038/nature03630. PubMed PMID: 15973413.
23. Wang C, Zhang X, Fan Y, Gao Y, Zhu Q, Zheng C, et al. XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. Mol Plant. 2014. doi: 10.1093/mp/ssu132. PubMed PMID: 25385701.
24. Tian D, Wang J, Zeng X, Gu K, Qiu C, Yang X, et al. The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. Plant Cell. 2014;26(1):497-515. doi: 10.1105/tpc.113.119255. PubMed PMID: 24488961; PubMed Central PMCID: PMC3963592.
25. Romer P, Hahn S, Jordan T, Strauss T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. Science. 2007;318(5850):645-8. doi: 10.1126/science.1144958.
26. Strauss T, van Poecke RM, Strauss A, Romer P, Minsavage GV, Singh S, et al. RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. Proc Natl Acad Sci U S A. 2012;109(47):19480-5. doi: 10.1073/pnas.1212415109. PubMed PMID: 23132937; PubMed Central PMCID: PMC3511116.
27. Antony G, Zhou J, Huang S, Li T, Liu B, White F, et al. Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os-11N3*. Plant Cell. 2010;22(11):3864-76. doi: 10.1105/tpc.110.078964. PubMed PMID: 21098734; PubMed Central PMCID: PMC3015117.
28. Hopkins CM, White FF, Choi SH, Guo A, Leach JE. Identification of a family of avirulence genes from *Xanthomonas oryzae* pv. *oryzae*. Mol Plant-Microbe Interact. 1992;5(6):451-9.
29. Ji Z, Ji C, Liu B, Zou L, Chen G, Yang B. Interfering TAL effectors of Xanthomonas oryzae neutralize R-gene-mediated plant disease resistance. Nature Communications. 2016;7:13435. doi: 10.1038/ncomms13435 https://www.nature.com/articles/ncomms13435 - supplementary-information.
30. Read AC, Rinaldi FC, Hutin M, He YQ, Triplett LR, Bogdanove AJ. Suppression of Xo1-Mediated Disease Resistance in Rice by a Truncated, Non-DNA-Binding TAL Effector of Xanthomonas oryzae. Front Plant Sci. 2016;7:1516. doi: 10.3389/fpls.2016.01516. PubMed PMID: 27790231; PubMed Central PMCID: PMCPMC5062187.
31. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science. 2009;326(5959):1509-12. doi: 10.1126/science.1178811. PubMed PMID: 19933107.
32. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science. 2009;326(5959):1501. doi: 10.1126/science.1178817.
33. Noel LD, Denance N, Szurek B. Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. Front Plant Sci. 2013;4:333. doi: 10.3389/fpls.2013.00333. PubMed PMID: 24027573; PubMed Central PMCID: PMC3759964.
34. Wilkins KE, Booher NJ, Wang L, Bogdanove AJ. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. Front Plant Sci. 2015;6:536. doi: 10.3389/fpls.2015.00536. PubMed PMID: 26257749; PubMed Central PMCID: PMCPMC4508525.
35. Bogdanove AJ, Koebnik R, Lu H, Furutani A, Angiuoli SV, Patil PB, et al. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. J Bacteriol. 2011;193(19):5450-64. doi: 10.1128/JB.05262-11. PubMed PMID: 21784931; PubMed Central PMCID: PMC3187462.
36. Grau J, Reschke M, Erkes A, Streubel J, Morgan RD, Wilson GG, et al. AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. Sci Rep. 2016;6:21077. doi: 10.1038/srep21077. PubMed PMID: 26876161; PubMed Central PMCID: PMCPMC4753510.

37. Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, et al. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. BMC Genomics. 2008;9:204. doi: 10.1186/1471-2164-9-204. PubMed PMID: 18452608; PubMed Central PMCID: PMC2432079.
38. Triplett LR, Hamilton JP, Buell CR, Tisserat NA, Verdier V, Zink F, et al. Genomic analysis of *Xanthomonas oryzae* isolates from rice grown in the United States reveals substantial divergence from known *X. oryzae* pathovars. Appl Environ Microbiol. 2011;77(12):3930-7. doi: 10.1128/AEM.00028-11. PubMed PMID: 21515727; PubMed Central PMCID: PMC3131649.
39. Verdier V, Triplett LR, Hummel AW, Corral R, Cernadas RA, Schmidt CL, et al. Transcription activator-like (TAL) effectors targeting *OsSWEET* genes enhance virulence on diverse rice (*Oryza sativa*) varieties when expressed individually in a TAL effector-deficient strain of *Xanthomonas oryzae*. New Phytol. 2012;196(4):1197-207. doi: 10.1111/j.1469-8137.2012.04367.x. PubMed PMID: 23078195.
40. Yu Y, Streubel J, Balzergue S, Champion A, Boch J, Koebnik R, et al. Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. oryzae depends on a new TAL effector that induces the rice nodulin-3 *Os11N3* gene. Mol Plant Microbe Interact. 2011;24(9):1102-13. doi: 10.1094/MPMI-11-10-0254. PubMed PMID: 21679014.
41. Booher NJ, Carpenter SC, Sebra RP, Wang L, Salzberg SL, Leach JE, et al. Single molecule real-time sequencing of Xanthomonas oryzae genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. Microb Genom. 2015;1(4). doi: 10.1099/mgen.0.000032. PubMed PMID: 27148456; PubMed Central PMCID: PMCPMC4853030.
42. Quibod IL, Perez-Quintero A, Booher NJ, Dossa GS, Grande G, Szurek B, et al. Effector Diversification Contributes to *Xanthomonas oryzae* pv. *oryzae* Phenotypic Adaptation in a Semi-Isolated Environment. Sci Rep. 2016;6:34137. doi: 10.1038/srep34137. PubMed PMID: 27667260; PubMed Central PMCID: PMCPMC5035989.
43. Huguet-Tapia JC, Peng Z, Yang B, Yin Z, Liu S, White FF. Complete Genome Sequence of the African Strain AXO1947 of *Xanthomonas oryzae* pv. *oryzae*. Genome Announc. 2016;4(1). doi: 10.1128/genomeA.01730-15. PubMed PMID: 26868406; PubMed Central PMCID: PMCPMC4751330.
44. Rodriguez-R L, KT K. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ. 2016;Preprints 4:e1900v1.
45. Meier-Kolthoff JP, Auch AF, Klenk H, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;21(14):60.
46. Ferreira RM, de Oliveira AC, Moreira LM, Belasque J, Jr., Gourbeyre E, Siguier P, et al. A TALE of transposition: *Tn3*-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. MBio. 2015;6(1):e02505-14. doi: 10.1128/mBio.02505-14. PubMed PMID: 25691597; PubMed Central PMCID: PMCPMC4337579.
47. Perez-Quintero AL, Lamy L, Gordon JL, Escalon A, Cunnac S, Szurek B, et al. QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. Front Plant Sci. 2015;6:545. doi: 10.3389/fpls.2015.00545. PubMed PMID: 26284082; PubMed Central PMCID: PMCPMC4522561.
48. 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.
49. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central PMCID: PMCPMC2796818.
50. Huber W, Fagan P, Borg R. Flann O'Brien : contesting legacies. Cork, Ireland: Cork University Press; 2014. xv, 278 pages p.
51. Tran TT, Doucoure H, Hutin M, Szurek B, Cunnac S, Koebnik R. Efficient enrichment cloning of TAL effector genes from Xanthomonas. bioRxiv. 2018. doi: 10.1101/256057.
52. Hutin M, Sabot F, Ghesquiere A, Koebnik R, Szurek B. A knowledge-based molecular screen uncovers a broad-spectrum *OsSWEET14* resistance allele to bacterial blight from wild rice. Plant J. 2015;84(4):694-703. doi: 10.1111/tpj.13042. PubMed PMID: 26426417.
53. Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the *ERF* gene family in Arabidopsis and rice. Plant Physiol. 2006;140(2):411-32. doi: 10.1104/pp.105.073783. PubMed PMID: 16407444; PubMed Central PMCID: PMCPMC1361313.
54. Schandry N, de Lange O, Prior P, Lahaye T. TALE-Like Effectors Are an Ancestral Feature of the *Ralstonia solanacearum* Species Complex and Converge in DNA Targeting Specificity. Front Plant Sci. 2016;7:1225. doi: 10.3389/fpls.2016.01225. PubMed PMID: 27582755.

55. Mackiewicz P, Mackiewicz D, Kowalczuk M, Cebrat S. Flip-flop around the origin and terminus of replication in prokaryotic genomes. Genome Biology. 2001;2(12):interactions1004.1-interactions.4. PubMed PMID: PMC138987.
56. Hutin M, Perez-Quintero AL, Lopez C, Szurek B. MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. Front Plant Sci. 2015;6:535. doi: 10.3389/fpls.2015.00535. PubMed PMID: 26236326; PubMed Central PMCID: PMCPMC4500901.
57. Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang C, He X, et al. Tomato transcription factors Pti4, Pti5, and Pti6 activate defense responses when expressed in Arabidopsis. Plant Cell. 2002;14(4):817-31. PubMed PMID: 11971137; PubMed Central PMCID: PMCPMC150684.
58. He P, Warren RF, Zhao T, Shan L, Zhu L, Tang X, et al. Overexpression of Pti5 in tomato potentiates pathogen-induced defense gene expression and enhances disease resistance to *Pseudomonas syringae* pv. *tomato*. Mol Plant Microbe Interact. 2001;14(12):1453-7. doi: 10.1094/MPMI.2001.14.12.1453. PubMed PMID: 11768541.
59. Journot-Catalino N, Somssich IE, Roby D, Kroj T. The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in *Arabidopsis thaliana*. Plant Cell. 2006;18(11):3289-302. doi: 10.1105/tpc.106.044149. PubMed PMID: 17114354; PubMed Central PMCID: PMCPMC1693958.
60. Yun KY, Park MR, Mohanty B, Herath V, Xu F, Mauleon R, et al. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. BMC Plant Biol. 2010;10:16. doi: 10.1186/1471-2229-10-16. PubMed PMID: 20100339; PubMed Central PMCID: PMCPMC2826336.
61. Wang L, Rinaldi FC, Singh P, Doyle EL, Dubrow ZE, Tran TT, et al. TAL Effectors Drive Transcription Bidirectionally in Plants. Mol Plant. 2017;10(2):285-96. doi: 10.1016/j.molp.2016.12.002. PubMed PMID: 27965000.
62. Perez-Quintero AL, Rodriguez RL, Dereeper A, Lopez C, Koebnik R, Szurek B, et al. An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. PLoS One. 2013;8(7):e68464. doi: 10.1371/journal.pone.0068464. PubMed PMID: 23869221; PubMed Central PMCID: PMC3711819.
63. Grau J, Wolf A, Reschke M, Bonas U, Posch S, Boch J. Computational predictions provide insights into the biology of TAL effector target sites. PLoS Comput Biol. 2013;9(3):e1002962. doi: 10.1371/journal.pcbi.1002962. PubMed PMID: 23526890; PubMed Central PMCID: PMC3597551.
64. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M. Next Generation Sequence Assembly with AMOS. Curr Protoc in Bioinformatics. 2012;11:11.08.
65. Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic inference. Genome Biol. 2008;9(10):R151. doi: 10.1186/gb-2008-9-10-r151. PubMed PMID: 18851752; PubMed Central PMCID: PMCPMC2760878.
66. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7. doi: 10.1093/nar/gkh340. PubMed PMID: 15034147; PubMed Central PMCID: PMCPMC390337.
67. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007;56(4):564-77. doi: 10.1080/10635150701472164. PubMed PMID: 17654362.
68. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21. doi: 10.1093/sysbio/syq010. PubMed PMID: 20525638.
69. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36(Web Server issue):W465-9. doi: 10.1093/nar/gkn180. PubMed PMID: 18424797; PubMed Central PMCID: PMCPMC2447785.
70. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007;23(8):1026-8. doi: 10.1093/bioinformatics/btm039. PubMed PMID: 17309896.
71. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394-403. doi: 10.1101/gr.2289704. PubMed PMID: 15231754; PubMed Central PMCID: PMCPMC442156.
72. Herbig A, Jäger G, Battke F, Nieselt K. GenomeRing: alignment visualization based on SuperGenome coordinates. Bioinformatics. 2012;28(12):i7-i15. doi: 10.1093/bioinformatics/bts217. PubMed PMID: PMC3371849.

73. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:402. doi: 10.1186/1471-2164-12-402. PubMed PMID: 21824423; PubMed Central PMCID: PMCPMC3163573.

74. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 2011;12(3):R30. doi: 10.1186/gb-2011-12-3-r30. PubMed PMID: 21443786; PubMed Central PMCID: PMCPMC3129680.

75. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20(3):307-15. doi: 10.1093/bioinformatics/btg405. PubMed PMID: 14960456.

76. Smyth GK. Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor (pp 397-420) Springer New York. 2005.

77. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80. doi: 10.1186/gb-2004-5-10-r80. PubMed PMID: 15461798; PubMed Central PMCID: PMCPMC545600.

78. Perez-Quintero AL, Lamy L, Zarate CA, Cunnac S, Doyle EL, Bogdanove AJ, et al. daTALbase: a database for genomic and transcriptomic data related to TAL effectors. Mol Plant Microbe Interact. 2017. doi: 10.1094/MPMI-06-17-0153-FI. PubMed PMID: 29143556.

877 **Table 1. Common targets in the rice genome for TAL effectors from strains MAI1 and BAI3.**

TAL effector group[a]	Target locus ID[b]	EBE distance to TSS[c]	EBE[d]	EBE rank MAI1[e]	EBE rank BAI3[f]	Annotation[g]	Fold change MAI1 vs. H20[h]	Fold change BAI3 vs. H20[i]	EBE strand[j]
TalA	Os05g40890	-139	**T**ACACCATCCATCCATCCATCCATCGT	1	2	expressed protein	3,62	2,12	+strand
TalA	Os05g43770	-728	**T**ATAACTTTCAAAAATATCTCAATAAC	261	138	NLI interacting factor-like phosphatase, putative, expressed	0,79	1,33	-strand
TalA	Os08g18974	-498	**T**CTCTCACACACGCATATATAAATTGT	274	146	expressed protein	3,91	3,20	+strand
TalA	Os04g03850	-436	**T**ATAACAATAATGAAAATACTAATCAT	423	213	OsSub39 - Putative Subtilisin homologue, expressed	2,29	2,88	+strand
TalB	Os09g39810	-347	**T**GCGATGCGTTTCCCACCTCCCACCTC	30	6	AP2 domain containing protein, expressed	6,05	4,73	+strand
TalB	Os09g29820	-134	**T**AAAAGGCCCTCACCAACCCATCGCCT	570	135	bZIP transcription factor domain containing protein, expressed	4,37	4,55	+strand
TalC	Os11g31190	-319	**C**ATGCATGTCAGCAGCTGGTCAT	3	3	nodulin MtN3 family protein, putative, expressed	6,46	4,31	+strand
TalC	Os03g58790	-88	**T**TTCCATATCCATATCCACCCAT	387	387	ATPase, putative, expressed	1,97	4,01	+strand
TalD	Os03g61760	-537	**T**ACGCGTCCCCATCCCATCCCCC	3	3	OsSPL6 - SBP-box gene family member, expressed	3,73	2,81	+strand
TalD	Os01g72009	-614	**T**ACAACTTTCCATCACATCAAAC	306	306	expressed protein	2,00	1,15	+strand
TalD	Os07g48760	-966	**T**ACGGGCCCCTGTCTCTTCCCCT	421	421	CAMK_KIN1/SNF1/Nim1_like.32, expressed	3,06	5,70	+strand
TalE	Os02g48740	-352	**C**ACCACTTCCCCTCCCCTCT	22	22	fimbrin-like protein 2, putative, expressed	2,57	1,86	+strand
TalE	Os07g46340	-45	**T**GCCGATGCTCCTCCCACCT	35	35	methyltransferase, putative, expressed	2,50	2,49	+strand
TalE	Os06g04540	-108	**T**AACCCTTCTCCTCTCCTCT	95	95	DNA binding protein, putative, expressed	3,77	3,99	+strand
TalE	Os12g24580	-272	**C**AGCAATGCTCCTCCCCTCT	192	192	Lg106, putative, expressed	1,20	1,64	+strand
TalE	Os09g39810	-365	**T**GCCGAATCTTCCCCCCCCT	488	488	AP2 domain containing protein, expressed	6,05	4,73	-strand
TalF	Os11g31190	-235	**T**AAGCTCATCAAGCCTTCA	4	>2000	nodulin MtN3 family protein, putative, expressed	6,46	4,31	+strand
TalF	Os08g06210	-158	**T**GCGCCCCCCAGGCCCCCA	98	72	expressed protein	1,94	1,18	+strand
TalF	Os02g44880	-632	**T**ACCATCACCAAAACTTCA	366	414	expressed protein	1,18	1,13	-strand

TalG	Os01g53420	-73	**T**AGAAAACATCATCTCAT	153	153	anthocyanidin 5,3-O-glucosyltransferase, putative, expressed	4,59	4,05	+strand
TalH	Os09g28650	-330	**T**ACACATAGACCATATT	6	6	gibberellin receptor, putative, expressed	3,08	1,12	+strand

[a]Identifier of the TAL effector group; corresponds to both the MAI1 and the BAI3 allele.
[b]Nipponbare MSU7 gene ID; genes that are significantly induced by both MAI1 and BAI3 strains and contain a predicted EBE in their promoters for TAL effectors from both strains are shown. Prefix "LOC_Os" is omitted.
[c]Distance from the leftmost base of the EBE to the annotated translation start site based on cDNA evidence in the Rice Genome Annotation Project Release 7 (http://rice.plantbiology.msu.edu/).
[d]Predicted TAL effector binding sequence for each TAL effector group. The 5' terminal nucleotide of each target box indicated in bold.
[e]Rank (according to prediction score) of the EBE among the predicted binding sites in all the Nipponbare MSU7 genome for the MAI1 allele of the corresponding TAL effector group.
[f]Rank (according to prediction score) of the EBE among the predicted binding sites in all the Nipponbare MSU7 genome for the BAI3 allele of the corresponding TAL effector group.
[g]Gene annotation in the Rice Genome Annotation Project Release 7.
[h]Log2 fold change of the corresponding gene in microarray data when comparing plants inoculated with *X. oryzae* pv. oryzae strain MAI1 vs water control at 24 hpi.
[i]Log2 Fold change of the corresponding gene in RNAseq data when comparing plants inoculated with *X. oryzae* pv. oryzae BAI3 vs water control at 24 hpi.
[j]Strand on the promoter where the EBE was predicted "+ strand" indicates the EBE is in 5' to 3' orientation in the sense strand of the gene"

Figure legends

Fig 1. Genomes and TALomes of African *Xoo* compared with other *X. oryzae* strains. A. MLSA-based phylogenetic tree of *X. oryzae* strains, built based on alignment of the amino acid sequences of 33 concatenated housekeeping genes using MUSCLE v.3.8.31 [66] and PhyML v. 3.1 [68]. Numbers in blue indicate branch support as calculated by approximate likelihood ratio tests (aLRT) in PhyML v. 3.1 [68]. The *X. campestris* pv. *leersiae* strain NCPPB4346 was included with the *X. oryzae* strains examined. *X. translucens* strain CFBP2541 was used as an out-group (left). Numbers in black indicate in silico DNA-DNA hybridization (DDH) as calculated using Genome-to-Genome Distance Calculator [45] with the *X. oryzae* pv. *oryzae* strain BAI3 as a reference. **B.** Circular representation of three African *X. oryzae* pv. *oryzae* strains showing positions of TAL effector genes and Tn*Xax1*-like inverted repeats. Features are not drawn to scale and position is approximate. Dashed lines highlight a 434kb inversion containing the replication origin in *Xoo* strain CFBP1947 in comparison to strains BAI3 and MAI1. **C.** Neighbour-joining tree built using the program DisTAL [47] based on alignments of repeat regions from TAL effectors from selected *X. oryzae* strains with fully sequenced TALomes. Strains used were (number of TAL effectors in parentheses): *Xoc* = BLS256 (28), CFBP7331 (22), *Xoo* = BAI3 (9), CFBP1947 (9), MAI1 (9), PXO83 (18), PXO86 (18) and PXO99^{A} (18). Each tip represents a TAL effector. Tip colors indicate the pathovar and geographic group of the strain; TalDP1 from *X. translucens* CFBP2541 was used as outgroup (accession number WP 039006168.1). S, T and E indicate that the TAL effectors can induce *OsSWEET11 or OsSWEET14*, *OsTFX1* or *OsERF#123* respectively. Asterisks indicate instances where the induction of the corresponding target is hypothesized. Scale bar represents branch lengths as calculated by the nj function of the package ape [69] on a distance matrix of alignments generated by DisTAL [47].

Fig 2. Variation in central repeat sequences and RVDs among African *X. oryzae* pv. *oryzae* TAL effectors. Neighbour-joining tree built using the program DisTAL based on alignment of repeats of TAL effectors from African *Xoo* strains BAI3, MAI1 and CFBP1947. Nine groups are shown and correspond to nine genomic loci (left). RVD sequences are shown for each TAL effector variant from each strain. Background colors for the RVDs represent unique nucleotide

repeat sequences, and similar colors indicate similar repeat sequences at the nucleotide level as determined by Needleman-Wunsch alignment. Red stars indicate RVD polymorphisms within one locus. Solid-line rectangles highlight the repeats shared by $TalF_{MAI1}$ and $TalG_{MAI1}$. Dashed-line rectangles highlight those in common to $TalF_{BAI3}$ and $TalH_{BAI3}$.

Fig 3. Virulence function of African *X. oryzae* pv. *oryzae* TAL effectors delivered by *X. oryzae* X11-5A. Mean lesion length was measured 15 days post leaf-clip inoculation of the rice variety Azucena with *X. oryzae* X11-5A derivatives carrying the empty vector (EV), p*talC*$_{BAI3}$ used as positive control, or each of the nine individual TAL effector genes from MAI1 and four of those from BAI3 that display RVD polymorphisms relative to their MAI1 counterparts. An asterisk denotes a significant difference from X11-5A(EV) ($P < 0.0001$). Values represent the averages of at least 10 inoculated leaves. This experiment was repeated four times with similar results.

Fig 4. $TalB_{MAI1}$ targets *OsTFX1* and a novel *S* gene *OsERF#123.* (A) DNA sequence of the promoter regions of $TalB_{MAI1}$-induced genes *Os09g39810* (aka *OsERF#123*) and *Os09g29820* (aka *OsTFX1*) in rice cv. Nipponbare. The effector binding elements (EBEs) for $TalB_{MAI1}$ and PthXo6 from the Asian *Xoo* strain $PXO99^{A}$ are depicted in bold and highlighted in gray, respectively. The EBEs for gene-specific dTALEs used in panels B and C are underlined and labeled. Italics indicate translational start sites, based on the rice Nipponbare v. MSU7 genome (http://rice.plantbiology.msu.edu/). (B) dTALEs targeted to *OsERF#123* and dTALEs targeted to *OsTFX1* increase the virulence of Xo strain X11-5A. Rice leaves of Azucena were inoculated by leaf-clipping with derivatives of X11-5A carrying an empty vector (*EV*) or expressing $talB_{MAI1}$, $dTALE_{ERF-1}$, $dTALE_{ERF-2}$, $dTALE_{TFX-1}$, $dTALE_{TFX-2}$ or *pthXo6*. Lesion length was measured at 15 days after inoculation (dai). Data are the mean of at least eight measurements. Error bars represent +/- SD. Bars with same letters are not statistically different based on a Tukey's multiple comparisons test ($\alpha = 0.05$). Data were plotted and statistical analyses were performed using the Graphpad prism 6 software. (C) $dTALE_{ERF-1}$ and $dTALE_{ERF-2}$ specifically activate *OsERF#123* but not *OsTFX1*. The expression levels of *OsERF#123* and *OsTFX1* were evaluated by quantitative reverse transcription polymerase chain reaction (qRT-PCR) 24 hours after syringe infiltration of Azucena rice leaves with X11-5A strains carrying an empty vector (*EV*) or

derivatives expressing *talB*$_{MAI1}$, *dTALE*$_{ERF-1}$ or *dTALE*$_{ERF-2}$. Gene expression was normalized against X11-5A (*EV*). Error bars represent +/- SD. (D) Population size and distribution of X11-5A(*EV*), X11-5A(p*talB*$_{MAI1}$), X11-5A(p*dTALE*$_{ERF-1}$) and X11-5A(p*dTALE*$_{ERF-2}$) in leaves of cultivar Azucena following leaf-clipping inoculation. Four-week-old plants were inoculated and bacterial populations in 5cm leaf segments were measured at 0, 4, 8, and 12 days after inoculation. Values represent averages of three inoculations. The experiment was repeated three times with similar results. Error bars represent +/- SD. Bars with the same letters are not statistically different, based on a Tukey's multiple comparisons test ($\alpha = 0.05$). Data were plotted and statistical analyses were performed using the Graphpad prism 6 software. Left panel: representative lesions photographed at 14 dai. This experiment was reproduced three times with similar results.

Fig 5. Mutation of *talB* in BAI3 results in a decrease of virulence partially complemented by the induction of *OsERF#123*

(A) BAI3 mutant strain BAI3Δ1 carrying plasmid-borne *talB*$_{MAI1}$, *dTALE*$_{ERF-1}$ or *dTALE*$_{ERF-2}$ but not BAI3Δ1 carrying the empty vector (*EV*) induces *OsERF#123*. The expression level of *OsERF#123* was evaluated by quantitative reverse transcription polymerase chain reaction (qRT-PCR) 24 hours after syringe infiltration of Nipponbare rice leaves. The wild type strain BAI3 was used as control. Gene expression was normalized against water inoculated leaves. Error bars represent +/- SD. (B) TalB and dTALEs$_{ERF}$ act as virulence factors for *Xoo* strain BAI3. Rice leaves of IR24 were inoculated by leaf-clipping with BAI3Δ1 strains carrying an empty vector (*EV*) or derivatives carrying *talB*$_{MAI1}$, *dTALE*$_{ERF-1}$, or *dTALE*$_{ERF-2}$. Lesion length was measured at 15 days after inoculation (dai). Data are the mean of at least 16 measurements. Error bars represent +/- SE. Bars with same letters are not statistically different based on a Tukey's multiple comparisons test ($\alpha = 0.05$).

Fig 6. Functional convergence and dual targeting activity among African *X. oryzae* pv. *oryzae* TAL effectors. Targeting of three susceptibility genes by multiple TAL effectors is shown. Not included in the figure, *OsERF#123* is also induced by Tal3c from *Xoc* strain BLS256 upon binding to two EBEs, different from the one bound by TalB [61]. Genes are indicated as arrows, black horizontal lines correspond to the promoter region of the gene, and the distance of

the targeted EBEs to the annotated translation start site is shown. EBE sequences are shown color-coded by nucleotide, and RVD sequences of TAL effectors are shown color-coded according to their nucleotide specificity. A vertical line between a nucleotide and an RVD highlights a match. To the left is indicated by YES instances where there is evidence that the corresponding gene is induced by the TAL effector, and by NO instances where there is no such evidence.

Supplementary Information

Figure S1. Whole genome alignments of *X. oryzae* strains. (A) Paired dot plots showing whole genome alignments between *Xo* genomes. Dotplots were made using Gepard [70] with word-size = 30 and window= 30. For visualization, the upper color limit was set to 50%, and the lower color limit and grayscale to zero. (B) Mauve multiple alignment of selected *Xo* genomes. (C) Heatmap shows average nucleotide identity (ANI) values for all pairs of genomes calculated using the ANI tool from the enveomics suite [44], top shows hierarchical clustering based on these values.

Figure S2. Insertion sequences (IS) in *X. oryzae* genomes. Frequency (total instances) of insertion elements identified in the genomes of African (MAI1, BAI3 and CFBP1947) and Asian (PXO99^{A}) *Xoo* strains, and in *Xoc* strain BLS256. IS were identified using IS-Saga, and family names used in the IS-Saga database are shown [74]. Tn*Xax1*-like sequences were identified using BLAST.

Figure S3. Shared predicted targets for groups of African *X. oryzae* pv. *oryzae* TAL effectors. Heatmap showing the percentage of shared predicted targets for all pairs of African *Xoo* TAL effectors. Predictions were made using TALVEZ v.3.1 [62], and the top 500 predicted targets based on locus IDs were used for each comparison.

Figure S4. *Os09g39810* and *Os09g29820* are induced by TalB$_{\mathbf{MAI1}}$. sqRT-PCR analysis of *Os09g39810* (aka *OsERF#123*), *Os09g29820* (aka *OsTFX1*) and actin expression levels 24 hours post infiltration of Azucena rice leaves with *Xoo* strain MAI1, and *Xo* X11-5A strains carrying *talB*$_{\mathrm{MAI1}}$ or an empty vector (*EV*). RNA quality and quantity was estimated using a ND-1000 Nanodrop spectrophotometer. cDNA was synthesized with 1µg of total RNA by means of the SuperScript first-strand synthesis system (Invitrogen, Carlsbad, CA, U.S.A.).

Figure S5. *OsERF#123* transcription start sites observed in response to TalB. RNA collected 24 hours post-infiltration of rice leaves with *Xoo* strains MAI1 and X11-5A (p*talB*) were subjected to 5' RACE. Obtained cDNA sequences were aligned above (MAI1) and below (X11-

5A (p*talB*)) the predicted 5'-UTR of *OsERF#123*. The effector binding element (EBE) for TalB is highlighted in gray. Numbering starts downstream of the 3' end of the EBE. The total number of sequences obtained (N) is given on the left.

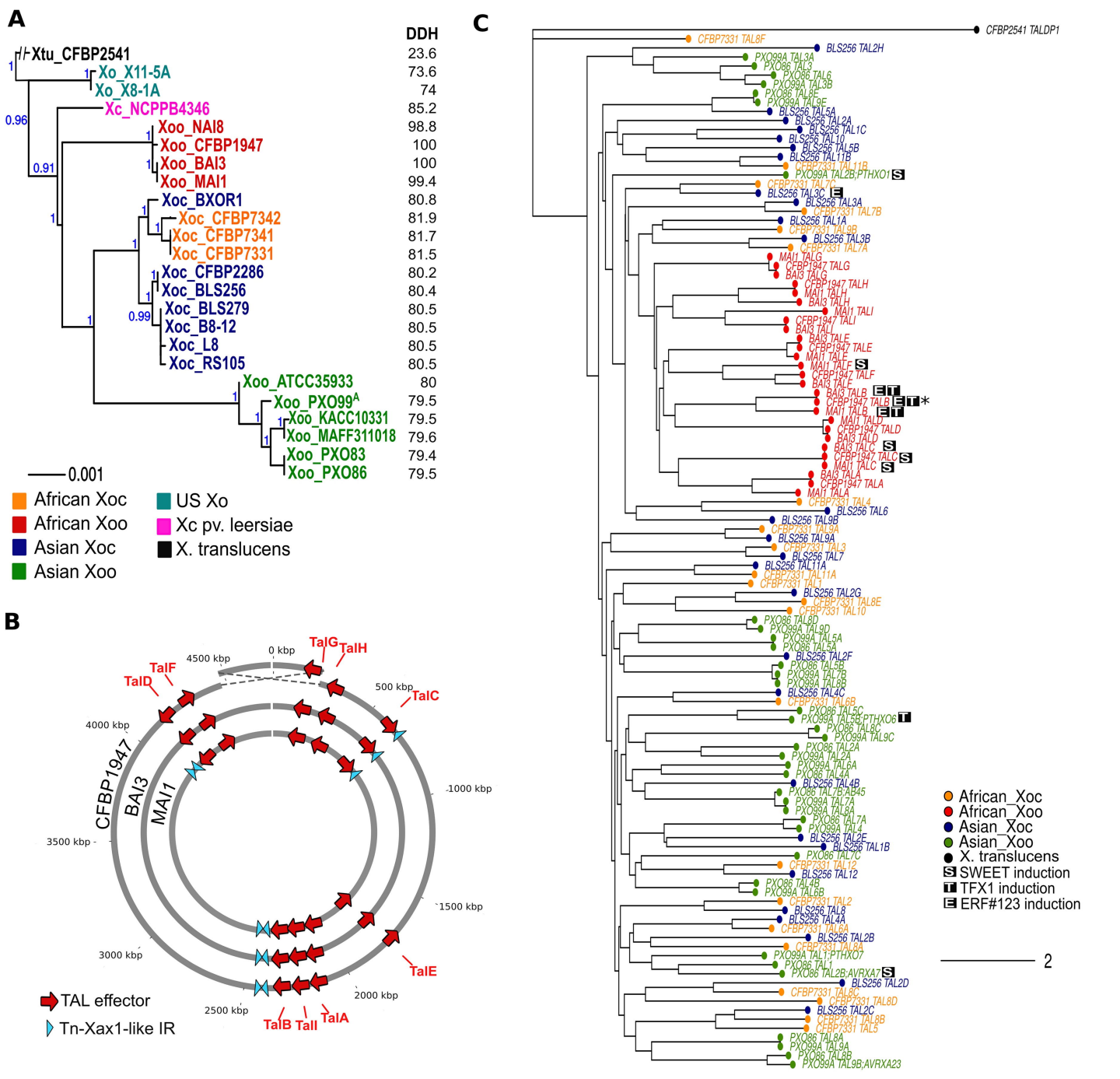
Table S1. African *X. oryzae* pv. *oryzae* TAL effector nomenclature.

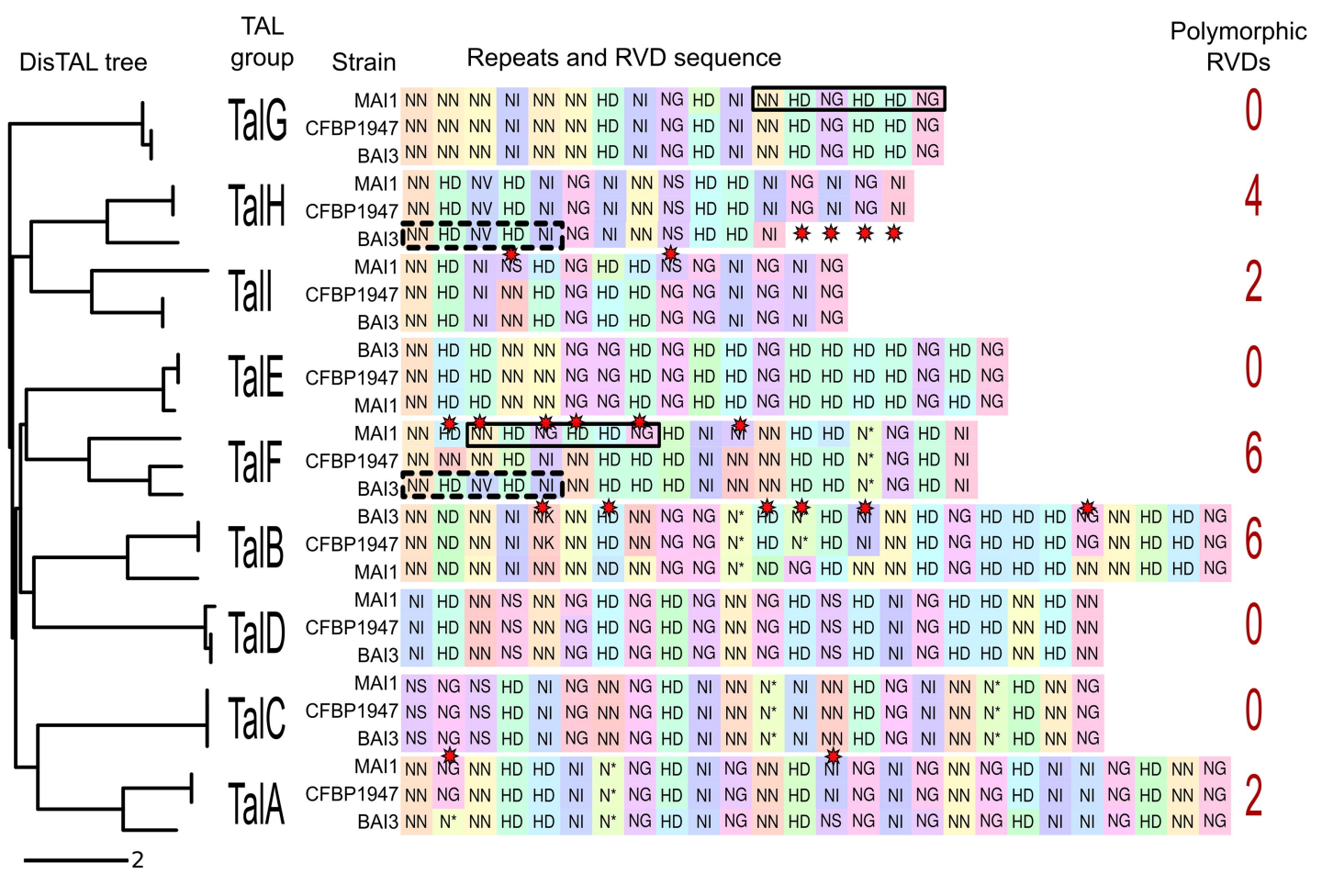
Table S2. Candidate targets for MAI1 TAL effectors in the rice genome.

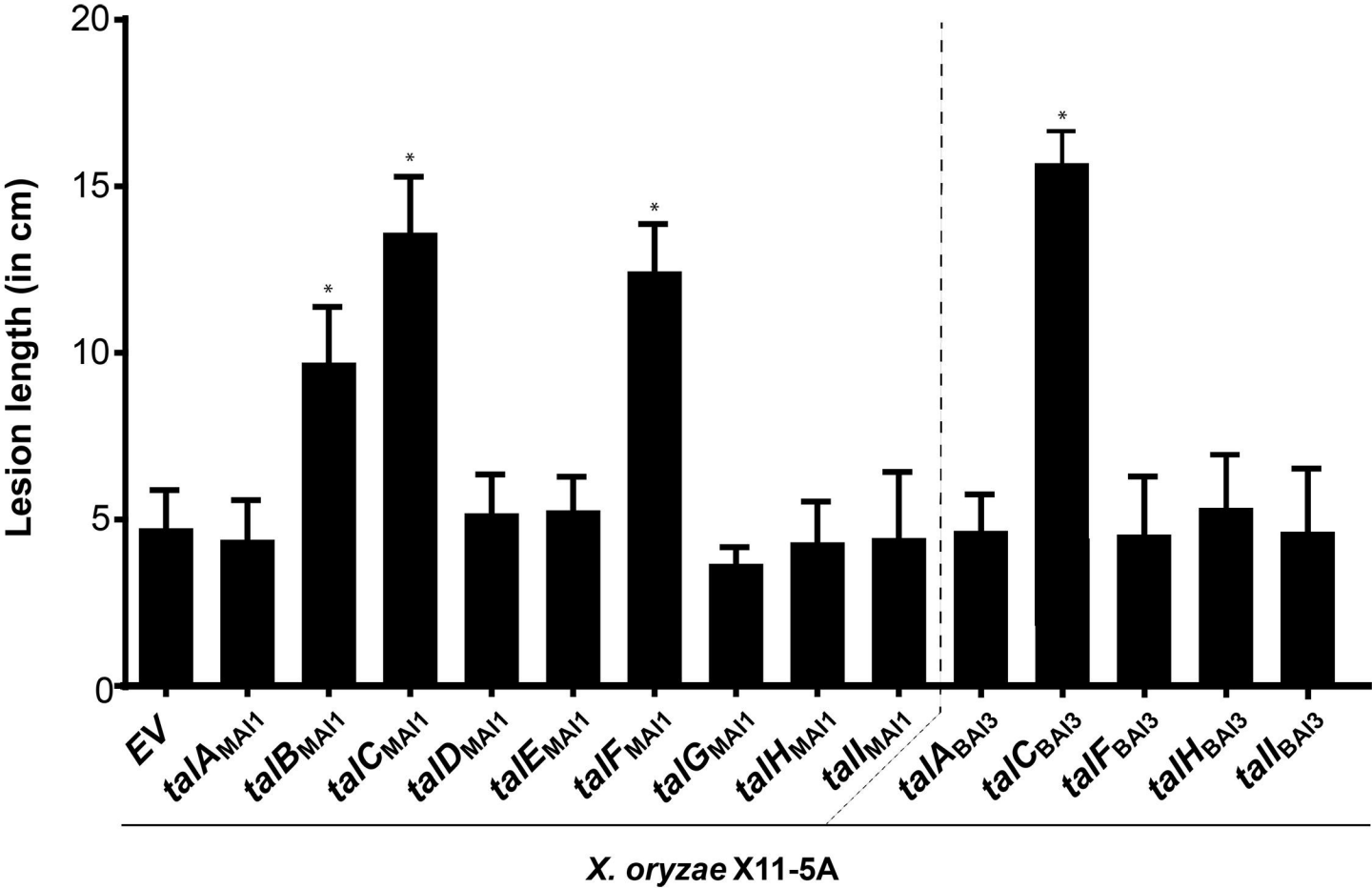
Table S3. Candidate targets for BAI3 TAL effectors in the rice genome.

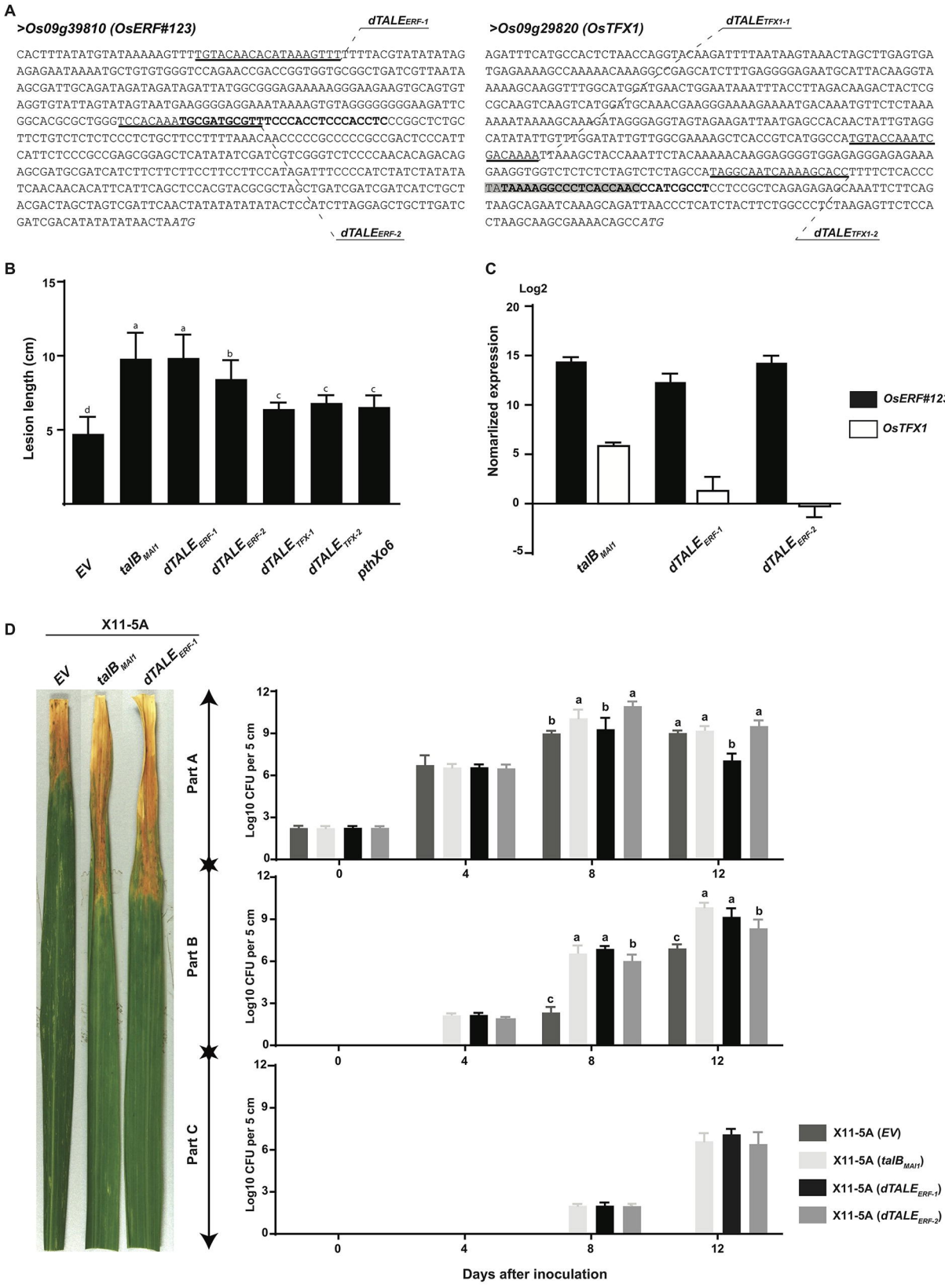
Table S4. Primer sequences used in this study.

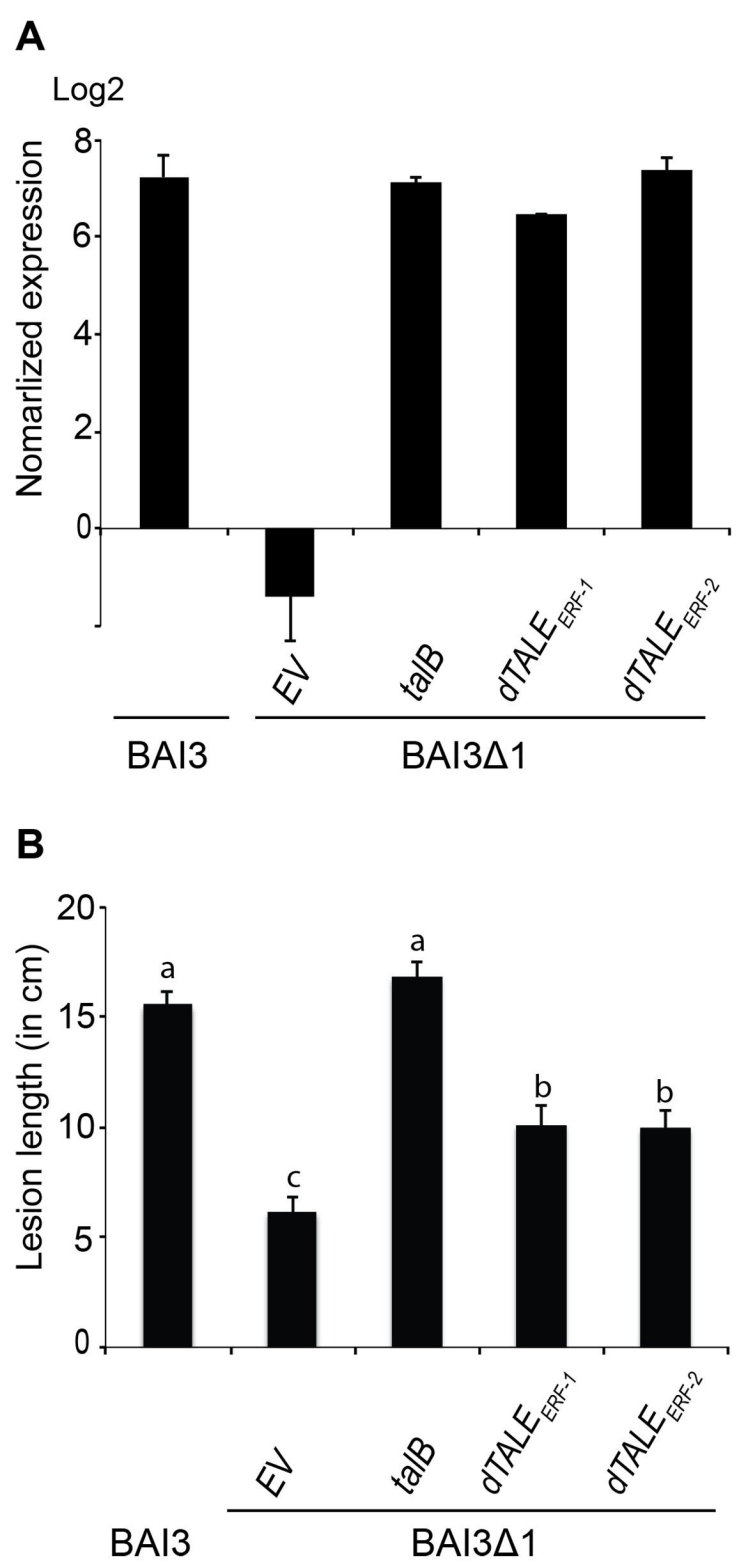
Table S5. RVD sequences of the dTALEs used in this study and their target sequence.

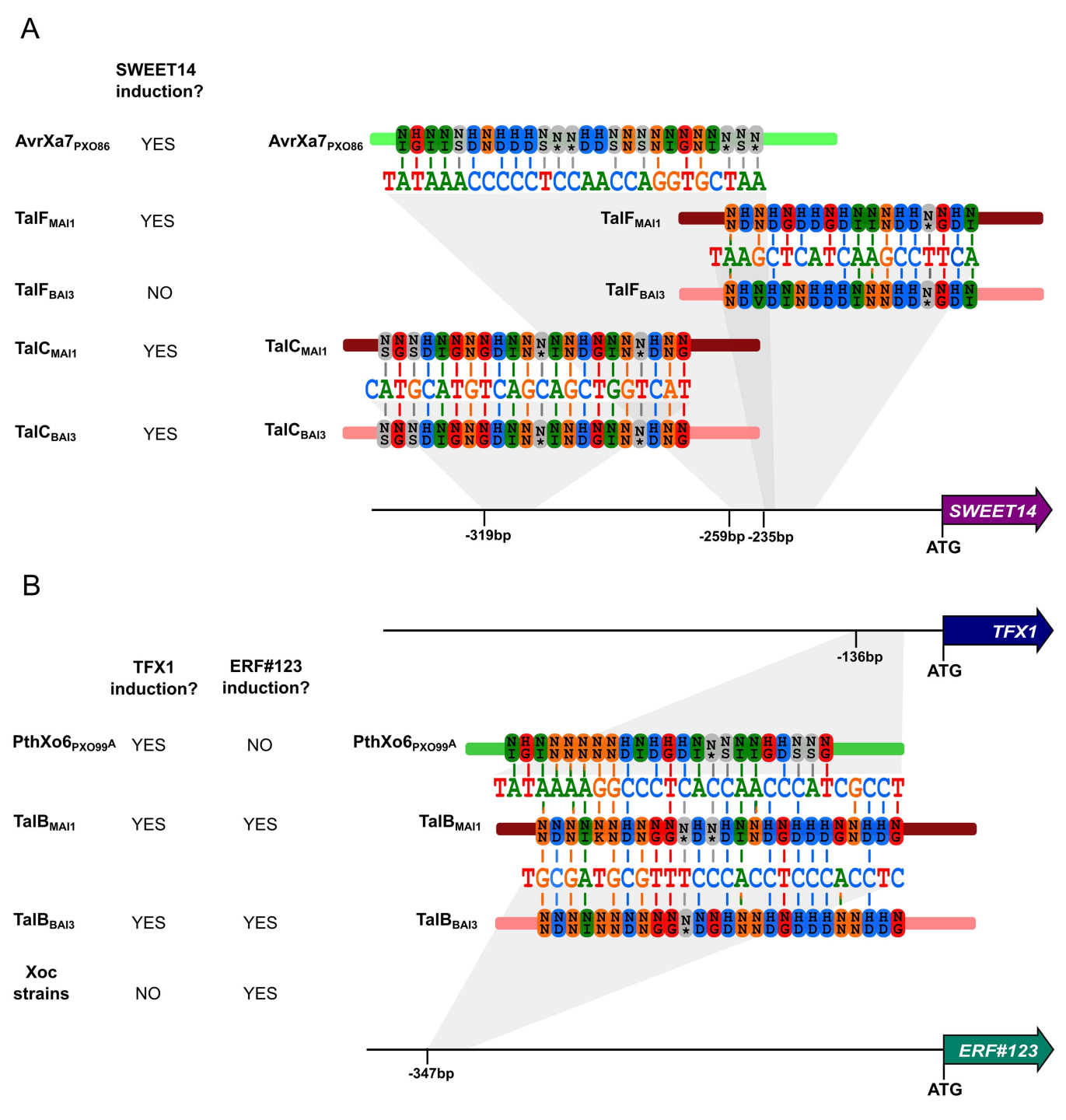

A
DDH
Xtu_CFBP2541 23.6
Xo_X11-5A 73.6
Xo_X8-1A 74
Xc_NCPPB4346 85.2
Xoo_NAI8 98.8
Xoo_CFBP1947 100
Xoo_BAI3 100
Xoo_MAI1 99.4
Xoc_BXOR1 80.8
Xoc_CFBP7342 81.9
Xoc_CFBP7341 81.7
Xoc_CFBP7331 81.5
Xoc_CFBP2286 80.2
Xoc_BLS256 80.4
Xoc_BLS279 80.5
Xoc_B8-12 80.5
Xoc_L8 80.5
Xoc_RS105 80.5
Xoo_ATCC35933 80
Xoo_PXO99A 79.5
Xoo_KACC10331 79.5
Xoo_MAFF311018 79.6
Xoo_PXO83 79.4
Xoo_PXO86 79.5
0.001
African Xoc
African Xoo
Asian Xoc
Asian Xoo
US Xo
Xc pv. leersiae
X. translucens
B
TalG
TalH
TalF
TalD
TalC
TalE
TalB
TalI
TalA
0 kbp
500 kbp
1000 kbp
1500 kbp
2000 kbp
2500 kbp
3000 kbp
3500 kbp
4000 kbp
4500 kbp
CFBP1947
BAI3
MAI1
TAL effector
Tn-Xax1-like IR
C
CFBP2541 TALDP1
African_Xoc
African_Xoo
Asian_Xoc
Asian_Xoo
X. translucens
SWEET induction
TFX1 induction
ERF#123 induction
2


DisTAL tree
TAL group
Strain
Repeats and RVD sequence
Polymorphic RVDs
TalG
MAI1 NN NN NN NI NN NN HD NI NG HD NI NN HD NG HD HD NG
CFBP1947 NN NN NN NI NN NN HD NI NG HD NI NN HD NG HD HD NG
BAI3 NN NN NN NI NN NN HD NI NG HD NI NN HD NG HD HD NG
0
TalH
MAI1 NN HD NV HD NI NG NI NN NS HD HD NI NG NI NG NI
CFBP1947 NN HD NV HD NI NG NI NN NS HD HD NI NG NI NG NI
BAI3 NN HD NV HD NI NG NI NN NS HD HD NI
4
TalI
MAI1 NN HD NI NS HD NG HD HD NS NG NI NG NI NG
CFBP1947 NN HD NI NN HD NG HD HD NG NG NI NG NI NG
BAI3 NN HD NI NN HD NG HD HD NG NG NI NG NI NG
2
TalE
BAI3 NN HD HD NN NN NG NG HD NG HD HD NG HD HD HD HD NG HD NG
CFBP1947 NN HD HD NN NN NG NG HD NG HD HD NG HD HD HD HD NG HD NG
MAI1 NN HD HD NN NN NG NG HD NG HD HD NG HD HD HD HD NG HD NG
0
TalF
MAI1 NN HD NN HD NG HD HD NG HD NI NI NN HD HD N* NG HD NI
CFBP1947 NN NN NN HD NI NN HD HD HD NI NN NN HD HD N* NG HD NI
BAI3 NN HD NV HD NI NN HD HD HD NI NN NN HD HD N* NG HD NI
6
TalB
BAI3 NN ND NN NI NK NN HD NN NG NG N* HD N* HD NI NN HD NG HD HD HD NG NN HD HD NG
CFBP1947 NN ND NN NI NK NN HD NN NG NG N* HD N* HD NI NN HD NG HD HD HD NG NN HD HD NG
MAI1 NN ND NN NI NN NN ND NN NG NG N* ND NG HD NN NN HD NG HD HD HD NN NN HD HD NG
6
TalD
MAI1 NI HD NN NS NN NG HD NG HD NG NN NG HD NS HD NI NG HD HD NN HD NN
CFBP1947 NI HD NN NS NN NG HD NG HD NG NN NG HD NS HD NI NG HD HD NN HD NN
BAI3 NI HD NN NS NN NG HD NG HD NG NN NG HD NS HD NI NG HD HD NN HD NN
0
TalC
MAI1 NS NG NS HD NI NG NN NG HD NI NN N* NI NN HD NG NI NN N* HD NN NG
CFBP1947 NS NG NS HD NI NG NN NG HD NI NN N* NI NN HD NG NI NN N* HD NN NG
BAI3 NS NG NS HD NI NG NN NG HD NI NN N* NI NN HD NG NI NN N* HD NN NG
0
TalA
MAI1 NN NG NN HD HD NI N* NG HD NI NG NN HD NI NG NI NG NN NG HD NI NI NG HD NN NG
CFBP1947 NN NG NN HD HD NI N* NG HD NI NG NN HD NI NG NI NG NN NG HD NI NI NG HD NN NG
BAI3 NN N* NN HD HD NI N* NG HD NI NG NN HD NS NG NI NG NN NG HD NI NI NG HD NN NG
2
2


Lesion length (in cm)
20
15
10
5
0
*
*
*
*
EV
talAMAI1
talBMAI1
talCMAI1
talDMAI1
talEMAI1
talFMAI1
talGMAI1
talHMAI1
talIMAI1
talABAI3
talCBAI3
talFBAI3
talHBAI3
talIBAI3
X. oryzae X11-5A


A
>Os09g39810 (OsERF#123)
dTALEERF-1
CACTTTATATGTATAAAAAGTTTTGTACAACACATAAAGTTTTTTTACGTATATATAG
AGAGAATAAAATGCTGTGTGGGTCCAGAACCGACCGGTGGTGCGGCTGATCGTTAATA
AGCGATTGCAGATAGATAGATAGATTATGGCGGGAGAAAAAGGGAAGAAGTGCAGTGT
AGGTGTATTAGTATAGTAATGAAGGGGAGGAAATAAAAGTGTAGGGGGGGGAAGATTC
GGCACGCGCTGGGTCCACAAATGCGATGCGTTTCCCACCTCCCACCTCCCGGCTCTGC
TTCTGTCTCTCTCCCTCTGCTTCCTTTTAAACAACCCCCGCCCCCGCCGACTCCCATT
CATTCTCCCGCCGAGCGGAGCTCATATATCGATCGTCGGGTCTCCCCAACACAGACAG
AGCGATGCGATCATCTTCTTCTTCCTTCCTTCCATAGATTTCCCCATCTATCTATATA
TCAACAACACATTCATTCAGCTCCACGTACGCGCTAGCTGATCGATCGATCATCTGCT
ACGACTAGCTAGTCGATTCAACTATATATATATACTCCATCTTAGGAGCTGCTTGATC
GATCGACATATATATAACTAATG
dTALEERF-2
>Os09g29820 (OsTFX1)
dTALETFX1-1
AGATTTCATGCCACTCTAACCAGGTACAAGATTTTAATAAGTAAACTAGCTTGAGTGA
TGAGAAAAGCCAAAAACAAAGGCCGAGCATCTTTGAGGGGAGAATGCATTACAAGGTA
AAAAGCAAGGTTTGGCATGGATGAACTGGAATAAATTTACCTTAGACAAGACTACTCG
CGCAAGTCAAGTCATGGATGCAAACGAAGGGAAAAGAAAATGACAAATGTTCTCTAAA
AAAATAAAAGCAAAGATAGGGAGGTAGTAGAAGATTAATGAGCCACAACTATTGTAGG
CATATATTGTTTGGATATTGTTGGCGAAAAGCTCACCGTCATGGCCATGTACCAAATC
GACAAAATTAAAGCTACCAAATTCTACAAAAACAAGGAGGGGTGGAGAGGGAGAGAAA
GAAGGTGGTCTCTCTCTAGTCTCTAGCCATAGGCAATCAAAAGCACCTTTTCTCACCC
TATAAAAGGCCCTCACCAACCCATCGCCTCCTCCGCTCAGAGAGAGCAAATTCTTCAG
TAAGCAGAATCAAAGCAGATTAACCCTCATCTACTTCTGGCCCTCTAAGAGTTCTCCA
CTAAGCAAGCGAAAACAGCCATG
dTALETFX1-2
B
Lesion length (cm)
15
10
5
d
a
a
b
c
c
c
EV
talBMAI1
dTALEERF-1
dTALEERF-2
dTALETFX-1
dTALETFX-2
pthXo6
C
Log2
Nomarlized expression
20
15
10
5
0
-5
talBMAI1
dTALEERF-1
dTALEERF-2
OsERF#123
OsTFX1
D
X11-5A
EV
talBMAI1
dTALEERF-1
Part A
Part B
Part C
Log10 CFU per 5 cm
12
9
6
3
0
0
4
8
12
X11-5A (EV)
X11-5A (talBMAI1)
X11-5A (dTALEERF-1)
X11-5A (dTALEERF-2)
Days after inoculation


A
Log2
Nomarlized expression
8
6
4
2
0
EV
talB
dTALE ERF-1
dTALE ERF-2
BAI3
BAI3Δ1
B
Lesion length (in cm)
20
15
10
5
0
a
c
a
b
b
EV
talB
dTALE ERF-1
dTALE ERF-2
BAI3
BAI3Δ1


A
SWEET14 induction?
AvrXa7PXO86 YES
TalFMAI1 YES
TalFBAI3 NO
TalCMAI1 YES
TalCBAI3 YES
AvrXa7PXO86
TATAAACCCCCTCCAACCAGGTGCTAA
TalFMAI1
TAAGCTCATCAAGCCTTCA
TalFBAI3
TalCMAI1
CATGCATGTCAGCAGCTGGTCAT
TalCBAI3
SWEET14
-319bp
-259bp -235bp
ATG
B
TFX1
-136bp
ATG
TFX1 induction?
ERF#123 induction?
PthXo6PXO99A YES NO
TalBMAI1 YES YES
TalBBAI3 YES YES
Xoc strains NO YES
PthXo6PXO99A
TATAAAAGGCCCTCACCAACCCATCGCCT
TalBMAI1
TGCGATGCGTTTCCCACCTCCCACCTC
TalBBAI3
ERF#123
-347bp
ATG