Transcription Activator-Like (TAL) Effectors of the Cassava Bacterial Blight Pathogen *Xanthomonas axonopodis* pv. manihotis

By

Mikel Everett Shybut

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Brian J. Staskawicz, Chair
Professor Jennifer Lewis
Professor Suzanne M. Fleiszig

Fall 2015

Abstract

Transcription Activator-Like (TAL) Effectors of the Cassava Bacterial Blight Pathogen *Xanthomonas axonopodis* pv. manihotis

by

Mikel Everett Shybut

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Brian J. Staskawicz, Chair

Cassava is an essential food crop relied on by hundreds of millions of people worldwide. *Xanthomonas axonopodis* pv. manihotis (Xam) is the causal agent of cassava bacterial blight (CBB) and the leading destructive bacterial pathogen of cassava. Xam utilizes a unique class of type three effectors known as transcription activator-like (TAL) effectors (TALEs) to activate specific host genes that contribute to virulence and bacterial growth. In Xam, TALEs are often localized to plasmids. TALE-containing plasmids of other *Xanthomonas* ssp. have been shown to be conjugative, providing a mechanism for the horizontal transfer of TALE virulence components. Here, I characterize the two-TALE containing plasmid pXam46 of Xam isolate CIO151, providing a full draft sequence, TALE virulence assays, and evidence for its mobilizing ability. The potential horizontal transfer of TALEs suggests that TALEs which confer a strong virulence phenotype may be well conserved amongst Xam communities. I screened a subset of 22 global Xam isolates spanning 28 years of evolution from 5 countries over 3 continents for their TALE repertoires. I identified one pair of highly conserved TALEs, including a single repeat variable diresidue (RVD) variant of a pXam46 localized TALE, and 3 additional well-conserved TALEs. Of the two highly conserved TALEs, both contribute to bacterial growth in planta and one is associated with a water soaking disease phenotype. The remaining 3 well-conserved TALEs did not show any measurable contributions to virulence. Some plants contain an evolutionary mechanism to defend against TALEs, carrying executor resistance (R) genes containing TALE binding elements upstream of disease resistance genes. I found that the highly conserved TALE of pXam46 triggers a specific, transcriptionally dependent HR-like phenotype in the non-host *Nicotiana benthamiana*. Employing RNA-seq, I have identified a list of candidate TALE-upregulated genes that may be involved in the defense response of *N. benthamiana*. Identifying TALE-triggered R genes as well as conserved TALEs and their susceptibility targets can assist in the design of durable resistance strategies against Xam. Successful strategies may include stacking promoters of multiple conserved TALEs in front of R genes or modifying cassava susceptibility gene promoters to abrogate TALE binding. It is my hope that the basic biology described herein may assist in those efforts.

TABLE OF CONTENTS

FIGURES INDEX ..... **IV**

TABLES INDEX ..... **V**

DEDICATION ..... **VI**

1 AN INTRODUCTION TO PLANT PATHOGEN INTERACTIONS AND CASSAVA BACTERIAL BLIGHT **1**

1.1 CASSAVA BACTERIAL BLIGHT 2

1.2 *XANTHOMONAS* SSP.: MORE THAN JUST CASSAVA 4

1.3 TYPE THREE SECRETED EFFECTORS: DISRUPTING PLANT IMMUNITY 4

1.4 TRANSCRIPTION ACTIVATOR-LIKE EFFECTORS (TALES) 5

1.5 PLANTS CATCH UP: EFFECTOR TRIGGERED IMMUNITY AND EXECUTOR GENES 6

1.6 SUMMARY AND DISSERTATION OUTLOOK 7

2 STANDARD AND UNIQUE MATERIALS, METHODS, AND PROTOCOLS **8**

2.1 BACTERIAL MEDIUM AND GROWTH CONDITIONS 8

2.2 PLANT GROWTH CONDITIONS 9

2.2.1 Cassava (Manihot esculenta Crantz) *9*

2.2.2 Nicotiana benthamiana *9*

2.3 MOLECULAR BIOLOGY 9

2.3.1 Polymerase chain reaction (PCR) *9*

2.3.2 Cloning *10*

2.4 XAM TRANSFORMATIONS 10

2.4.1 Electroporation *10*

2.4.2 Triparental matings *11*

2.5 SEQUENCING 11

2.5.1 Sanger and Illumina Next-Generation Sequencing *11*

2.5.2 TALE sequencing *12*

2.6 ASSAYS OF VIRULENCE 12

2.6.1 Growth assays and symptom development *13*

2.6.2 Luciferase assays *13*

2.6.3 Effector expression in minimal medium and TALE western blots *14*

2.7 SUMMARY 16

3 SEQUENCING, ASSEMBLY, AND CHARACTERIZATION OF A MULTIPLE TALE-CONTAINING PLASMID **17**

3.1 INTRODUCTION 17

3.1.1 Importance of plasmids to pathogenic bacteria *17*

3.1.2 Identification of the TALE-containing plasmid in Xam *18*

3.1.3 TALE target prediction *18*

3.1.4 Summary *19*

3.2 METHODS 19

3.2.1 Plasmid DNA extraction via CsCl gradient *19*

3.2.2 Restriction enzyme mapping *20*

3.2.3 Separation and sequencing of TAL effectors *20*

3.2.4 Final assembly and analysis of remaining open reading frames *20*

3.2.5 TALE knockouts and cassava growth assays ... *21*
3.2.6 Analytical assessment of conjugative ability ... *23*
3.3 RESULTS/DISCUSSION ... 23
3.3.1 Plasmid assembly and ORF annotation ... *23*
3.3.2 TAL effector sequences ... *26*
3.3.3 $TAL14_{CIO151}$ contributes to virulence ... *28*
3.3.4 TALE target prediction ... *31*
3.3.5 pXam46 not likely to be conjugative but it may be mobilizable ... *32*
3.3.6 Summary ... *34*
4 GLOBAL PROFILE OF XAM TALES AND THE IDENTIFICATION OF CONSERVED TALES ... 35
4.1 INTRODUCTION ... 35
4.1.1 Exploring TAL effector profiles of global isolates ... *36*
4.1.2 The significance of cassava to Africa ... *37*
4.1.3 From TALE profiles to TALE targets–the search for susceptibility genes ... *38*
4.1.4 Summary ... *38*
4.2 METHODS ... 39
4.2.1 Isolation of Xam from Ug-2013 leaves ... *39*
4.2.2 Quantification of TALEs ... *39*
4.2.3 Preparation of large plasmids ... *39*
4.2.4 TALE cloning and sequencing ... *40*
4.2.5 TALE knockouts ... *40*
4.2.6 Growth assays and symptom development ... *40*
4.3 RESULTS/DISCUSSION ... 40
4.3.1 A pattern emerges: The TALE profiles of different Xam isolates are similar ... *41*
4.3.2 Plasmid profiles of Xam isolates with same TALEs can be distinct ... *43*
4.3.3 Obtaining new TALE sequences ... *45*
4.3.4 TALE knockouts and the identification of 'major' TALEs ... *47*
4.3.5 Summary ... *52*
5 THE HUNT FOR AN EXECUTOR R GENE IN *NICOTIANA BENTHAMIANA* ... 54
5.1.1 An 'avirulent' role for TALEs ... *54*
5.1.2 Executor gene discoveries ... *54*
5.1.3 Novel functions of the E genes, a new class of R genes ... *54*
5.1.4 Summary ... *55*
5.2 METHODS ... 55
5.2.1 Plant inoculations and AAD mutant ... *55*
5.2.2 Virus induced gene silencing (VIGS) in N. benthamiana ... *56*
5.2.3 RNA-sequencing of N. benthamiana cDNA ... *57*
5.3 RESULTS AND DISCUSSION ... 57
5.3.1 Screen for TAL-dependent HR ... *57*
5.3.2 Determination of transcriptional dependence ... *59*
5.3.3 VIGS to determine role of EDS1 and SGT1 ... *60*
5.3.4 RNA-seq and filtering of candidate genes ... *61*
5.3.5 Cloning and screening of candidate executor genes ... *62*
5.3.6 Summary ... *64*

6 SUMMARY: AN INTERESTING PAST, A PROMISING FUTURE 67
7 REFERENCES..... 68
8 APPENDIX..... 80
8.1 pXam46 sequence: 80

FIGURES INDEX

FIGURE 1-1 GLOBAL CASSAVA PRODUCTION IN TONS. 2
FIGURE 1-2 CASSAVA GROWTH AND XAM SYMPTOMS. 3
FIGURE 1-3 TRANSCRIPTION ACTIVATOR-LIKE (TAL) EFFECTORS (TALES)..... 6
FIGURE 2-1 VISUALIZATION OF VASCULAR XAM MOVEMENT IN CASSAVA LEAVES. 14
FIGURE 3-1 SINGLE TALE KNOCKOUTS AND PLASMID CURING. 22
FIGURE 3-2 RESTRICTION ENZYME MAPPING OF THE CIO151 PLASMID PXAM46. 24
FIGURE 3-3 PXAM46 ANNOTATIONS AND SEQUENCING. 28
FIGURE 3-4 CONFIRMATION OF A CIO151 PXAM46 CURED VARIANT 151ΔP. 29
FIGURE 3-5 APOPLAST GROWTH ASSAY OF PXAM46 CURED CIO151 AND INDIVIDUAL TALE KNOCKOUTS. 30
FIGURE 4-1 MAP AND TIMELINE OF XAM ISOLATES IN THIS STUDY. 36
FIGURE 4-2 TALE WESTERN BLOT PROFILE OF XAM668. 41
FIGURE 4-3 SYMPTOM PHENOTYPES AND TALE PROFILES OF UG2013 ISOLATES 1 – 10. 42
FIGURE 4-4 TALE PROFILES OF THE REMAINING XAM ISOLATES. 42
FIGURE 4-5 CHARACTERIZATION OF PLASMID PROFILES OF SOME XAM ISOLATES. 44
FIGURE 4-6 A PIPELINE FOR THE AMPLIFICATION, CLONING, AND SEQUENCING OF TALES. 46
FIGURE 4-7 SOUTHERN BLOT OF BAMHI AND ECORI PLASMID DNA OF AFRICAN XAM ISOLATES. 46
FIGURE 4-8 CASSAVA SYMPTOM PHENOTYPE AND WESTERN BLOT OF XAM668 TALE KNOCKOUT STRAINS. 47
FIGURE 4-9 APOPLAST GROWTH ASSAY OF XAM668, 668ΔT14, AND 668ΔT20 KNOCKOUT STRAINS. 49
FIGURE 4-10 APOPLAST GROWTH ASSAY OF XAM668 KNOCKOUT STRAINS 668ΔT13, 668ΔT15, AND 668ΔT22. 50
FIGURE 4-11 LATE SYMPTOMS OF XAM WILD TYPE AND KNOCKOUT STRAINS. 51
FIGURE 5-1 XAM TRIGGERS A NON-HOST HR IN *N. BENTHAMIANA*. 58
FIGURE 5-2 *AGROBACTERIUM*-MEDIATED TRANSIENT EXPRESSION OF TALES SHOWS TAL14-SPECIFIC RESPONSE. 58
FIGURE 5-3 WILD TYPE XG AND XG(TAL20::TAL14 PVSP) IN *N. BENTHAMIANA*. 59
FIGURE 5-4 MUTATIONS IN THE ACTIVATION DOMAIN OF $TAL14_{CIO151}$ ABROGATE *N. BENTHAMIANA* HR. 60
FIGURE 5-5 *AGROBACTERIUM*-MEDIATED TRANSIENT EXPRESSION OF TAL14 IN EDS1 AND SGT1 KNOCKDOWNS. 61
FIGURE 5-6 A SIMPLE RNA-SEQ PIPELINE FOR IDENTIFYING THE TAL14-TARGETED HR INDUCER IN *N. BENTHAMIANA*. 62
FIGURE 5-7 WESTERN BLOT OF TAL14 EXPRESSION AT 8, 24, AND 48 HPI. 62
FIGURE 5-8 PCR PRODUCTS OF *N. BENTHAMIANA* TAL14 CANDIDATES #1 - #8. 64
FIGURE 5-9 INOCULATIONS OF THE TOP 8 *N. BENTHAMIANA* TAL14 CANDIDATE TARGETS. 65
FIGURE 5-10 PCR PRODUCTS OF OTHER *N. BENTHAMIANA* TAL14 CANDIDATES. 66

TABLES INDEX

TABLE 2-1 BACTERIAL GROWTH MEDIUM. 8
TABLE 2-2. BACTERIAL GROWTH CONDITIONS. 8
TABLE 2-3 ANTIBIOTICS USED AND FINAL CONCENTRATIONS. 8
TABLE 2-4 COMMONLY USED VECTORS AND THEIR PRIMARY FEATURES. 9
TABLE 2-5 PRIMERS FOR SEQUENCING XAM TALES. 12
TABLE 2-6 RECIPE FOR 5X MINIMAL MEDIUM A (5X MMA; 1000 ML). 15
TABLE 2-7 RECIPE FOR 1X MINIMAL MEDIUM A (1X MMA; 50 ML). 15
TABLE 2-8 RECIPE FOR LAEMMLI LOADING BUFFER (LLB; 3X, 20 ML). 16
TABLE 2-9 RECIPE FOR TBST (1 L). 16
TABLE 3-1 LYSIS BUFFER FOR CELL PREPARATION OF CSCL DNA EXTRACTIONS AND PLASMID DNA. 20
TABLE 3-2 PRIMERS FOR MAKING TALE KNOCKOUTS USING PLVC18. 21
TABLE 3-3 TRANSLATED AND ANNOTATED ORF RESULTS FOR PXAM46 BY BLASTX. 24
TABLE 3-4 TAL EFFECTOR RVD SEQUENCES OF XAM ISOLATES CIO151 AND XAM668. 27
TABLE 3-5 PREDICTED CASSAVA TARGETS OF THE TWO CIO151 TALES. 32
TABLE 4-1 XAM STRAINS USED IN THIS STUDY 37
TABLE 4-2 PRIMERS FOR TOPO COMPATIBLE AMPLIFICATION OF TALE GENES. 45
TABLE 4-3 NEWLY OBTAINED AFRICAN XAM TALE RVD SEQUENCES. 47
TABLE 5-1. INOCULATION MEDIUM (IM) BUFFER FOR *AGROBACTERIUM*-MEDIATED TRANSIENT EXPRESSION IN *N. BENTHAMIANA*. 56
TABLE 5-2 TMV VIGS TARGETS IN *N. BENTHAMIANA*. 56
TABLE 5-3 AVR AND R GENE CLONES FOR VIGS CONTROLS IN *N. BENTHAMIANA*. 56
TABLE 5-4 TALE CONSTRUCTS TO BE TESTED IN *N. BENTHAMIANA* VIGS ASSAY 57
TABLE 5-5 CANDIDATE $TAL14_{CIO151}$ TARGETS IN *N. BENTHAMIANA*. 63

Dedication

To my mother and father, Sandra Breach and John Shybut.

To my siblings, Natalia, Nicholas, and Aleksis.

To the tireless efforts of my elementary, middle, and high school teachers.

And to my Grinnell College research advisor, Dr. Ben DeRidder, who inspired me to take this journey.

1 An introduction to plant pathogen interactions and cassava bacterial blight

Where there is life, there is disease. This is as true for plants as it is for humans and other forms of life. In fact, plant life on Earth doesn't come easy. In addition to confronting the abiotic environmental extremes of the planet's shifting climate–drought, heat, flooding, frost–plants are challenged by numerous classes of pathogens including bacteria, viruses, fungi, and oomycetes. I spent much of my college scientific career researching how plants respond and adjust when exposed to prolonged heat stress, where we characterized a stabilized version of a key photosynthesis regulatory transcript that is only expressed at elevated temperatures (DeRidder et al., 2012). I have spent the past several years investigating the other side of the plant-stress dynamic, examining how microorganisms with flexible, rapidly evolving genomes have adapted to infect plants. Just as plants have evolved mechanisms that allow them to tolerate some environmental pressures, they have also evolved a robust immune system that allows them to defend themselves from most pathogens. It was the fundamental observation in the early 20th century that pathogens are highly specialized, only able to infect a limited number of species, that has driven much of the field of plant-pathogen interactions (Martin and Kamoun, 2012). Early studies by Flor in the flax rust fungus *Melampsora lini* led to the establishment of a gene-for-gene hypothesis, where resistance or susceptibility is dependent on the interaction of a single host gene and its corresponding pathogen gene (Flor, 1946; 1947). This hypothesis resulted from the curious observation that a fungal mutant deficient in only one gene suddenly acquired the ability to infect its host. Such genes were coined 'avirulence' genes due to their triggering the plant immune system, rendering the pathogen avirulent. Why would pathogens retain such avirulence genes if their loss led to disease? Little was known at the time, other than that single observation, with more soon to follow, of the gene-for-gene hypothesis.

Today we have a much clearer and more elaborate picture on both sides of the plant-pathogen interaction, expanding upon the gene-for-gene hypothesis and identifying the determinants of disease specificity, characterizing several main components of the plant immune system (Dodds and Rathjen, 2010; Jones and Dangl, 2006; Spoel and Dong, 2012), pathogen effector proteins (Lindeberg et al., 2012), and factors that determine plant susceptibility to disease (Lapin and Van den Ackerveken, 2013). Additionally, the recent availability of affordable next-generation sequencing techniques has allowed us to obtain and compare genomes of model and non-model organisms, pathogenic and non-pathogenic strains, hosts and non-hosts alike. For phytopathogens, such comparative genomics approaches have led to major progress in tracking the events of horizontal gene transfer and characterizing pathogenic virulence components (Thynne et al., 2015). These new and now affordable technologies have opened the door to using advanced techniques such as RNA-sequencing for studying pathogens of non-model organisms like cassava. While our techniques and technologies for studying plant pathogens have advanced remarkably, our approaches to preventing and defeating disease need to improve just the same. By identifying and characterizing the weaponry used by pathogens to cause disease and their mode of action as well as the immunity approaches plants have adapted to prevent disease, the plant-pathogen community will be equipped with the tools necessary to develop precise and durable resistance strategies to protect crops from pathogen devastation. What follows is a multi-year investigation into the disease causing agents of cassava bacterial blight, using old techniques and new to uncover the specifics of this compatible interaction.

1.1 Cassava bacterial blight

Cassava (*Manihot esculenta* Crantz), also known as mandioca or yuca, is a starchy, perennial tuber crop grown by millions of farmers worldwide. While its name may be unfamiliar to most Americans, cassava is the third largest source of calories in the tropics behind rice and maize (FAO, 2008) and a staple crop in Africa, Asia, and Latin America that well over 800 million people rely on. Cassava is one of the oldest agricultural crops, predicted to have been cultivated 9,000 years ago in the south Brazilian Amazon forest (Howeler et al., 2013). Overall, it is the 8th largest commodity produced in the world ahead of soybeans, barley, and tomatoes as of 2013 and is the most produced crop in Africa reaching an annual average of 145 million tons since 2010 (**Figure 1-1**), more than double the production of maize on the continent (FAOSTAT, 2015). As a crop, cassava is known for its versatility–able to grow in drier climates with as little as 400 mm of annual rainfall (Howeler et al., 2013) and in relatively poor, acidic soil with limited inputs. However, it was not considered a major crop for intensification practices due to the laborious cultivation and its slow propagation due to its vegetative reproduction (**Figure 1-2AB**). These barriers to larger production combined with the limited inputs required resulted in mostly poorer, smallholder farmers growing cassava and depending on it both to provide food and revenue for their families (FAO, 2008). Cassava is also known for its characteristic production of cyanogenic glycosides in the roots and leaves, an evolutionary mechanism to thwart pests but potentially poisonous as hydrocyanic acid if not cooked properly (El-Sharkawy, 2004). Despite being a sturdy crop well-adapted to withstand poor environmental conditions such as dry weather and low-quality soil, cassava is susceptible to several major pathogenic diseases: cassava brown streak virus (CBSV), cassava mosaic virus (CMV), and cassava bacterial blight. Both CBSV and CMV, spread by the whitefly (*Bemisia tabaci*) pose major threats to cassava, especially in Africa where whitefly abundance has increased leading to both pest and disease threats (FAO, 2009). Cassava bacterial blight (CBB) is the widespread, destructive disease caused by the Gram-negative bacteria *Xanthomonas axonopodis* pv. manihotis (Xam). Hallmark CBB symptoms include angular leaf spots, shoot wilting, blight, gum exudation, dieback, and vascular necrosis (Restrepo et al., 2000) as seen in **Figure 1-2C**. Since Xam is a vascular pathogen and cassava is vegetatively propagated

Figure 1-1 Global cassava production in tons. Average annual production from 2010 – 2013 of cassava in countries worldwide. Africa grows 55% of the global cassava supply. Map and data repurposed from FAOSTAT (2015).

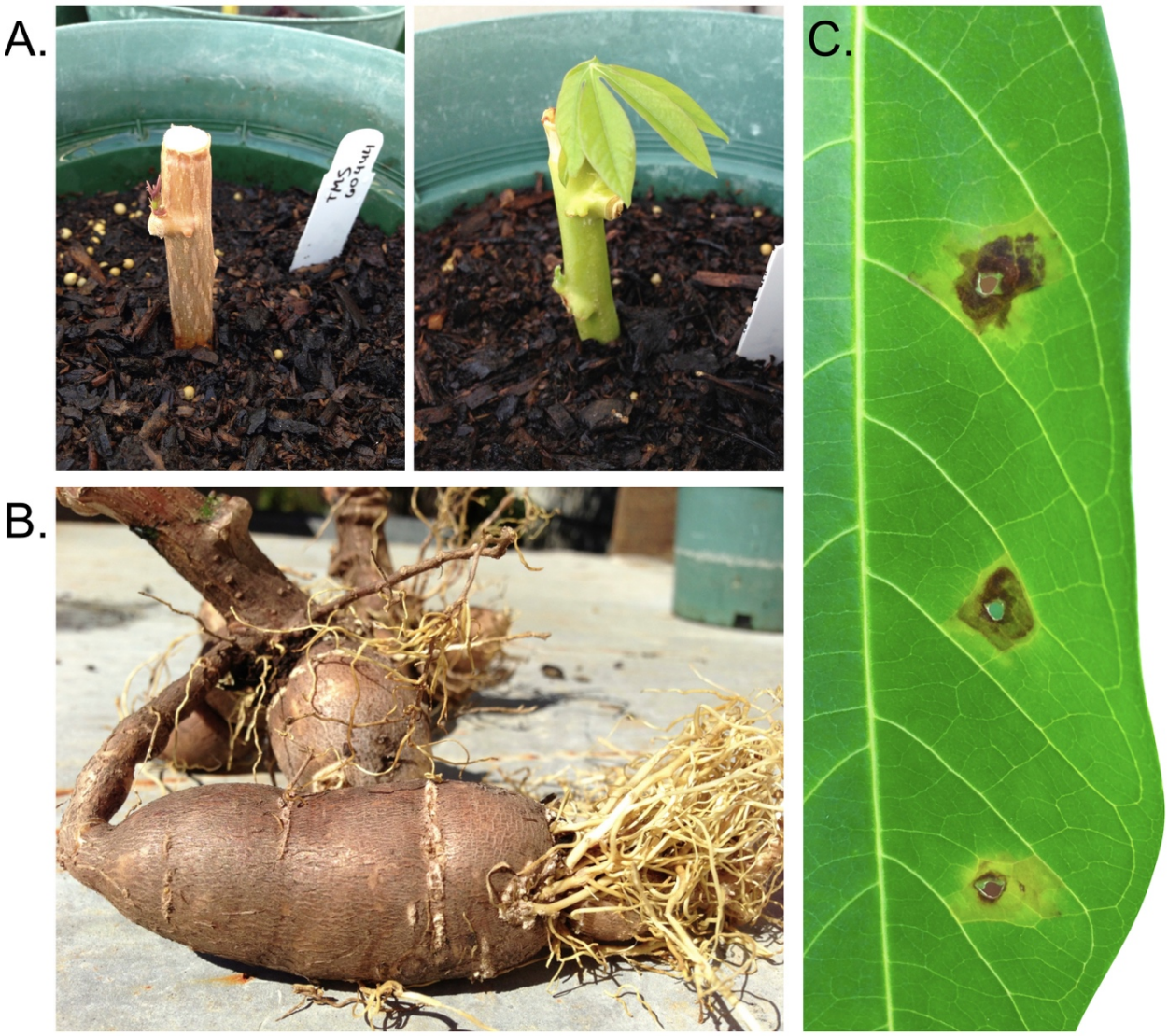

Figure 1-2 Cassava growth and Xam symptoms. New cassava plants are propagated from the stem cuttings of older plants. **A.** For each cutting we preserve two to three leaf nodes and dip the bottom in a root hormone powder to promote growth. **B.** Cassava is grown primarily for its starchy root tuber and used in food processing for products like tapioca or cooked and served on its own. **C.** Disease symptoms of a cassava leaf inoculated with Xam, exhibiting signature phenotypes of necrosis and water soaking.

through stem cuttings, spread of CBB can be widely spread with severe yield losses. When an infected stem is transferred to a new field, yield loss approaches 30% and can exceed 80% under proper environmental conditions (Lozano, 1986). Cassava provides a good example of the difficult balancing act plants must maintain in the environment defending against both abiotic and biotic pressures, yet research on cassava is limited relative to other major crops such as rice and maize. However, there has been a renewed interest in cassava over the past 16 years. Cassava starch has gained greater use in food processes and there is a multimillion-dollar interest in converting cassava to biofuel, presenting poor, developing nations an opportunity to expand their economies with further production and export of cassava (FAO and IFAD, 2004). This renewed interest has

led to increased and more intensified production of cassava, which highlights the importance of finding resistance strategies against Xam to limit yield losses due to the pathogen.

1.2 *Xanthomonas* ssp.: more than just cassava

The genus *Xanthomonas* includes 27 species of bacteria that are associated with over 400 host plants including many agriculturally important cash crops (Ryan et al., 2011). Members of the *Xanthomonas* ssp. cause a wide range of diseases with a significant agricultural impact, including leaf blight and streak in rice (*X. oryzae* pv. oryzae, Xoo; *X. oryzae* pv. oryzicola, Xoc), leaf spot in pepper and tomato (*X. campestris* pv. vesicatoria; Xcv), leaf rot in crucifers (*X. campestris* pv. campestris; Xcc), banana wilt and rot (*X. vasicola* pv. musacearum) and fruit canker in citrus (*X. axonopodis* pv. citri; Xac). The rod-shaped, Gammaproteobacteria are often characterized by their excess production of extracellular polysaccharide as well as their production of yellow outer membrane-bound xanthomonadin pigments (Subramoni et al., 2006), except the white-colored Xam attributed to a mutation in the xanthomonadin cluster gene (Arrieta-Ortiz et al., 2013). Xam is both a foliar and vascular pathogen, meaning it can live as a seemingly innocuous epiphyte when not causing disease. When conditions are warm and humid enough to promote growth, it enters cassava leaves through wounds, stomata or hydathodes, where it multiplies in the intercellular mesophyll space before invading the vascular system and trafficking systemically through the leaves and stems (López and Bernal, 2012). Xam can spread to neighboring plants through water dispersal via rain drops and is primarily introduced to other cassava fields unintentionally by either stem propagation of infected plants or the use of contaminated tools (Lozano, 1986). A coordinated effort by the food and agriculture organization of the United Nations (FAO) has been undertaken called "Save and Grow" to provide smallholder farmers with an incredibly in-depth resource on pest management and sustainability practices, including specific recommendations for the prevention of Xam infection in cassava (Howeler et al., 2013).

1.3 Type three secreted effectors: disrupting plant immunity

Like many other Gram-negative bacterial pathogens, Xam has evolved a set of effector proteins that it delivers into cassava via the type III secretion system (T3SS) to overcome host innate immunity and promote disease. A central component of many bacterial pathogens of plants (*Xanthomonas, Pseudomonas, Erwinia, Ralstonia*) and animals (*E. coli*, *Pseudomonas*, *Salmonella*, *Shigella*, *Yersinia*), the needle-like T3SS provides a means of translocating effector proteins from the bacterial cytoplasm to the eukaryotic host cell (Buttner and He, 2009). The T3SS of plant pathogens was identified by mutations in *hrp* (hypersensitive reaction and pathogenicity) genes as reviewed. (Tampakaki et al., 2010). Transposon insertion mutagenesis of *P. syringae* pv. phaseolicola (Lindgren et al., 1986) and *P. solanacearum* (Boucher et al., 1987) resulted in altered HR and virulence phenotypes. In *X. campestris* pv. vesicatoria (Xcv), a cluster of *hrp* genes necessary for either HR or virulence was identified and characterized as a pathogenicity determinant (Bonas et al., 1991). The *hrp* genes of *P. solanacearum* were later linked to similar virulence components of the human pathogen genus *Yersinia* (Gough et al., 1992), demonstrating conserved functionality of effector-based pathogen strategies. In *Xanthomonas* ssp., secretion of effectors was established by identifying *hrp*-dependent secretion of avirulence genes of both animal and plant pathogens using minimal media induction of effector expression (Rossier et al., 1999).

Plant pathogen type III effectors have been well characterized for their diverse and sometimes redundant roles in disrupting the plant immune system (Grant et al., 2006). Effectors can manipulate and disrupt host defense responses through direct protein or DNA interactions, contributing to successful bacterial infection. Type III secreted effector proteins (T3Es) have been shown to alter gene expression, cytoskeleton dynamics, endocytic trafficking, and programmed cell death (Coburn et al., 2007; Kosarewicz et al., 2012). Extensive work has been done to characterize and classify effector repertoires, even the complete disassembly and reassembly of the 28 effectors of *Pseudomonas syringae*, identifying determinants of host range and demonstrating the dynamic nature of effectors as part of the bacterial genome (Lindeberg et al., 2012). *Xanthomonas* effectors have been classified into 39 *Xanthomonas* outer protein (Xop) groups based on homology, including the AvrBs3-style transcription activator-like (TAL) effectors (TALEs), and many individual effectors have been characterized by their ability to suppress or trigger resistance (White et al., 2009). In Xam, there are at least 18-23 T3Es including TALEs depending on the isolate (Bart et al., 2012). TALE number varies with many strains having upwards of 5, which in Xam are usually localized on plasmids. Using next-generation sequencing technologies to study the presence and diversity of type III effectors in different *Xanthomonas* species and their contributions to both host specificity and host range expansion has been a major focus of our lab recently (Bart et al., 2012; Schwartz et al., 2015). Unfortunately, short-read sequencing limitations did not allow for reliable assembly of TALEs, a unique class of effector proteins that I have taken several alternative approaches to characterizing.

1.4 Transcription activator-like effectors (TALEs)

TALEs represent a functionally unique class of T3Es found in the genus *Xanthomonas*. Secreted via the T3SS, these effectors contain a nuclear localization signal (NLS) that targets them for transport to the nucleus, where they bind mRNA promoter effector binding elements (EBEs) and activate gene expression in the host, leading to plant susceptibility or defense (Kay et al., 2007; Römer et al., 2009b). By mimicking eukaryotic transcription factors, TALEs effectively expand their gene repertoire to include the entire genome of their host, targeting in a very specific manner those genes that will assist in bacterial invasion. All TALEs have in common similar domains: an N-terminal dependent type III secretion (T3S) signal and, much like transcription factors, a C-terminal NLS and an acidic activation domain (AAD; **Figure 1-3**). Between the T3S signal and NLS, each TAL effector contains a region of nearly identical tandem amino acid repeats 33 – 35 residues long–referred to as the central repeat region (CRR) (Wilkins et al., 2015). Within each repeat, at the 12^{th} and 13^{th} position, are two adjacent variable amino acids that determine the nucleotide binding specificity. Comparing these repeat variable diresidues (RVDs) to their nucleotide targets, a simple code has been created to predict TAL EBEs based on the amino acids that comprise each RVD (Boch et al., 2009; Moscou and Bogdanove, 2009). Among the RVDs, only seven pairs of amino acids, HD, NG, NI, NN, NS, HG and an RVD missing the 2^{nd} residue N* comprise almost 90% of all RVDs thus far and bind nucleotides C, T, A, G/A, A/C/T/G, T, and C/T, respectively. Together, the repeat regions form a superhelix that intertwines with the DNA major helix and exposes the RVDs to the nucleotides (Deng et al., 2012; Mak et al., 2012). Their modular structure and predictable binding code has led to development of powerful TALE-derived tools for genomic engineering (Richter et al., 2016) and known and predicted EBEs have been used to engineer TALE-trap resistance strategies for affected plants (Boch et al., 2014; Schornack et al., 2012).

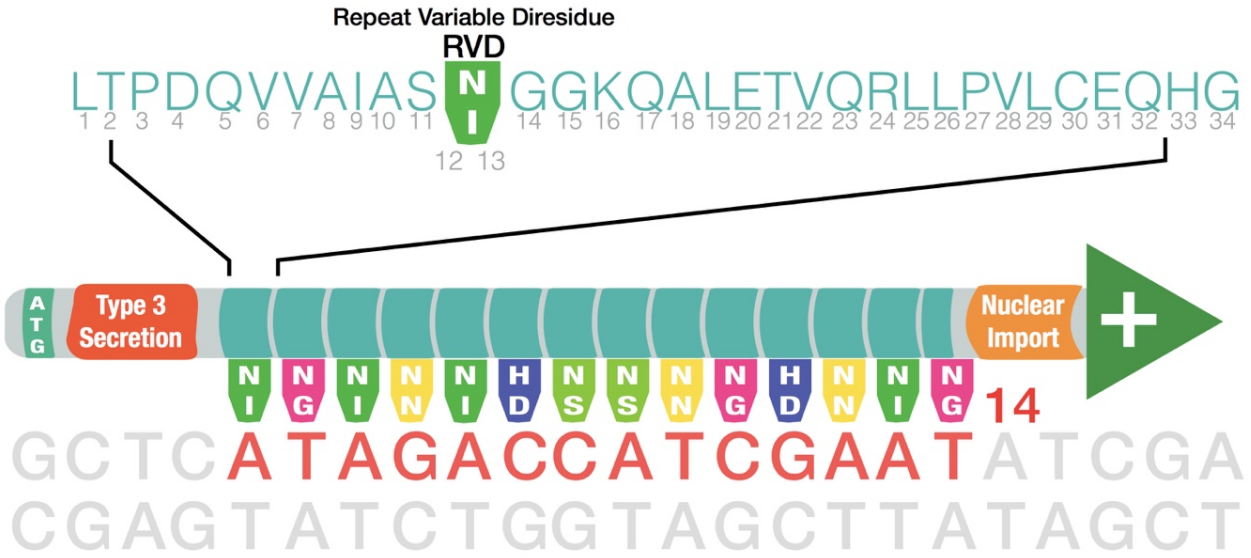

Figure 1-3 Transcription activator-like (TAL) effectors (TALEs). TAL effectors have a characteristic N-terminal type 3 secretion signal (T3SS) as well as both a C-terminal nuclear localization signal (NLS) and an acidic activation domain (+) involved in transcriptional activation. In the middle of the effector is the central repeat region (CRR) (Wilkins et al., 2015). The CRR contains a variable number of tandem repeats each 34 amino acids long and nearly identical except for the 12th and 13th AA in each repeat, known as the repeat variable diresidue (RVD). The RVD sequence determines the corresponding nucleotide that the TALE will interact with. Here, the 14 RVDs of TAL14$_{CIO151}$ interact with 14 consecutive nucleotides.

1.5 Plants catch up: effector triggered immunity and executor genes

While we now know that many type III effectors have biochemical roles inside the plant cell often involved in disrupting innate immunity, T3Es were originally discovered due to their 'avirulence' activity, triggering HR in incompatible interactions. Early on in the molecular era, the first avirulence gene was cloned from *Pseudomonas syringae* pv. glycinea using cosmid libraries and was shown to confer race incompatibility when inoculated into previously compatible cultivars of soybean (Staskawicz et al., 1984). This avirulence response represents a host evolutionary adaption to the pathogenic onslaught of type III secreted effectors. Plants have evolved specialized resistance (R) proteins that recognize the presence of specific effectors either directly through binding in the gene-for-gene manner (Flor, 1971) or indirectly according to the guard hypothesis (Jones and Dangl, 2006) by recognizing the modification of another host protein (Spoel and Dong, 2012). Traditionally, recognition is often carried out by nucleotide-binding leucine-rich repeat (NLR) proteins which upon recognition activate a localized cell death response known as a hypersensitive response (HR) (Qi and Innes, 2013). This effector-recognition form of resistance is known as effector triggered immunity (ETI). At least one TALE, AvrBs3, is subject to NLR recognition by the tomato (*Lycopersicon esculentum*) R protein Bs4 (Schornack et al., 2004). Plants have also evolved a unique mechanism for TALE-dependent resistance in a non-NLR mediated way that takes advantage of TALEs unique functionally as transcription factors. In rice and pepper, both hosts to *Xanthomonas* spp. pathogens, a class of resistance genes termed the executor (E) genes has been identified that are activated in a TALE-dependent manner (Zhang et al., 2015), suggesting that host plants have evolved TALE EBEs in promoter regions of novel R

genes to defend themselves from TALE-wielding pathogens. One E gene example is the *AvrBs3* TALE from Xcv, which induces hypertrophy of mesophyll cells in susceptible hosts (Kay et al., 2007) but triggers a resistance response in pepper plants that carry the *Bs3* R gene. AvrBs3 binds a specific promoter element of Bs3, eliciting an HR (Römer et al., 2009b). While plant immunity can be robust, ever-evolving pathogens are creative in finding new ways to circumvent R genes to thrive in the host (Lapin and Van den Ackerveken, 2013). In the field, this selective pressure should select for strains that have lost or never had *AvrBs3* and the pressure will be back on the plant to evolve a new R gene promoter. This continual co-evolution is a fundamental element of plant-pathogen interactions. The predictable binding code of TALEs and functionality as transcription factors could allow the identification of both new susceptibility and executor-style resistance genes, providing key insights into this very specific relationship.

1.6 Summary and dissertation outlook

To thrive in their ever-changing environment, plants must respond to both abiotic and biotic stresses. Plants have evolved several mechanisms to recognize conserved components of phytopathogens in order to thwart them from causing disease. Early observations implicated 'avirulence' genes of pathogens in a resistance response that depend on the presence of a corresponding R gene in the plant host. Since then, 'avirulence' genes, now effector genes, have been discovered to be important tools for many plant pathogens in circumventing or disrupting the innate immune system of susceptible hosts. Xam, a vascular pathogen and the causal agent of cassava bacterial blight, utilizes TAL effectors to promote disease. TALEs are a unique class of effector that when delivered into the plant cell acts as a eukaryotic transcription factor. TAL effectors often contribute to virulence in the form of in planta bacterial growth or symptom development. Xam TAL effectors are usually found on plasmids, which in other *Xanthomonas* ssp. have been shown to be self-transmissible or conjugative in planta. Resistant plants have mechanisms to defend against TALE-coding pathogens. Several TALEs have been found that activate a resistance response upon upregulating executor R genes of resistant hosts, where TALE EBEs are located in R gene promoter regions.

Over the course of my research, I have studied many different aspects of the TAL effectors of Xam and their role in causing cassava bacterial blight. I have identified new TALE sequences, studied their genetic context on plasmids as a possible mode for horizontal gene transfer, and characterized their roles in pathogenicity. I hypothesized that TALE mobilization via plasmids with type IV secretion system components would lead to a shared, conserved set of TALEs with significant virulence functions. To test this hypothesis and to gain a better understanding of the TALE gene pool, I screened a subset of 22 global Xam isolates spanning 28 years of evolution from 5 countries over 3 continents for their TALE repertoires. I identified a set of conserved TALEs and assessed their individual contributions to virulence by creating knockout strains. Notably, the highly conserved TALEs contributed to virulence and symptom development while the variable TALEs had no measureable impact on bacterial fitness in planta. In search of new modes of resistance, I identified one conserved Xam TALE which triggers a transcriptionally dependent HR-like response in a non-host plant and sought its target. Along the way I have used many techniques, old and new, and have developed some simple and efficient protocols to work with Xam to study TALEs. It is my hope that basic research such as this will support and lead to the development of current and new durable resistance strategies against plant pathogens.

2 Standard and unique materials, methods, and protocols

This chapter contains many of the basic methods I used for growing both *E. coli* and *Xanthomonas* ssp. bacterial cultures and plant growth conditions. Additionally, I've included several protocols I developed or used that may be generally useful for assaying Xam and other bacteria. Included are cloning, sequencing, and transformation strategies as well as in depth methods for expressing and extracting type III secreted effector proteins. In addition, multiple methods for assaying virulence of Xam in cassava are described.

2.1 Bacterial medium and growth conditions

Bacteria were grown in standard liquid medium Luria-Bertani (LB) and nutrient yeast extract glycerol broth (NYGB) or on medium supplemented with agarose (**Table 2-1**). Unless otherwise noted, bacterial strains were grown at standard conditions (**Table 2-2**). Antibiotic selection was used at the listed concentrations (**Table 2-3**) unless noted. A list of vectors used throughout my research are shown along with their antibiotic resistance markers and notable features (**Table 2-4**).

Table 2-1 Bacterial growth medium. Recipes for standard mediums used for bacterial growth. For solid medium, supplement LB and NYGB with 1.5% bacto-agar.

Medium	pH	Components
LB / LA*	7	1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl, 1.5% bacto-agar*
NYGB / NYGA*	7	0.5% bacto-peptone, 0.3% bacto-yeast extract, 2% glycerin, 1.5% bacto-agar*

Table 2-2. Bacterial growth conditions. Plated bacteria were grown in constant temperature incubators. Liquid cultures were grown at listed temperatures with constant shaking.

	Medium	Temperature
E. coli	LB	37 ºC
Agrobacterium tumefaciens GV3101	LB	30 ºC
Xanthomonas axonopodis pv. manihotis	NYGB	28 ºC

Table 2-3 Antibiotics used and final concentrations. Antibiotics used for bacterial selection in solid and liquid medium. Antibiotics were made from stock solutions at the listed concentrations unless otherwise noted.

Antibiotic	Concentration
Rifampicin (Rf)	100 µg / mL
Kanamycin (Km)	25 µg / mL
Tetracycline (Tc)	10 µg / mL
Chloramphenicol (Cm)	25 µg / mL
Gentamycin (Gm)	10 µg / mL
Carbenicillin (Carb)	100 µg / mL

Table 2-4 Commonly used vectors and their primary features. Antibiotics were used at concentrations previously listed (**Table 2-3**).

Vector	Size (bp)	Antibiotic	Feature
pBluescript KS(+)	2,961	Carb	Blue/white screen
pUC19	2,686	Carb	Blue/white screen
pLAFR3	20, 774	Tc	Lambda packaging, broad host range
pVSP61	11,442	Km	Stability in planta, broad host range
pLVC18 Tc SacBR	10,301	Tc	Suicide vector, ccdB cassette
pBRneo	5,508	Km	Km cassette
pE Gm Link 2	3,536	Gm	Gateway compatible with MCS
pEarleyGate100	11,649	Km	35s, Gateway, *Agrobacterium*
pE1776	12,475	Km	OCS promoter overexp., Gateway

Carb = carbenicillin; Tc = tetracycline; Km = Kanamycin; Gm = gentamycin; MCS = multiple cloning site; 35s = CaMV constitutive promoter; OCS = octopine synthase promoter

2.2 Plant growth conditions

2.2.1 Cassava (*Manihot esculenta* Crantz)

All cassava plants used were the model susceptible cultivar TMS60444 unless otherwise noted and grown in greenhouses with natural lighting. New plants were propagated from stem cuttings. Older, healthy greenhouse plants were pruned and then cut into stem pieces with each stem piece containing at least three in tact leaf nodes. The bottom of each stem piece was dipped in rooting hormone powder and placed into a pre-packed pot with 2 – 3 leaf nodes exposed. They were kept on a 15-minute cycling mist cycle between 6:00 am – 6:00 pm at an average temperature of 27ºC (24 – 30ºC range). Ideal experimental plants are 8 – 12 weeks old. After inoculation with bacterial cultures, cassava plants were kept under artificial light at room temperature (25 – 26ºC).

2.2.2 Nicotiana benthamiana

Nicotiana benthamiana was used for *Agrobacterium*-induced transient expression of TALE constructs and virus induced gene silencing (VIGS) assays. All *Nicotiana benthamiana* plants were grown in a walk-in growth chamber on a 16 hour light/8 hour dark cycle at 24ºC. Experimental plants were returned to the chamber after inoculation, transformation, or gene silencing.

2.3 Molecular Biology

2.3.1 Polymerase chain reaction (PCR)

Working with repetitive sequences like TALEs requires a high fidelity polymerase. We have used several successfully in the lab. More recently, our primary polymerase has been the Phusion high-fidelity (HF) polymerase (New England Biolabs), which comes as a 5x MM with both HF buffer and a separate buffer for GC rich templates. It requires a denaturation temperature of 98ºC with a 72ºC extension temperature. We have also used Klentaq LA (long and accurate; DNA Polymerase Technology, St. Louis, MO) which I have also used for amplifying TALEs. It also comes as a 5x MM but requires a denaturation temperature of 94ºC and a lower extension temperature at 68ºC.

OneTaq 2x MM (New England Biolabs) also proved very reliable where others failed, but was used primarily for amplification screening-only methods rather than high fidelity sequencing methods.

2.3.2 Cloning

Several cloning methods were used and favored during this research. Initially, standard restriction enzyme cloning was used with two sites being preferred. More modern techniques were later adopted. Gibson cloning is very reliable and easy to set up. We use the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) in 10 µL reactions. For Gibson cloning, I add on 24 bp of overlap to my PCR primers, cut and column purify the vector, and gel purify the PCR product. Incubation at 50ºC for as little as 15 minutes suffices before transformation. I found Golden Gate cloning (Engler et al., 2008) to be useful for assembling multiple fragments in a single incubation. By simply incorporating cut sites of various type II restriction enzymes with variable bases, several fragments can be cut with the same enzyme together while also being ligated in the same tube. Cloning with pENTR/D-TOPO (Invitrogen) followed up by LR clonase (Invitrogen) into a Gateway compatible vector was also common, though less so after the adoption of Gibson cloning.

2.4 Xam transformations

2.4.1 Electroporation

Electroporation of competent Xam cells offers the cleanest and fastest way to transform a vector into Xam. Involving only plasmid DNA and the recipient Xam cells, there is no mixing of bacterial species required, meaning less opportunity for obtaining *E. coli* contaminants. We use a competent cell protocol that largely adheres to the standard protocol (White and Gonzalez, 1995). While some prefer to create fresh competent cells in small batches for each electroporation, it is easiest to make them in a large batch and store them at -80ºC for future use. To prepare a large prep of competent Xam cells, start a 5 mL overnight culture of Xam cells. The next day, use the 5 mL culture to inoculate 200 mL of NYGB without antibiotics and shake at 225 RPM at 28ºC for approximately 6 hours or until the OD_{600} = 0.6. Then, transfer the cells to ice and keep them on ice for the duration of the procedure. Collect cells at 6,000 RPM at 4ºC for 10 minutes. Pour off the supernatant and resuspend with ice cold dH_2O. Collect the cells again and decant the water. Wash the cells three more times, gently resuspending them each time. Remove any remaining water with a pipette and resuspend the cells in cold 15% glycerol. Collect the cells and remove the glycerol. Then resuspend the cells in 400 µL 15% glycerol. Make 40 µL aliquots into 1.5 mL cold tubes. Freeze the cells in liquid nitrogen and store at -80ºC.

For use in electroporation, thaw competent Xam cells along with the DNA to be used on ice for 10 minutes. Pre-set the electroporator to 12.5 kV/cm (equivalent of 2.5 kV when using a 2 mm gap cuvette), at 50 Ω resistance and 50 µF capacitance. Add 3 µL of 30 ng/µL plasmid DNA to the competent cells and mix gently by tapping the tube. Do not pipette or otherwise disturb the cells. Leave the combined cells and DNA on ice another 10 – 20 minutes. Then, transform in the electroporator at the pre-set settings and recover in 800 µL NYGB. Shake the resuspended cells at 28ºC for two hours to express the transformed plasmid. Then plate in serial dilutions on selective medium (NYGA plus appropriate antibiotics). We have used this technique with some success and seems to complement the alternate approach, triparental matings, well. If competent cells are

pre-made, it is worth trying the transformation while also setting up a triparental mating to save time and eliminate the need for mixed bacteria.

2.4.2 Triparental matings

Often in Xam, it proved more reliable to deliver a vector or plasmid into the cells by triparental mating with two *E. coli* strains than by electroporation of competent Xam cells, especially when attempting to deliver large vectors. Though we have had some success with electroporation in Xam, it seems to be less consistent. Triparental matings are commonly used to transfer a mobilizable plasmid from a donor strain, in our case often *E. coli* DH5a, to a recipient strain, Xam. The third strain is known as the helper and carries a plasmid encoding all the requirements for conjugation, including a type IV secretion system. In our case we use the plasmid RK600 from *E. coli* HB101 carrying chloramphenicol (Cm^r) resistance and the ColE1 replicon (Figurski and Helinski, 1979; Kessler et al., 1992). One barrier to successful conjugation is the recipient restriction-modification (R-M) system, designed to selectively prevent uptake of foreign DNA (Raleigh and Brooks, 1998). We've adjusted a step in our triparental protocol to potentially mediate these effects by conducting the mating and single colony purification at an intermediate temperature of 33ºC, between the 28ºC growth temperature for Xam and the 37ºC for *E. coli*. We can't say whether this indeed affects the Xam R-M system and plasmid uptake but anecdotally it seems to help with the success rate. Additionally, adjusting the ratios of donor, helper, and recipient can have an effect. In my matings I use more Xam than donor or helper.

For a triparental mating, streak out all three strains on appropriate medium: *E. coli* donor strain (LA, 37ºC), *E. coli* RK600 helper (LA, Cm^r, 37ºC), and Xam strain (NYGA, Rf, 28ºC 2 days). When grown up, resuspend an equal amount of each plated culture into 1 mL of LB/NYGB in a 1.5 mL tube without antibiotics. Shake for 2 hours at either 37ºC or 28ºC. In a new tube, combine 150 µL of the donor culture, 150 µL of the helper, and 300 µL of the recipient strain and mix. Add this mixed culture to an NYGA plate without antibiotics in 50 µL portions. Leave plates uncovered in a sterile hood until liquid is completely dry. Cover the plates, bag them and place in a 33ºC incubator overnight. The next day, use a pipette tip to scrape up a streak from each 50 µL bacterial spotting and resuspend in 1 mL NYGB. Plate a series of 100 µL dilutions on selective media (NYGA, Rf + plasmid resistance), place at 33ºC, and check for colonies after 2 – 3 days. Once grown, pick colonies and plate out again for single colony purification. After obtaining a single colony, strains may be screened by presence of the donor plasmid and/or by colony PCR of the target of interest.

2.5 Sequencing

2.5.1 Sanger and Illumina Next-Generation Sequencing

All Sanger sequencing was conducted at the UC Berkeley DNA Sequencing Facility according to their listed requirements. Illumina RNA-sequencing was carried out at the Genomics Sequencing Laboratory at UC Berkeley. Initially, RNA was extracted using the Spectrum plant total RNA kit (Sigma-Aldrich) with on-column DNase digestion. RNA quality was confirmed before sequencing by Bioanalyzer (RNA pico chip for eukaryotic RNA) at the Functional Genomics Laboratory at UC Berkeley. RNA-seq libraries were then made and barcoded using the Illumina TruSeq RNA sample preparation kit (version 2.0). The libraries were checked by Bioanalyzer (DNA chip, high sensitivity) and submitted to the Genomics Sequencing Laboratory where

samples were quantified by qRT-PCR and pooled for 100 bp paired end sequencing on the Illumina HiSeq 2000.

2.5.2 TALE sequencing

To date, there is no accurate way to assemble TALE sequences using short-read-based next-generation sequencing techniques like Illumina. New technologies like the single molecule, real-time (SMRT) sequencing of Pacific Biosciences (PacBio) show promise (Wilkins et al., 2015) and have been effective in assembling TALEs and using the assembled TALEs to produce accurate genome assemblies, even correcting some Sanger produced genomes (Sebra et al., 2015). However, at the moment, the most cost effective way to sequence a limited number of individual TALEs is still Sanger sequencing. The non-repetitive regions of TALEs are highly conserved, allowing us to use universal primers to sequence in from the transcriptional start and stop sites as well as internal primers to sequence into the repeat region from either end. These universal primers, however, require that a TALE be cloned or isolated in another fashion to ensure clean sequencing. Once isolated, a short TALE can be sequenced in one run using a set of four primers (**Table 2-5**). We've been able to sequence through TALEs with fewer than 18 RVDs completely. In TALE genes with long repeat regions like $TAL21_{CIO151}$ and $TAL22_{Xam668}$, the limitations of Sanger with 1 kb reads (often shorter than that through repetitive DNA) and the lack of unique sites within the repeat region make it difficult to sequence through the RVDs. To address this, we have successfully employed the Takara Deletion Kit for Kilo Sequencing. Once cloned, the vector is opened at one end of the insert and a unidirectional exonuclease is used to digest the insert the desired amount and then religated. In the case of the TALEs, we could digest into one end of the repeat region, ligate to the vector, and then sequence in from the M13 vector priming sight directly into the repeat region. We have also found some success for using PCR to amplify single, short TALEs. In the presence of multiple TALEs, the shortest will be preferentially amplified though not cleanly. PCR can be error prone especially for repetitive genes like TALEs. Most TALEs sequenced were off subclones, but I will report a few sequences from PCR amplified TALEs as noted.

Table 2-5 Primers for sequencing Xam TALEs. These primers were used universally to sequence entire TALE genes including the RVD region. For TALEs containing more than 18 RVDs, further steps were required to obtain the full RVD sequence**.**

Primer	Sequence (5' to 3')	Region
TALE Fw ATG	ATGGATCCCATTCGTCCGCGC	5' of RVDs
TALE 744 Fw	AGATTGCAAAACGTGGCGGCG	Fw into RVDs
TALE 2409 Rev	TGCGGCAATCCCTTTCTCACT	Rev into RVDs
TALE Rev Stop	TCACTGAGGAAATAGCTCCAT	3' of RVDs

Fw = forward; Rev = reverse; ATG = TALE start codon; Stop = TALE stop codon; RVD = repeat variable diresidue

2.6 Assays of virulence

There are many ways to characterize the virulence of a pathogen on its host. In addition to assessing leaf symptom phenotypes such as water soaking and necrosis, we've developed a series of standard and unique protocols to assay the virulence of Xam on cassava by looking at in planta bacterial growth, bacterial spread throughout the plant vascular system, and TALE profiles. We've

previously detailed the growth and luciferase assays, including a unique stem inoculation method (Cohn et al., 2015b). Together, these assays allow us to distinguish and characterize strain-level variations in virulence and the individual contribution of type III effectors to bacterial fitness and the promotion of disease. Despite our best efforts to replicate environmental conditions, in lab assays may not accurately portray the conditions and results in the field, where variation in temperature, light, humidity, bacterial populations, and other variable conditions may affect host-pathogen relationships on a daily basis. We've attempted to develop assays to be consistent and to measure tangible phenotypes of Xam once inside cassava, including apoplastic growth, traversal to the vascular system, and growth inside the vascular system.

2.6.1 Growth assays and symptom development

A standard measure of phytopathogen virulence is its ability to proliferate once inside the leaf, where it faces off with the plant innate immune system. Individual type III effectors can help a pathogen grow to a higher abundance within the leaf by disrupting the plant immune response, altering cell wall dynamics, or by other means. Therefore, especially when assessing phenotypes of effector knockouts, the growth assay is the go-to procedure. Plated Xam cultures were resuspended in 10 mM $MgCl_2$ to an ultimate $OD_{600} = 0.00005$, unless otherwise noted. If only assaying for symptom development, a higher $OD_{600} = 0.01$ was used. Abaxial nicks were made with the tip of a razor blade and then inoculated with a needleless syringe to a surface area of 0.25 cm^2. For symptom development, plants were observed daily and scored for characteristic phenotypes such as water soaking, chlorosis, or necrosis. For growth assays, at zero and six days post inoculation leaf punches of 0.6 cm^2 were taken, pulverized with a bead beater, diluted in 10 mM $MgCl_2$, and plated for single colonies. Single colonies were counted and converted to $\log_{10}$ colony forming units (CFUs) per mL based on the dilution and plotted.

2.6.2 Luciferase assays

A notable feature of Xam is its function as vascular pathogen, meaning once it has inhabited a plant leaf it makes its way into the vascular components of the leaf, spreading throughout and clogging up essential nutrient transport tunnels for the plant. Therefore, in addition to measuring bacterial growth at the site of inoculation, we also wanted to track and characterize the ability of Xam to inhabit an entire leaf once inside the leaf tissue. As seen in **Figure 2-1**, we used a bioluminescent reporter to track the vascular movement of Xam after inoculation into cassava leaves (Cohn et al., 2015b). Competent Xam cells were made and transformed (Amaral et al., 2005) with an expression vector containing the *lux* operon genes *luxCDABEG* from *Vibrio fischeri* (pLAFR-lux, unpublished, from S. Schornack via F. Thieme). Liquid cultures of transformants were screened for luciferase expression using an EnVision microplate reader. Luminescent Xam strains were grown on NYGA and for two days and resuspended in 10 mM $MgCl_2$ at $OD_{600} = 0.01$. On the abaxial side of a cassava leaf, three small nicks were made with the corner of a razor blade, avoiding major veins. The leaves were then inoculated using a needleless syringe with inoculum to an area of approximately 0.25 cm^2 or, alternatively, punctured with a toothpick dipped in the Xam inoculum. After 5 days, leaves could be non-destructively imaged by direct exposure to x-ray film, placing the whole plant in a dark room or imaging chamber. Through this technique we have demonstrated that even small amounts of Xam inoculated in various ways, though puncture by a culture-dipped toothpick or forced infiltration by needleless syringe, can develop locally before inhabiting nearby veins and spreading to the midstem and other parts of the leaf.

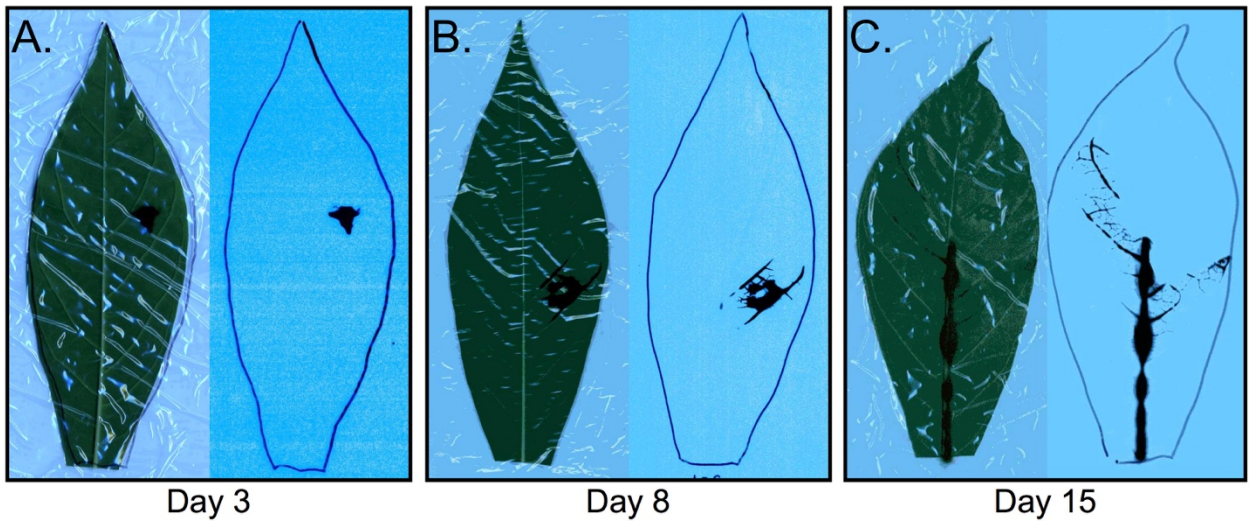

Figure 2-1 Visualization of vascular Xam movement in cassava leaves. Three cassava leaves were inoculated via needeless syringe infiltration with a strain of Xam carrying the luciferase reporter operon at an $OD_{600} = 0.01$. Leaves were removed and exposed to x-ray film for up to 60 minutes. **A.** At 3 dpi Xam is seen multiplying at the site of inoculation. **B.** By 8 dpi Xam begins to colonize the nearby vasculature. **C.** By 15 dpi Xam has colonized the leaf midstem and is no longer seen at the sight if inoculation, where leaf tissue is dying. (dpi = days post inoculation)

2.6.3 Effector expression in minimal medium and TALE western blots

A simple, quick, and accurate pipeline was developed for extracting effector proteins from Xam without plant infiltration to ultimately identify the number and size of expressed TALEs. This pipeline was essential to screening dozens of putative TALE knockouts and obtaining individual knockouts in strains that carried upwards of five TALEs (Cohn et al., 2014). The pipeline takes advantage of the nearly identical nature of TAL effectors, save for their RVDs. We used a polyclonal antibody as a universal probe for TALEs in Xam. The antibody was made from a clone of $TALE_{Xam1}$ (Castiblanco et al., 2012) that was tagged and expressed in pETDuet for antibody production in rabbit (Covance; Feng, B., unpublished). Use of this antibody has allowed for the distinction of multiple TALEs in individual strains and as evidence of TALE expression in infected plant tissue, including transient expression by *Agrobacterium*. Its accuracy has been supported by our restriction mapping and sequencing efforts. It is the best way to get a count of the size and number of expressed TALEs in any given Xam strain.

2.6.3.1 Minimal medium expression and protein extraction protocol

Resuspend overnight liquid cultures of Xam ($OD_{600} = 0.6 – 1.0$) in 125µL minimal medium A (**Table 2-7**) at pH 5.14 (Ausubel et al., 1994) to an $OD_{600} = 0.2$ along with appropriate antibiotics for each Xam strain. Incubate at 28ºC with constant shaking for 4 hours. The low pH and minimal nutrients in the medium replicate the acidic environment of the apoplast, leading to effector secretion (Cornelis and Van Gijsegem, 2000). Add an equal volume of 3x Laemmli buffer (Laemmli, 1970) to each tube (**Table 2-8**) and heat samples for 10 minutes at 98ºC, vortex, collect at 12,000 RPM for 10 minutes and load 15 – 20 µL into the wells of a NuPAGE Novex gel (Life Technologies). Store unused samples as 50 µL aliquots at -80ºC.

Protein lysates were separated via SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using one of two setups: NuPAGE Novex 3-8% Tris-Acetate in NuPAGE Tris-Acetate SDS running buffer or NuPAGE Novex 4-12% Bis-Tris precast gels in NuPAGE MOPS SDS buffer (Life Technologies). Gels were run for 10 minutes at 50V and then either 2 hours at 125V for Tris-Acetate gels or 4.5 hours at 120V for Bis-Tris gels. The Tris-Acetate gels allow for a shorter run time with no compromise in final blot quality. After SDS-PAGE, TALE content was analyzed using standard immunoblotting techniques using the anti-TALE antibody.

2.6.3.2 Minimal medium and western blot recipes

Recipes for the buffers required to induce T3E expression in minimal medium and extract proteins as described (**2.6.3**).

Table 2-6 Recipe for 5x minimal medium A (5x MMA; 1000 mL). A concentrated stock solution for making MMA.

5x Minimal Medium A (MMA; 1000 mL)	
$(NH_4)_2SO_4$	5 g
KH_2PO_4	22.5 g
K_2HPO_4	52.5 g
Sodium Citrate $2H_2O$	2.5 g
–	Autoclave

Table 2-7 Recipe for 1x minimal medium A (1x MMA; 50 mL). The working solution of MMA for bacterial growth and induction of effector secretion.

1x MMA (50 mL)	
5x MA Stock	10 mL
$MgSO_4$ $7H_2O$ (1 mM)	50 µL (1 M stock)
Sucrose (10 mM)	5 mL (100 mM stock)
Casamino Acids (0.3%)	5 mL (3% stock)
BSA (50 µg/mL)	250 µL (10 mg/ML stock)
Sterile H_2O	To 50 mL
	– Bring pH to 5 - 5.4
	– Filter Sterilize

Table 2-8 Recipe for Laemmli loading Buffer (LLB; 3x, 20 mL). A buffer for lysing protein and loading it into wells for polyacrylamide gel electrophoresis (PAGE).

3x Laemmli Loading Buffer (20 mL)	
Tris-Cl (0.24 M, pH 6.8)	4.8 mL (1 M stock)
SDS (6%)	6 mL (20% stock)
Glycerol (30%)	6 mL (100% stock)
B-mercaptoethanol FW78 (16%)	3.2 mL
Bromophenol blue (0.006%)	120 µL (1% stock)
Urea	12 g
–	Aliquot into 1.5 mL tubes
–	Store at -20º C

Table 2-9 Recipe for TBST (1 L). Tris-buffered saline with Tween-20 (TBST) for blocking.

TBST (1 L)	
H_2O	900 mL
TBS (1x)	100 mL (10x)
Tween-20 (0.05%)	0.5 mL

2.7 Summary

TALEs are unique proteins that occasionally require unique techniques to study. Additionally, culturing conditions and molecular techniques of working with Xam require enough small differences to warrant some specialized protocols for Xam. I have included in detail several protocols and recipes for culturing and transforming Xam and assaying its virulence in cassava. The use of a luciferase reporter construct is an especially powerful tool for studying vascular pathogens and would be valuable for genetic studies of the requirements for Xam traversal through the plant leaf environment. Additionally, a method for measuring local Xam growth in the apoplast and a reference to the leaf midvein assay have been included. The luciferase assay demonstrates that growth at the initial site of inoculation is important before the bacteria enter the vascular system, therefore other genes may be involved in adapting to the plant apoplast, allowing for the local replication. Type III effectors are critical components of many Gram-negative bacterial pathogens, including Xam. I included a detailed approach to inducing bacterial secretion of T3Es into a minimal medium and the immediate loading onto a gel and visualization by Western blot. This technique has allowed us to assess TALE profiles in many Xam strains using a polyclonal antibody that binds all TALEs in any given strain.

3 Sequencing, assembly, and characterization of a multiple TALE-containing plasmid

3.1 Introduction

Horizontal gene transfer (HGT) has played a major role in bacterial acquisition of DNA and as a driver of bacterial speciation (Wiedenbeck and Cohan, 2011). Unlike in sexually reproducing eukaryotes, bacteria, which clonally replicate, do not acquire most of their variation through proliferation. Rather, a major portion of variation comes from accessory genetic elements such as prophages, transposons, integrons, and plasmids (Levin and Bergstrom, 2000). Conjugative plasmids are a major component of HGT, allowing bacteria to obtain and 'try out' genes from the horizontal gene pool (Norman et al., 2009). These plasmids often confer selective advantages to the bacterial community, conferring adaptation to the local environment and specific hosts (Tauch et al., 2002). Adaptations can be common depending on the environment, such as copper resistance in crop pathogens or antibiotic resistance in human pathogens. Examples like this show how mobile genetic elements contribute to the fluidity and flexibility of the bacterial genome. Plasmids or transposed genes that confer bacterial fitness can be maintained or even incorporated into the chromosome. If a plasmid contains genes that are detrimental to the bacteria by triggering a defense response in the host for example, the plasmid can be lost without any fundamental alterations to the bacterial genome. For pathogens, HGT provides a mechanism to alter host susceptibility, where successful pathogens in a community could share common mobile genetic elements.

3.1.1 Importance of plasmids to pathogenic bacteria

Plasmids have been found that carry a wide range of virulence factors, including genes for plant hormones like auxin and cytokinin, the coronatine toxin, UV resistance, and even the complete T3SS, though the biology of *Xanthomonas* plasmids is lesser-known than those of the well studied *Pseudomonas syringae* (Sundin, 2007). For phytopathogenic bacteria like Xam, one such selective advantage of maintaining and exchanging genes on mobile elements such as plasmids would be obtaining an enhanced effector repertoire for increasing or expanding virulence. Therefore, it could be hypothesized that in pathogenic bacteria, plasmids would be found that contain both type III effectors to disrupt their host immune system as well as genes involved in the transfer of the plasmids to promote rapid HGT. Such examples have been identified. In the citrus pathogen *Xanthomonas citri* pv. *aurantifolii* strain B69, the 37.1 kb plasmid pXcB was sequenced and characterized (Yacoubi et al., 2007). It was shown to contain a TAL effector called *pthB* (unrelated to the Xam pthB) as well as a group of 12 genes that encode a complete *virB* type IV secretion system (T4SS). The VirB T4SS is involved in the transfer of proteins and DNA from cell to cell. It is often found on plasmids but can also but found on the chromosome, such as in *X. axonopodis* pv. *citri*, where it's found both chromosomally and on the plasmid pXAC64 (Alegria et al., 2005). The presence of the T4SS on pXcB played a important role in HGT. When combined on a leaf surface, pXcB was shown to conjugatively transfer in planta at very high-efficiencies from *X. citri* pv. *aurantifolii* to a *pthA* mutant *X. citri* pv. *citri* in a redundant, *virB*-dependent manner (Yacoubi et al., 2007). This system represented a minimal, mobile arsenal of pathogenicity. The importance of acquiring even just a single effector can be significant, where transfer of *pthB* on pXcB restored citrus canker or in the Columbian Xam strain CFBP1851, where deletion of a region containing $TALE1_{Xam}$ (formerly $pthB_{Xam}$) resulted in a less virulent variant (López and Bernal, 2012). Our understanding of effectors has been increased by studying plasmids. Early studies showed that

plasmid curing could result in non-pathogenic strains or sometimes race change, where each was often attributable to a single plasmid-localized effector (Murillo et al., 2001). While single effectors can have significant effects, plasmids often contain dozens of genes, many of unknown function, and can harbor multiple TAL effectors.

3.1.2 Identification of the TALE-containing plasmid in Xam

Type III effectors are not limited to plasmids. They are often encoded chromosomally and can be recombined from plasmid to chromosome. In *Xanthomonas* spp., TAL effector localization varies by species. In *Xanthomonas oryzae* pv. *oryzicola*, the causal agent of bacterial leaf streak in rice, over 250 TALEs have been identified, all encoded chromosomally (Wilkins et al., 2015). However, in other *Xanthomonas* varieties it seems more common to find TALEs on plasmids, such as in *X. citri.* pv. *aurantifolii* and *X. citri* pv. *citri*, where plasmid-born TALEs were identified on pXcB and pXAC64, respectively (Alegria et al., 2005; Yacoubi et al., 2007). In Xam, it was observed nearly thirty years ago that the presence of plasmids in geographically distinct isolates of Xam, then termed *Xanthomonas campestris* pv. manihotis, was associated with pathogenic traits (Verdier, 1988). This work was followed up with experiments using restriction fragment length polymorphism (RFLP) analysis along with ribotyping using a *pthB* TALE probe and showed the widespread presence of *pthB* fragments on Xam plasmids, where the plasmid ribotypes often mirrored the genomic results (Restrepo and Verdier, 1997). We have observed that TAL effectors are found on large plasmids at or greater than 40 kb in size via southern blot of many different strains. Most Xam strains we have obtained carry TAL effector-containing plasmids, some even having multiple plasmids with differing number of TAL effectors on each. One strain, CIO151 isolated from Colombia in 1995, contains a single plasmid that harbors two TAL effectors. Based on the observed virulent nature of CIO151 and the presence of only one plasmid we chose CIO151 as our reference strain for the plasmid studies. Additionally, its limited number of only two TAL effectors and their localization on the sole plasmid in CIO151 made it an ideal reference strain to study with simple genetic approaches including knockout and plasmid curing strategies. To date, no plasmids from Xam have been fully sequenced and annotated. At least eight plasmids from five other *Xanthomonas* species have been sequenced, but most of them lack any functional characterization (Leplae et al., 2010). Often, in next-generation sequencing data, plasmid contigs are predicted as leftover contigs that don't assemble to the genome and have slightly different GC percentages. The CIO151 genome has been sequenced and characterized in-depth, but no full plasmid sequences have been reported (Arrieta-Ortiz et al., 2013). I have assessed both functional and non-functional aspects of the CIO151 plasmid pXam46 in Xam via Sanger sequencing, annotation, TALE knockouts and plasmid curing.

3.1.3 TALE target prediction

Along with their explanation of the RVD cipher, Moscou and Bogdanove (Moscou and Bogdanove, 2009) published a Python script that uses experimentally confirmed TALE-target interactions to predict the most likely promoter binding sequence and host gene of any given TALE. The script, when combined with other patterns they've noticed, such as a T preceding each host binding site, can be used to accurately narrow down potential targets. Moscou and Bogdanove were able to use this strategy, along with rice microarray data showing upregulation by TALEs, to correctly predict four out of five targets. They speculated that the fifth TALE, AvrXa27, which targets the R gene Xa27 could represent a weak host adaptation to confer resistance that fits the code imperfectly. Their predictions were aided by the availability of rice promoter sequences.

Fortunately for this study, a draft of the estimated 770 Mb cassava genome has been completed that spans 96% of the protein coding space and a library of 30,666 predicted genes has been made available on the webserver Phytozome (Prochnik et al., 2012). Additionally, multiple TAL target predictors can be combined to improve predictive accuracy (Szurek, 2013). I have used these prediction tools to identify potential targets of both TAL effectors on the CIO151 plasmid.

3.1.4 Summary

Plasmids are a major component of bacterial diversity and a primary contributor to horizontal gene transfer. Whether just mobilizable, encoding a minimum set of components including a relaxase that allows it to transfer to another cell, or fully conjugative, encoding the entire set of components needed to transfer, the movement of plasmids offers bacteria the opportunity to quickly adapt to their environments. The presence of TALEs on plasmids represent a mobile arsenal, potentially capable of providing pathogenic tools to *Xanthomonas* spp. for increased virulence and disease. Sequencing the TALE-containing plasmid of CIO151 should provide a better understanding of how TALEs are obtained by Xam and assessing the functional contributions of the TALEs should increase our knowledge of TALE pathogenicity. By combining techniques old and new, I assembled the sequences of the two TAL effectors and analyzed several genetic components of the plasmid, here called pXam46. I report the sequence of pXam46 and an assessment of its contribution to virulence and potential conjugation. I provide an annotated version of pXam46, the contributions to virulence of each TALE by creating single and double knockouts of the effectors, and analyzed its potential involvement in horizontal gene transfer by assessing the presence of transfer gene repertoires.

3.2 Methods

3.2.1 Plasmid DNA extraction via CsCl gradient

Plasmid DNA from CIO-151, CFBP1851 (exhibits same banding patterns as CIO151), and from strain ORST4, which contains a spontaneous 8 kb deletion of the smaller of the two TALEs, was purified from chromosomal DNA using a cesium chloride equilibrium gradient method. Bacterial cultures were grown in 200 mL of medium with constant shaking. Cells were pelleted 5 min at 5,000 RPM, decanted, rinsed with H_2O then resuspended in 250 µL H_2O and transferred to a 50 mL conical tube. Lysis buffer (16 mL) was added (**Table 3-1**) and mixed until cells lysed, then incubated at 65ºC 2 minutes. Then 25 mL of phenol:chloroform:isoamyl alcohol (1:1:24) was added and mixed 2 minutes. Cells were pelleted 15 minutes at 8,000 RPM at 4ºC. Only 4 mL of supernatant was carefully collected from the top of the upper phase, transferred to a new tube, and 1 g CsCl per 1 mL DNA was added. The following was added to a 5.5 mL heat seal Vti80 rotor tube: 0.4 mL EtBr (10 mg/mL), 1 mL of DNA/CsCl mixture, and 3.6 mL of 1 g CsCl/1 mL TE (Tris/ethylenediaminetetraacetic acid) stock. The tubes were heat sealed and the gradient was run at 55,000 RPM 20 – 24 hours at 15ºC, deceleration 9. Tubes were punctured with a 1 mL syringe and 18-gauge needle and the plasmid bands were extracted and collected into a 15 mL conical tube. The EtBr was extracted 4 times with 1 volume of water saturated butanol, then dialyzed in TE buffer at 4ºC overnight. If necessary, DNA volume was reduced with PEG 35,000 crystals on the outside of the dialysis tubing for 1 hour. DNA was then collected into a new 1.5 mL tube.

Table 3-1 Lysis buffer for cell preparation of CsCl DNA extractions and plasmid DNA.

Lysis Buffer (10 mL)	
H2O	6.25 mL
NaOH (125 mM)	62 µL (10 N stock)
NaCl (30 mM)	60 µL (5 M stock)
Tris (50 mM)	0.5 mL (1 M, pH 8 stock)
EDTA (5 mM)	100 µL (0.5 M stock)
SDS (3%)	3 mL (10% stock)

3.2.2 Restriction enzyme mapping

Purified plasmid DNA was then cut with EcoRI (which cuts outside of Xam TALE ORFs), BamHI (which cuts just inside both ends of the TALE ORF), and XhoI individually and in paired combinations. Banding patterns between CIO-151, CFPB1841, and ORST4 were compared to discern location within the plasmid. TALE fragments were identified via southern blot analysis by running these digests on an agarose gel, imaging them, then transferring to Hybond N+ nylon membrane and probing with a P-32 labeled TALE BamHI fragment. Band size patterns between single and double digestions along with the location of the TALEs as identified with the southern analysis was analyzed and used to create a simple restriction map containing the used enzymes. Further mapping was done with the enzymes PstI and HindIII.

3.2.3 Separation and sequencing of TAL effectors

Xam strain plasmids were cut with BamHI and run on an agarose gel to predict the number and size of TALEs present. RVD number was predicted from the total size of each TALE since each RVD is 102 bp. TALE-containing fragments were confirmed via southern blot analysis. In order to obtain sequencing outside of the TALE regions, multiple methods of subcloning were utilized. To obtain larger subclones, the plasmid was cut with EcoRI to produce fragments of 27 kb, 14 kb, and 5 kb and then ligated into the EcoRI-cut, broad-spectrum host vector pLAFR3 (Tc^R, ColE1) using T4 ligase. These pLAFR3 ligation reactions were then combined with lambda packaging extract and packaged into E. coli DH5a cells via bacteriophage packaging as described (White and Nichols, 2003). Plasmid DNA was extracted from successful cosmid clones and cut with EcoRI to verify and determine which insert was cloned. In an attempt to obtain smaller subclones to sequence, pXam46 was partially digested with TSP5091, an enzyme that cleaves the frequently occurring four base pair sequence AATT to make a subclone library. Fragments of 5 kb were extracted and subcloned into the pBluescript vector pBKs for blue/white screening on X-gal containing LB plates. All obtained subclones were then sequenced across, employing Sanger sequencing to obtain no more than 1 kb long reads per sample. For repetitive regions within subclones, direct sequencing from gel eluted restriction fragments was utilized.

3.2.4 Final assembly and analysis of remaining open reading frames

The majority of the CIO151 Xam plasmid pXam46 was Sanger sequenced to obtain 3 large contigs, however several regions proved too repetitive to sequence and difficult to clone. The region of most difficulty was the large section of the 27 kb EcoRI pLAFR3 subclone. I had no smaller subclones and could only sequence into a limited portion of the insert. This was likely largely due to the sheer size of the clone and its 20 kb pLAFR3 backbone for a total vector and insert size of 47bkb. Using Illumina short-read sequencing data from genomic DNA of CIO151 from Bart et

al. (Bart et al., 2012), the plasmid regions that were difficult to sequence with Sanger methods were bridged by mapping the reads to the Sanger-sequenced CIO151 fragments as a reference. These bridged regions have been partially confirmed via Sanger sequencing. All AUG ORFs were predicted using a combination of the default parameters in CLC Genomics Workbench 6.0 as well as Glimmer (Delcher et al., 2007). Annotations were assigned using the top predicted gene from BLASTx (NCBI BLAST).

3.2.5 TALE knockouts and cassava growth assays

Prior attempts at making markerless TALE knockouts often led to plasmid curing during selection for the 2nd crossover event. Therefore, we developed a universal TALE suicide-vector construct to make TALE knockouts while also being able to select for them with antibiotics (**Figure 3-1A**). This involved disrupting the TALE start codon by recombining in a TALE region missing the start site. Using PCR, primers were designed (**Table 3-2**) to amplify a 667 bp TALE fragment 5' of the RVD region–a highly conserved region among all Xam TALEs–but excluding the ATG start codon of *TAL13*$_{Xam668}$. This fragment was amplified and directionally cloned into the entry vector pENTR/D-TOPO (Km^r). I then used an LR clonase reaction to put each insert into the suicide vector, pLVC18 (Tc^r) and transformed this vector into chemically competent *E. coli* DH5a. I conjugated this pLVC18-based suicide vector construct into the Xam strain CIO151 via triparental mating with the pRK600 *E. coli* helper strain (chloramphenicol, Cm^r). I then cycled the conjugants several times on NYGB (Rf, Tc) plates to select for the first crossover. This first crossover event was sufficient to knockout the targeted TALE due to the modified ATG site, whether deleted or mutated. Further, a pXam46-cured CIO151 strain was created by attempting a second crossover event (**Figure 3-1B**). Colonies that had been confirmed to contain the suicide vector pLVC18 (Tc) were cycled on Rf and then transferred to Rf/Tc plates. All Tc sensitive colonies were screened for plasmid loss by agarose gel electrophoresis.

Table 3-2 Primers for making TALE knockouts using pLVC18. These primers were used to amplify a 665 bp region 5' of the RVDs and excluded the TALE start codon. They are TOPO compatible for cloning into pENTR/D-TOPO and subsequent LR cloning into the suicide vector pLVC18. They were designed using *TAL13*$_{Xam668}$ as a template but should be universally compatible with most Xam TALEs.

Primer	Sequence	Amplicon size
TALE Fw ΔATG TOPO	caccCGAACCCGTCTCCCGTCTC	667 bp
TALE Rev ΔATG	ACCGCGGTCACGCCGCCACGT	-

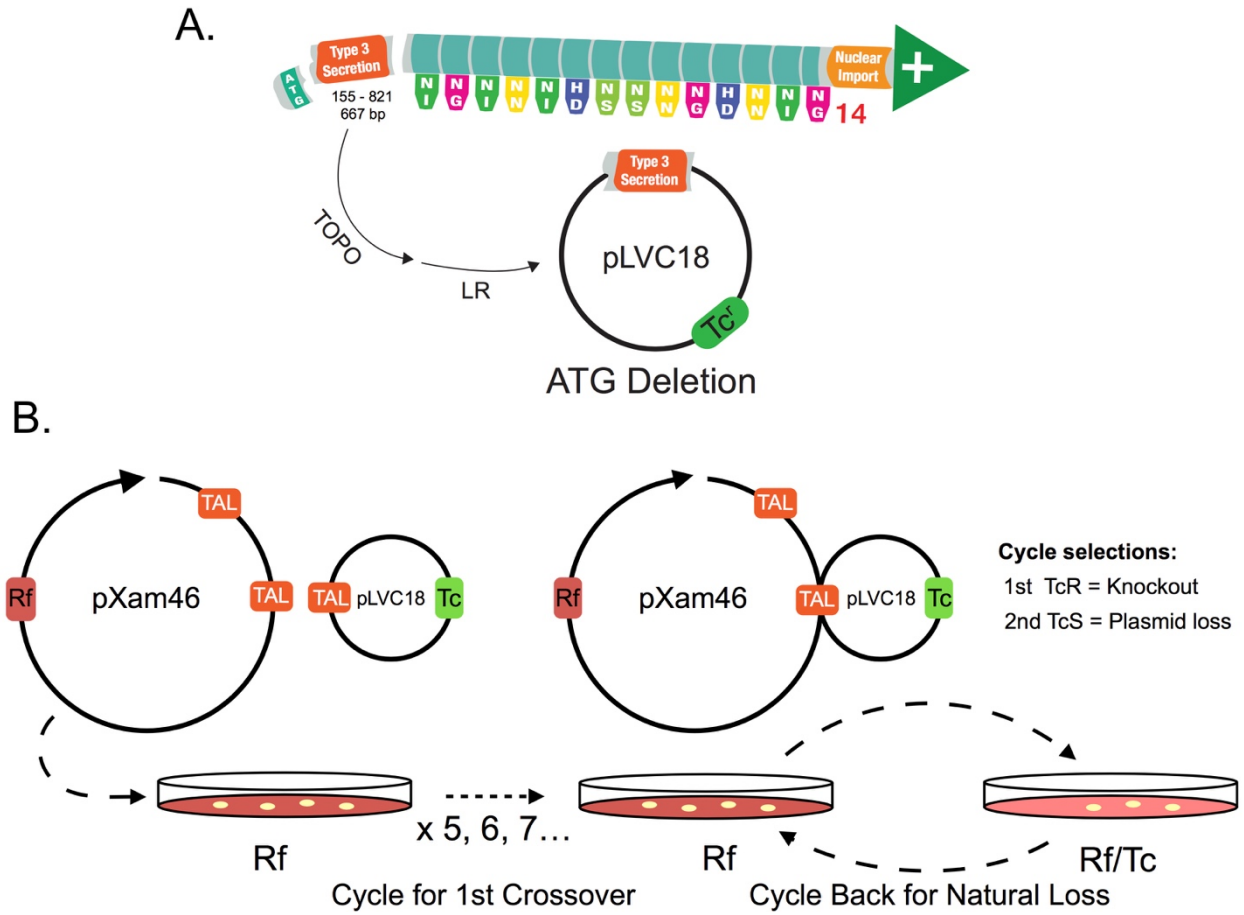

Figure 3-1 Single TALE knockouts and plasmid curing. A. A 667 bp region 3' of the ATG start codon but 5' of the RVD region was cloned into the suicide vector pLVC18 Tc^r. **B.** The pLVC18 TALE constructs were introduced into Xam via triparental mating, cycled on Rf to select for a 1st crossover event, and screened by Western blotting for TALE knockout confirmation. Spontaneous plasmid curing was obtained by cycling pLVC18 containing strains back on Rf and then screening by DNA plasmid prep for plasmid loss.

After obtaining the single and double TALE deletion constructs, I determined the impacts on virulence by measuring bacterial growth in inoculated leaves over the course of 6 days as described (**2.6.1**). Cultures were inoculated at OD_{600} = 0.0005 to an area of approximately 0.25 cm^2. For each sample to be taken, 6 inoculations were made, accounting for 2 leaf punches per sample and 3 reps per sample, in triplicate. In this case with 7 strains (CIO151, 151Δp, 151ΔTAL14, 151ΔTAL21, Xam668, 668ΔTAL14, and 668ΔTAL20), 2 inoculations were required for 2 leaf punches per sample times 3 in triplicate times two time points, equaling 84 required inoculations. Immediately after inoculation (Day 0) and 6 days after inoculation, two 0.5 cm^2 punches were combined in triplicate per sample. At day 0, the two punches were ground using a bead beater and resuspended in 500 µL $MgCl_2$ (10 mM) and plated (50 µL) onto selective NYGA plates at 28ºC. Colonies were counted 3 days later. Day 6 samples were serially diluted and plated at 10^{-4}, 10^{-5}, and 10^{-6} dilutions. After counting, CFU per mL of ground tissue was back calculated from the initial suspension and subsequent dilutions.

3.2.6 Analytical assessment of conjugative ability

In order to assess whether pXam46 may be in involved in HGT, I analyzed the plasmid sequence for ORF clusters related to transmissibility and conjugation. I compared the annotated genes of pXam46 to those of other *Xanthomonas* and non-*Xanthomonas* self-transmissible and conjugative plasmids as a reference, including other TALE-containing plasmids. The genes on pXam46 were then assessed for type, number and function. Specifically, in requirement for a conjugative plasmid, pXam46 was assessed for the ability to encode a relaxase involved in the initial nicking to prepare the plasmid for transfer, a type IV coupling protein (T4CP) to receive and prepare the plasmid DNA for transport, and a T4SS involved in the actual delivery of the plasmid to a neighboring cell. Attempts at assessing conjugative ability were made through attempted incorporation of an antibiotic resistance marker in order to track mobilization of the plasmid to another strain.

3.3 Results/Discussion

3.3.1 Plasmid assembly and ORF annotation

In order to characterize the size, restriction sites, and the location of TALEs on the CIO-151 plasmid without any sequencing data, the plasmid was cut with various restriction enzymes individually and in combination and then probed using a P-32 labeled BamHI internal fragment from of $TALE_{Xam1}$ shown to hybridize to all Xam TALEs (**Figure 3-2**). The number of fragments created identified the number of cut sites for any given enzyme and the fragments that lit up with the probe indicated those with TALEs. Using these data allowed the assembly of a restriction enzyme map and the placement of the TALEs. For example, cutting with EcoRI produced three fragments (27 kb, 14 kb, and 5 kb). The 14 kb and 5 kb fragments hybridized with the probe, suggesting they each contained TALEs or TALE fragments. When cut in combination with BamHI which we've observed to cut just inside the ends of the TALE ORF, both the 14 kb and 5 kb fragments disappeared, giving way to smaller 3.2 and 3.7 kb TALE hybridized fragments. The mapping of the TALEs within the plasmid and identification of restriction enzyme sites provided context for developing a cloning and sequencing strategy. In addition, I used the fragment sizes created by the different enzymes to predict that the the plasmid was approximately 46 kb in size. Given the size and the restriction sites, I used a combination of subcloning strategies guided by my restriction map along with Sanger sequencing to slowly assemble the plasmid. I obtained four primary subclones of different sizes and containing different genes of interest in three different vectors. The smaller, individual TALE clones were essential for sequencing the TALEs, as the presence of more than one TALE on a plasmid leads to mixed peaks due to their high similarity. The larger pLAFR3 subclones that were obtained via lambda packaging of EcoRI fragments proved essential for sequencing the regions outside of the TALEs, especially the 27 kb fragment that does not contain any TALEs.

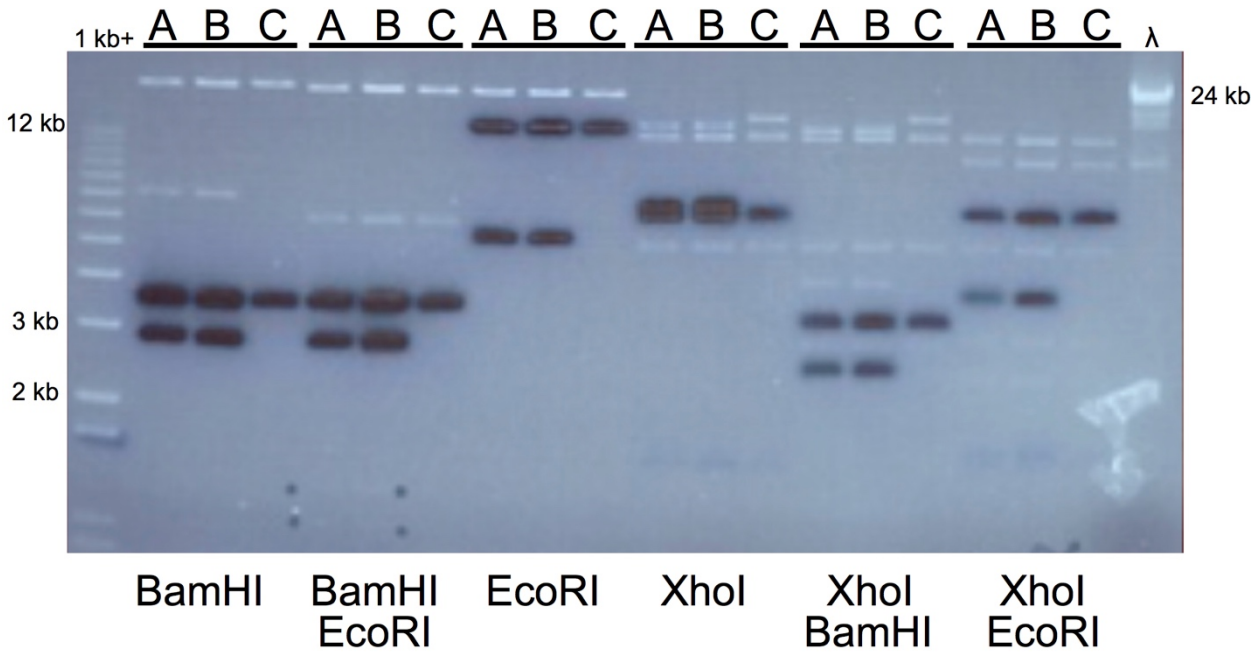

Figure 3-2 Restriction enzyme mapping of the CIO151 plasmid pXam46. The CIO151p pXam46 (**A**) as well as a plasmid from CFBP1851 that shows similar banding patterns (**B**) and the ORST4 3 kb TALE deletion plasmid (**C**) were cut with BamHI (previously demonstrated to cut just inside the TALE 5' and 3' ends), EcoRI (observed to cut outside of TALE ORFs), and XhoI both in single and double digests. Southern blot analysis was conducted using a ^{32}P labeled TALE BamHI fragment to identify TALE-containing fragments. Here, the Southern blot is overlayed on top of the restriction digest. The missing 3 kb TALE of the ORST4p can be seen in the BamHI digest. Based on these banding patterns, a restriction enzyme map of pXam46 was assembled.

The final assembly of the plasmid was constructed using a majority of Sanger sequenced reads combined with Illumina short-read data of genomic DNA from CIO151, which contains plasmid DNA as well as chromosomal DNA (**Figure 3-3**). Using genomic DNA was not ideal but because most of the plasmid had been subcloned and sequenced already, it proved still a valuable resource for assembling the plasmid and assemblies were supported by some further sequencing analysis and restriction digest patterns. Ultimately, the plasmid pXam46 (46,478 bp) was assembled. pXam46 was predicted to have 62 ORFs by Glimmer (Delcher et al., 2007) but through further analysis primarily using BLASTx (NCBI BLAST) to check each ORF I annotated just 55 genes (**Table 3-3**) for an average gene density of 1.18 ORFs per 1 kb. Of these 55 genes, 32 have been annotated to have some known or predicted function and 23 of them are conserved hypothetical or putative genes. Many of the genes are predicted to be involved in diverse functions like plasmid stability and maintenance, conjugative transfer, transposition, and secretion. Several genes may be involved in pathogenicity. A few genes are predicted to be type III secreted effectors and two of the annotated genes are the previously described TALEs, $TAL14_{CIO151}$ and $TAL21_{CIO151}$.

Table 3-3 Translated and annotated ORF results for pXam46 by BLASTx. ORFs as predicted by CLC Main Workbench 6.0 and Glimmer were annotated by closest BLASTx hit.

Name	Start	End	Size (nt)	% AA Id.	Top Match	Organism
TAL14	4320	1183	3138	99	ADM80412	*X. ax.* pv. manihotis
Hyp	4596	4856	261	96	WP_054318962	*Xanth. axonopodis*
Hyp	4853	5908	1056	100	WP_017154877	*Xanth. axonopodis*
Hyp	7943	7122	822	100	WP_033481311	*Xanth. axonopodis*
Hyp	8450	8268	183	100	WP_050544949	*Xanth. axonopodis*
ISXac2	8522	8788	267	100	WP_011051331	Multisp. *Xanth*. (Xac)
ISXac2 transp.	8782	9609	828	100	WP_011346757	Multisp. *Xanth*. (Xcv)
Hyp	10147	10548	402	100	WP_017154948	Multisp. *Xanth*.
T3E	10854	9652	1203	99	EFF46466	*X. fusc.* subsp. aurant.
Hyp transp.	10971	11315	345	96	EFF46465	*X. fusc.* subsp. aurant.
TAL21	11531	15337	3807	87	WP_044756818	*Xanth. oryzae*
ISXac2	16485	16751	267	100	WP_011051331	Multisp. *Xanth*. (Xac)
ISxac2 transp.	16745	17572	828	100	WP_011051331	Multisp. *Xanth*.
Hyp	16404	16078	327	98	WP_046934538	*Xanth. perforans*
Hyp	17779	17606	174	100	WP_017166921	*Xanth. axonopodis*
Hyp	18280	17888	393	99	CDF63654	*X. fusc.* subsp. fuscans
Hyp	18684	18292	393	100	WP_017154798	Multisp. *Xanth*.
Abi Family	20178	19156	1023	99	EKQ58643	*X. ax.* pv malvacearum
Invertase	20998	20432	567	100	WP_017159668	*Xanth. axonopodis*
Hyp T3E	22451	21249	1203	99	EFF46466	*X. fusc.* subsp aurant.
Hyp	22568	22858	291	89	WP_017158366	Multisp. *Xanth*.
Hyp	23247	22930	318	100	WP_007966095	Multisp. *Xanth*.
ParA1	23890	23255	636	100	WP_007966093	Multisp. *Xanth*.
Hyp	24794	24000	795	100	EFF42076	*X. fusc*. subsp aurant.
Hyp	26400	25582	819	100	WP_033480771	*Xanth. axonopodis*
Hyp	27333	26941	393	100	WP_007974941	Multisp. *Xanth*.
Hyp	27726	27472	255	100	WP_017154799	Multisp. *Xanth*.
ParA2	28352	27723	630	100	WP_017158739	*Xanth. axonopodis*
Endonuclease	28683	28381	303	98	CDF63650	*X. fusc.* subsp. fuscans
Hyp MobD	28770	28904	135	100	WP_011153914	*Xanth. citri*
ISXac2	28969	29235	267	100	WP_011051331	Multisp. *Xanth*. (Xac)
ISxac2 transp.	29229	30056	828	100	WP_011346757	Multisp. *Xanth*. (Xcv)
Hyp	30186	30317	132	100	WP_0171164192	*Xanth. axonopodis*
Acetyltransf.	30981	30484	498	100	WP_017157757	*Xanth. axonopodis*
Hyp	31456	31674	219	100	WP_033478260	Multisp. *Xanth*.
Hyp	31259	30978	282	100	WP_017157756	*Xanth. axonopodis*
PemK	31671	31997	327	100	WP_033475610	Multisp. *Xanth*.
TraI ATPase	33023	32067	957	100	WP_017160605	*Xanth. axonopodis*
ISxac2 transp.	33839	33024	816	98	WP_011346757	Multisp. *Xanth*. (Xcv)
ISxac2	34099	33833	267	100	WP_011051331	Multisp. *Xanth*. (Xac)
TraI ATPase	36229	34148	2082	100	WP_017158768	*Xanth. axonopodis*
TraB	37817	36243	1575	100	WP_017158767	Multisp. *Xanth*.
Hyp secreted	38626	38090	537	100	WP_007974814	Multisp. *Xanth*.
Transp.	39710	38910	801	62	WP_053842402	*Acidovorax avenae*
Transp.	40750	39710	1041	100	WP_033474620	Multisp. *Xanth*.
Hyp Secreted	41224	40814	411	98	WP_007974812	Multisp. *Xanth*.
RelE	41630	41337	294	99	WP_017158795	Multisp. *Xanth*.
Antitoxin	41920	41627	294	100	WP_033475978	Multisp. *Xanth*.
TraA	42389	41994	396	100	WP_017164031	*Xanth. axonopodis*
TraD	42746	43192	447	100	WP_017158798	Multisp. *Xanth*.
StbB	43896	43201	696	100	WP_017158799	Multisp. *Xanth*.
Hyp	43911	44297	387	100	WP_017158800	*Xanth. axonopodis*
Tn3 resolvase	45438	44476	963	100	WP_017158801	*Xanth. axonopodis*
Hyp	45605	45865	261	96	WP_054318962	*Xanth. axonopodis*
Hyp	45862	439	1056	100	WP_017154877	*Xanth. axonopodis*

Hyp. = hypothetical; transp. = transposase; T3E = type III effector; acetyltransf. = acetyltransferase; *X*. = *Xanthomonas*; *fusc*. = *fuscans*; *aurant*. = *aurantifolii*.

The plasmid has a 61.4% GC content overall, which compares to 66.6% GC content of the repetitive TAL effector genes and is lower than the 65.1% GC content of the CIO151 genome (Arrieta-Ortiz et al., 2013). Notably, the plasmid is full of genes involved in transposition, containing 14 transposition related genes, including 4 insertion sequences (IS) transposases of the *ISXac* gene family alone. Insertion sequences and transposable elements are widely associated with and abundant in Xanthomonads, suggesting a major role in their evolution and diversity (Ferreira et al., 2015). IS elements have also been shown to play a role in rapid pathogen evolution away from resistance, such as IS476 in Xcv, disrupting *avrBS1* and thwarting the pepper *Bs1*-dependent resistance (Kearney and Staskawicz, 1990). Additionally, transposases are the most abundant genes found in metagenomes (Aziz et al., 2010). Compared to other *Xanthomonas* plasmids, however, the 4 transposases on pXam46 represent an abnormally abundant level of IS elements. The Xac genome is predicted to have 44 IS elements, but only 3 are attributed to plasmids with 2 on pXac33 and 1 on pXac64 (Monteiro-Vitorello et al., 2005). A high percentage of IS elements is common in other *Xanthomonas* species including *X. oryzae pv oryzae* and *X. campestris* pv. *vesicatoria* where transposases comprise up to 11% and 1.9% of the genome, respectively (Varani et al., 2013), however the presence of so many plasmid-localized IS elements as in pXam46 is less so. Notably, the IS elements are associated with the few Sanger gaps in the assembly and even with assembled genomes and Illumina reads, de-novo assembled contigs of previous CIO151 genome assemblies do not fully span the plasmid (Arrieta-Ortiz et al., 2013; Bart et al., 2012). The repetitive nature of the IS elements and the inverted repeats that flank them likely contribute to difficulty in assembling, as does the RVD region of TALEs. In this case, pXam46 contains two TALEs with highly similar surrounding sequences, compounding assembly capabilities. These few low-confidence regions (6,000 – 7,000 and 27,000 – 34,000) are surrounded by IS elements and represent regions that could be corrected with longer reads from new sequencing technologies such as PacBio.

3.3.2 TAL effector sequences

TALEs can be difficult to sequence. The presence of highly repetitive sequence both in and outside of the TALE regions makes sequencing by any method very difficult. While we were able to assemble the CIO151 genome using Illumina next-generational sequencing short, paired-end reads, we were not able to assemble any large, high quality contigs of the plasmid via *de novo* assembly based on a comparison to the restriction map and were not able to assemble the individual TALE sequences (Bart et al., 2012). Additionally, when multiple TALEs are present PCR preferentially amplifies the smallest TALE. Digestion with BamHI, which cuts just inside both ends of the TALE ORF, can make it difficult to distinguish multiple TALEs of similar sizes. On the other hand, the highly conserved nature of TALEs (outside of the repeat region), including the sequence containing the T3S signal, the NLS, and the AAD affords us the convenient ability to identify and sequence any TALE using the same probes and primers. One strategy I used was to perform a partial digest of the plasmid with TSP509I in order to make a subclone library. By this method, a subclone of the 3 kb $TALE14_{CIO151}$ was isolated just outside of the BamHI sites, short enough to Sanger sequence. Shorter TALEs like $TALE14_{CIO151}$ can be sequenced through entirely using the TALE primers listed (**Table 2-5**).

The longer TALE was not as simple to sequence. Through all of our TALE sequencing experience, we have identified some restriction sites, such as EcoRI (GAATTC), which do not seem to be present in the TALE ORF, but are still present throughout the chromosome and plasmid sequences. Using the pLAFR3 subclones, I was able to further subclone a 14 kb EcoRI fragment that contained the 3.7 kb *TALE21*$_{CIO151}$. Having 21 RVDs at 102 bp each, the repeat region was too large to be sequenced from the unique primer sites outside of the RVD region (a limitation of the <1 kb Sanger reads). To gain access to this middle region, a SacII fragment and a BspMI/SalI fragment were removed from the 5' and 3' regions, respectively flanking the RVDs so the RVD region could be sequenced in from the M13 vector primer sites. Based on the sequence obtained from these subclones and gel electrophoresis, it was apparent that a second BspMI site existed 100 bp upstream of the known BspMI site in the repeat region. A primer with a 3' end unique to the known BspMI site was used to sequence through the second site. Mixed peaks resulted as the primer annealed to both BspMI sites but were clarified using the clean sequence from the BspMI/SalI digested subclone.

These two sequenced TALEs joined another set of five sequenced TALEs from Xam668 (**Table 3-4**), an Indonesian isolate that we have observed to be particularly virulent (Bart et al., 2012). The smaller of the two CIO151 TALEs, *TAL14*$_{CIO151}$, shares the same sequence as *TALE1*$_{Xam}$ formerly known as *pthB* from CFBP1851 (Castiblanco et al., 2012) and only differs from *TAL14*$_{Xam668}$ by one RVD. The larger of the two, *TAL21*$_{CIO151}$, has a unique RVD sequence. Other than the overlap of both versions of TAL14 in the Columbian strains, none of the sequences are exactly the same, which may be explained by their origins—CIO151 is from Columbia, whereas Xam668 is from Indonesia—though the two 14 RVD TALEs and the 21/22 RVD sequences are similar, suggesting that these TALEs may have a common ancestor or simply have evolved to target the same EBE.

Table 3-4 TAL effector RVD sequences of Xam isolates CIO151 and Xam668.

TALE	Repeat Variable Diresidue (RVD) Sequences
13$_{Xam668}$	NI NS NG HD NG HD NI NG HD NN NI NI NG
14$_{Xam668}$	NI NG NI NN NG HD NS NS NN NG HD NN NI NG
14$_{CIO151}$	NI NG NI NN NI HD NS NS NN NG HD NN NI NG
15$_{Xam668}$	NI NG NI NN HD HD NS NS NS HD HD NS HD NG NG
20$_{Xam668}$	NI NG NI NN NI HD NS NS NN NG HD NS HD NN HD HD HD NI NG NG
21$_{CIO151}$	NI NG HD NG HD N* NG NG HD HD HD NG N* NG HD NG NG NG HD NG NG
22$_{Xam668}$	NI NG HD NG NG NG HD HD NG NG HD NG HD HD NG NG HD NG NG HD NG NG
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Cipher:	**NI = A NG = T HD = C NN = G/A NS = A/C/T/G N* = C/T**

Cipher = TALE RVD binding code based on (Boch et al., 2009; Moscou and Bogdanove, 2009) and updated in (Richter et al., 2016).

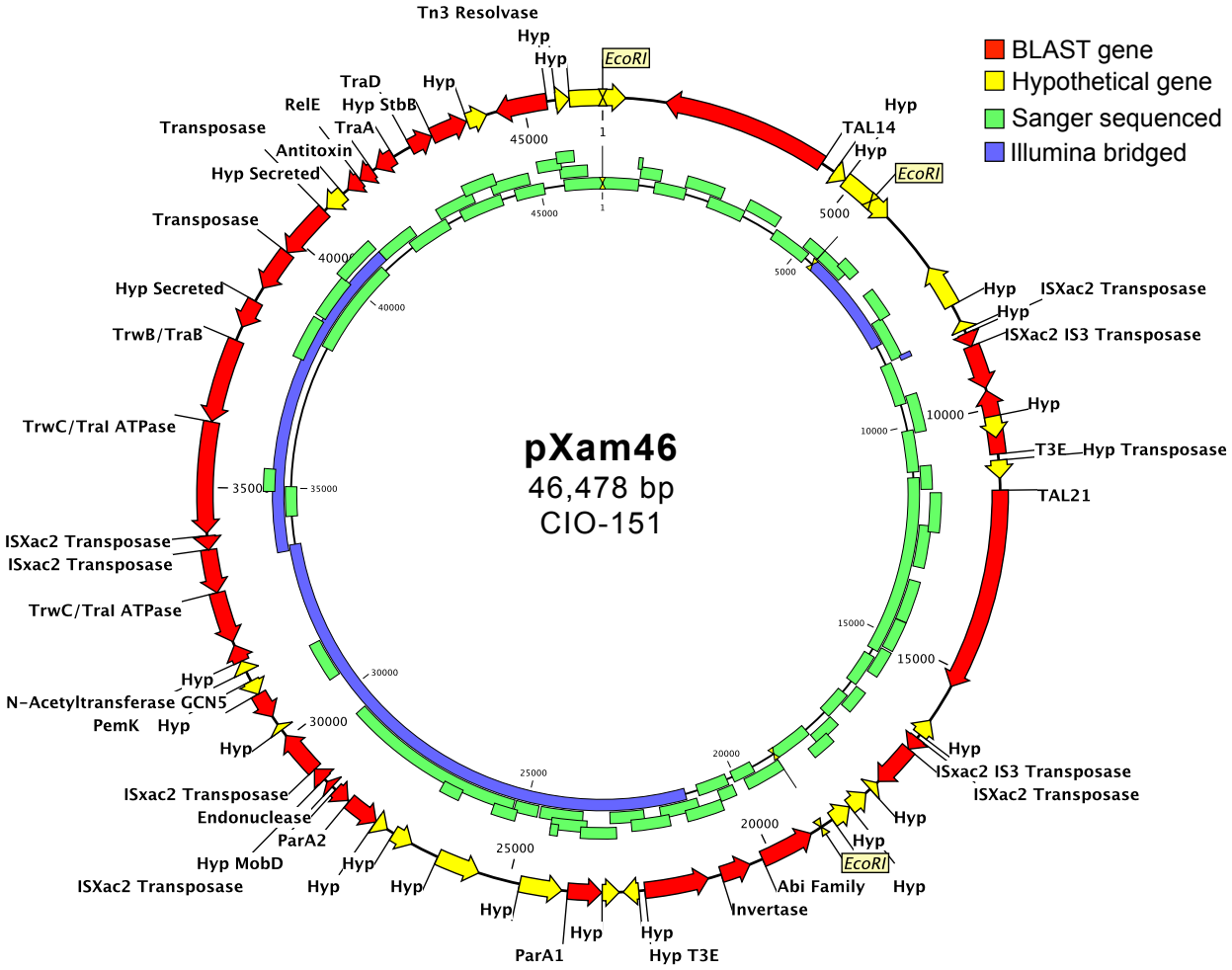

Figure 3-3 pXam46 annotations and sequencing. The single plasmid from Xam strain CIO151 and its gene annotations. Initially, pXam46 was mapped using restriction enzyme analysis and subsequently subcloned using both a 5 kb library strategy and in pLAFR3 subclones of the 3 EcoRI fragments shown here at 1 bp, 5470 bp, and 18928 bp. The two TALEs, *TAL14*$_{CIO151}$ and *TAL21*$_{CIO151}$ were Sanger sequenced and characterized first. Then, the remaining sequence was obtained mostly by Sanger (green). Some regions proved difficult to sequence and subclone, so the plasmid gaps were bridged by using Illumina data from genomic CIO151 assemblies (Bart et al., 2012). Contigs were created by mapping the 100 bp paired-end short reads against the existing Sanger sequence as a reference in CLC Genomics Workbench. Contigs were found that bridged the Sanger fragments (purple) and were then sequenced partially by Sanger to confirm. The plasmid was annotated for ORFs in CLC Main Workbench 6.0 and using the open access tool Glimmer (Delcher et al., 2007). Identified ATG ORFs were then annotated using blastn of NCBI BLAST. Thirty two genes with well-characterized functions were annotated (red) and 23 identified as conserved proteins with no known function were identified as hypothetical (yellow).

3.3.3 TAL14$_{CIO151}$ contributes to virulence

A beneficial aspect of having only two TALEs in CIO-151 is that single knockouts were easier to obtain than a strain with several TALEs and each knockout was more informative in terms of determining the individual roles of any single TALE in the context of other TALEs and their comparative impacts on virulence. In order to assess their contributions to virulence, TALE

knockouts were made in pXam46 via allelic exchange using the suicide vector pLVC18 (**Figure 3-1A**) and tested for growth in cassava. Typically for Gram-negative bacteria like Xam, we would use a suicide vector with the sucrose sensitive operon *sacBR*. The *sacBR* operon catalyzes sucrose hydrolysis and subsequent synthesis and accumulation of a toxic amount of the polymer levan in the periplasm, making it a useful tool in Gram-negative bacteria for positive selection of a second crossover for a markerless deletion (Gay et al., 1985; Tan et al., 2012). However, we discovered through our attempts to make markerless deletions in Xam that the *sacBR* gene would not be necessary and, surprisingly, could actually be a hindrance. First of all, the presence of sucrose in the plant and the demonstrated and predicted ability of *Xanthomonas* spp. to upregulate host SWEET sucrose transporters meant that the *sacBR* gene would likely kill the bacteria in planta and skew the results. Second, after obtaining single crossovers that were Km resistant and sucrose sensitive, additional cycling for the second crossover event would often result in complete plasmid curing, losing the suicide vector and taking the native plasmid with it. Therefore we developed a new knockout strategy where we utilized a single crossover event along with a modified ATG start site in the TALE gene, disrupting the start site and maintaining the pLVC18 suicide vector along with Tc resistance in the bacteria. This allowed for the continued selection of the knockout strain by Tc. Using this method I created individual knockouts of $TAL14_{CIO151}$ and $TAL21_{CIO151}$. In addition, to assess the overall contribution of plasmid genes to virulence relative to the individual TALE knockouts, I took advantage of the plasmid curing side effect of selecting for the second crossover to create a plasmid cured CIO-151 strain, 151Δp. The spontaneous loss of pXam46 in 151Δp was confirmed by plasmid mega preparation, PCR marker analysis, and western blot (**Figure 3-4**).

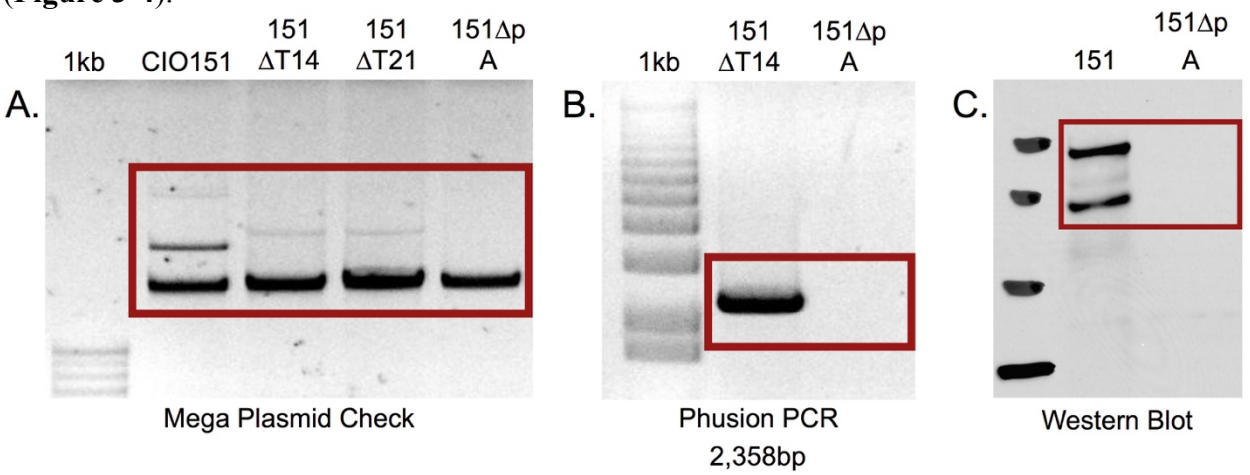

Figure 3-4 Confirmation of a CIO151 pXam46 cured variant 151Δp. A. CIO151 TALE knockout constructs containing a single crossover of pLVC18 Tcr were cycled on Rf and then screened for spontaneous plasmid curing by DNA mega plasmid preparation. **B.** A 2,358 bp pXam46 marker was PCR amplified to screen for plasmid cured candidates. **C.** Finally, the candidate plasmid cured strain was induced in minimal medium to express T3Es and checked by Western blot for TALE expression.

I tested the virulence of CIO151, 151Δp, 151ΔTAL14 and 151ΔTAL21 by measuring the ability of the Xam strains to grow in planta after leaf inoculation (**Figure 3-5**). As controls and for comparisons, I used the virulent, water-soaking strain Xam668 and two Xam668 TALE knockout strains I made that showed decreased growth, 668ΔTAL14 and 668ΔTAL20 (Cohn et al., 2014). All strains were inoculated at a low OD (OD_{600} = 0.005) resulting in an inoculation of

approximately three colony forming units per cm^2 leaf tissue. As expected, Xam668 showed significantly more growth according to a two-tailed t-test than 668ΔTAL14 ($p = 0.008$) and 668ΔTAL20 ($p = 0.007$), differing by over 0.5 log_{10} CFU per 1 cm^2 (0.57 and 0.50 difference in growth for 668ΔTAL14 and 668ΔTAL20, respectively). The differences for CIO151 were less but still significant for one TALE. Strain 151ΔTAL14 differed significantly ($p = 0.013$) by 0.29 log_{10} CFU per 1 cm^2 or 4.35 x 10^7 CFU per cm^2. Therefore, $TAL14_{CIO151}$ seems to play a role in bacterial growth that at 6 dpi appears minor but significant. The role of the other TALE, $TAL21_{CIO151}$, is less clear. While there is a minor difference in growth (0.16 log_{10} CFU per 1 cm^2), it is not significant ($p = 0.185$). I have conducted this experiment twice and it shows similar results with a significant decrease for 151ΔTAL14 but a moderate difference for 151ΔTAL21. It is possible and we have observed before that this difference may in fact be significant but requires a more sensitive and potentially more relevant assay such as a stem inoculation growth assay in order to clarify apparently minor differences (Cohn et al., 2015b).

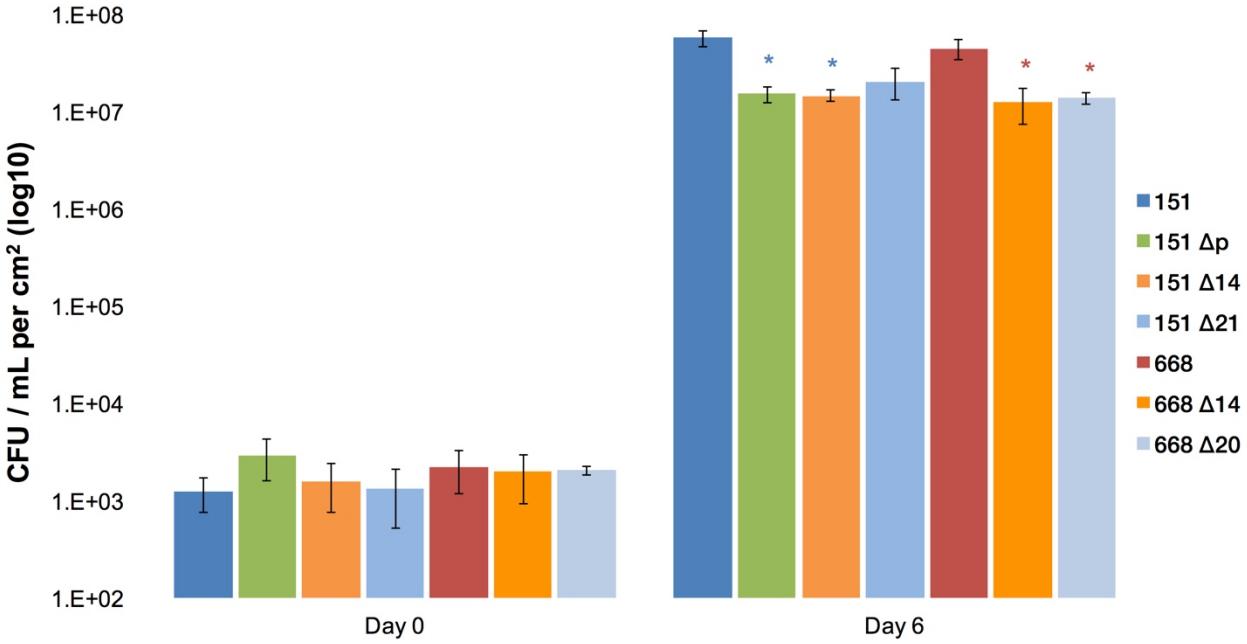

Figure 3-5 Apoplast growth assay of pXam46 cured CIO151 and individual TALE knockouts. Cassava leaves were inoculated with Xam cultures (WT CIO151, a plasmid cured variant 151Δp, individual TALE knockouts 151Δ14 and 151Δ21, WT Xam668 and 668Δ14 and 668Δ20) at $OD_{600} = 0.0005$ to an area of approximately 0.25 cm^2. At each time point and for each sample two 0.5 cm^2 punches were combined in triplicate, pulverized by bead beater, and plated in serial dilutions for single colonies. Error bars represent ±1SD.

The use of the pXam46 plasmid cured strain, CIO151Δp, in combination with the single knockouts allowed me to determine the overall role in virulence of pXam46 relative to the TALEs. Surprisingly, it appears that the sole contribution to virulence of pXam46, at least as measured by this apoplast growth assay, is $TAL14_{CIO151}$. I suggest this because the growth of 151Δp and 151ΔTAL14 were about equal. Both were significantly less than WT CIO151 ($p = 0.018$ for WT/151Δp). The plasmid cured strain differed by 0.28 log10 CFU per cm^2 from WT compared to the nearly identical difference of 0.29 log10 CFU per cm2 for 151ΔTAL14. In my other replicate, results were similar where both strains were significantly decreased and the plasmid cured strain grew to similar levels as 151ΔTAL14. This result, however, is not entirely surprising.

Based on the plasmid sequence of pXam46, it appears that most of the genes are involved in general plasmid biology and only a few, especially the TALEs, have a more apparent role in pathogenicity. These results suggest that the major role of this plasmid in pathogenicity is as a vessel for carrying TALEs, specifically $TAL14_{CIO151}$. That said, there are likely other roles of the plasmid for the biological benefit of Xam that are not apparent in an in planta growth assay. The presence of the Abi family gene (abortive infection system) suggests there is some mechanism of protection against phage where the infected cells can initiate cell death (Varani et al., 2013). The *RelE* toxin/antitoxin predicted genes suggests that once obtained this plasmid has mechanisms to ensure its maintenance in the gene pool. Toxin/antitoxin (TA) systems are clever evolutionary mechanisms for plasmids to ensure their maintenance. By encoding a stable toxin that is deadly to the bacterial cell as well as a less stable antitoxin, the plasmid ensures that upon errant replication where the plasmid is not copied, the lingering toxin will remain after the antitoxin degrades, leading to cell death in any cells lacking the plasmid. Another example of this is the CcdB/CcdA TA systems (Hayes, 2003). The *CcdB* cassette is used in the suicide vector pLVC18 as a mechanism to ensure it can only be selected for upon first crossover and disruption of the cassette. These other genes are part of what is known as the plasmid backbone, the area of the plasmid necessary for stabilization and maintenance (Garcillán-Barcia et al., 2011). The backbone genes suggest plasmid roles besides pathogenicity but important in their own right, providing mechanisms of bacterial defense against phage and ensured maintenance of the plasmid using the TA system. However, it is still notable that loss of the plasmid and the knockout of just a single TALE lead to a similar in planta growth phenotype. This suggests an important role for $TAL14_{CIO151}$, which is also known to be in other Xam strains (CFBP1851) and a very similar TALE in Xam668, differing by only a single RVD.

3.3.4 TALE target prediction

To identify potential targets of $TAL14_{CIO151}$ and $TAL21_{CIO151}$, I partnered with Andrew Kassen, a former computational biologist in our lab to use the TALE script and Phytozome database to predict the target sequences and their highest probability target genes in cassava for $TAL14_{CIO151}$ and $TAL21_{CIO151}$. I selected the top three targets for each strain and characterized their potential functions (**Table 3-5**). Interestingly, one top predicted target of $TAL21_{CIO151}$ is a pectate lyase, involved in cell wall degradation and also implicated in pollen development. In legumes, the *Lotus japonicus nodulation pectate lyase gene* (*LjNPL*) is necessary for rhizobial root infection (Xie et al., 2012). Similarly, the rice (*Oryza sativa*) S gene *Os8N3*/*OsSWEET11*, a member of the SWEET sucrose efflux transporter family, has been implicated in pollen development, possibly transporting glucose necessary for starch accumulation in pollen grains (Slewinski, 2011). Though these genes are all from different organisms, the overlap in both general and specific functions suggests that $TALE21_{CIO151}$ may target a common susceptibilty gene. However, target prediction is most effective when coupled with a technology to confirm the targets. In order to clearly assess the validity of these predicted targets and identify new targets, further analysis by RT-PCR, RNA-seq and chromatin immunoprecipitation sequencing (ChIP-seq) could be used to identify both the upregulated transcripts and the TALE binding site. While not employed for these TAL effectors, a similar TALE $TAL14_{Xam668}$, differing by only one RVD has been transcriptionally characterized via RNA-seq (Cohn et al., 2015a). This single RVD difference appears to contribute to differential targeting, with $TAL14_{CIO151}$ and $TAL14_{Xam668}$ sharing some targets but also having unique targets. The top three predicted $TAL14_{CIO151}$ cassava gene are not identified in the 52 $TAL14_{Xam668}$ upregulated genes, however there are functionally overlapping gene targets. $TAL14_{CIO151}$ is

predicted to target an oxidoreductase (007404) and $TAL14_{Xam668}$ targets two oxidoreductases (022871, 012090).

Table 3-5 Predicted cassava targets of the two CIO151 TALEs.

TALE	Cassava Gene	Score	BLAST Hit
$\mathbf{14}_{CIO151}$	4.1_032796m.g	0.405	Glutaredoxin, e- carrier, cell redox, SA signaling
$\mathbf{14}_{CIO151}$	4.1_000483m.g	0.405	Kinesin-like microtubule-based movement, motor activity
$\mathbf{14}_{CIO151}$	4.1_007404m.g	0.419	Pyruvate dehydrogenase subunit, metabolism, oxidoreductase activity
$\mathbf{21}_{CIO151}$	4.1_006540m.g	0.459	Pectate lyase, pollen development
$\mathbf{21}_{CIO151}$	4.1_012779m.g	0.478	Protein tyrosine kinase
$\mathbf{21}_{CIO151}$	4.1_012073m.g	0.514	Predicted membrane protein

Cassava gene = "cassava4.1_#m.g" assembly version 4.1 and gene number as found in Phytozome.net

3.3.5 pXam46 not likely to be conjugative but it may be mobilizable

The contribution of plasmids to bacterial genome plasticity and HGT is largely due to their ability to conjugatively transfer between cells. Conjugative plasmids are often organized in a somewhat structured manner, where different segments of the plasmid are dedicated to mainly four different roles–stability, replication, propagation, and adaptation–as seen in the *E. coli* IncX1 plasmid pOLA52 (Norman et al., 2009). This is somewhat observed in pXam46, where there appears to be a TALE-rich pathogenic or 'adaptation' region, a stability region containing *stb*, *RelE* (TA), *Abi* and the *par* partition genes, and a region dedicated to conjugative transfer. The presence of two potential virulence inducing TALEs on one plasmid would provide a high selective advantage for any recipient bacteria in a conjugative transfer. In addition to TALEs, annotation reveals that this plasmid also contains several members of the transfer (*tra*) family of type IV secretion system (T4SS) genes, implicated in conjugative plasmid transfer, as well as a putative origin-of-transfer (*oriT*) sequence, though it lacks the *vir* cluster of genes typical of the T4SS (Alvarez-Martinez, 2009). Further, plasmids of a similar size are seen in other Xam strains. At 46 kb, pXam46 is on the smaller scale of known conjugative plasmids. Some conjugative mega plasmids are well over 100 kb, such as the recently discovered pMP7017 from *Bifidobacterium breve* (Bottacini et al., 2015). At the other end, the smallest predicted conjugative plasmid pCRY from *Yersinia pestis* is predicted to be only 21.8 kb long and the median size of all conjugative plasmids is approximately 181 kb (Smillie et al., 2010). Therefore pXam46 is within the range of possibility for being a conjugative plasmid but it is firmly on the smaller side. The reason conjugative plasmids tend to be larger is that they carry with them the entire suite of genes encoding the necessary components for conjugation. This include a relaxase for opening up the plasmid and is found in mobilizable and conjugative plasmids alike, a type IV coupling protein (T4CP) which pairs the DNA to the transport machinery, and a T4SS which serves as the transport machinery (Smillie et al., 2010).

The pXam46 annotated sequence suggests that it is at the very least mobilizable, if not conjugative. It appears to have a relaxase gene. There is a hypothetical *MobD* gene that aligns with a similar gene in *X. citri* pv. *citri*. However, the pXam46 gene is much shorter at 135 bp compared to 345 bp and therefore may or may not be functional. Similarly, pXam46 contains what looks like *TrwC*, putative *TraI* relaxase domain, however it appears to be disrupted by an *ISxac2* transposase insertion, as the full length *TrwC* that both annotations align with is approximately the length of

the two combined. Mutations in *TrwC* have shown that it is necessary for mobilization (Llosa et al., 1994). There is a full length *TrwB* (*TraB*) protein, which is likely the T4CP, a transmembrane domain containing, nucleotide binding protein involved in initiating the shuttling of DNA and other components through the T4SS (Alvarez-Martinez, 2009). The full length *TraA*, which may be an F-plasmid like pilin, is present though somewhat distal from *TrwB*/*TrwC*, separated by two transposases among other genes. Both *TrwB* (*TraB*) and *TrwC* (*TraI*) appear to be in an operon, as they are often found. The *TraD* gene also appears to be fully intact and it may assist with the T4CP to initiate secretion. The small number of *tra* genes on pXam46 as well as the apparent interruptions of key genes such as the putative relaxase *TrwC* make me skeptical that pXam46 is conjugative.

It is possible that recent transpositions disrupted the *tra* gene operon but in other Xam isolates with similar plasmids it remains intact. These genes, if all functional, would likely be enough to mobilize the plasmid but there is one glaring omission–a T4SS. When compared to other plasmids that are conjugative, pXam46 appears to lack the entire set of genes required for T4SS assembly. A good comparison is the similarly sized though slightly smaller TALE-carrying plasmid pXcb (37,106 bp) found in the *X. citri* pv. *aurantifolii*. On pXcB, 38 genes were annotated and a portion of the plasmid shares gene organization with pXam46, containing a *tnpR* resolvase, *stbB* stability gene, *traD*, *traA*, then (without transposase insertion) *trwB* and *trwC*. These genes in combination and when fully functional likely cut the plasmid at the *oriT* site and engage it for shuttling through the T4SS. In addition to these *tra* genes, 12 genes are found in tandem that are dedicated to the *virB* cluster of type IV transfer genes. None of these *vir* genes are found on pXam46, suggesting it does not have the genes to assemble the components of the T4SS on the plasmid itself. By comparison, a conjugative plasmid pA506 found in the plant epiphyte *Pseudomonas fluorescens* contains a suite of *vir* genes encoded the T4SS along with a suite of *pil* genes encoding type IV pili for mating (Stockwell et al., 2013). Another conjugative plasmid, pD2RT of *Pseudomonas migulae* carries *mobABCDE* genes for initial nicking and DNA preparation for transport in addition to the *traMNOPQRTUWXY*, *traKJIH*, and *trbCBAN* groups of transfer genes, dedicating approximately 60,000 bp of the 129,894 bp plasmid to genes encoding conjugation related proteins (Jutkina et al., 2013). The lack of plasmid-encoded T4SS does not mean that pXam46 can not be conjugated. Often with mobilizable plasmids, they take advantage of conjugation associated genes located on other plasmids, as in triparental mating, or encoded chromosomally. Annotation of the CIO151 genome reports 22 genes located on two 11-gene clusters related to the T4SS with one cluster predicted to be plasmid-borne and another 24 genes encoding T4 pili (Arrieta-Ortiz et al., 2013).

Testing mobilization or conjugation ability would be the best way to determine the translocation abilities of pXam46. The most straightforward way would be to introduce an antibiotic resistance gene into pXam46 by way of a suicide vector such as pLVC18. My attempts to insert an antibiotic resistance marker at several sites were unsuccessful. There are several other techniques to measure or track conjugation, including more modern techniques based on fluorescence that are sensitive and quantitative. One such example is the use of a *gfp* gene driven by the *lacI* promoter on the plasmid combined with the insertion of a *lacI* repressor gene in the donor chromosome where the recipient strains lack this repressor. While in the donor cell, GFP expression will be repressed. However, upon secretion into the recipient cell lacking the repressor, the cells will luminesce with GFP. Thus, expression of GFP after mating indicates successful conjugation into the recipient

strain, where the *lacI* promoter is not repressed (Sorensen et al., 2005). Such a method also allows for separation of recipients via flow cytometry and cell sorting. There are even sensitive, qPCR-based methods for tracking the conjugation transfer rate, comparing chromosomal versus plasmid abundance (Wan et al., 2011). However, the easiest method is still recombining in an antibiotic resistance marker. This should be possible, especially given the success of making the TALE knockouts. However, there may be some difficulty in obtaining the second crossover event as that selection process can result in plasmid cured strains.

3.3.6 Summary

I identified and sequenced a 46 kb plasmid harboring two TALE genes, $TAL14_{CIO151}$ and $TAL21_{CIO151}$, in the Colombian Xam isolate CIO-151. Surprisingly, the entirety of the measurable virulence contribution of pXam46 to CIO151 was attributable mostly to a single gene, $TAL14_{CIO151}$. While $TAL21_{CIO151}$ had a more limited impact on Xam growth in planta over the course of a 6-day growth assay, it is possible that there are conditions in the field where each TALE would have a more significant role. The forced inoculation of Xam cells into cassava leaves is may not be directly reflective of field conditions, but speaks to the role of effectors once bacteria have inhabited the plant apoplast. Other conditions not measured with these knockouts include factors like varying temperatures or humidity levels, bacterial traversal throughout the leaf, and the initial ability of the bacteria to enter the plant and initiate disease. Bacterial conjugation is an essential component of HGT, however its unclear whether pXam46 maintains the ability to conjugate. It is clear that the machinery is there for basic mobilization, however some essential components of the machinery appear to be disrupted by transposase insertion, specifically the putative relaxase proteins required for initial nicking of the plasmid before transfer. Unlike other conjugative plasmids, pXam46 lacks the genes required for assembly of a T4SS. Therefore it is likely at most mobilizable, when in the presence of chromosomal or other plasmid encoded T4SS proteins. The number of transposases found in pXam46 is noteworthy, suggesting a number of insertion events and making it difficult to assess the full functional ability of the plasmid. It can be said, however, that there are 2 expressed TALEs and they appear to be part of separate insertions. Overall, we've gained insight into the plasmid-localized context of these two TALEs, observing that the conserved TALE, $TAL14_{CIO151}$, that is found in other strains contributes most to virulence.

4 Global profile of Xam TALEs and the identification of conserved TALEs

4.1 Introduction

The presence of TALEs on potentially mobilizable plasmids provides a mechanism for Xam strains to share TALEs through horizontal gene transfer. With their demonstrated contribution to virulence, some TALEs could prove strong additions to a Xam effector repertoire. In recent years we have gained a much better understanding of what TALEs are structurally and how they function (Mak et al., 2013), but on a broader scale we still don't have a portrait of the TALE diversity amongst different Xam strains. A simple case-study in our lab consisted of the characterization of CIO151, isolated in Colombia in 1995, and Xam668, isolated in Indonesia in 1978. As expected, the strains were quite different. They each had a different number of TALEs on a different number of plasmids and they varied in RVD sequences. Despite all their differences, however, they did share one TALE of the same size, $TAL14_{Xam668}$/$TAL14_{CIO151}$, and it differed by only a single RVD. Should the same be expected of other strains–high variability with some apparent overlap? Are there conserved TAL effectors that are widespread over time in varying populations of Xam? Are there only a few different TALEs or many? Are they each functionally significant? In the rice bacterial blight inducing pathogen *Xanthomonas oryzae* pv. *oryzicola* (Xoc), predicted to have upwards of 28 TALEs based on PacBio sequencing (Wilkins et al., 2015), it has been observed that most TALEs have no measured relevance to virulence and that of the 19 identified TALE-upregulated genes, only one was a true susceptibility gene (Cernadas et al., 2014). In Xam, the more limited number of TALEs in the genome and their common localization on plasmids might suggest a leaner repertoire with fewer extraneous TALEs. Additionally, their high similarity to each other outside of the RVDs, repetitive nature, and proximity to mobile elements would likely make them subject to higher levels of natural recombination and transposition, which may contribute to the formation of new TALEs (Ferreira et al., 2015). Given this combination of genetic mobility and increased likelihood of genetic recombination, I hypothesized that TALEs in different Xam strains would be numerous and varied. However, over the course of evolution and selection, some TALEs or TALE targets should be selected for as effective for enhancing disease susceptibility in their plant host. Therefore, in addition to wide TALE diversity I hypothesized that there would be some set of conserved TALEs and some TALEs that have varied RVD sequences but share the same host target.

To address some of the questions of TALE quantitative and functional diversity, I utilized a subset of a temporally and geographically diverse collection of Xam isolates obtained by the lab for phylogenetic studies. We had previously shown that these Xam strains could be geographically organized into clades based on their genomes and compared their type III effector repertoires to identify a core set of conserved effectors (Bart et al., 2012). However, this study was limited to non-TAL effector sequences and approximation of TALE count by Southern blot of an EcoRI genomic DNA digest. To date, no thorough comparison of TALE repertoires in Xam has been conducted. This is mostly due to the difficulty of assembling TAL effectors using next generational sequencing and even conventional Sanger sequencing because of their highly repetitive nature. More recent studies have ameliorated the sequencing situation, using PacBio to identify five conserved TALEs in Xoc and using RNA-seq to identify an upregulated gene *Os03g07540* which is part of the same bHLH family as the Xcv AvrBs3 target UPA20 in pepper (Wilkins et al., 2015). These recent studies show that surveying geographically distinct rice

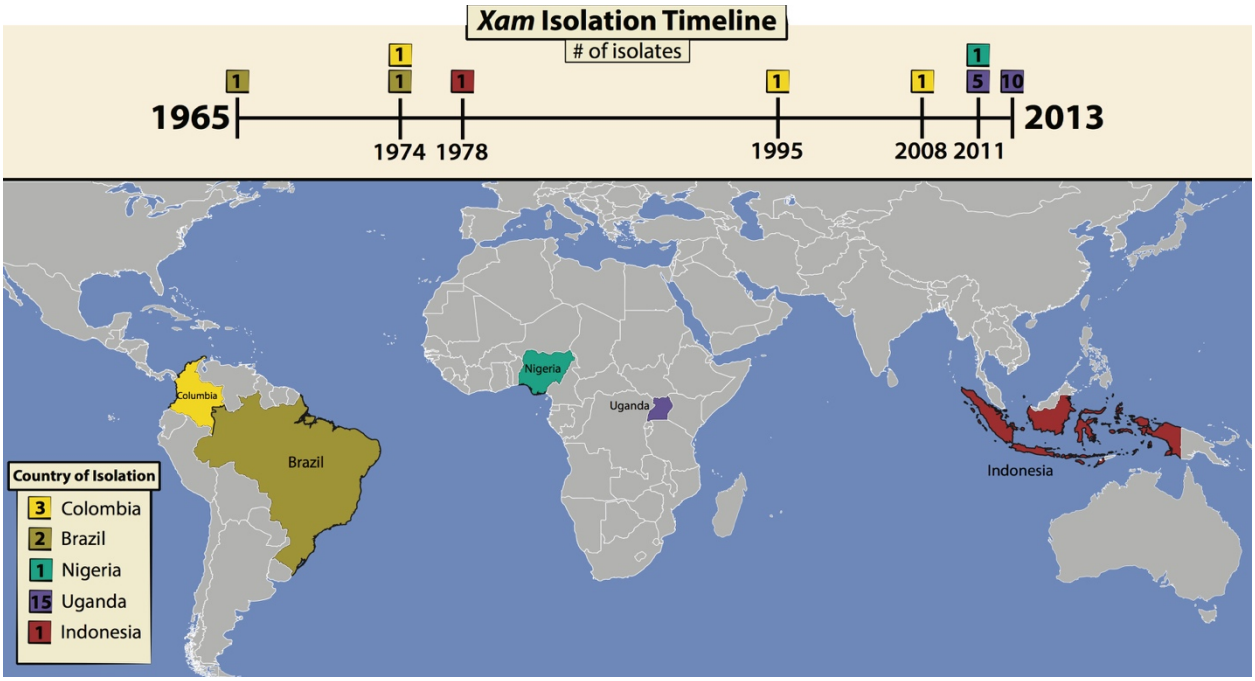

Figure 4-1 Map and timeline of Xam isolates in this study. Twenty-two strains isolated from 5 countries over the span of 48 years were selected to be profiled for their TAL effectors. Numbers inside boxes indicate number of isolates, colors indicate country of isolation.

pathogens identified conserved TALEs and led to the identification of S genes and suggest that a similar approach for Xam could yield similar results. With slightly fewer sequencing resources and before the advent of SMRT PacBio sequencing, I set out to do a global survey with a different, non-sequencing based method to compare the number, size, localization, and phenotypic functionality of TALEs in a set of global isolates of Xam. The results were somewhat surprising and informed our future directions in the lab.

4.1.1 Exploring TAL effector profiles of global isolates

My motivating questions for this chapter were simple. Do different Xam isolates have different TAL effectors? Do TAL effectors vary depending on geographical region? Are there any conserved TALEs that stood the test of time? If so, are these conserved TALEs major contributors to virulence? For my study, I selected 22 Xam strains that represented 5 countries over 48 years of evolution (**Figure 4-1**, **Table 4-1**), including a new set of African field isolates that I harvested from diseased leaves obtained by our lab from a recent Xam outbreak in Uganda in 2013 (Vincent Kyaligonza). Some of these strains had been observed to confer differential phenotypes such as water soaking and assigned differing disease scores (Bart et al., 2012). These differences could not be attributed to the presence/absence of any given effector, however. While TAL effectors represent only a few of the sometimes 30 or more effectors that may be secreted into a plant cell at one time, their unique mode of action as transcription factors offers a potential for novel and interesting routes to pathogenicity. Important targets of TALEs may be identified using modern sequencing techniques such as RNA-seq.

Table 4-1 Xam strains used in this study. Twenty two Xam strains from 5 countries, including 10 recently isolated from Uganda, sorted by year isolated.

Strain	Year	Origin
IBSBF278	1965	Brazil
CFBP1851	1974	Colombia
IBSBF321	1974	Brazil
Xam668	1978	Indonesia
CIO151	1995	Colombia
UA306	2008	Colombia
Ug12	2011	Uganda
Ug21	2011	Uganda
Ug23	2011	Uganda
Ug27	2011	Uganda
Ug28	2011	Uganda
NG1	2011	Nigeria
Ug1 - 10 '13	2013	Uganda

4.1.2 The significance of cassava to Africa

Cassava is particularly important to Africa, both agriculturally and industrially. Over 70 million people in Africa rely on cassava as a staple crop and calorie source (FAO, 2009). Thirteen of the top twenty cassava-producing countries are in Africa, including the largest producer, Nigeria, in total representing 55.7% of global cassava production between 2010 – 2013 (FAOSTAT, 2015). During that same time frame, intensification practices and increased interest in cassava has led the growth rate of cassava production in Africa to match that of maize and reach levels nearly triple those of rice in southeastern Asia (Howeler et al., 2013). Cassava bacterial blight (CBB) originated in South America where cassava was domesticated. It was only first reported in Africa in the early 1970s, leading to starvation for many in Zaire, but it is speculated to have been pervasive long before then, perhaps in less agriculturally relevant regions (Ogunjobi, 2006). African strains of Xam have been observed to be more virulent than Asian varieties but less diverse than South American strains (López and Bernal, 2012). We collected 17 African isolates of Xam spanning 7 countries and 40 years of evolution, 4 of which (NG1, ORST17, UG28, and UG45) had been predicted to contain 4, 5, 5, and 5 TALEs, respectively based on EcoRI digests and Southern blot analysis (Bart et al., 2012). In addition, I have harvested Xam from 10 diseased leaves taken from different plots of a recent CBB outbreak in Uganda in 2013. Here, I provide accurate counts of expressed TALEs by Western blot analysis including some of these African isolates and other South American and Asian strains. Identifying differences in TALE profiles may help to explain observed phenotypic differences in virulence of the African, Asian, and South American isolates. Characterization of these African TALEs would be incomplete without an understanding of the diversity of cassava in Africa. A group of scientists at the Joint Genome Institute have isolated germplasm from hundreds of African cassava cultivars with the intent of creating a full genome marker variation database of the different cultivars (Prochnik et al., 2012). In the context of conserved TALEs, cassava markers could be screened to identify cultivars containing an altered EBE which could lead to a less susceptible cultivar.

4.1.3 From TALE profiles to TALE targets–the search for susceptibility genes

One of the driving hypotheses of this TALE study is that there will be some conserved TALEs, or major TALEs, that will be found in multiple strains and contribute to virulence in a significant way by targeting a susceptibility (S) gene. Beyond their conserved nature, the pathogenic importance of any conserved TAL effectors can be tested by making individual TALE knockouts and looking for alterations in symptom development or in planta bacterial growth. Once major TALEs are characterized, their cognate S genes can be identified. The convenience of working with TAL effectors is their predictable and specific mode of action as transcription factors. The host gene targets of any TAL effector can be determined, allowing for a mechanistic assessment of its virulence role. Initially, the targets of these TALEs can be predicted based on their RVD sequences but ultimately identified and verified by RNA-seq, allowing for a better understanding of which plant components they may manipulate and leading to the discovery of potentially novel S genes. It is important to identify the target of conserved TALEs and to compare the Xam targets to known targets of other *Xanthomonas* species, looking for both conserved and novel gene family targets. Target identification will be essential to characterizing the full role of the TALE, but may not be enough to completely explain the contribution to virulence of any single TALE. Even in TALEs with confirmed S gene targets, it can be initially unclear what role the S gene plays in promoting or allowing disease.

The TALEs of rice pathogens, especially, have been well characterized and show the promise and utility of identifying conserved TALEs. It has been observed that some strains carry multiple TALEs that have different targets, which may be of the same gene family, and some TALEs seem to be redundant, targeting the same gene (AvrXa7, PthXo3; *Os11N3*). AvrXa7 and PthXo3 are both deemed 'major' TAL effectors because they exhibit major contributions to virulence and knockdown of their *Os11N3* target results in loss of susceptibility (Antony et al., 2010). Further, *Os11N3* is induced by 32 of 40 Xoo strains worldwide (Li et al., 2012a). These studies in rice suggest that in cassava there may also be a family of genes that are common targets of Xam TALEs and that multiple TALEs can evolve to target the members of that family that confer the greatest susceptibility. Identifying common, conserved TALEs can also lead to durable resistance strategies in cassava. In rice, the promoter of the TALE-activated R gene *Xa10* was modified to include EBEs of five TALEs in a stable transgenic line that demonstrated full resistance to Xoo strains with the corresponding TALEs (Zeng et al., 2015).

4.1.4 Summary

Next-generation sequencing has allowed for wide scale studies of bacterial populations for comparative genomics. Such wide TALE studies for Xam do not exist. Thus, investigating the nature of variation of the Xam TAL effector profiles found throughout the world and comparing them to each other as well as to our current collection of TALEs from Columbia and Indonesia offers a significant contribution to the understanding of TALE diversity in Xam and its interaction with cassava. Further, I have assessed individual contributions of TAL effectors by creating single knockouts of each TALE in a given strain. Given that TALEs have often been shown to contribute more to virulence than other *Xanthomonas* effectors, I predicted that making knockouts especially of conserved TALEs would identify several TALEs that contribute significantly to virulence, at least in our quantitative assays. However, I also expected that some of these identified TALEs may not have any observed functional relevance or may be redundant with other TALEs.

Therefore, individual deletions of some TALEs may not produce any measurable result and for redundant TALEs, multiple targeted deletions based on shared predicted EBEs may be necessary to achieve a true deletion phenotype, as observed in rice (Bai et al., 2000). This work has helped identify potential families of S genes in cassava that are common targets of TALEs. Once identified, any conserved TALEs can be tested for contribution to virulence by making deletion constructs of each TALE in select strains that contain some of these conserved or suggestive targets and assaying these deletion strains for virulence in order to identify major TALEs and major TALE repertoires.

4.2 Methods

4.2.1 Isolation of Xam from Ug-2013 leaves

We obtained a set of 10 leaves from different cassava fields in Uganda infected with cassava bacterial blight. From each leaf, plant tissue encompassing dry and water soaked areas was harvested, placed in 200 µL 10 mM MgCl2 along with 2 beads and pulverized with a bead beater. Tissue was diluted with 800 µL 10 mM $MgCl_2$ and plated in serial dilutions for single colonies. Colonies were selected for characteristic Xam qualities (primarily the production of exopolysaccharides) and then single colony purified. These colonies were then screened for spontaneous Rf resistance on NYGA (Rf) plates. These 10 Ugandan isolates were then inoculated into cassava at OD_{600} = 0.1 and assessed for water soaking symptoms.

4.2.2 Quantification of TALEs

While we had previously primarily used BamHI restriction digests and Southern blotting to identify the number of TALEs in any given Xam strain, I developed a Western blotting strategy to identify expressed TALEs, utilizing a polyclonal antibody that binds all TALEs (Feng, B. *unpublished*). Xam cultures were grown overnight in rich medium and then transferred to minimal medium to induce expression of T3E proteins into the medium as described (**2.6.3**). Cells were boiled in Laemmli loading buffer, collected by centrifugation, and directly loaded into the wells of a pre-cast NuPAGE Novex gel. After transfer to HyBond N+ membrane, TALEs were probed using the anti-TALE antibody, followed by the secondary anti-HRP rabbit antibody and luminescence using Super Signal West Femto Maximum Sensitivity Substrate (Thermo Fisher). Probed membranes were then exposed to x-ray film for seconds or minutes depending on the level of expression.

4.2.3 Preparation of large plasmids

To harvest DNA of large plasmids (30 kb+), 1 mL of cells at OD600 = 0.25 was resuspended in H2O. Then 175 µL of lysis buffer (**Table 3-1**) was added and vortexed. Cells were incubated at 65ºC for 5 minutes. I then added 400 µL phenol:chloroform (1:1) and vortexed 30 seconds. Cells were pelleted 5 minutes at room temperature and then 20 µL from the top of the upper phase was collected and loading for gel electrophoresis. Alternatively, to avoid the use of phenol:chloroform, buffers 1 – 3 of any miniprep kit (Qiagen, Bioneer) can be used followed by an isopropanol (IPA) precipitation. Add 150 µL of buffer 1, 250 µL of buffer 2, and 350 µL of buffer 3. Spin 10 minutes at 4ºC, decant into new tube, and add 700 µL IPA. Shake vigorously, spin 2 minutes at 13,000 RPM, and pour off the supernatant. Spin again and remove any supernatant. Allow the pellet to air dry 5 minutes and resuspend the pellet in 40 µL buffer 5.

4.2.4 TALE cloning and sequencing

A new PCR-based method to obtain TALE sequences was attempted in order to quickly isolate and sequence novel TALEs (**2.5.2**). High quality chromosomal DNA was obtained by CTAB extraction. Liquid cultures of African Xam isolates were resuspended in TE and lysed in an SDS/CTAB buffer, followed by an extraction with phenol:chloroform:isoamyl alcohol (25:24:1), precipitation in IPA and a 70% EtOH wash. Strains were cut with BamHI or EcoRI and probed with a ^{32}P labeled BamHI TALE fragment. The BamHI digests were used to estimate the number of TALEs and the EcoRI digests were used to determine whether those TALEs could be discretely separated by gel electrophoresis. The TALEs of strains with separable TALEs were then gel extracted, PCR amplified using Phusion HF polymerase and a touchdown PCR protocol. After amplification, TALEs were TOPO cloned and then sequenced. TOPO cloning ensures that even if multiple TALEs are amplified, each clone should only contain a single TALE.

4.2.5 TALE knockouts

In order to assess virulence and ultimately characterize any major TALEs, individual and combinatorial TALE knockouts were made by using the suicide vector pLVC18 with the 5' TALE region containing a modified start codon as described (**3.2.5**). One of the disadvantages to this approach, however, is that it is non-preferential. That is, in strains that contain multiple TALEs, individual TALEs could not be specifically targeted for knockouts because of the conserved nature of the TALEs outside of the RVD region. Instead, successful crossover events based on Tc antibiotic resistance and single colony purification were screened using the western blot technique described (**4.2.1**). Some multiple knockouts were obtained spontaneously using this method. In order to intentionally obtain multiple knockouts, plasmid cured strains were obtained by selecting for the second crossover event and then screening by gel electrophoresis.

4.2.6 Growth assays and symptom development

A growth assay strategy was used to assess the effect of individual TALE knockouts as described (**2.6.1**). Both WT and TALE knockout strains were inoculated at OD_{600} = 0.01 into the apoplast of cassava leaves in triplicate and measured for in planta growth over the course of 5 or 7 days as noted. A single 0.6 cm^2 tissue punch was harvested for each sample, placed in 200 µL 10 mM $MgCl_2$ with two glass beads and ground in a bead beater. Ground tissue was diluted with 800 µL 10 mM $MgCl_2$ and plated in serial dilutions for single colonies. Growth was also measured by a novel mid-vein inoculation method, where 5 µL of Xam culture at OD_{600} = 0.2 was placed in a 3 mm bored hole on the midvein and 'downstream' midvein sections were removed, ground, and plated at each time point (Cohn et al., 2015b).

4.3 Results/Discussion

When scanning global TALE profiles and looking for major TALEs, I expected to find a few varieties of TALE categories, including those that are widely conserved and target a major S gene but also those with variable RVDs that target that same S gene and those with no clear virulence contribution at all. It has been shown that two distantly related TALEs, AvrBs3 and AvrHahI, activate the same gene, Bs3, by binding the promoter region just one RVD away from each other (Schornack et al., 2008). Thus, I hypothesized that I would find some TALEs that have similar RVDs and therefore similar predicted EBEs. Also, some RVDs have been shown to be associated with multiple nucleotides (Mak et al., 2012), so I expect to find some TALEs with slightly different RVDs to be associated with the same target. On the other hand, TALEs have been shown to be

highly variable due to intragenic recombination (Yang and Gabriel, 1995) and to unequally contribute to virulence in Xam (Bart et al., 2012) as well as in rice, where targeted mutations of chromosomal and plasmid-borne copies of avrXa TALEs in Xoo led to different levels of virulence phenotypes (Bai et al., 2000). Therefore, I expected that some isolates would have TALEs that may not be main contributors to virulence. Most of these predictions were accurate.

4.3.1 A pattern emerges: The TALE profiles of different Xam isolates are similar

TALE profiles of 22 Xam strains representing 5 countries and 48 years of evolution were evaluated to compare the size (estimated number of RVDs) and quantity of TALEs present. Unlike previous studies in our lab based on restriction enzyme analysis combined with Southern blotting using a ^{32}P labeled TALE fragment probe, I used a radiation independent method of Western blotting with our universal TALE antibody to identify expressed TALEs. Xam strains were grown in minimal medium to induce type III effector expression and protein was prepared and blotted as described (**2.6.3**). I used Xam668 for comparison as our lab uses it as a reference strain. It is notably virulent and encodes 5 TALEs with 3 located close together at a molecular weight of 115 kDa and 2 larger TALEs around 150 kDa (**Figure 4-2**). A single RVD difference (34 AAs) is enough to distinguish each TALE in the Western blot.

The 10 new Ugandan isolates, Ug1 – Ug10 2013, that I harvested from freshly infected leaves sent from African field plots were inoculated into cassava and checked for symptoms four days later. They were also prepared and blotted for TALEs. When compared to CIO151 and Xam668, all of the Ug2013 isolates showed water soaking symptoms similar to Xam668 (**Figure 4-3A**) and contained identical TALE profiles (**Figure 4-3B**), also matching with Xam668. It was remarkable to observe that strains isolated 35 years apart and from different continents harbored five similarly-sized TALEs. A similar pattern emerged with the other isolates. For the most part, the TALE profiles fit into the Xam668 grid, where strains contained TALEs with either 13, 14, 15, 20, or 22 RVDs, nothing more, nothing less (**Figure 4-4**). Five of the African isolates (Ug 23, Ug28, NG1, UA306, IBSBF278) had all 5 TALEs of similar sizes. Other strains were missing a few TALEs (Ug12, Ug21, Ug27). The only strains that had characteristically different TALE profiles were the

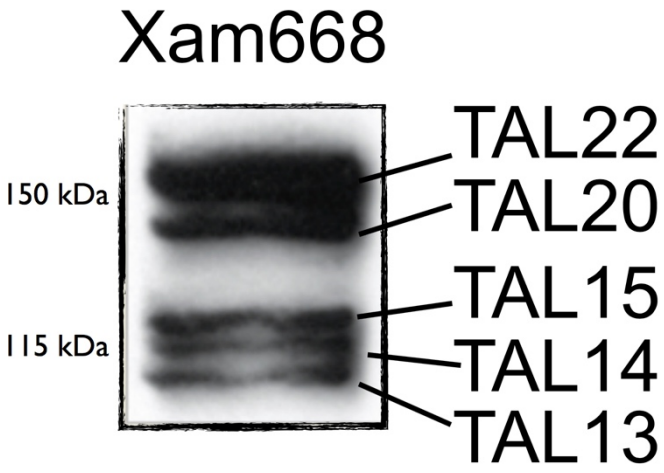

Figure 4-2 TALE Western blot profile of Xam668. The reference strain Xam668 was used to compare the profiles of other Xam isolates. TAL# represents the number of RVDs.

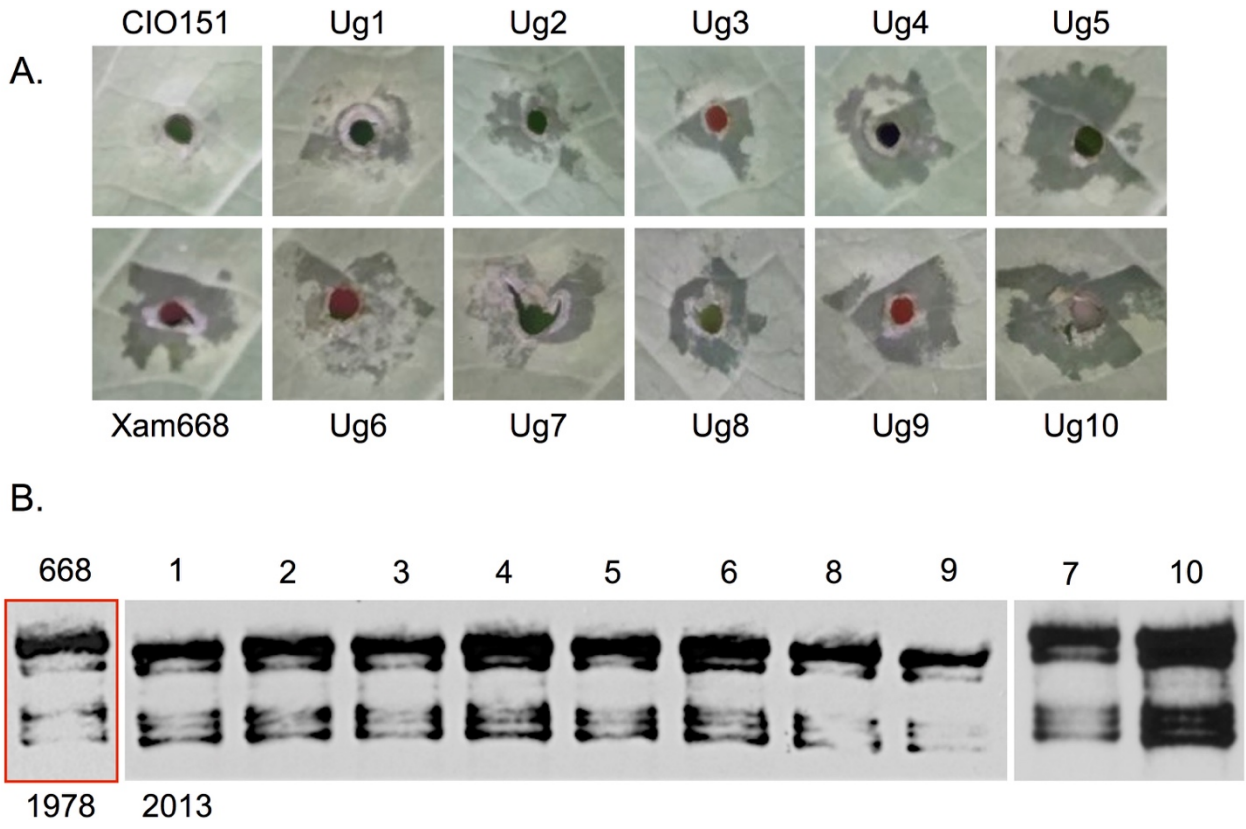

Figure 4-3 Symptom phenotypes and TALE profiles of Ug2013 isolates 1 – 10. A. Reference strains CIO151 and Xam668 along with the 2013 Ugandan Xam isolates were infiltrated into cassava leaves at OD600 = 0.1 and assessed for symptoms 6 days later. **B.** All isolates cause characteristic water soaking similar to Xam668 and in contrast to CIO151. According to Western blot analysis each isolate appears to have exactly the same TALE profile, each containing 5 TALEs of equivalent size to the set found in Xam668.

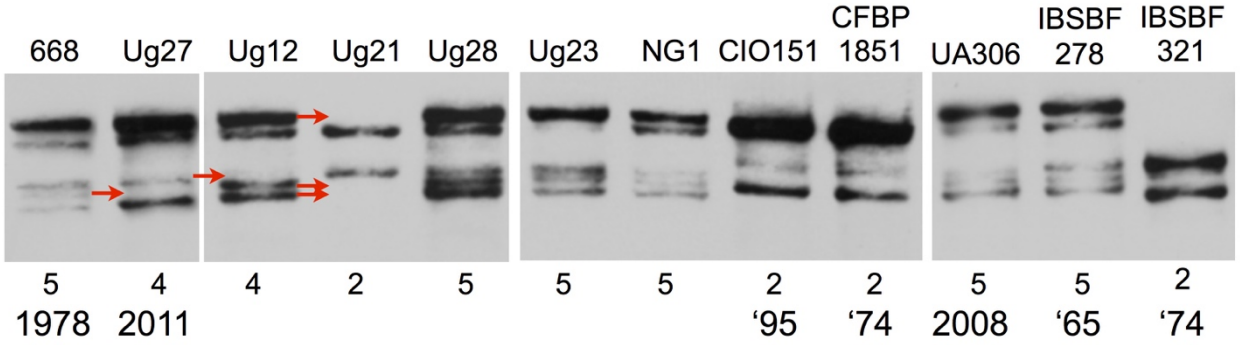

Figure 4-4 TALE profiles of the remaining Xam isolates. TALE expression was induced in minimal medium and imaged by western blot. Numbers below indicate number of TALEs and year isolated. Many isolates seem to share the profile of Xam668 or at least share several TALEs of the same size.

Colombian isolates CIO151 and CFBP1851, which are thought to share the same plasmid with the same two TALEs: $TALE14_{CIO151}$ and $TALE21_{CIO151}$. Notably, all of the 668-like profiles have a TAL20-sized TALE. Two of them are missing a TALE14 (Ug21, Ug27), and one of them is

missing a TALE15 and TALE22 (Ug21). These results suggest that the Xam668 TALE profile contains widely conserved TALEs, especially $TAL20_{Xam668}$ which may be a major TALE.

4.3.2 Plasmid profiles of Xam isolates with same TALEs can be distinct

It is a remarkable observation that geographically and temporally diverse Xam isolates contained similar TALE-profiles. As we know in CIO151 and Xam668, their entire set of TALEs are plasmid localized. In CIO151, both TALEs are located on a single plasmid, pXam46. In Xam668, they are spread out onto 3 different plasmids: TAL20 alone on the smallest, TAL13/14/22 on the intermediate sized plasmid, and TAL15 on the largest one. To determine whether strains with similar TALE profiles shared similar plasmid profiles, plasmid DNA from each strain was extracted and separated by gel electrophoresis. The plasmid profile diversity is much more notable than the TALE diversity (**Figure 4-5A**). There is no obvious plasmid that seems to be conserved and Xam668 has one of the more distinct profiles with a unique TAL20-containing smaller plasmid and a larger TAL-15 containing plasmid. Even in the Uganda 2013 isolates there is variation, though many contain several plasmids within a similar size range. There doesn't seem to be a correlation between TALE copy number and plasmid profile (**Figure 4-5B**). I attempted to type the plasmids by amplifying a marker from pXam46 (**Figure 4-5C**). While none of them had the same amplicon size as pXam46, several of them shared different amplicon signatures, suggesting they might at least be sharing one similar plasmid.

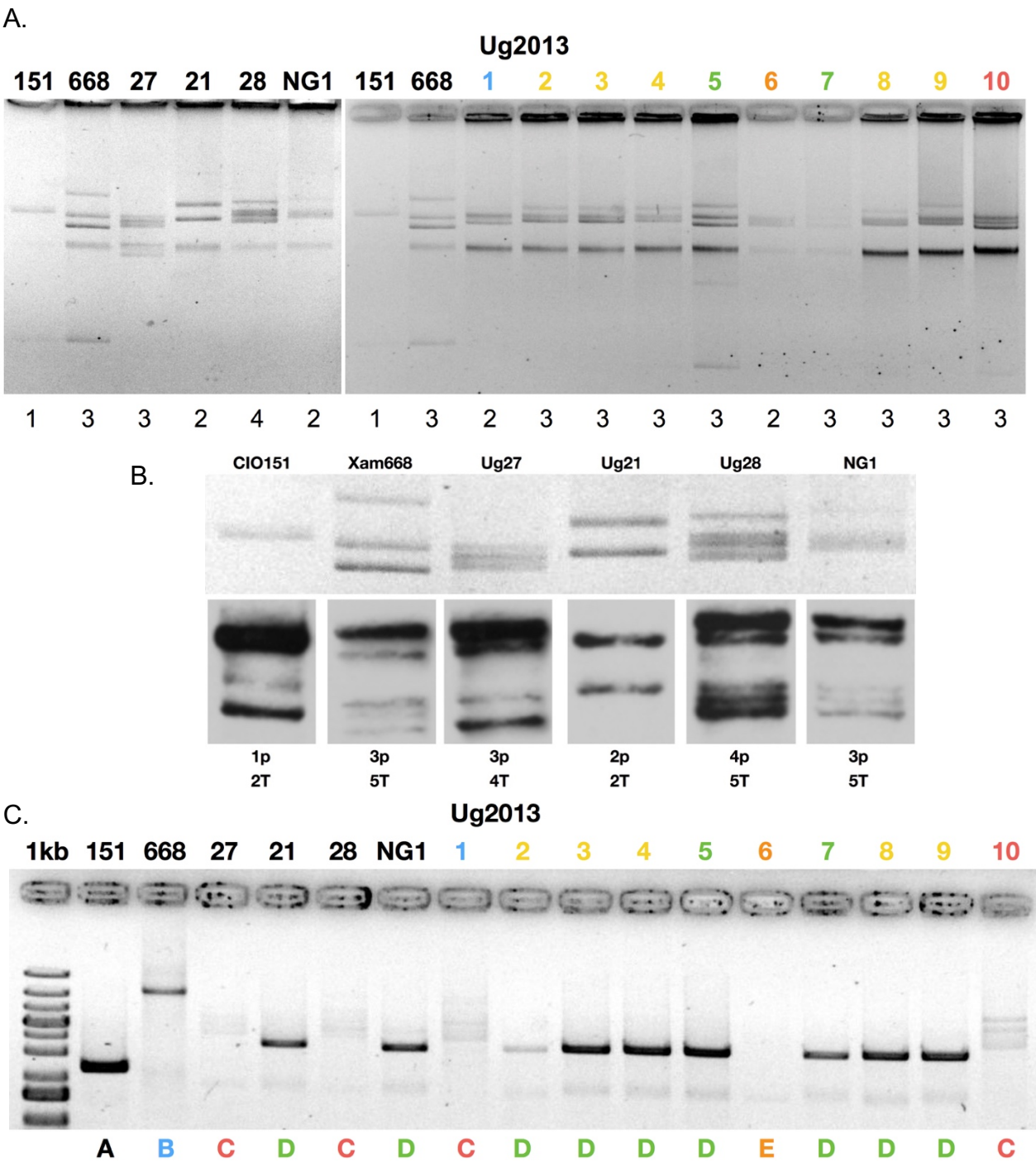

Figure 4-5 Characterization of plasmid profiles of some Xam isolates. A. Xam mega plasmid preps identify an array of plasmid profiles more diverse than the Western blotting profiles. **B.** Plasmid profiles compared to a subset of Xam protein profiles. **C.** Amplification of a pXam46 marker to type the plasmids.

4.3.3 Obtaining new TALE sequences

With a large set of Xam strains, subcloning or making libraries of each one and then identifying TALE-containing clones and sequencing in would be laborious. Instead, I used a gel extraction, PCR and TOPO cloning technique to isolate and sequence TALEs that I had previously identified with chromosomal DNA (**Figure 4-6**). For the African strains, I took advantage of the fact that most TALEs are plasmid-localized in Xam and did southern blot analysis on BamHI to quantify TALEs and EcoRI to look for separation (**Figure 4-7**). I identified EcoRI bands that hybridized to the ^{32}P labeled TALE fragment. I then gel extracted the regions identified with the probe, amplified them using modified TOPO-compatible TALE primers (**Table 4-2**) and a touchdown PCR protocol, and TOPO cloned them into the pENTR/D-TOPO vector and Sanger sequenced. While this method may not be reliable for TALEs with many RVDs, it seemed reliable for shorter TALEs. I was able to obtain 6 new TALE RVD sequences with 15 RVDs or fewer and 2 incomplete TALE RVD sequences each containing at least 18 RVDs (**Table 4-3**). Interestingly, many of these TALEs are similar to known Xam TALEs, some are identical and some are unique. Both $TAL13_{Ug39/Ug43}$ are only one RVD different from $TAL13_{Xam668}$. $TAL14_{Ug23}$ is identical to $TAL14_{Xam668}$ while $TAL14_{Ug12}$ is missing one RVD compared to $TAL15_{Xam668}$. $TAL15_{Ug21}$ is identical to $TAL15_{Xam668}$ but $TAL15_{Ug44}$ is quite divergent. Finally, the two incomplete TALEs, $TAL18+_{Ug21/Ug39}$ are identical to each other and very similar to $TAL20_{Xam668}$ but have at least 3 observable RVD differences and at least one unknown RVD. This PCR approach did prove effective in obtaining new sequences, but if possible should likely be avoided for longer TALEs as even high fidelity polymerases may be error-prone when amplifying the RVD region. Even so, it's clear that several of these TALE RVDs are conserved while others differ in mostly small but also some big ways. The question remaining from these differences is do they impact target activation? How many RVD differences can any one TALE withstand before activation is lost?

Table 4-2 Primers for TOPO compatible amplification of TALE genes. These primers were used to amplify TALE genes from African Xam isolates to directionally clone into pENTR/D-TOPO.

Primer	Sequence (5' to 3')
$TAL14_{CIO151}$ Fw ATG TOPO	caccATGGATCCCATTCGTCCGCGC
$TAL14_{CIO151}$ Rev Stop	TCACTGAGGAAATAGCTCCAT

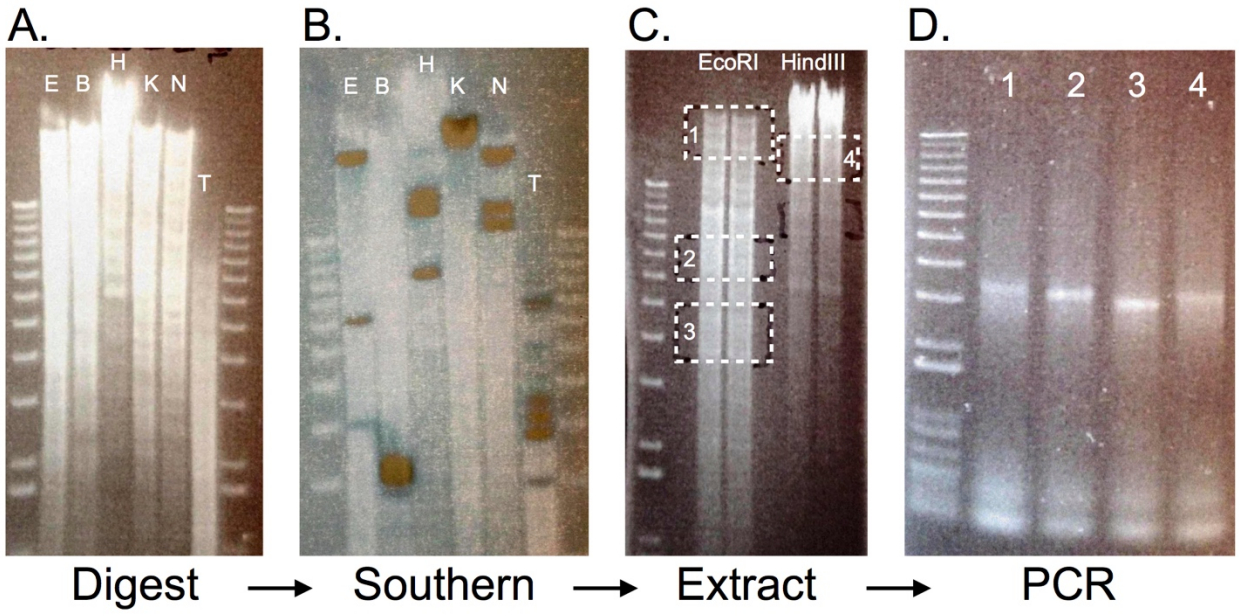

Figure 4-6 A pipeline for the amplification, cloning, and sequencing of TALEs. **A.** 2 µg of Xam669 chromosomal DNA was cut with EcoRI HF, BamHI, HindIII HF, KpnI HF, NotI HF, and the four bp cutter TSP509. **B.** Southern blot analysis of A was conducted using ^{32}P labeled TALE BamHI fragment to identify TALE-containing fragments. **C.** Cut chromosomal regions that contained separable TALEs were gel extracted (white boxes indicate TALE-containing regions to be extracted). **D.** The gel extracted DNA was then PCR amplified using Phusion polymerase and subsequently TOPO cloned for sequencing (bands at ~3.5 kb represent amplified TALEs).

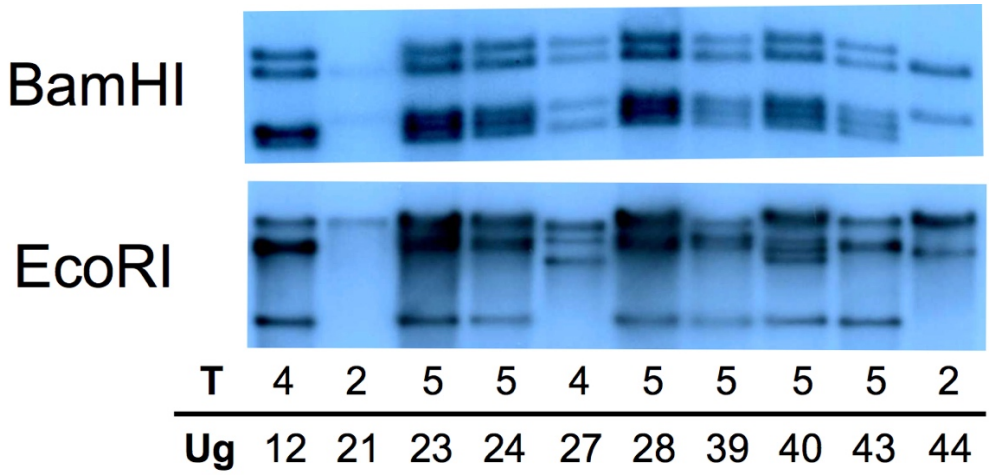

Figure 4-7 Southern blot of BamHI and EcoRI plasmid DNA of African Xam isolates. BamHI and EcoRI digests of plasmid DNA were analyzed for TALE-containing fragments. EcoRI fragments containing TALEs were gel extracted, PCR amplified with a 5' CACC tag, and TOPO cloned into an pENTR/D-TOPO for sequencing.

Table 4-3 Newly obtained African Xam TALE RVD sequences. Previously obtained TALEs of CIO151 and Xam668 included (in color) for comparison. Several additional TALE RVD sequences were obtained using a PCR-based method, amplifying TALEs from gel-extracted fragments and directionally cloning into TOPO-compatible entry vectors.

TALE	Repeat variable diresidue (RVD) sequence
13_{Xam668}	NI NS NG HD NG HD NI NG HD NN NI NI NG
13_{Ug39}	NI NS NN HD NG HD NI NG HD NN NI NI NG
13_{Ug43}	NI NS NN HD NG HD NI NG HD NN NI NI NG
14_{Xam668}	NI NG NI NN NG HD NS NS NN NG HD NN NI NG
14_{CIO151}	NI NG NI NN NI HD NS NS NN NG HD NN NI NG
14_{Xam669}	NI NG NI NN NI HD NS NS NN NG HD NN NI NG
14_{Ug12}	NI NG NI NN HD HD NS NS HD HD NS HD NG NG
14_{Ug23}	NI NG NI NN NG HD NS NS NN NG HD NN NI NG
15_{Xam668}	NI NG NI NN HD HD NS NS NS HD HD NS HD NG NG
15_{Ug21}	NI NG NI NN HD HD NS NS NS HD HD NS HD NG NG
15_{Ug44}	NI HD NN NG HD NS HD NN HD NG HD NI HD NG NG
20_{Xam668}	NI NG NI NN NI HD NS NS NN NG HD NS HD NN HD HD HD NI NG NG
$18+_{Ug21}$	NI NG NI NN NG HD NN HD NN… NS HD NN HD NG HD NI NG NG
$18+_{Ug39}$	NI NG NI NN NG HD NN HD NN… NS HD NN HD NG HD NI NG NG

4.3.4 TALE knockouts and the identification of 'major' TALEs

From the TALE profiles and sequence analysis, it is clear that some TALEs are present in several strains, despite geographic and temporal separation, and that the virulent strain Xam668 contains five TALEs, many of which are found in other strains. To determine whether the conservation of TALEs indicates an important role in pathogenicity, I created individual knockouts of each TALE by indiscriminately recombining the suicide vector pLVC18 with a TALE fragment containing a modified start codon into Xam668. To obtain a full set I scanned dozens of transformants, maintaining selection at the single crossover event. Once all knockouts were obtained, double knockouts were created by using the same setup in plasmid cured strains. This produced the double knockout 668ΔTal14ΔTal18p. These knockouts were then assessed for disease phenotypes and in planta growth.

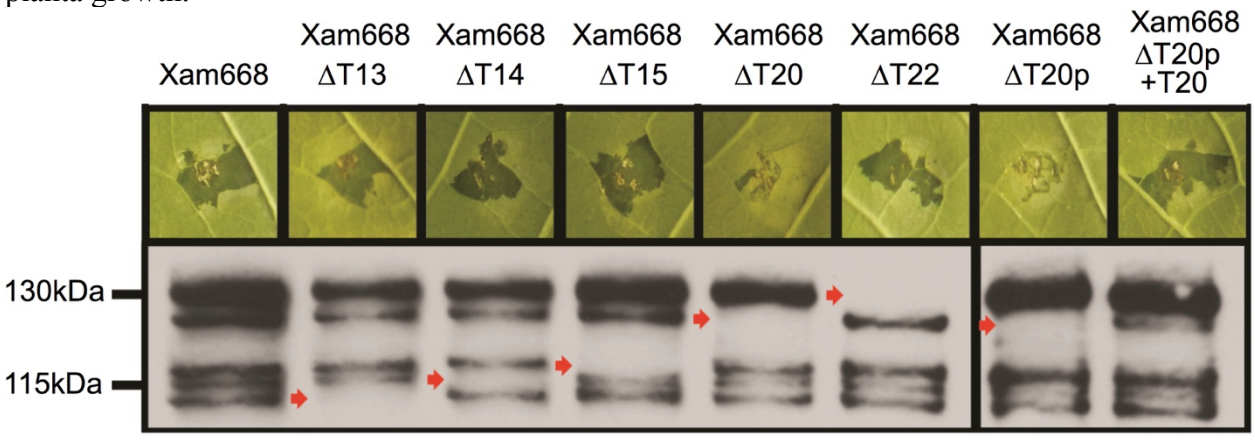

Figure 4-8 Cassava symptom phenotype and western blot of Xam668 TALE knockout strains. All five TALEs of Xam668 were individually knocked out using pLVC18 containing a 647 bp TALE fragment 5' of the RVD region. One knockout, 688ΔT20 along with its corresponding plasmid loss variant 668ΔT20p displayed decreased levels of water-soaking. This

water-soaking phenotype could be restored upon complementation with TAL20Xam668 driven by its native promoter (688ΔT20+T20). Figure reprinted with permission (Cohn et al., 2014).

To assess the phenotypic variation of the knockout strains, all Xam668 knockouts (668ΔT13, ΔT14, ΔT15, ΔT20, ΔT22, ΔT20p, ΔT14/ΔT20p) were infiltrated into cassava leaves at OD_{600} = 0.1 and monitored for disease symptoms, including a characteristic water soaking phenotype (**Figure 4-8**). Of all the knockouts, one varied notably in its phenotype–668ΔTAL20 and other ΔTAL20 knockout strains. The absence of TAL20 correlated with an absence of typical water soaking. Water soaking could be restored by complementing ΔTAL20 with its native promoter. This suggested that TAL20 of Xam668, the only TAL conserved in all of my profiled isolates, activates a characteristic symptom of Xam, water soaking. Further, Cohn *et al.* showed that water soaking ability could be bestowed upon CIO151 when transformed with $TAL20_{Xam668}$ (Cohn et al., 2014).

The knockout strains were also assessed for in planta growth via both apoplast and midvein growth assays. For apoplast growth assays, strains were infiltrated at OD_{600} = 0.01 and measured after 6 days for bacterial growth at the site of inoculation. These growth assays consistently indicated that the TALEs contribute differentially to in planta bacterial growth. Some TALEs affected Xam growth by a similar amount and while others appeared to have no measurable effect. I noticed small but consistent decreases in growth in knockout strains 668ΔTAL14 and 668ΔTAL20 as well as an additive effect in the double knockout 668ΔTAL14ΔTAL18p (**Figure 4-9**). There were no measurable differences for the other conserved TALE knockouts, 668ΔTAL13, TALΔ15, or TALΔ22 (**Figure 4-10**). These results were corroborated in a recently developed midvein growth assay (Cohn et al., 2015b), designed to better represent the role of Xam as a vascular pathogen. The assay showed clear and significant effects for Xam, a vascular pathogen, when growing inside the midvein (Cohn et al., 2014).

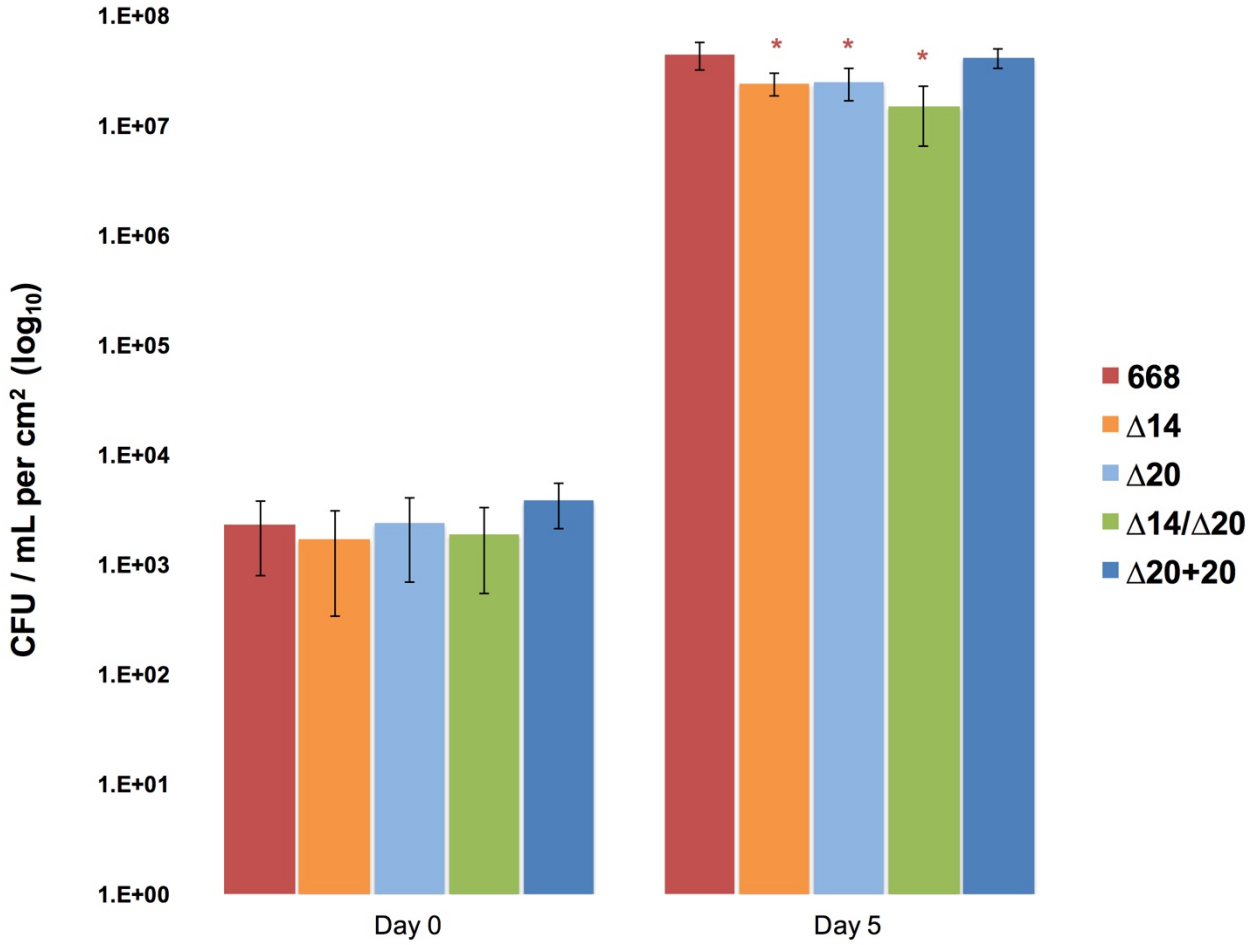

Figure 4-9 Apoplast growth assay of Xam668, 668ΔT14, and 668ΔT20 knockout strains. Xam wild type and knockout strains were infiltrated into cassava leaves at OD_{600} = 0.01 and, here, measured for Xam growth 5 days post inoculation. Both ΔT14 and ΔT20 showed decreased growth in planta. The ΔT14/ΔT20p double knockout displayed an additive effect on growth. * indicates significant difference ($p < 0.02$) to Xam668. Error bars represent ±1SD.

Together these results suggest that of the five 668 TALEs which seem to be present in other Xam isolates, only two of them could qualify as major TALEs. One of them, $TAL20_{Xam668}$ appears to be conserved in all but the Colombian isolates CIO151 and CFBP1851. This highly conserved TALE triggers an enhanced water soaking phenotype and also affects in planta Xam growth by approximately 0.5 $\log_{10}$ CFU/cm^2. The other major TALE knockout, 668ΔTAL14 displays no discernable symptom phenotype but also shows a similar decrease in in planta growth at 668ΔTAL20. Symptoms and virulence are not always correlated. A well characterized TALE *avrb6* from the cotton blight pathogen *X. campestris* pv. *malvacearum* was found to induce extreme water soaking in cotton lines but did not contribute to an increase in in planta bacterial growth. Rather, it was found that more bacteria were released from the symptomatic region onto the leafs surface than in a strain lacking *avrb6* (Yang et al., 1994). This is one example of a measure of bacterial virulence that might escape our traditional growth assay protocols. It's also

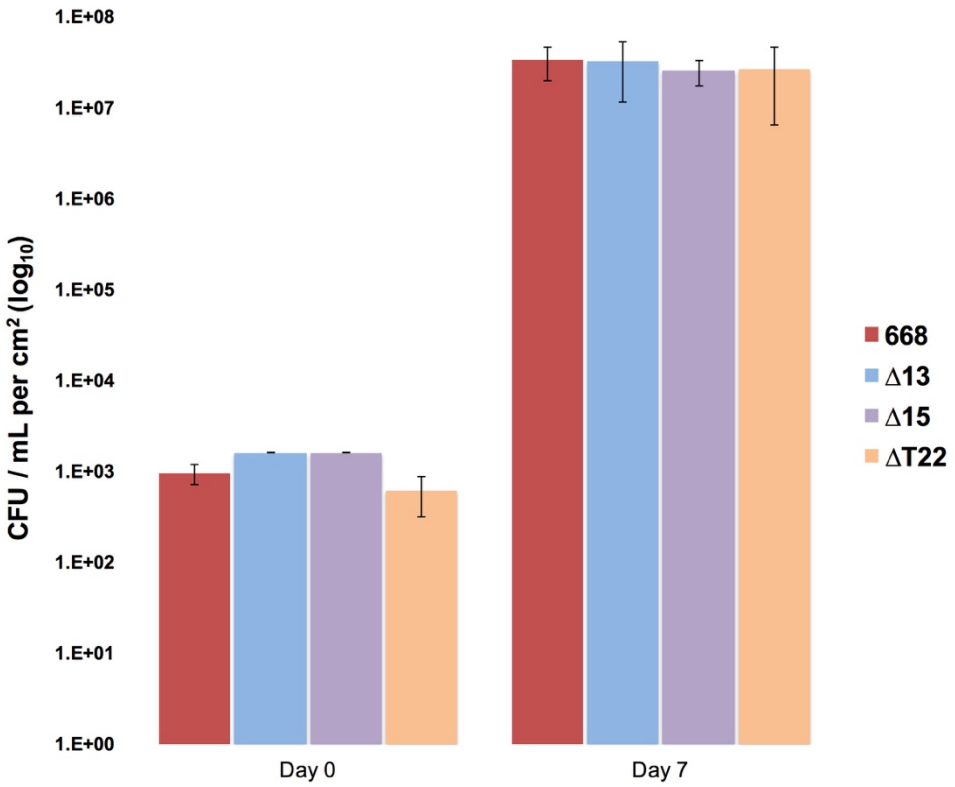

Figure 4-10 Apoplast growth assay of Xam668 knockout strains 668ΔT13, 668ΔT15, and 668ΔT22. This set of Xam668 TALEs did not show any significant different to wild type Xam668 7 days after inoculation. Leaves were inoculated with Xam at OD_{600} = 0.01. Error bars represent ±1SD. For Day 0 ΔT13 and ΔT15, accurate error bars could not be created due to low replicates.

important to note that, while some TALEs are conserved and contribute to growth and symptom development, strains without these major TALEs still cause disease and eventually catch up to the symptoms (**Figure 4-11**), resulting in chlorosis and necrosis in the infected area and eventually leaf senescence. This is another reason why approaches to assaying Xam phenotypes may be needed that are more physiologically relevant to field conditions.

Through RNA-sequencing, Cohn *et al*. demonstrated that $TAL20_{Xam668}$ targets and activates the cassava *MeSWEET10a* gene, functionally characterized to be a sucrose transporter (Cohn et al., 2014). The TALE-targeting and activation of *MeSWEET10a* by $TALE20_{Xam668}$ was the first identification of a SWEET class susceptibility gene in cassava. This class of S genes has been found to be a common target of *Xanthomonas* ssp. TALEs. In rice, several TALEs target genes belonging to the SWEET family. One example is the *Os8N3*/*OsSWEET11* (*Oryza sativa*, Chromosome 8, NODULIN3) S gene in rice and its recessive allelic variant, *xa13*, which confers race-specific resistance to Xoo PXO99 due to promoter mutations in the TALE EBE necessary for *Xa13* expression (Chu et al., 2006; Yuan et al., 2011). Resistance by *xa13* is defeated in the presence of other TALEs, including AvrXa7 and PthXo3 which both target another SWEET gene, *Os11N3*/*OsSWEET14* (Antony et al., 2010). Another example of an N3-targeted gene is *UPA16* in pepper, targeted by AvrBs3 of Xcv (Kay et al., 2009). The activation of N3/SWEET genes by

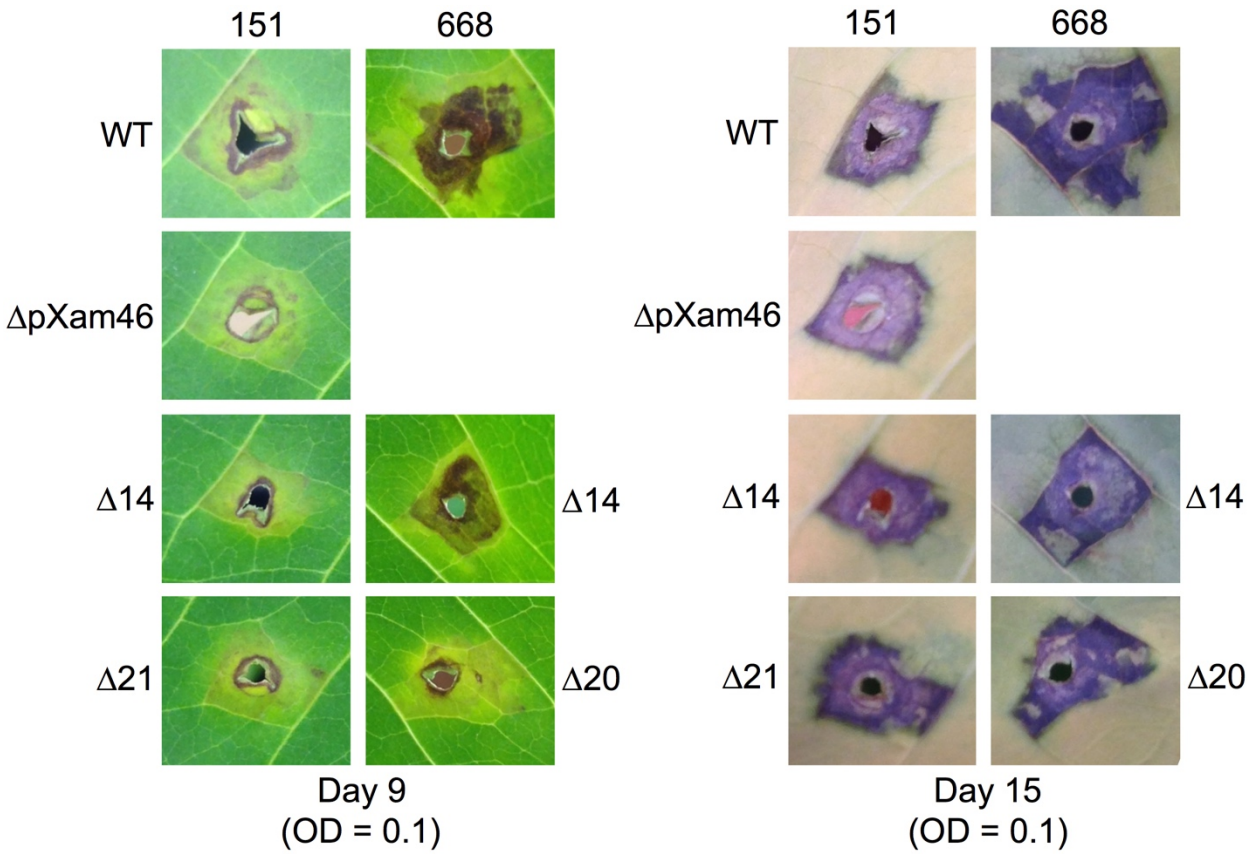

Figure 4-11 Late symptoms of Xam wild type and knockout strains. Symptoms of Xam strains shown to differ in in planta bacterial growth and symptom development display similar phenotypes when inoculated at higher levels and assessed at a later stage, here 9 dpi and 15 dpi.

many different TALEs in different pathogens is suggestive of an important role for this class of targets in bacterial virulence. Beyond its involvement in starch accumulation in pollen grains (Chu et al., 2006) it is speculated that Os8N3/OsSWEET11 may be involved in removing inhibitory amounts of copper from the xylem (Yuan et al., 2010) and/or, more likely, that these *N3*/SWEET factors may be transporting sucrose into the xylem for bacterial access (Chen et al., 2010). TALE-dependent activation of SWEET genes does not always indicate their role as S genes, however. In two strains of *Xanthomonas citri* ssp. citri (Xcc), the similar TALEs pthA4 and pthAW both activate two targets in citrus, *CsLOB1* and *CsSWEET1* (Hu et al., 2014). However, only *CsLOB1* was shown to be an S gene, promoting both pustule formation and increased in planta bacterial growth. This result is somewhat surprising given the characterization of other TALE-targeted SWEET genes as S genes and while it could simply reflect coincidental off-target effects, it may also have a functional role in field conditions that are not represented in the lab.

The cassava *MeSWEET10a* gene is upregulated by each strain with a TAL20-sized TALE (although the Ug2013 strains were not tested) according to semi quantitative reverse-transcriptase PCR (qRT-PCR). Interestingly, Ug21 shows a weaker activation of *MeSWEET10a* in semi qRT-PCR. This may be explained by the divergent RVD sequence where at least three internal RVDs differed from $TAL20_{Xam668}$ while most of outer RVDs were identical. Some of the RVDs that varied, however, represent flexible RVDs that have been found to associate with multiple

nucleotides. Therefore, it is entirely likely that $TAL20_{Ug21}$ binds *MeSWEET10a* but that it may be a weaker interaction than $TAL20_{Xam668}$. This hypothesis could be validated with a luciferase promoter assay with the $TAL20_{Xam668}$ EBE, comparing $TAL20_{Xam668}$ activation to $TAL20_{Ug21}$ activation. This example demonstrates the flexibility of the TALE cipher and shows that TALEs with slightly different RVD sequences can target the same S gene in Xam. It also suggests that, even with weaker promoter activity, the different $TAL20_{Ug21}$ is still worth maintaining.

4.3.5 Summary

TALEs are unfortunately left out of our next-generational sequencing strategies to study effector repertoires and pathogen determinants of host specificity. Using a focused approach, I characterized the TALE protein profiles of 22 Xam strains with distinct geographical and temporal origins. To my surprise, many of the profiles appeared identical to our reference strain Xam668, isolated in 1978, containing five TALEs of the same size or varying by only one or two TALEs. This included a more recent set of Xam strains isolated from active outbreaks in Uganda fields in 2013, 35 years apart from Xam668. To determine whether a conserved TALE protein profile indicated conserved RVD sequences, TALEs from a subset of strains were amplified via PCR of gel excised digested Xam plasmid DNA. These TALE amplicons were cloned and sequenced, revealing some identical RVD sequences, some similar sequences, and only a few divergent sequences. One TALE size, TAL20, was conserved in all but three isolates, including two Colombian isolates, CFBP1851 and CIO151 from 1974 and 1995, respectively, and one Brazilian isolate, IBSBF321 from 1974. Knockout strains identified $TAL20_{Xam668}$ as both a contributor to leaf water-soaking symptoms and in planta apoplastic and midvein bacterial growth. Further analysis identified the $TAL20_{Xam668}$ S gene at a SWEET family sucrose transporter, a common target among TALEs from different *Xanthomonas* ssp. Another conserved TALE, $TAL14_{Xam668}$, was present in almost all isolates except for two Ugandan strains, Ug21 and Ug27 from 2011, and once again the Brazilian isolate IBSBF321. Like $TAL20_{Xam668}$, $TAL14_{Xam668}$ also conferred a decreased growth phenotype in the knockout line, which is consistent with the results of the single RVD variant $TAL14_{CIO151}$ previously obtained (**3.3.3**), but it did not contribute to any observable differences in symptom development, demonstrating that in planta bacterial growth and the development of characteristic symptoms can be separable.

Much like the two TALEs of pXam46 in CIO151, all five of the Xam668 TALEs are accounted for on three plasmids, including $TAL14_{Xam668}$. The smallest and largest plasmids contain a single TALE, $TAL20_{Xam668}$ and $TAL15_{Xam668}$, respectively. The intermediate plasmid contains the remaining 3 TALEs: $Tal13_{Xam668}$, $TAL14_{Xam668}$, and $TAL22_{Xam668}$. The presence of tranposases and other IS elements flanking the TALEs of pXam46 suggest that introduction of new TALEs to plasmids may be conducted via mobile elements, potentially explaining why we see plasmids with a differing number of TALEs. Combined with a hypothesized mechanism for establishing novel TALE RVDs by polymerase slippage (Ferreira et al., 2015), there is a mechanism for the emergence of new TALEs and their incorporation onto plasmids. The plasmid profiles of different Xam isolates do not appear to reflect the TALE profiles, other than a correlation that strains with more plasmids tend to have more TAL effectors. Overall, these results are similar to many of the patterns found in other *Xanthomonas* spp. and their hosts, where a small set of TALEs is conserved among diverse field isolates but only a few of the conserved TALEs seem to have any observable contributions to virulence (Cernadas et al., 2014; Wilkins et al., 2015). The conserved nature of the other TALEs, however, suggests that they may likely have some yet unknown function that

our assays currently don't account for. While the work described here has identified several conserved TALEs of Xam which have been identified to upregulated the same cassava gene targets, our understanding of the RVD diversity and conservation is still limited. A study on a wider scale complete with SMRT PacBio sequencing to obtain complete TALE sequences could inform the evolutionary history of TALEs and the emergence of TAL14 and TAL20. It is interesting that the three divergent isolates in terms of TALE profile are all from South America. This alone does not explain their difference, as one Brazilian isolate IBSBF278 has a Xam668-like profile, but it does suggest that among some populations there are unique TALE profiles. Also, while these conserve TALEs contribute to phenotypes in the lab, it would be interesting to better replicate field conditions to determine relevant effects of any given TALE.

5 The hunt for an executor R gene in *Nicotiana benthamiana*

5.1.1 An 'avirulent' role for TALEs

Much of my research has focused on the destructive nature of TAL effectors as weapons in the Xam effector arsenal to successfully infect cassava. However, the very first TAL effector to be cloned, the plasmid-borne *AvrBS3*, was initially identified based on the plasmid-dependent HR in a cured strain (Bonas et al., 1989). This is consistent with the zig-zag model of plant-pathogen interactions (Jones and Dangl, 2006)–just as plants have developed a robust immune system for recognizing other type III effectors and activating a hypersensitive response (Greenberg and Yao, 2004), some plants have evolved TALE EBEs in promoter sequences upstream of a new class of R gene. These TALE-dependent R genes were coined executor R genes because they execute a resistance response (Bogdanove et al., 2010). While executor R genes may be relatively new in terms our detailed functional knowledge–they were among the earliest of the then termed 'avirulence' genes discovered–it would be more than 15 years before the first TALE/executor gene pair was characterized. This mechanism represents an evolutionary advantage in the host where a modified promoter sequence EBE upstream of R genes allow the specific binding of a TAL effector and activation of an otherwise transcriptionally off R gene. Their discovery has prompted a new strategy for obtaining durable resistance in host crops by creating TALE traps, modifying the promoters of executor genes to include EBEs of multiple TAL effectors and has been demonstrated effectively in rice (Boch et al., 2014; Schornack et al., 2012; Zeng et al., 2015).

5.1.2 Executor gene discoveries

To date, only five executor genes have been identified and cloned. The first was characterized in rice, where the TALE avrXa27 from the pathogen *Xanthomonas oryzae* pv. *oryzae* (Xoo) activated the executor gene *Xa27* (Gu et al., 2005), which had previously been identified by fine genetic mapping of rice introgression lines (Gu et al., 2004). Two years later, the pepper (*Capsicum annum*) gene *Bs3* was identified through an AvrBs3 dependent HR screen of bacterial artificial chromosome (BAC) clones of *Xanthomonas campestris* pv. vesicatoria (Xcv) in *Nicotiana benthamiana* (2007). Both of these discoveries used classical approaches to identify the TALE targeted executor genes. However, new technologies and the decreasing costs of sequencing allowed for cutting edge, cloning-independent approaches to find TALE targets. For example, just five years after the characterization of the *Bs3* gene, Strauß *et al.* used RNA sequencing (RNA-seq) to identify the third known executor, *Bs4C*, the R protein target of AvrBs4 from Xcv (Strauß et al., 2012). The next executor to be cloned had been characterized years before in a similar fashion as *Xa27* by using isogenic lines and fine mapping to identify six *Xa10* R gene candidates (Gu et al., 2008), but it wasn't until six years later that the AvrXa10 target was cloned and characterized (Tian et al., 2014). Finally, sub-libraries of BAC clones were used to identify the AvrXa23 target *Xa23*, allowing for its functional characterization as a TALE-dependent resistance gene (Wang et al., 2014; 2015).

5.1.3 Novel functions of the E genes, a new class of R genes

An intriguing aspect of the currently known executor genes is their diversity and lack of similarity to R genes involved in a typical innate response. A comparison of all known executor genes showed no sequence homology to any known R genes and the authors suggested that the five executors can be broken down into two groups, those genes that likely have a developmental function within the plant and are co-opted by the TALE and those of no known function that may

be solely built for defense (Zhang et al., 2015). This grouping separates BS3, which has been categorized as a flavin monooxygenase (FMO) protein and the remaining E genes, which are shorter in length and of no known function. XA23 appears to be a homolog of XA10, sharing 50% AA identity and 64% similarity, and both containing similar transmembrane domains (Wang et al., 2015).

The search for executor genes provides an opportunity for discovering novel components of resistance pathways. Innate immunity in plants is well studied. The main actors in PAMP and effector triggered immunity are usually leucine rich receptor kinases (LRR-RKs) and nucleotide-binding LRRs (NB-LRRs), respectively (Boller and He, 2009; Tameling and Takken, 2008). While much is known about these recognition events that start the resistance pathway and the structure and function of NB-LRR R proteins involved in those interactions (Bai et al., 2012; Takken and Goverse, 2012), there is still much to discover in the downstream events leading to cell death. There are many ways for a plant to trigger programmed cell death as observed in an HR (van Doorn and Woltering, 2005). Given their dependence on TALE transcriptional activation rather than direct or indirect recognition by an NB-LRR protein to trigger a signaling cascade, executor genes could represent downstream components of a resistance pathway or be part of a yet uncharacterized pathway to resistance.

5.1.4 Summary

To date, only TALEs from Xoo and Xcv have been identified with their cognate executor genes in rice and pepper, respectively. Executor genes may be a widespread, robust defense mechanism against TALE containing bacteria. We have observed that Xam causes a resistance response in *N. benthamiana* that is consistent with type II non-host resistance, triggering a hypersensitive response (Gill et al., 2015; Senthil-Kumar and Mysore, 2013). Using the TALE knockouts I developed previously (**4.2.5**), I set out to determine whether this resistance response to Xam was TALE-dependent. I identified a specific $TAL14_{CIO151/Xam668}$ transcriptionally-dependent HR and characterized its dependence to known NB-LRR resistance pathway proteins and used RNA-seq to identify putative targets. While resistance strategies against TALEs can be deployed using known R genes and modified promoters, identifying more executor genes will shed light on the various mechanisms of HR and the genes involved in addition to adding to the database of R genes available for heterologous expression and engineering durable resistance against quickly evolving pathogens.

5.2 Methods

5.2.1 Plant inoculations and AAD mutant

The TALEs $TAL14_{CIO151}$, $TAL14_{Xam668}$, $TAL20_{Xam668}$ and $TAL22_{Xam668}$ were directionally cloned into the Gateway compatible pEarleyGate100 vector (Earley et al., 2006), driven by the CaMV 35S constitutive promoter and the nopaline synthase terminator (NOS-T) and transformed into competent *A. tumefaciens*. Overnight *Agrobacterium* GV3101 liquid cultures were resuspended in IM buffer (**Table 5-1**) at OD_{600} = 0.1. A 20 µL pipette tip was used to punch holes into *N. benthamiana* leaves and the Xam cultures were hand infiltrated into these holes with a 1 mL needless syringe. Plants were placed in a growth chamber at 24ºC on a 16 hour light/8 hour dark cycle. After three days, the plant leaves were assessed for visual symptoms of HR.

Table 5-1. Inoculation medium (IM) buffer for *Agrobacterium*-mediated transient expression in *N. benthamiana*.

Inoculation medium (IM) Buffer (50 mL)	
H2O	49 mL
10 mM MES pH 5.6	0.5 mL (1 M stock)
10 mM MgCl2	0.5 mL (1 M stock)
150 µM acetosyringone	75 µL (0.1 M stock)

5.2.2 Virus induced gene silencing (VIGS) in *N. benthamiana*

Virus induced gene silencing (VIGS) was used to knockdown *N. benthamiana* genes involved in innate immunity responses (**Table 5-2**) using a *Tobacco rattle virus* bipartite system (Liu et al., 2002). A series of Avr/R gene combinations were used as a positive control (**Table 5-3**). They were tested against one $TAL14_{CIO151}$, known to trigger HR, an activation domain (AD) mutant $TAL14AD_{CIO151}$ unable to transcriptionally activate its target gene, and $TAL20_{Xam668}$ as a negative control (**Table 5-4**). The TMV VIGS protocol was carried out as previously described (Leister, 2005). *Agrobacterium* GV2260 strains containing vectors representing both strains of the RNA1/RNA2 bipartite setup were grown overnight on selective media. Cultures were resuspended to OD_{600} = 2.0 in 10 mL of IM buffer and incubated at room temperature for 3 hours. Then, the pYL192 (TRV-1) construct was mixed with each pYL279 (TRV-2)-target containing construct in a 1:1 ratio for a final OD_{600} = 1.0 each. Celite was added to 1 mg per 1 mL. Sterile cotton stick applicators were dipped into each mixed culture and vigorously rubbed onto the leaves of at least three different plants. Plants targeted for different knockdowns were carefully separated from each other during and after the application process. Plants were covered and incubated in the growth chamber overnight at 24ºC. The cover was removed and the plants were grown for 2 – 3 weeks in the growth chamber, until the phytoene desaturase (PDS) control leaves turned white.

Table 5-2 TMV VIGS targets in *N. benthamiana*.

TMV VIGS targets	Vector	Strain
-	pYL192 (TRV-1, RNA1, Km)	GV2260
GUS (Negative control)	pYL279 (TRV-2, RNA2, Km)	GV2260
NbPDS (Positive control)	pYL279 (TRV-2, RNA2, Km)	GV2260
NbSGT1	pYL279 (TRV-2, RNA2, Km)	GV2260
NbEDS1	pYL279 (TRV-2, RNA2, Km)	GV2260

Table 5-3 Avr and R gene clones for VIGS controls in *N. benthamiana.* Here, positive controls for the HR. AvrBS2 and BS2 when coexpressed trigger an SGT1-dependent HR.

R gene controls	Vector	Strain
AvrBS2	pMA-1	GV3101
BS2	pMA-1	GV3101
RPS2 (CC-NBS)	pE1776	C58C1

Table 5-4 TALE constructs to be tested in *N. benthamiana* VIGS assay. Here, TAL14AD and $TAL20_{Xam668}$ were used as negative controls to demonstrate transcriptional dependence and RVD-dependent specificity, respectively.

Tested effectors	Vector	Strain
$TAL14_{CIO151}$	pEG100	GV3101
$TAL14AD_{CIO151}$	pEG100	GV3101
$TAL20_{Xam668}$	pEG100	GV3101

5.2.3 RNA-sequencing of *N. benthamiana* cDNA

RNA-seq was conducted as described (**2.5.1**) (Cohn et al., 2014). Leaves of *N. benthamiana* were inoculated with either *Agrobacterium* alone or with *Agrobacterium* delivering a 35S-driven $TAL14_{CIO151}$. Leaves of two plants were either inoculated with *A. tumefaciens* GV3101 or with *A. tumefaciens* GV3101 (pEG100 $TAL14_{CIO151}$) at OD_{600} = 0.5. Per sample, 6 leaf punches from inoculated plant tissue from three separate leaves of two plants were harvested using a 0.6 cm^2 cork borer at two time points (24 and 48 hours post inoculation) in duplicate. Separately, extra samples were taken at each time point and at 8 hpi for western blot analysis to confirm TALE expression. For RNA-seq, RNA was extracted using the Spectrum plant total RNA kit (Sigma-Aldrich) with on-column DNase digestion. RNA quality was confirmed before library preparation by Bioanalyzer (RNA pico chip for eukaryotic RNA) at the Functional Genomics Laboratory at UC Berkeley. Eight RNA-seq libraries (two strains at two time points in duplicate) were then made and barcoded (indexed adapter sequences set B) using the Illumina TruSeq RNA sample preparation kit (version 2.0). The libraries were checked by Bioanalyzer (DNA chip, high sensitivity) and submitted to the Genomics Sequencing Laboratory where samples were quantified by qRT-PCR and pooled for 100 bp paired-end sequencing on the Illumina HiSeq 2000. Illumina reads were imported into CLC Genomics 8 and trimmed using default settings. All reads were combined to assemble a *de novo N. benthamiana* transcriptome using default settings. Reads from each experimental condition were mapped to the new transcriptome using the RNA-seq tool in CLC Genomics 8 with default parameters, resulting in expression levels of reads per kilobase of transcript per million mapped reads (RPKM). Each value represents reads mapped to one of the thousands of contigs assembled in the transcriptome. Candidate genes were annotated using BLASTX. For some contigs the assemblies appear incomplete based on BLAST analysis. To identify putative EBEs in the genome using TALE-NT 2.0 (Doyle et al., 2012), assemblies were compared to a previously compiled *N. benthamiana* transcriptome (Nakasugi et al., 2013) and genome (Naim et al., 2012) assemblies.

5.3 Results and discussion

5.3.1 Screen for TAL-dependent HR

Two Xam strains were chosen for the screen based on their characterization in our previous studies and their differing sets of TAL effectors: CIO151 isolated from Colombia (TAL14, TAL21) and Xam668 from Indonesia (TAL13, TAL14, TAL15, TAL20, TAL22). Both Xam strains trigger a non-host HR in *N. benthamiana*. To assess the contributions of each strain's TAL effectors to the HR, strains with individual knockouts of each TALE as well as some dual and triple knockouts obtained via plasmid loss were screened for HR and compared to WT (**Figure 5-1**). Notably, knockouts of TAL14 in each strain, despite their single RVD difference, lead to an apparent weaker response. This was observed in single $TAL14_{CIO151}$ and $TAL14_{Xam668}$ knockouts as well as the

double knockout plasmid cured strain in CIO151 and the triple knockout plasmid cured strain of Xam668. To determine the roles of individual TALEs independent of the numerous other type III effectors secreted by Xam, each TALE was expressed in a 35S-driven binary vector via *Agrobacterium*-mediated transient expression in *N. benthamiana*. *Agrobacterium* alone does not result in an HR. Expression of each TALE validated what was observed in the knockouts. Induction of both $TAL14_{CIO151}$ and $TAL14_{Xam668}$ individually was sufficient to trigger an HR while induction of other Xam TALEs, including $TAL20_{Xam668}$ and $TAL22_{Xam668}$ did not trigger any observable resistance response (**Figure 5-2**). Additionally, a $TAL14_{Xam668}$ construct driven by the $TAL20_{Xam668}$ promoter in pVSP was introduced into the pepper and tomato infecting pathogen and *N. benthamiana* virulent *Xanthomonas gardneri* (Xg) 153. When infiltrated at low levels (OD_{600} = 0.00001), the Xg(TAL20::TAL14) strain displayed a delayed phenotype that appeared consistent with an oxidative burst relative to the compatible Xg WT interaction (**Figure 5-3**). The TAL20 promoter is likely weaker than the 35S promoter, which may explain the delayed phenotype observed in the Xg(TAL14) strain. The phenotype is likely dosage dependent and, in this case, only activated when adequate growth levels of Xg are reached.

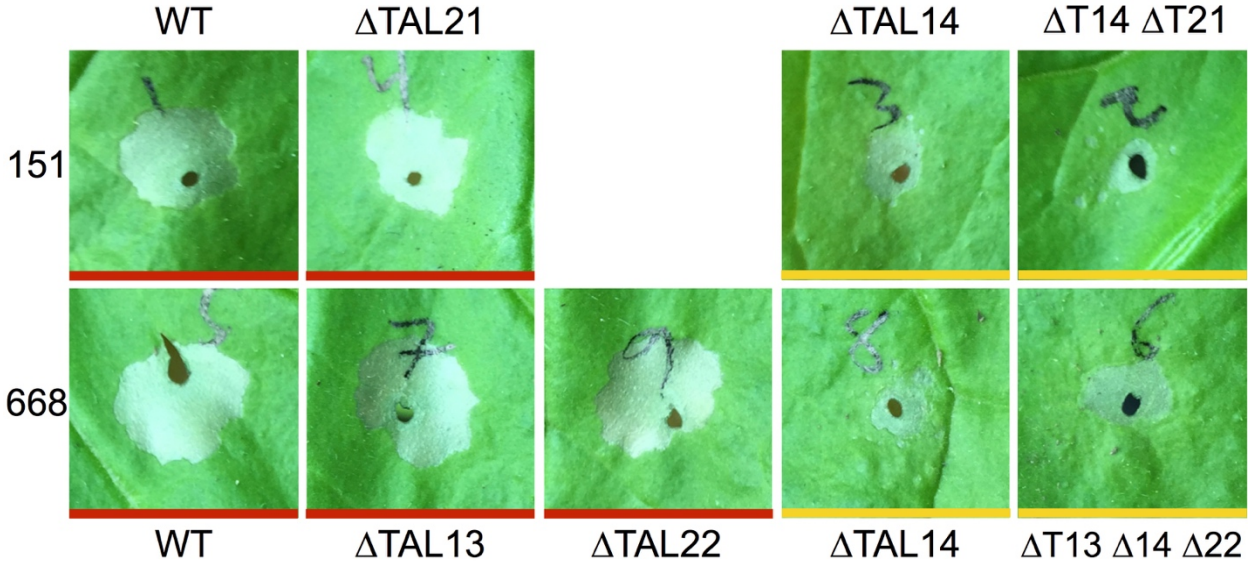

Figure 5-1 Xam triggers a non-host HR in *N. benthamiana*. Xam strains CIO151 and 668 elicit HR in *N. benthamiana* (red). Single and combined TAL effector knockouts of TAL14 show a lessened response to Xam (yellow).

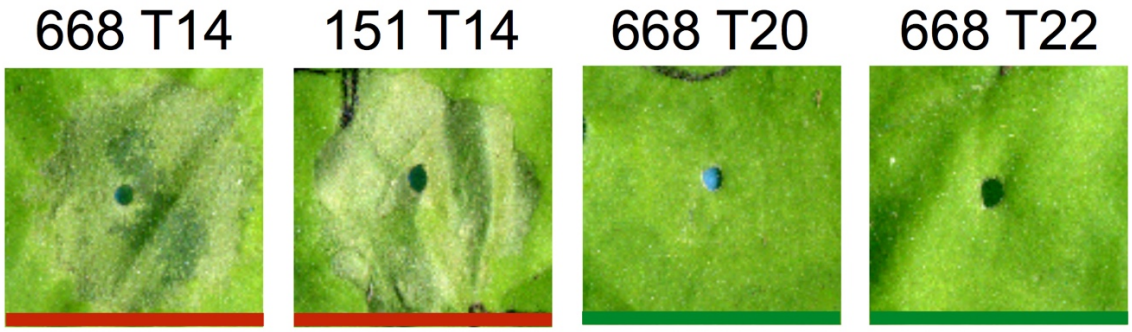

Figure 5-2 *Agrobacterium*-mediated transient expression of TALEs shows TAL14-specific response. Transient overexpression of various Xam668 TALEs demonstrates that the HR response observed is specific to TAL14 of both Xam668 and CIO151. Photos: Alex Schultink.

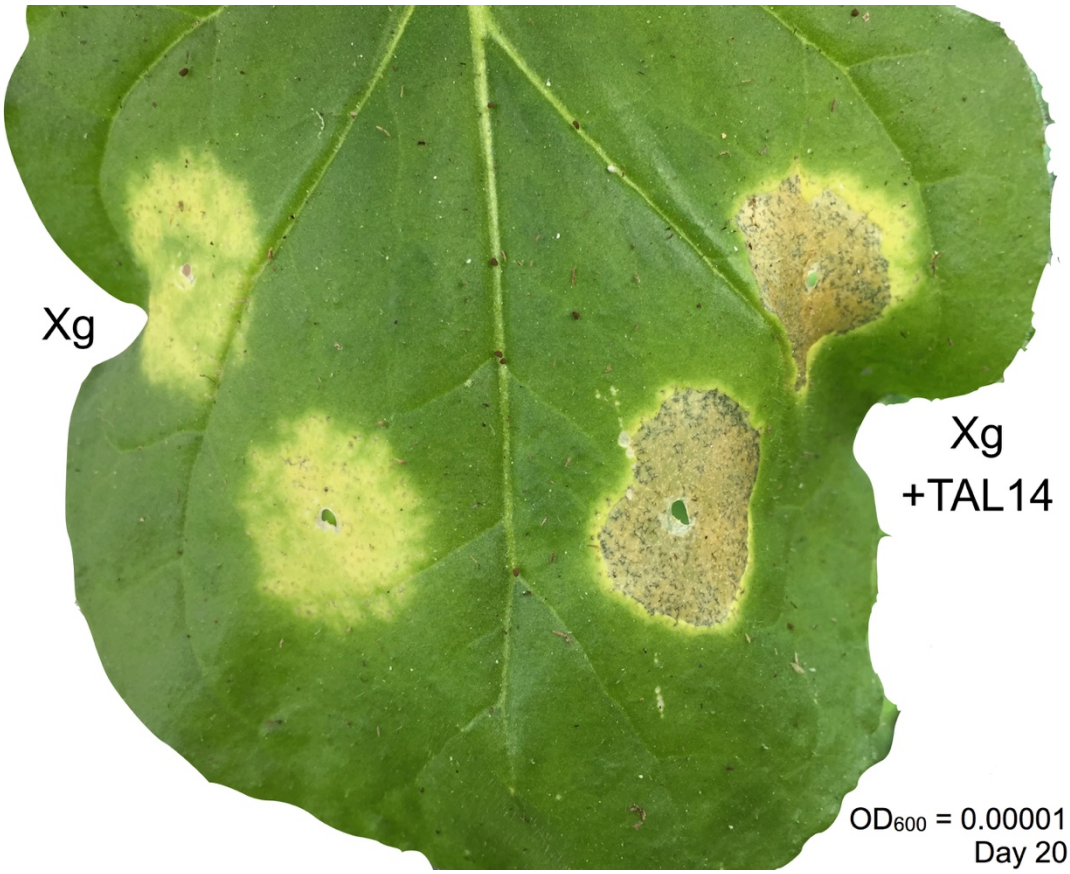

Figure 5-3 Wild type Xg and Xg(TAL20::TAL14 pVSP) in *N. benthamiana.* After inoculation at very low levels (OD_{600} = 0.00001), *Xanthomonas gardneri* (Xg) expressing a $TAL14_{Xam668}$ construct driven by a $TAL20_{Xam668}$ promoter induces a response that appears similar to HR, likely involving release of reactive oxygen species (ROS).

5.3.2 Determination of transcriptional dependence

The observation that $TAL14_{ClO151}$ and $TAL14_{Xam668}$ are each sufficient to induce an HR in *N. benthamiana* while other Xam TAL effectors do not suggests that this interaction is specific. However, a specific interaction could involve recognition by other means, such as direct recognition by an NB-LRR of the RVD region of TAL14. This has been observed in tomato (*Solanum lycopersicum*), where the AvrBs4 NLS is not necessary to induce a resistance response by Bs4 recognition (Ballvora et al., 2001), which was later characterized as a Toll/interleukin-1-receptor (TIR)-NB-LRR (Schornack et al., 2004). To be a true executor gene, the TAL14 target must be transcriptionally dependent on TAL14, rather than involved in a post-translational interaction. Therefore, to determine whether the observed HR can be attributed to transcriptional induction by TAL14, a mutant, $TAL14_{AD}$, was created with 3 AA mutations in the acidic activation domain (AAD) to abrogate transcription activation ability (**Figure 5-4A**). Upon inoculation into *N. benthamiana,* TAL14AD shows a much reduced response and often no response is observed (**Figure 5-4B**). This demonstrates that the TAL14 triggered HR is transcriptionally dependent and is not likely due to recognition by an NB-LRR protein.

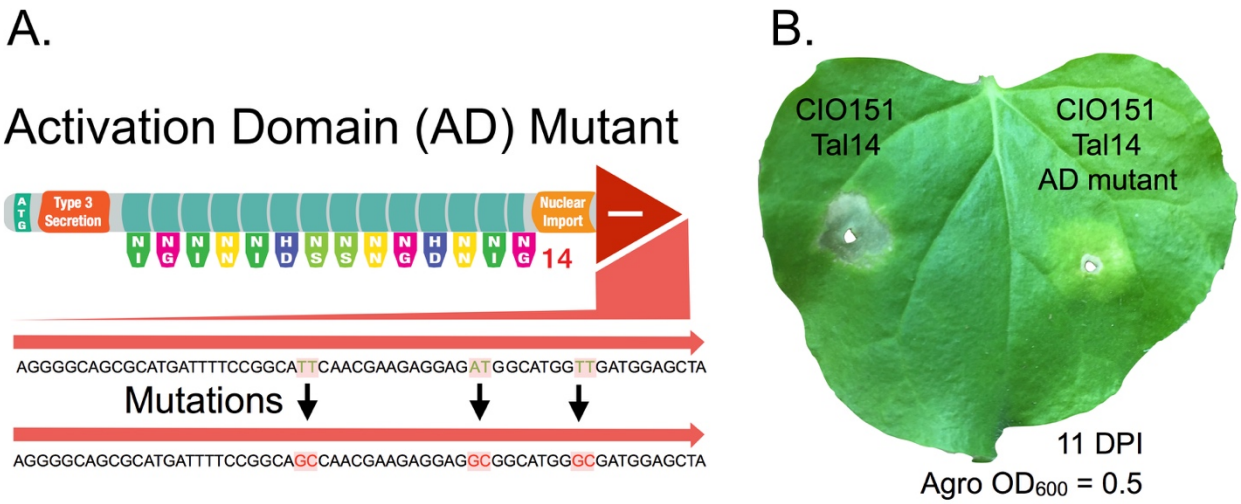

Figure 5-4 Mutations in the activation domain of TAL14$_{ClO151}$ abrogate *N. benthamiana* HR. **A.** The acidic activation domain of TAL14$_{ClO151}$, required for transcription factor recruitment, was mutated in 3 locations. **B.** Transcriptional dependence of the TAL14 HR was verified upon *Agrobacterium*-mediated expression of the AD mutant TALE (image background removed).

5.3.3 VIGS to determine role of EDS1 and SGT1

To further characterize the TAL14 response in *N. benthamiana* and its independence from traditional NLR recognition, I assessed the dependence of TAL14 on two proteins known to be gatekeepers to immunity, the SGT1 (suppressor of the G2 allele of the S-phase kinase-associated protein 1) and EDS1 (enhanced disease susceptibility 1). For example, the non-executor style recognition of AvrBs4 by Bs4 was shown to be both SGT1 and EDS1 dependent (Schornack et al., 2004). The plant SGT1 protein is known to interact with two other defense genes, HSP90 and RAR1, and may be involved in the formation of NLR complexes (Innes, 2004). Functional mutants of SGT1 lead to a loss of resistance by R proteins such as RPP1 of Arabidopsis (Austin et al., 2002). Similarly, EDS1 has been shown to be essential to TIR-NB-LRR responses in Arabidopsis (Falk et al., 1999) and is a common target of bacterial effectors (Bhattacharjee et al., 2011). In *N. benthamiana*, virus induced gene silencing (VIGS) was used to knockdown both SGT1 and EDS1 in addition to PDS and GFP as positive and negative controls, respectively. Leaves of each knockdown line were then infiltrated with *Agrobacterium* delivering TAL14$_{ClO151}$, TAL14$_{AD}$ mutant, TAL20$_{Xam668}$, AvrBs2, AvrBs2 and Bs2, or a shortened RPS2$_{CC\text{-}NBS}$ missing the LRR domain that we've observed to cause an SGT1/EDS1 independent HR. In both SGT1 and EDS1 knockdown lines, the TAL14 dependent response was delayed but observable within 9 days post inoculation (**Figure 5-5**). The eventual response, however, appeared much weaker relative to the strong HR of the R protein controls and that observed in untreated plants.

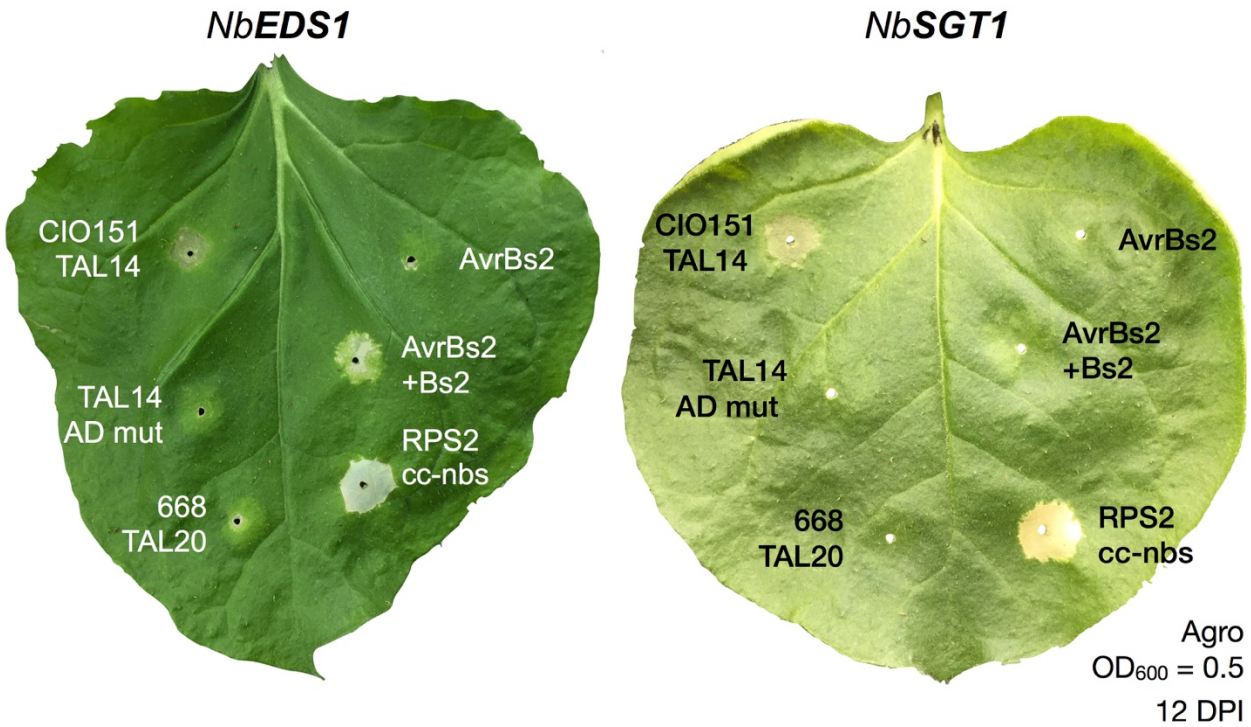

Figure 5-5 *Agrobacterium*-mediated transient expression of TAL14 in EDS1 and SGT1 knockdowns. Using VIGS, EDS1 and SGT1 of N. benthamiana were knocked down to determine their role in the TAL14-triggered HR. In both knockdowns, the TAL14 HR appeared later and weaker than in WT plants, but did appear after 9 days, here shown at 12 dpi. Neither the AD mutant nor Xam668$_{TAL20}$ showed any phenotype after the same amount of time. Image backgrounds removed.

5.3.4 RNA-seq and filtering of candidate genes

A simple RNA-seq pipeline was designed to identify the putative executor target of TAL14$_{CIO151}$ in *N. benthamiana* (**Figure 5-6**), requiring no need for complex genomic mapping as previously shown (Cohn et al., 2014; Strauß et al., 2012). Before sequencing, I confirmed TAL14$_{CIO151}$ expression at 8, 24, and 48 hpi to confirm protein levels (**Figure 5-7**). In my screen I wanted to identify genes that were off unless activated by the presence of TAL14. For greater coverage, I combined reads from both experimental conditions (*Agrobacterium* only and Agro TAL14) to assemble a transcriptome. Combining the experimental reads resulted in the transiently expressed TAL14 reads being incorporated in the transcriptome. I used the levels of TALE transcript in the *Agrobacterium*-only experimental results as an indicator of zero or off, since those reads should not be found or map in the absence of TAL14. In the results I filtered for very low RPKM in the presence of Agro at both 24 and 48 hpi but elevated RPKM in TAL14 treated plants, either by fold change or raw numbers. I focused on results where the *Agrobacterium*-only value remained at or below the zero value as identified by TALE expression, but some values were too high in the treatment and still relatively low in the untreated to ignore, such as for the proteinase inhibitors. I then aligned the reads to contigs, verified their increased mapping under TAL14 conditions, and used CLC to map all AUG open reading frames. In some cases where open reading frames appeared incomplete, I used an alternate draft *N. benthamiana* transcriptome to identify any additional nucleotides. I identified a top set of 28 candidate TAL14$_{CIO151}$ gene targets and grouped

them based on expression level into genes expressed at some point over 1,000 RPKM (1-8) and those that stay under 1,000 RPKM (9-28).

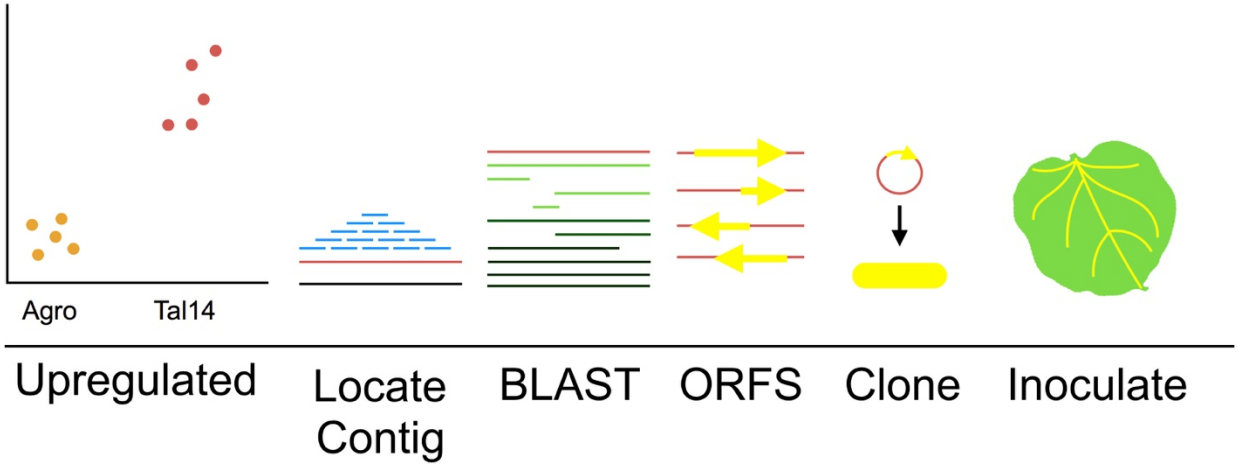

Figure 5-6 A simple RNA-seq pipeline for identifying the TAL14-targeted HR inducer in *N. benthamiana*. Using RNA-seq, I've identified genes that are upregulated only in the presence of $TAL14_{CIO151}$. These genes are then aligned to their assembled contig and then annotated using BLASTx. Identified ORFs are cloned into pE1776 and then expressed in *N. benthamiana* for the HR screen.

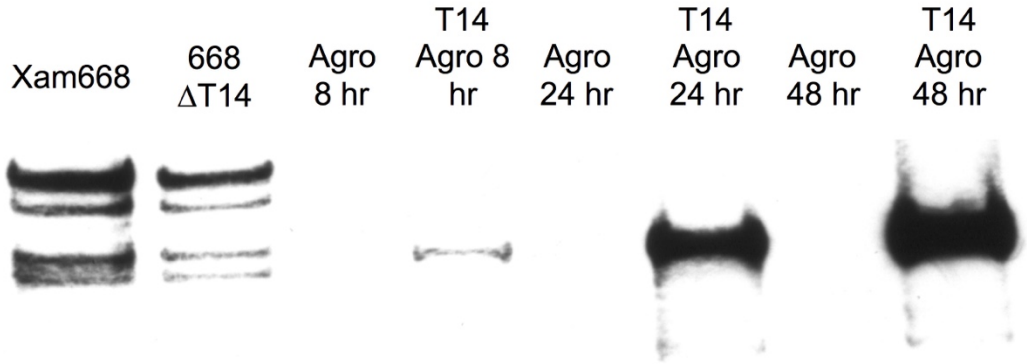

Figure 5-7 Western blot of TAL14 expression at 8, 24, and 48 hpi. Each *Agrobacterium* protein isolate represents 2 x 1.2 cm^2 punches from separate inoculations on the same leaf. Here, TAL14 is driven by 35s promoter. Protein is clearly already expressed at 8 hpi but only at small levels. I chose 24 and 48 hpi for my RNA-seq experiment. Novex NuPAGE 3-8% Tris Acetate gel. 4 min exposure.

5.3.5 Cloning and screening of candidate executor genes

A convenient aspect of studying the HR is that it provides an easy way of screening phenotypes. Upon identification of the top 28 candidate genes (**Table 5-5**), I began cloning them from TAL14-treated *N. benthamiana* cDNA (**Figure 5-8**) into the binary vector pE1776 (Leister, 2005). Successful clones were then transformed into *Agrobacterium* for *Agrobacterium*-mediated transient expression in *N. benthamiana*. If $TAL14_{CIO151}$ truly activates an executor gene, then sole expression of that gene should result in the HR. I have screened the top 8 candidates to date, but none of them result in a TAL14 similar response (**Figure 5-9**). While additional screening is still required and additional candidates have been amplified (**Figure 5-10**), there are a few other

possibilities that could make this screen more difficult. One such possibility is that $TAL14_{CIO151}$ activates multiple genes in *N. benthamiana* and it is only when they are expressed together that the plant exhibits an HR. We know this to be the case in cassava where another RNA-seq experiment (Cohn et al., 2015a) shows that $TAL14_{Xam668}$, a demonstrated contributor to Xam growth in planta, seems to be more promiscuous than others such as $TAL20_{Xam668}$, which was shown to target only a single SWEET susceptibility gene (Cohn et al., 2014). My RNA-seq data does suggest that there are many genes activated by $TAL14_{CIO151}$ in *N. benthamiana* and many of them do have predicted TAL14 EBEs in their promoter region. Some classes of genes are upregulated both in cassava and *N. benthamiana* such as the serine protease inhibitors.

Table 5-5 Candidate $TAL14_{CIO151}$ targets in *N. benthamiana*. The top 28 candidate $TAL14_{CIO151}$ targets in *N. benthamiana* as predicted by RNA-seq. The top EBE target scores were predicted using TAL Targeter 2.0 and the distance from the ATG start site is listed. ORFs were annotated using BLASTx. The candidates were arbitrarily separated into two groups, genes expressed at some point close to or over 1,000 RPKM (1-8) considered highly expressed and those mostly that stay under 1,000 RPKM (9-28). Here, two heat maps apply to the two groupings.

	Annotation	Ag 24 RPKM	Ag 48 RPKM	TAL14 24 RPKM	TAL14 48 RPKM	Fold Change	EBE Score	Distance	Size
–	TALE fragment	1	2	992	6179	2563	-		
1	Serine protease inhibitor	21	29	6469	8744	304	16	496 bp	447
2	Gamma thionin	3	2	1813	3673	1051	13	126 bp	318
3	Serine endopeptidase inhibitor	10	10	3261	2887	295	13	197 bp	720
4	Uncharacterized thionin-like	2	2	426	1177	423	14	65 bp	360
5	Carbonic anhydrase	6	10	2534	1156	229	18	150 bp	771
6	Methionine sulfide reductase	19	22	663	1029	42	8	43 bp	591
7	Premnaspirodiene oxygenase	0	1	9	972	1271	17	327 bp	1518
8	Cationic peroxidase	0	0	118	607	1666	16	618 bp	456

	Annotation	Ag 24 RPKM	Ag 48 RPKM	TAL14 24 RPKM	TAL14 48 RPKM	Fold Change	EBE Score	Distance	Size
9	Polyphenol oxidase	0	0	999	72	1277	17	454 bp	1788
10	Lipoxygenase	2	2	944	160	281	13	123 bp	2706
11	UDP glucoronosyl transferase	2	1	819	15	344	15	474 bp	341
12	Glyoxalase	1	0	394	5	408	19	29 bp	639
13	Acetyltransferase	1	1	356	14	270	11	207 bp	1329
14	Alpha dioxygenase	2	1	307	132	196	14	399 bp	1932
15	Transparent testa 12 like	1	0	296	24	237	16	427 bp	1581
16	Uncharacterized Na/Ca Channel	1	1	171	82	182	19	310 bp	1707
17	Uncharacterized	2	1	167	91	98	18	426 bp	906
18	Glycine rich cell wall protein	0	0	102	316	5855	14	291 bp	423
19	LRR serine/threonine kinase	1	1	81	208	140	14	423 bp	3573
20	Legumin B Like Partial	0	0	74	175	1955	17	302 bp	1452
21	Serine/threonine kinase	0	2	63	425	210	13	381 bp	1320
22	Hydroxylase-like	2	1	46	299	117	14	487 bp	1506
23	Lipoxygenase	0	0	44	272	1304	9	203 bp	552
24	Cytosolic sulfotransferase	0	0	32	280	335	8	451 bp	984
25	Cannabidiolic acid synthase	0	3	25	791	257	17	48 bp	1605
26	Vetispiradiene synthase	0	0	5	496	2786	13	69 bp	1674
27	Cadmium resistance protein	0	0	1	234	1843	14	230 bp	447
28	PPO polyphenol oxidase	0	2	1	537	304	15	102 bp	1833
						Best possible:	16		

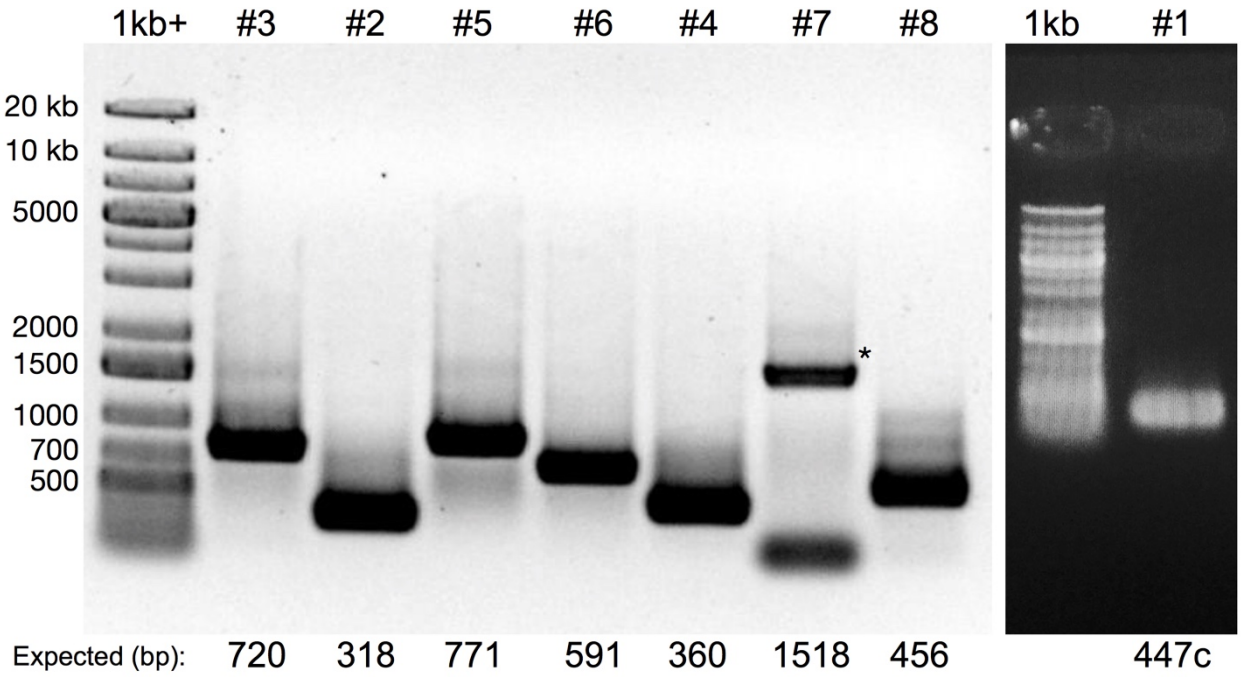

Figure 5-8 PCR products of *N. benthamiana* TAL14 candidates #1 - #8. The top 8 most highly expressed genes identified by RNA-seq in $TAL14_{CIO151}$ transiently delivered *N. benthamiana* were amplified from cDNA using Gibson primers for cloning into pE1776 with 5' Fw (acaccaaatcgactctaggggtacc-) and 3' Rev (-catccacctctgttaattcgagctc) overlapping sequences added to each primer.

The non-host resistance response in *N. benthamiana* has been well characterized and many genes required or involved for resistance have been identified (Senthil-Kumar and Mysore, 2013). A comparison of my top 28 candidate genes and the known genes yields at least one potential overlap, candidate #8 the putative cationic peroxidase, involved in the oxidative burst of HR. Many of the genes appear to be involved in activating or mediating the release of reactive oxygen species (ROS), a common response to abiotic and biotic stress in plants. Some of the candidate genes have been implicated in disease resistance, such as polyphenol oxidase (PPO) which increased resistance of tomato (*Lycopersicon esculentum*) to *Pseudomonas syringae* DC3000 potentially due to the production of quinones and ROS which could ultimately lead to cell death (Thipyapong et al., 2007). It's difficult to speculate whether TALE-dependent activation of one of these candidate genes leads to the general defense response or whether $TAL14_{CIO151}$ activates many of these genes with apparent EBEs, collectively leading to cell death. In cassava, $TAL14_{Xam668}$ upregulated 52 genes (Cohn et al., 2015a), suggesting many of these targets in *N. benthamiana* may be $TAL14_{CIO151}$ activated and it is possible that activation of multiple genes leads to the HR phenotype. Further screening of candidate genes and expressing multiple candidate genes will be required to identify the HR-inducing gene. Another strategy would be to use VIGS to knockdown the candidate genes and then screen for the abrogation of the TAL14-triggered HR.

5.3.6 Summary

Xam elicits a non-host resistance response in *N. benthamiana*. Part of this response appears to be activated in a TALE-dependent manner. Knockouts of both $TAL14_{CIO151}$ and $TAL14_{Xam668}$ show a lessened response and *Agrobacterium*-mediated transient delivery of each effector triggers an

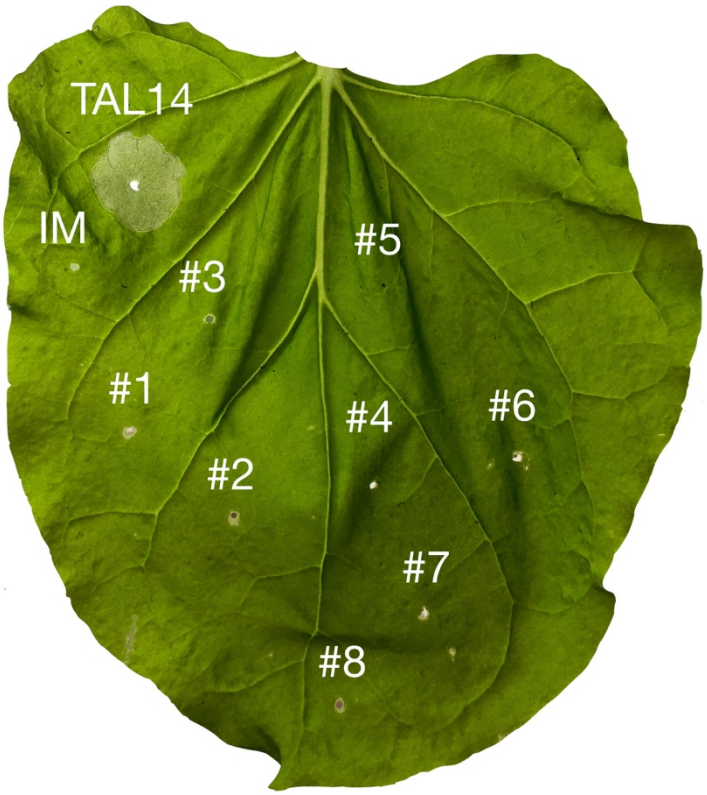

Figure 5-9 Inoculations of the top 8 *N. benthamiana* TAL14 candidate targets. The top 8 candidates cloned into pE1776 were transformed into *A. tumefaciens* 3101 for transient expression in *N. benthamiana*. Inoculations were screened for the HR-like phenotype as in the positive control $TAL14_{CIO151}$ (pEG100).

apparent HR. The response is dependent on and specific to the transcriptional activity of TAL14 as a $TAL14_{CIO151}$ activation domain mutant strain mostly abrogated the response and delivery of other Xam TALEs did not trigger the response. An RNA-seq based pipeline was designed to identify *N. benthamiana* genes upregulated by $TAL14_{CIO151}$, clone and transform them for transient expression, and screen for HR. A list of the top 28 upregulated candidates was compiled and a subset of them with predicted TAL14 EBEs has been screened but did not identify the elicitor of the phenotype. A conserved effector, found in many isolates and shown to contribute to bacterial growth in planta, $TAL14_{Xam668}$ was characterized in an RNA-seq experiment in cassava which did not identify a singular target for the effector, but identified 52 upregulated genes (Cohn et al., 2015a). This result suggests that TAL14 may be a promiscuous TALE, activating a wide swath of genes, perhaps targeting a common plant promoter element. Therefore, in the TAL14-dependent HR in *N. benthamiana*, there may be a combination of genes leading to the response or there may be just a single gene amongst the many that leads to the HR phenotype. A wide screen of all the potential candidates could be conducted through individual cloning or via VIGS assay, knocking down targets and screening for suppressed HR.

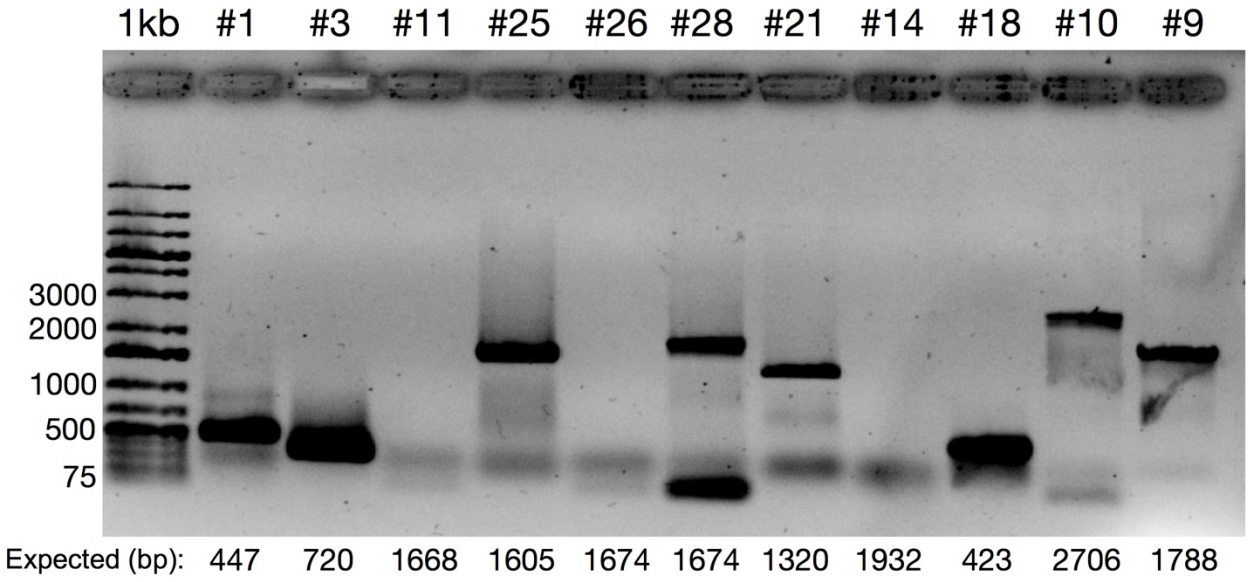

Figure 5-10 PCR products of other *N. benthamiana* TAL14 candidates. Additional candidates identified by RNA-seq in $TAL14_{CIO151}$ transiently delivered *N. benthamiana* were amplified from cDNA using Gibson primers for cloning into pE1776 with 5' Fw (acaccaaatcgactctaggggtacc-) and 3' Rev (-catccacctctgttaattcgagctc) overlapping sequences added to each primer.

6 Summary: an interesting past, a promising future

First discovered 26 years ago in the context of the avirulence capacity of *avrBs3* (Bonas et al., 1989), it would be 17 years until TALEs were called TAL effectors (Yang et al., 2006) and 20 years until the RVD cipher was decoded (Boch et al., 2009; Moscou and Bogdanove, 2009). Now, the structure of a TALE bound to its target has been solved (Deng et al., 2012; Mak et al., 2012) and TALE-like genes are being found in other genera such as *Ralstonia*-injected proteins (Rip)TALs of *Ralstonia solanacearum* (Lange et al., 2013) and the MOrTLs mined from metagenomes of marine organisms (de Lange et al., 2015). TALEs made the headlines in recent years with their application in genome editing. Armed with our understanding of both the predictable binding code and their modular structure, TALEs can be designed to target desired genes for activation or modified with other features, such as TALE nucleases (TALENs) or TALE-SRDX repressors (Bogdanove and Voytas, 2011; Li et al., 2012a; 2011; Mahfouz et al., 2011; Zhang et al., 2011) with efficient construction using Golden Gate cloning (Zhang et al., 2011). TALENs have been used in a wide range of organisms and cell types (Kühn et al., 2016). The availability of TALEN and CRISPR/Cas9-based systems offer a promising future for genome editing with each having their advantages. One advantage of TALENs is they allow fine tuning of target site specificity (Richter et al., 2016) and do not require PAM sites, but different tools will be appropriate for different goals.

Our growing understanding of the TALE code combined with TALE sequencing and EBE identification can lead to new durable resistance strategies against Xam. In one case, TALENs were used to modify the EBE site of the S gene *Os11N3* (*OsSWEET14*), creating a Xoo resistant rice (Li et al., 2012b). Obtaining durable resistance is a major struggle against quickly evolving pathogens, however, and will likely require a multi-pronged approach targeting several conserved effectors. Designing executor R genes and stacking EBEs of known TALEs in front of a host or heterologous R gene has already shown to be effective in the lab (Römer et al., 2009a; Schornack et al., 2012) and could be a promising approach in the field. Basic research into both the conserved and diverse effector repertoires as well as the functional roles of effectors will essential to support these strategies and to innovate new ones. Identifying bacterial components that are widely conserved, either because they are essential to the bacterial biology or are major contributors to virulence, will provide new targets for engineering resistance. My dissertation work studying TALEs of *Xanthomonas axonopodis* pv. manihotis has uncovered a set of TALEs that appear to be widely conserved in Xam and would lend themselves to the durable EBE-stacking executor R gene approach. On the other hand, my creation of TALE-cured strains through plasmid curing demonstrated that while some TALEs contribute to virulence in Xam, they do not appear to be essential for successful infection and bacterial growth in planta. Therefore, the road to durable resistance may be a long one as Xam could lose any avirulence-associated TALEs by losing a plasmid or obtain new ones through recombination. Regardless of the difficulties in achieving durable resistance, engineered resistance strategies are a promising method of providing relief for farmers against potentially devastating pathogens. Basic research will continue to inform applied research in developing new strategies.

7 References

Alegria, M.C., Souza, D.P., Andrade, M.O., Docena, C., Khater, L., Ramos, C.H.I., Da Silva, A.C.R., and Farah, C.S. (2005). Identification of new protein-protein interactions involving the products of the chromosome-and plasmid-encoded type IV secretion loci of the phytopathogen *Xanthomonas axonopodis* pv. citri. Journal of Bacteriology *187*, 2315–2325.

Alvarez-Martinez (2009). Biological Diversity of Prokaryotic Type IV Secretion Systems. Microbiol Mol Biol Rev *73*, 775–808.

Amaral, A.M.D., Toledo, C.P., Baptista, J.C., and Machado, M.A. (2005). Transformation of *Xanthomonas axonopodis* pv. citri by electroporation. Fitopatoloogia Brasileira *30*, 292–294.

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., and Yang, B. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. The Plant Cell Online *22*, 3864–3876.

Arrieta-Ortiz, M.L., Rodriguez-R, L.M., Perez-Quintero, A.L., Poulin, L., Díaz, A.C., Arias Rojas, N., Trujillo, C., Restrepo Benavides, M., Bart, R., Boch, J., et al. (2013). Genomic Survey of Pathogenicity Determinants and VNTR Markers in the Cassava Bacterial Pathogen *Xanthomonas axonopodis* pv. Manihotis Strain CIO151. PLoS ONE *8*, e79704.

Austin, M.J., Muskett, P., Kahn, K., Feys, B.J., Jones, J.D.G., and Parker, J.E. (2002). Regulatory role of SGT1 in early R gene-mediated plant defenses. Science *295*, 2077–2080.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1994). Current protocols in molecular biology (New York, NY: John Wiley & Sons).

Aziz, R.K., Breitbart, M., and Edwards, R.A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Research *38*, 4207–4217.

Bai, J., Choi, S.H., Ponciano, G., Leung, H., and Leach, J.E. (2000). *Xanthomonas oryzae* pv. oryzae avirulence genes contribute differently and specifically to pathogen aggressiveness. Mol. Plant Microbe Interact. *13*, 1322–1329.

Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., and Wang, Q. (2012). Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance.

Ballvora, A., Pierre, M., van den Ackerveken, G., Schornack, S., Rossier, O., Ganal, M., Lahaye, T., and Bonas, U. (2001). Genetic mapping and functional analysis of the tomato Bs4 locus governing recognition of the *Xanthomonas campestris* pv. vesicatoria AvrBs4 protein. Mol. Plant Microbe Interact. *14*, 629–638.

Bart, R., Cohn, M., Kassen, A., McCallum, E.J., Shybut, M., Petriello, A., Krasileva, K., Dahlbeck, D., Medina, C., Alicai, T., et al. (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. Proceedings of the National Academy of Sciences *109*, E1972–E1979.

Bhattacharjee, S., Halane, M.K., Kim, S.H., and Gassmann, W. (2011). Pathogen effectors target Arabidopsis EDS1 and alter its interactions with immune regulators. Science *334*, 1405–1408.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science *326*, 1509–1512.

Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors - pathogen strategies and plant resistance engineering. New Phytologist *204*, 823–832.

Bogdanove, A.J., and Voytas, D.F. (2011). TAL Effectors: Customizable Proteins for DNA Targeting. Science *333*, 1843–1846.

Bogdanove, A.J., Schornack, S., and Lahaye, T. (2010). TAL effectors: finding plant genes for disease and defense. Current Opinion in Plant Biology *13*, 394–401.

Boller, T., and He, S.Y. (2009). Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. Science *324*, 742–744.

Bonas, U.U., Stall, R.E.R., and Staskawicz, B.B. (1989). Genetic and structural characterization of the avirulence gene avrBs3 from *Xanthomonas campestris* pv. vesicatoria. Mol Gen Genet *218*, 127–136.

Bonas, U., Schulte, R., Fenselau, S., Minsavage, G.V., Staskawicz, B.J., and Stall, R.E. (1991). Isolation of a gene-cluster from *Xanthomonas campestris* pv. vesicatoria that determines pathogenicity and the hypersensitive response on pepper and tomato. Mol. Plant Microbe Interact. *4*, 81–88.

Bottacini, F., Motherway, M.O., Casey, E., McDonnell, B., Mahony, J., Ventura, M., and van Sinderen, D. (2015). Discovery of a conjugative megaplasmid in *Bifidobacterium breve*. Applied and Environmental Microbiology *81*, 166–176.

Boucher, C.A., Vangijsegem, F., Barberis, P.A., Arlat, M., and Zischek, C. (1987). *Pseudomonas solanacearum* genes controlling both pathogenicity on tomato and hypersensitivity on tobacco are clustered. Journal of Bacteriology *169*, 5626–5632.

Buttner, D., and He, S.Y. (2009). Type III Protein Secretion in Plant Pathogenic Bacteria. Plant Physiology *150*, 1656–1664.

Castiblanco, L.F., Gil, J., Rojas, A., Osorio, D., Gutiérrez, S., Muñoz-Bodnar, A., Perez-Quintero, A.L., Koebnik, R., Szurek, B., Lopez, C., et al. (2012). TALE1 from *Xanthomonas axonopodis* pv. manihotis acts as a transcriptional activator in plant cells and is important for pathogenicity in cassava plants. Molecular Plant Pathology.

Cernadas, R.A., Doyle, E.L., Niño-Liu, D.O., Wilkins, K.E., Bancroft, T., Wang, L., Schmidt, C.L., Caldo, R., Yang, B., White, F.F., et al. (2014). 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 *10*, e1003972.

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.-Q., Guo, W.-J., Kim, J.-G., Underwood, W., Chaudhuri, B., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. Nature *468*, 527–532.

Chu, Z., Yuan, M., Yao, L., Ge, X., Yuan, B., Xu, C., Li, X., Fu, B., Li, Z., Bennetzen, J., et al. (2006). Promoter mutations of an essential gene for pollen development result in disease resistance in rice. Genes & Development *20*, 1250–1255.

Coburn, B., Sekirov, I., and Finlay, B.B. (2007). Type III secretion systems and disease. Clin. Microbiol. Rev. *20*, 535–549.

Cohn, M., Bart, R.S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., Hou, B.-H., Frommer, W.B., Lahaye, T., and Staskawicz, B.J. (2014). *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. Mol. Plant Microbe Interact. *27*, 1186–1198.

Cohn, M., Morbitzer, R., Lahaye, T., and Staskawicz, B.J. (2015a). Comparison of gene activation by two TAL effectors from *Xanthomonas axonopodis* pv. manihotis reveals candidate host susceptibility genes in cassava. Molecular Plant Pathology.

Cohn, M., Shybut, M., Dahlbeck, D., and Staskawicz, B. (2015b). Assays to Assess Virulence of *Xanthomonas axonopodis* pv. manihotis on Cassava. Bio-Protocol.

Cornelis, G.R., and Van Gijsegem, F. (2000). Assembly and function of type III secretory systems. Annu. Rev. Microbiol. *54*, 735–774.

de Lange, O., Wolf, C., Thiel, P., Krüger, J., Kleusch, C., Kohlbacher, O., and Lahaye, T. (2015). DNA-binding proteins from marine bacteria expand the known sequence diversity of TALE-like repeats. Nucleic Acids Research gkv1053–16.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics *23*, 673–679.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., and Yan, N. (2012). Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. Science.

DeRidder, B.P., Shybut, M.E., Dyle, M.C., Kremling, K.A.G., and Shapiro, M.B. (2012). Changes at the 3'-untranslated region stabilize Rubisco activase transcript levels during heat stress in Arabidopsis. Planta –.

Dodds, P.N., and Rathjen, J.P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. Nat Rev Genet *11*, 539–548.

Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K., and Bogdanove, A.J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Research *40*, W117–W122.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006).

Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal *45*, 616–629.

El-Sharkawy, M.A. (2004). Cassava biology and physiology. Plant Mol Biol *56*, 481–501.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS ONE *3*, e3647.

Falk, A., Feys, B.J., Frost, L.N., Jones, J.D., Daniels, M.J., and Parker, J.E. (1999). EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis has homology to eukaryotic lipases. Proceedings of the National Academy of Sciences *96*, 3292–3297.

FAO (2008). Why cassava? Fao.org/Ag/Agp/Agpc/Gcds/Index_en.Html.

FAO (2009). Cassava Diseases in Africa: A major threat to food security. Fao.org/Ag/Agp/Agpc/Gcds/Index_en.Html.

FAO, and IFAD (2004). Global cassava market study: Business opportunites for the use of cassava. Http://Www.Fao.org/Docrep/007/Y5287e/Y5287e00.HTM.

FAOSTAT (2015). FAOSTAT. Faostat3.Fao.org.

Ferreira, R.M., de Oliveira, A.C.P., Moreira, L.M., Belasque, J., Gourbeyre, E., Siguier, P., Ferro, M.I.T., Ferro, J.A., Chandler, M., and Varani, A.M. (2015). A TALE of transposition: Tn3-like transposons play a major role in the spread of pathogenicity determinants of Xanthomonas citri and other xanthomonads. MBio *6*, e02505–e02514.

Figurski, D.H., and Helinski, D.R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proceedings of the National Academy of Sciences *76*, 1648–1652.

Flor, H. (1946). Genetics of pathogenicity in Melampsora lini. J. Agric. Res *73*, 335.

Flor, H. (1947). Inheritance of reaction to rust in flax. J. Agric. Res *74*, 241.

Flor, H. (1971). Current Status of Gene-for-Gene Concept. Annu. Rev. Phytopathol. *9*, 275–&.

Garcillán-Barcia, M.P., Alvarado, A., and la Cruz, de, F. (2011). Identification of bacterial plasmids based on mobility and plasmid population biology. FEMS Microbiol. Rev. *35*, 936–956.

Gay, P., Lecoq, D., Steinmetz, M., Berkelman, T., and Kado, C.I. (1985). Positive Selection Procedure for Entrapment of Insertion-Sequence Elements in Gram-Negative Bacteria. Journal of Bacteriology *164*, 918–921.

Gill, U.S., Lee, S., and Mysore, K.S. (2015). Host versus nonhost resistance: distinct wars with similar arsenals. Phytopathology *105*, 580–587.

Gough, C.L., Genin, S., Zischek, C., and Boucher, C.A. (1992). Hrp Genes of Pseudomonas-

Solanacearum Are Homologous to Pathogenicity Determinants of Animal Pathogenic Bacteria and Are Conserved Among Plant Pathogenic Bacteria. Mol. Plant Microbe Interact. *5*, 384–389.

Grant, S.R., Fisher, E.J., Chang, J.H., Mole, B.M., and Dangl, J.L. (2006). Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. Annu. Rev. Microbiol. *60*, 425–449.

Greenberg, J.T., and Yao, N. (2004). The role and regulation of programmed cell death in plant-pathogen interactions. Cell. Microbiol. *6*, 201–211.

Gu, K., Tian, D., Yang, F., Wu, L., Sreekala, C., Wang, D., Wang, G.L., and Yin, Z. (2004). High-resolution genetic mapping of Xa27(t), a new bacterial blight resistance gene in rice, *Oryza sativa* L. Theor Appl Genet *108*, 800–807.

Gu, K., Sangha, J.S., Li, Y., and Yin, Z. (2008). High-resolution genetic mapping of bacterial blight resistance gene Xa10. Theor Appl Genet *116*, 155–163.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.-L., White, F.F., et al. (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. Nature *435*, 1122–1125.

Hayes, F. (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science *301*, 1496–1499.

Howeler, R., Lutaladio, N., and Thomas, G. (2013). Save and Grow: Cassava (full version). Fao 1–142.

Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W.B., Yang, B., White, F.F., Wang, N., and Jones, J.B. (2014). Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. Proceedings of the National Academy of Sciences *111*, E521–E529.

Innes, R.W. (2004). Guarding the goods. New insights into the central alarm system of plants. Plant Physiology *135*, 695–701.

Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. Nature *444*, 323–329.

Jutkina, J., Hansen, L.H., Li, L., Heinaru, E., Vedler, E., Jõesaar, M., and Heinaru, A. (2013). Complete nucleotide sequence of the self-transmissible TOL plasmid pD2RT provides new insight into arrangement of toluene catabolic plasmids. Plasmid *70*, 393–405.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science *318*, 648–651.

Kay, S., Hahn, S., Marois, E., Wieduwild, R., and Bonas, U. (2009). Detailed analysis of the DNA recognition motifs of the Xanthomonas type III effectors AvrBs3 and AvrBs3Deltarep16. The Plant Journal *59*, 859–871.

Kearney, B., and Staskawicz, B.J. (1990). Characterization of Is476 and Its Role in Bacterial

Spot Disease of Tomato and Pepper. Journal of Bacteriology *172*, 143–148.

Kessler, B., de Lorenzo, V., and Timmis, K.N. (1992). A general system to integrate lacZ fusions into the chromosomes of gram-negative eubacteria: regulation of the Pm promoter of the TOL plasmid studied with all controlling elements in monocopy. Mol Gen Genet *233*, 293–301.

Kosarewicz, A., Königsmaier, L., and Marlovits, T.C. (2012). The blueprint of the type-3 injectisome. Philosophical Transactions of the Royal Society B: Biological Sciences *367*, 1140–1154.

Kühn, R., Wurst, W., and Wefers, B. eds. (2016). TALENs (New York, NY: Springer New York).

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

Lange, O., Schreiber, T., Schandry, N., Radeck, J., Braun, K.H., Koszinowski, J., Heuer, H., Strauß, A., and Lahaye, T. (2013). Breaking the DNA-binding code of *Ralstonia solanacearum* TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. New Phytologist *199*, 773–786.

Lapin, D., and Van den Ackerveken, G. (2013). Susceptibility to plant disease: more than a failure of host immunity. Trends in Plant Science *18*, 546–554.

Leister, R.T. (2005). Molecular Genetic Evidence for the Role of SGT1 in the Intramolecular Complementation of Bs2 Protein Activity in Nicotiana benthamiana. Plant Cell *17*, 1268–1278.

Leplae, R., Lima-Mendez, G., and Toussaint, A. (2010). ACLAME: A CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Research *38*, D57–D61.

Levin, B.R., and Bergstrom, C.T. (2000). Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. Proceedings of the National Academy of Sciences *97*, 6981–6985.

Li, C., Wei, J., Lin, Y., and Chen, H. (2012a). Gene silencing using the recessive rice bacterial blight resistance gene xa13 as a new paradigm in plant breeding. Plant Cell Rep *31*, 851–862.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012b). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat Biotechnol *30*, 390–392.

Li, Y.R., Che, Y.Z., Zou, H.S., Cui, Y.P., Guo, W., Zou, L.F., Biddle, E.M., Yang, C.H., and Chen, G.Y. (2011). Hpa2 Required by HrpF To Translocate *Xanthomonas oryzae* Transcriptional Activator-Like Effectors into Rice for Pathogenicity. Applied and Environmental Microbiology *77*, 3809–3818.

Lindeberg, M., Cunnac, S., and Collmer, A. (2012). *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. Trends in Microbiology *20*, 199–208.

Lindgren, P.B., Peet, R.C., and Panopoulos, N.J. (1986). Gene cluster of *Pseudomonas syringae* pv. phaseolicola controls pathogenicity of bean plants and hypersensitivity of nonhost plants. Journal of Bacteriology *168*, 512–522.

Liu, Y.L., Schiff, M., Marathe, R., and Dinesh-Kumar, S.P. (2002). Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. Plant J *30*, 415–429.

Llosa, M., Bolland, S., and la Cruz, de, F. (1994). Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. Journal of Molecular Biology *235*, 448–464.

Lozano, J.C. (1986). Cassava bacterial blight: a manageable disease. Plant Disease *70*, 1089–1093.

López, C.E., and Bernal, A.J. (2012). Cassava Bacterial Blight: Using Genomics for the Elucidation and Management of an Old Problem. Tropical Plant Biol. *5*, 117–126.

Mahfouz, M.M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J.K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. (National Acad Sciences), pp. 2623–2628.

Mak, A.N.S., Bradley, P., Cernadas, R.A., Bogdanove, A.J., and Stoddard, B.L. (2012). The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. Science.

Mak, A.N.-S., Bradley, P., Bogdanove, A.J., and Stoddard, B.L. (2013). TAL effectors: function, structure, engineering and applications. Current Opinion in Structural Biology *23*, 93–99.

Martin, F., and Kamoun, S. (2012). Effectors in Plant-Microbe Interactions (Oxford, UK: John Wiley & Sons).

Monteiro-Vitorello, C.B., De Oliveira, M.C., Zerillo, M.M., Varani, A.M., Civerolo, E., and Van Sluys, M.-A. (2005). Xylella and Xanthomonas Mobil'omics. Http://Dx.Doi.org/10.1089/Omi.2005.9.146 *9*, 146–159.

Moscou, M.J., and Bogdanove, A.J. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. Science –.

Murillo, J., Vivian, A., and Jackson, R.W. (2001). The roles of plasmids in phytopathogenic bacteria: mobile arsenals? Microbiology *147*, 763–780.

Naim, F., Nakasugi, K., Crowhurst, R.N., Hilario, E., Zwart, A.B., Hellens, R.P., Taylor, J.M., Waterhouse, P.M., and Wood, C.C. (2012). Advanced engineering of lipid metabolism in *Nicotiana benthamiana* using a draft genome and the V2 viral silencing-suppressor protein. PLoS ONE *7*, e52717–e52717.

Nakasugi, K., Crowhurst, R.N., Bally, J., Wood, C.C., Hellens, R.P., and Waterhouse, P.M. (2013). De novo transcriptome sequence assembly and analysis of RNA silencing genes of Nicotiana benthamiana. PLoS ONE *8*, e59534–e59534.

Norman, A., Hansen, L.H., and Sorensen, S.J. (2009). Conjugative plasmids: vessels of the communal gene pool. Philosophical Transactions of the Royal Society B: Biological Sciences *364*, 2275–2289.

Ogunjobi, A. (2006). Molecular variation in population structure of *Xanthomonas axonopodis* pv manihotis in the south eastern Nigeria. African Journal of Biotechnology *5*.

Prochnik, S., Marri, P.R., Desany, B., Rabinowicz, P.D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., Harkins, T., et al. (2012). The Cassava Genome: Current Progress, Future Directions. Tropical Plant Biol. *5*, 88–94.

Qi, D., and Innes, R.W. (2013). Recent advances in plant NLR structure, function, localization, and signaling. Front Immunol *4*.

Raleigh, E.A., and Brooks, J.E. (1998). Restriction modification systems: where they are and what they do. Bacterial Genomes.

Restrepo, S., and Verdier, V. (1997). Geographical differentiation of the population of *Xanthomonas axonopodis* pv. manihotis in Colombia. Applied and Environmental Microbiology *63*, 4427–4434.

Restrepo, S., Duque, M., and Verdier, V. (2000). Resistance spectrum of selected Manihot esculenta genotypes under field conditions. Field Crop Res *65*, 69–77.

Richter, A., Streubel, J., and Boch, J. (2016). TAL Effector DNA-Binding Principles and Specificity. In TALENs, R. Kühn, W. Wurst, and B. Wefers, eds. (New York, NY: Springer New York), pp. 9–25.

Rossier, O., Wengelnik, K., Hahn, K., and Bonas, U. (1999). The *Xanthomonas* Hrp type III system secretes proteins from plant and mammalian bacterial pathogens. Proceedings of the National Academy of Sciences *96*, 9368–9373.

Römer, P., Recht, S., and Lahaye, T. (2009a). A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. Proceedings of the National Academy of Sciences *106*, 20526–20531.

Römer, P., Strauß, T., Hahn, S., Scholze, H., Morbitzer, R., Grau, J., Bonas, U., and Lahaye, T. (2009b). Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. Plant Physiology *150*, 1697–1712.

Ryan, R.P., Vorhölter, F.-J., Potnis, N., Jones, J.B., Van Sluys, M.-A., Bogdanove, A.J., and Dow, J.M. (2011). Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. Nat. Rev. Microbiol. *9*, 344–355.

Schornack, S., Ballvora, A., Gürlebeck, D., Peart, J., Ganal, M., Baker, B., Bonas, U., and Lahaye, T. (2004). The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defense responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. The Plant Journal *37*, 46–60.

Schornack, S., Minsavage, G.V., Stall, R.E., Jones, J.B., and Lahaye, T. (2008). Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. New Phytol. *179*, 546–556.

Schornack, S., Moscou, M.J., Ward, E.R., and Horvath, D.M. (2012). Engineering Plant Disease Resistance Based on TAL Effectors. Annu. Rev. Phytopathol. *51*, 130531172133008.

Schwartz, A.R., Potnis, N., Timilsina, S., Wilson, M., PatanÃ, J., Martins, J., Minsavage, G.V., Dahlbeck, D., Akhunova, A., Almeida, N., et al. (2015). Phylogenomics of *Xanthomonas* field strains infecting pepper and tomato reveals diversity in effector repertoires and identifies determinants of host specificity. Front. Microbiol. *6*, 1–17.

Sebra, R.P., Salzberg, S.L., Carpenter, S.C.D., Wang, L., Booher, N.J., Bogdanove, A.J., and Leach, J.E. (2015). Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex TAL (transcription activator-like) effector gene relationships. Microbial Genomics *1*, 1–22.

Senthil-Kumar, M., and Mysore, K.S. (2013). Nonhost Resistance Against Bacterial Pathogens: Retrospectives and Prospects. Http://Dx.Doi.org/10.1146/Annurev-Phyto-082712-102319.

Slewinski, T.L. (2011). Diverse functional roles of monosaccharide transporters and their homologs in vascular plants: a physiological perspective. Molecular Plant … *4*, 641–662.

Smillie, C., Garcillán-Barcia, M.P., Francia, M.V., Rocha, E.P.C., and la Cruz, de, F. (2010). Mobility of plasmids. Microbiol Mol Biol Rev *74*, 434–452.

Sorensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005). Studying plasmid horizontal transfer in situ: a critical review. Nat. Rev. Microbiol. *3*, 700–710.

Spoel, S.H., and Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. Nat Rev Immunol *12*, 89–100.

Staskawicz, B.J., Dahlbeck, D., and Keen, N.T. (1984). Cloned Avirulence Gene of *Pseudomonas syringae* pv glycinea Determines Race-Specific Incompatibility on Glycine-Max (L) Merr. Proceedings of the National Academy of Sciences *81*, 6024–6028.

Stockwell, V.O., Davis, E.W., Carey, A., Shaffer, B.T., Mavrodi, D.V., Hassan, K.A., Hockett, K., Thomashow, L.S., Paulsen, I.T., and Loper, J.E. (2013). pA506, a Conjugative Plasmid of the Plant Epiphyte Pseudomonas fluorescens A506. Applied and Environmental Microbiology *79*, 5272–5282.

Strauß, T., van Poecke, R.M.P., Strauß, A., Römer, P., Minsavage, G.V., Singh, S., Wolf, C., Strauß, A., Kim, S., Lee, H.-A., et al. (2012). RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. Proc. Natl. Acad. Sci. U.S.a. *109*, 19480–19485.

Subramoni, S., Jha, G., and Sonti, R. (2006). Virulence functions of xanthomonads. Plant-Associated Bacteria 535–571.

Sundin, G.W. (2007). Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. Annu. Rev. Phytopathol. *45*, 129–151.

Szurek, B. (2013). Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. 1–4.

Takken, F.L., and Goverse, A. (2012). How to build a pathogen detector: structural basis of NB-LRR function. Current Opinion in Plant Biology *15*, 375–384.

Tameling, W.I.L., and Takken, F.L.W. (2008). Resistance proteins: Scouts of the plant innate immune system. Eur J Plant Pathol *121*, 243–255.

Tampakaki, A.P., Skandalis, N., Gazi, A.D., Bastaki, M.N., Panagiotis F, S., Charova, S.N., Kokkinidis, M., and Panopoulos, N.J. (2010). Playing the "Harp": Evolution of Our Understanding of hrp/hrcGenes 1. Annu. Rev. Phytopathol. *48*, 347–370.

Tan, Y., Xu, D., Li, Y., and Wang, X. (2012). Construction of a novel sacB-based system for marker-free gene deletion in Corynebacterium glutamicum. Plasmid *67*, 44–52.

Tauch, A., Schneiker, S., Selbitschka, W., Pühler, A., van Overbeek, L.S., Smalla, K., Thomas, C.M., Bailey, M.J., Forney, L.J., Weightman, A., et al. (2002). The complete nucleotide sequence and environmental distribution of the cryptic, conjugative, broad-host-range plasmid pIPO2 isolated from bacteria of the wheat rhizosphere. Microbiology *148*, 1637–1653.

Thipyapong, P., Stout, M.J., and Attajarusit, J. (2007). Functional Analysis of Polyphenol Oxidases by Antisense/Sense Technology. 1–27.

Thynne, E., McDonald, M.C., and Solomon, P.S. (2015). Phytopathogen emergence in the genomics era. Trends in Plant Science *20*, 246–255.

Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., Zhou, Z., Goh, M., Luo, Y., Murata-Hori, M., et al. (2014). The Rice TAL Effector-Dependent Resistance Protein XA10 Triggers Cell Death and Calcium Depletion in the Endoplasmic Reticulum. Plant Cell *26*, 497–515.

van Doorn, W.G., and Woltering, E.J. (2005). Many ways to exit? Cell death categories in plants. Trends in Plant Science *10*, 117–122.

Varani, A.M., Monteiro-Vitorello, C.B., Nakaya, H.I., and Van Sluys, M.-A. (2013). The role of prophage in plant-pathogenic bacteria. Annu. Rev. Phytopathol. *51*, 429–451.

Verdier, V. (1988). Studies on the variability of *Xanthomonas campestris* pv. manihotis , the causal agent of the bacterial blight disease of cassava (Manihot esculenta Crantz.).

Wan, Z., Varshavsky, J., Teegala, S., McLawrence, J., and Goddard, N.L. (2011). Measuring the Rate of Conjugal Plasmid Transfer in a Bacterial PopulationUsing Quantitative PCR. Biophysj *101*, 237–244.

Wang, C., Fan, Y., Zheng, C., Qin, T., Zhang, X., and Zhao, K. (2014). High-resolution genetic

mapping of rice bacterial blight resistance gene Xa23. Mol. Genet. Genomics *289*, 745–753.

Wang, C., Zhang, X., Fan, Y., Gao, Y., Zhu, Q., Zheng, C., Qin, T., Li, Y., Che, J., Zhang, M., et al. (2015). XA23 Is an Executor R Protein and Confers Broad-Spectrum Disease Resistance in Rice. Molecular Plant … *8*, 290–302.

White, F.F., Potnis, N., Jones, J.B., and Koebnik, R. (2009). The type III effectors of Xanthomonas. Molecular Plant Pathology *10*, 749–766.

White, M.J.A., and Nichols, W.A. (2003). Cosmid packaging and infection of E. coli. Methods Mol. Biol. *235*, 67–73.

White, T.J., and Gonzalez, C.R. (1995). Electroporation of *Xanthomonas*. Methods in Molecular Biology 1–7.

Wiedenbeck, J., and Cohan, F.M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. *35*, 957–976.

Wilkins, K.E., Booher, N.J., Wang, L., and Bogdanove, A.J. (2015). 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 *6*, 536.

Xie, F., Murray, J.D., Kim, J., Heckmann, A.B., Edwards, A., Oldroyd, G.E.D., and Downie, J.A. (2012). Legume pectate lyase required for root infection by rhizobia. Pnas *109*, 633–638.

Yacoubi, El, B., Brunings, A.M., Yuan, Q., Shankar, S., and Gabriel, D.W. (2007). In planta horizontal transfer of a major pathogenicity effector gene. Applied and Environmental Microbiology *73*, 1612–1621.

Yang, B., Sugio, A., and White, F.F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences *103*, 10503–10508.

Yang, Y., and Gabriel, D.W. (1995). Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. Journal of Bacteriology *177*, 4963–4968.

Yang, Y., De Feyter, R., and Gabriel, D.W. (1994). Host-specific symptoms and increased release of *Xanthomonas citri* and *X. campestris* pv. malvacearum from leaves are determined by the 102-bp tandem …. MPMI-Molecular Plant Microbe ….

Yuan, M., Chu, Z., LI, X., Xu, C., and Wang, S. (2010). The bacterial pathogen *Xanthomonas oryzae* overcomes rice defenses by regulating host copper redistribution. The Plant Cell Online *22*, 3164–3176.

Yuan, T., LI, X., XIAO, J., and Wang, S. (2011). Characterization of *Xanthomonas oryzae*-responsive cis-acting element in the promoter of rice race-specific susceptibility gene Xa13. Molecular Plant … *4*, 300–309.

Zeng, X., Tian, D., Gu, K., Zhou, Z., Yang, X., Luo, Y., White, F.F., and Yin, Z. (2015). Genetic engineering of the Xa10 promoter for broad-spectrum and durable resistance to Xanthomonas oryzae pv. oryzae. Plant Biotechnol J *13*, 993–1001.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nature Publishing Group *29*, 149–153.

Zhang, J., Yin, Z., and White, F. (2015). TAL effectors and the executor R genes. Front Plant Sci *6*, 641.

8 Appendix

8.1 pXam46 sequence

GAATTCCTACTCCACCAGCGATTACAACGCGTGCTATAGCAACAGCAATGGGGTGAT
TGTGCCGCTGGTTCCAATCTCGCAGATCGGCAGCATCAAGATGAGCGCTTCGGCGAC
CGCGGGCGGCGTGGACACCGTGACGTTCTCGATCGACGGCAGGGCCTACAGCGTCA
GTCAAAGCGCCAACACTCTGAACATCAACAGGATTTGGCGGCGCTCCGAGTTCAAT
ATATTCGGCAACGGCAGTGACCACCCGCTGGTGTTGTTCAATTCGGGCTCGCATGTC
ACCGTGAACGTGGCCGTCAAAGATGGAACCTCCAATGCTCCGAGGTGCCTGGGTCC
CAACGCCGGTTATTCGGGGCAACAGAACAACCTCACGCTCGGCAAATGCACCGCGT
CTGGTGGTGCCTCGCCTTCCATCACCTTTACCGAAAGCAACTGACGCCCGGGCCTTT
TCGTTGCCGTGGCTACCGGGCGCTGTCTCAGCGTGAGTTGGCGCGTGCGCAATGCCC
AATCTGCGTGCGCAACATCACTGAAACCTTCAAACAGAACATCCCGACGCTGCACA
GTTCCGGCTTACATCCCCAATAAGCAGGCAGGGTGCGCAGCTTGACTTAAGCGGTCG
CGCTCGGTTGCCCGTAGCTTTTTCGGGGGCATTGCCGGCGATCACTCGCTGTGTTTGG
CCCCGGCCACGCTGTGCCCTGGTCCTTTCCGAAGCTCGATCAACGGCCACAGGGCAA
TTGCCTTGCCAACGTCCGACTAGTAACCCCCTGCTGACCCGCTTTTGGGGTGAGGGA
GCCGCCGGCTCTGCCATCGGAGGTGCTTAGAGTCCGCGGGGCAGTGAGGGCAGAGC
CGGTGGGAGAGGGGCGCTTGCGCCCCTCTCCGCGGGGTCGAAGTTGCAACCGTAGC
GCCATAGGCGCGCAGGGAGGACTAGGGGCGCGCAGCCCCCGCAACTGGATAAAGCT
CTTAACGCTGTCGGCGTCGCGTAGCGGCGGCGACACCGATTTGCTGTTGATTTTCCA
CTGACTGTGCTTTCAACCAAAGGTTGACCTGGACGCGCTGCTCGACGGTCTGGACCT
GGAGTGACGGAAATTTCGGCGGTTTGTGGTTCAACCCCCTAACTCATTGAGTGAGAG
GTTTATTACCCCCATCACTGCCAACGCTCAGGTAATCTCCGATTAAGCCTCACTGAG
GAAATAGCTCCATCAACCATGCCATCTCCTCTTCGTTGAATGCCGGAAAATCATCCG
CTGCCCCTGCGAAGGGGTCAGCATCTTGTTCCCACATCACGGTGCTGGACGCTGCCA
GGTCGGCGTCCATGGGTGTACCAGGATCCGGGAGGCCGCCCCCGATCCTGGTACGC
GGGCGTTTTACACCCCAGCTGAGGGGGGGTAAATGCAGCGCATCGCGCTGTTCGGG
AACGCGCACCTCGACAGCTTGCTGTGCGAAGGAGCCGGTGACAGAACTCTCCGATC
GGGACCGTTTACGGCTGCTGCTTGCCAGCGCCTGGCCCGCCCGGTCTATTGGGCTGG
GCGCATCCAGCTCACGCTCCAGCGCATCGGCGAATGCATGCAAAGACGCCTGGTCC
GGCGTTTGAGCCGAAGCACCGGACGGTTCGGCCCTTTTCATCCCTGATGCCTGGAGG
ATACGGTGCCAACGCTGCGGGGCTGGGGGGAGCGTTCCACTGCGGGCCTCGAGTCC
GGTGACGCCCACTCTGCGAAATAGCTGTAACAACCCGTGCCTGCTCATCCCGAACTG
CGTCATGGCTTCATCAAATGCTTGCGCTGGGTGGGAGTGGCACTGGAAAAAACCCA
GCACGCGAGCCACTTGCGCGTGGTCGGCAACGCGATGGGACGTGCGTTCGGGAAGA
CGGCGATTGGTTCTTTTGATCAAGGTCGGCGCGTGCGGCAATCCCTTTCTCACTGCCT
CCAGCGCAGGACGCCCGCCGAGGCAGGCCAAGGCGACGAGGTGGTCGTTGGTCAAC
GCGGCCAACGCCTGATCAGGGCGAGATAACTGGGCAAAAGTGCTCTCCAGCGCCGG
CTTGCCGCCGCCATTGCTGGCGATGGCCACCACCTGGTCCGGGGTCAGGCCATGTTG
CTCGCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGCCAAT
ATTGCTGGCGATAGCCACCACCTGGTCCGGGGTCAGGCCATGTTGCTCGCACAGCAC
CGGCAACAGCCGCTGCACCGTCTCCAGCGCCGGCTTGCCGCCGTTATTGCTGGCGAT
GGCCACCACCTGCGCCGGGGTCAGGCCATGGGCCTGGCGCAGCACCGGCAACAGCC
GCTGCACCGTCTCCAGCGCCTGCTTGCCGCCATCGTGGCTGGCGATGGCCACCACCT
GGTCCGGGGTCAGGCCATGTTGCTCGCACAGCACCGGCAACAGCCGCTGCACCGTC

TCCAGCGCCTGCTTGCCGCCGCCATTGCTGGCGATGGCCACCACCTGCGCCCGGGTC
AGGCCATGTTGCTCGCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGC
TTGCCGCCGTTATTGCTGGCGATGGCCACCACCTGGTCCGGGGTCAGACCATGTTGC
TCGCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCGGCTTGCCGCCGCTA
TTGCTGGCGATGGCCACCACCTGGTCCGGGGTCAGGCCATGGGCCTGGCGCAGCAC
CGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGCCGCTATTGCTGGCGAT
GGCCACCACCTGCGCCGGGGTCAGGCCATGTTGCTCGCACAGCACCGGCAACAGCT
GCTGCACCGTCTCCAGCGCCTGCTTGCCGCCATCGTGGCTGGCGATGGCCACCACCT
GCGCCGGGGTCAGGCCATGGGCCTGGCGCAGCACCGGCAACAGCCGCTGCACCGTC
TCCAGCGCCTGCTTGCCGCCAATATTGCTGGCGATGGCCACCACCTGGTCCGGGGTC
AGACCATGTTGCTCGCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGC
TTGCCGCCGTTATTGCTGGCGATGGCCACCACCTGGTCCGGGGTCAGACCATGTTGC
TCGCACAGCACCGGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGCCAATA
TTGCTGGCGATGGCCACCACCTGGTCCGGGGTCAGACCATGTTGCTCGCACAGCACC
GGCAACAGCCGCTGCACCGTCTCCAGCGCCTGCTTGCCGCCGCCATTGCTGGCGATG
GCCACCACCTGGTCCAGGGTCAGGCCATGTTGCTCGCACAGCACCGGCAACAGCCG
CTGCACCGTCTCCAGCGCCTGCTTGCCGCCAATATTGCTGGCGATGGCCACCACCTG
GTCCGGGGTCAGGTTCAGGGGAGCGCCCGTCAGTGCATTGCGCCATGCATGCACTG
CCTCCACCGCGGTCACGCCGCCACGTTTTGCAATCTTGAGAAGTTGACCTGTGTCCA
ACTGTAACGGTGGACCTCTCAACTCTCCCGACACCGTGAGCAATGCTTCCAGGGCGC
GTGCGCCGGACCACTGTTTGCCGCCGCCAACGATGTCTTCGTGTGTCGCCTCCGGCA
ACGCCGCGATCATGGCCTGGTACTTGACAGCGACGGTCCCTAAGGCTGCCGGGTGTT
GGCTGAGCGCAACGATGTGCGCACGTGTAAACCCATGGCCGATCAGTGCCTCGTGG
TGCTGTGCTACTGTCGAACGCGCCTTCAGTTTGATCTTCTCTTGCTGCTGCTGGCTGT
AGCCGAGCGTGCATAGATCGACCTGCCCGGCCGGCGAAGCGTCAGAGGTGTGCGCA
GCACGCCGTCGCGGCGCCGGCTTGGCGCGCGGCGGTCGCGCGGCCGTCACAGCGAC
CTGCACGGTGGCTTGCGGGTCATCGGCTGCACGCAGACCCGATTGCACCTCATCCCC
CTCTCCTGAGGCAGCCTCTGTATGAGGAGCGCCGAAGGCAGACATCGAATTAAAAA
GCGATGTATCAAGAAGCGACGGATCGAACTGACGGAGCAGATCGCTGAAACTGCCC
GCTGAGAACGCAGGCAAGGGTGCAGGGGGAGACGGGAGACGGGTTCGGGACATCG
TCCGTCGAGCGGGCAAGCCATCCAGGGGGCTGCCGGCAGGCGGAGCCCCCCCACGA
TCTGCAGTCGGCTGCGGCTGAACCCTATCCGGCTGGGGTCCGGCCAGAAGTTCGTGG
GCAGGACTTGGCGTGCGCGGACGAATGGGATCCATCAGGCATACCTGTTTTATAGA
AAAGATGGTCGCCCAAGACTAAAAGATCTCTGGTCCCTCCACTCCTTGCCGAATCCG
CACGGCAGCACACAACTTCGGGGGATCCTGGCAACCCTCGACGACCGGGGGCGAGG
CGCAGACATGAAGGCGGTCCTGGACCCGTCTGGAGGGCCGGCAAGAAGTTGTGAGA
GATTTCTTTTCTGACAGCAAGTTAGCTCACTTTCGGCTGTTTTTTACACGAATCCCTG
CCGACCCTCATCCTGCCCTGCCAGGATGCTCAACGTGGATGCACGCGTGGCCGGACT
GGAGCTGGGGTGACGGAAATTTCGGCGGTTCCGGCTTACCCCCCCTTTATCGCCGTC
GATAGCCGCATCGCTCCGATATCGGTGCCGCAGTCCGGACCCAGCGGGTCGCAGAT
TGCACTGCATCGCGGCACGCTTGTGATGGGAGCCGACTTGTACATACCGCCTTATCG
GCGCCTTCGCCAACGCGCAGATCGCTTCCGGCTTGGCCGTCGTGCTGCGTCGCTGAC
GAGCATGATCCAGCGCTCCACCCATCACCAGGGACGCAAATGATGAATATCGGAGC
CAAGCTCTTGCTTGCCACTTGCATTATCGCAAGCACGGCCAATGCAGCGGAAACCGG
GCAAACTGCGGCGACCGCATCGGCGGGCCGCACTGCGGCGATCTCGTCCTGGGAAA
CGACGATCCGGCAGGAAGCTCCCACTATGGAGGGATGCTTCCGTTCGTCGTTCCCCA

ACAATGGCTGGCAGGCAACCGCCTGCGTGCCCTCGCAAAAGCTGCTTTCGAGGCCG
CCGGCCATCACTGTCCACAAGGATGGCGCCATCACCGCGCGCACCGATGCGGCGCT
GACCGCCGGTAACGGTGTCGACTACGTTGCAGGCACACGCCAGCTTATCCAATCGA
CGTTGGGAACATTCCCGAGCGTGACTGGCGTGGCCACTGGCGTGGCCGACTATTCGC
TACAGATCAACACCAATCTCGATAGGAACGCTGTGACCTGCGCACAGTTCGGATATT
CCTCGTGCTGGACCTGGCAGCAGTTCTTCTACTCCACCGACTACGACCGCGACCCGA
TCAACGGCCGCACGCCGGTCATCTTTATCGAGAACTGGTTCTATGCCGGCAATGCCG
AGGAATTTGAGGCAAAGGGCTGCCCGTCTGGTTGGAATTCCTACTCCACCAGCGATT
ACAACGCGTGCTATAGCAACAGCAATGGGGTGATTGTGCCGCTGGTTCCAATCTCGC
AGATCGGCAGCATCAAGATGAGCGCTTCGGCGACCGCGGGCGGCGTGGACACCGTG
ACGTTCTCGATCGACGGCAGGGCCTACAGCGTCAGTCAAAGCGCCAACACTCTGAA
CATCAACAGGATTTGGCGGCGCTCCGAGTTCAATATATTCGGCAACGGCAGTGACCA
CCCGCTGGTGTTGTTCAATTCGGGCTCGCATGTCACCGTGAACGTGGCCGTCAAAGA
TGGAACCTCCAATGCTCCGAGGTGCCTGGGTCCCAACGCCGGTTATTCGGGGCAACA
GAACAACCTCACGCTCGGCAAATGCACCGCGTCTGGTGGTGCCTCGCCTTCCATCAC
CTTTACCGAAAGCAACTGACGCCCGGGCCTTTTCGTTGCCGTGGCTACCGGGCGCTG
TCTCAGCGTGAGTTGGCGCGTGCGCAATGCCCAATCTGCGTGCGCAACATCACTGAA
ACCTTCAAACAGAACATCCCGACGCTGCACAGTTCCGGCTTACATCCCCAATAAGCA
GGCAGGGTGCGCAGCTTGACTTAAGCGGTCGCGCTCGGTTGCCCGTAGCTTTTTCGG
GGGCATTGCCGGCGATCACTCGCTGTGTTTGGCCCCGGCCACGCTGTGCCCTGGTCC
TTTCCGAAGCTCGATCAACGGCCACAGGGCAATTGCCTTGCCAACGTCCGACTAGTA
ACCCCCTGCTGACCCGCTTTTGGGGTGAGGGAGCCGCCGGCTCTGCCATCGGAGGTG
CTTAGAGTCCGCGGGGCAGTGAGGGCAGAGCCGGTGGGAGAGGGGCGCTTGCGCCC
CTCTCCGCGGGGTCGAAGTTGCAACCGTAGCGCCATAGGCGCGCAGGGAGGACTAG
GGGCGCGCAGCCCCCGCAACTGGATAAAGCTCTTAACGCTGTCGGCGTCGCGTAGC
GGCGGCGACACCGATTTGCTGTTGATTTTCCACTGACTGTGCTTTCAACCAAAGGTT
GACCTGGACGCGCTGCTCGACGGTCTGGACCTGGAGTGACGGAAATTTCGGCGGTTT
GTGGTTCAACCCCCTATTAGCTCTTTTTGTGCGCGTAGCGCACGGCGACCGCTCTTCA
CCGCAGACCACCCGAAGCCGGCAGCCGTTCCAGGGGCATGCCCTGCTTCGAAGTCT
GGTTAATCACTCCCTGCTGACCCGCTGTTGTTCTTTGCGGGAGGGAACCGCCGCCAC
TGCGCGCTCATCGGGTTTGGTTTCCCGATCTGGCTGATGCGCAGTGGCGGTGGGAGG
GGCGCAGCCCCTCTATCGAGCTTCTAACGTTGTCGGCACCGCGAAGCGGCGGCGAC
GCAGATTTGCTGTTGAGCGTTCTACTGACTGTGATTTCAACCAGAGGTTGACTGGTT
GGCCGGTCAATTAAGGTGCACAAATAGCCTTACTTCCCCACAGGCTTATCCCCCGAA
GCGGCCGGGGACTGAGGGTGATGGGTAGCGTTGAGGGCGTTGGGGCAGTGATCGAG
GCCGCCGTAGGTGGGGCGGCTCGCCTTCTCGCGGCAGATTTGCCGCCGTGTGCGCCT
GTTTTTGCGGCATCGGGCAGGGTCATGCCTGCCGGGACGGCTATTTGAACAGGGCGA
GGATGTCGGAGATGTGCTGCACGGCGCTCTGTACGCCGCCTGGCGGGTCGCCCGTG
GCTCGGTGCGCTGCAGGGCGGGCAGCGCCCGTGCGCGGCGTAGTGCCCGCGTCCTG
GGGCGATTGCCGGCGGGCGGGCGCCTGGGCGAAGGACGCAACACCGGTCCCACGCG
CCTGGTCGGCGTGCTTGCGGTCGCGCTTGAGTAACCAGCTGGTGCCAGCCATGTCGC
ACAGCCGTTGGAAGTTGAGCCGGCGCACGCCCGGCTCGCTGCTCCACCCGTCGCGG
CCTTCGGCGTCTTTGGTGTGCCGGCGCACCTGCTTGGTGGGGAATGCCCAGCCAAGG
CTCACCATCATGTTGATCGAGCGTTTGGCACGTTCCAGCTCGGCGCCGGTCTGTCCG
CCATAGGCGAGCTGGGCGAGGTCCTTGAGGGTGTAGCGGTGCCAACGGCCGCCGCC
ATGGCGCGGCTTGCCCACAAACCCGGTGCGCACGTCGGCCAGGTAGAGCAGGGCAA

CGGACATGGCCAGCAGTGATTCGGCGCCATCGGCGCGCAGCGCCCGCACCCGCGTC
AGTGCCTTGTCCAGGGCATAGCCGGCCACCATGACGTTGCGGCTAATCCGCATCCAC
AGCGAGCCGCGCCGTTGTTCCAGGCTGCGCACAAGCCCGTCGCGGGTCGCGTCCAG
GACACCCGGCCGGCGCTTCTTGCCGGCCAGCGGCTTGGGCTGACGCCAGTGCACAC
GCTTGTGCGACCGACCGCTGGGAGCGGTCGGTTCGGGGCCTGGCGTGTTGGAGCCTG
GATCTACTGCTGGCATGTCCCGCGCCGATGAACGGCAGCGGGGCAAGGTGGTCGGC
GTTTCGGCCTTTGCATCCTTTTCCCTCTTGCACGGCTTCGCCGGCGGGATTAGGATGG
GGGTACGTTAATGCCCCATTGGTCTGGTTTTGTCCCGCCAGTCCAGTGATGAAAAAA
CCGCCTCGCCAAAGGCGGTTTTTTCGTTTTCGGGGGTCTACTGCGCTAGCAGCTGGC
CCACCCGCGTTGTCATATCGCTGCCAAGTATCTGCGCTCGGGCCTTAAGATGCCAGC
AGCGCCATCATCTGCGCTCCGTTGAGCGTAGCTGTAGACGCTCACGCGCGATCGATC
AGTGCAGCGTCGCCAATGCTGCCGGCATGATGTGCCGCCGCACTTCGCGCAGCAGC
CGCGCGAAGCGTTCGGCCGTCTCCGTCCCCTGCATTGAGCCAGCGAAGCCGGGATC
GCTCGCAATGCGCCGGAGATCGCCGCTCTGCATCGCAAGGTAGTCGAGTGCGTAGG
CAGCAGCGATCAAGTGGTAATCCCCCCGCCCATTAGCAGACGCCAGAAGTGGAATT
TTCTCGTATCCTTTCTCGAGGAGGTTCCATGAAGAAGTCCCGCTTTACCGACAGCCA
GATCATCGCCGTGCTCAAGCAGGCCCAGGCCGGTGCGCCCGTGCCGGAGCTGTGCC
GCGAGCACGGCATCAGCTCGGCCACGTTCTACAAGTGGCGCAGCAAGTTCGGCGGC
ATGGACGTGTCCATGGTCGCGCGCATGAAGGAGCTGGAGGAGGAGAACCGCCGGCT
CAAGAAGATGTACGCCGAGGCGCAGCTCAGTACCGACCTGCTGAAGGAAGCGCTCG
CAAAAAAATGGTGAGGCCATCTCAGCGACGCGAGATGGCCCAATCGGCAGTCACGA
GCGGGCGTACGAACATCCGCCACGCCTGCCAGGCCTTCGCGGTGAGCGAGACCTGC
TTCCGCTACCAGGCCAAGGCCAGCGAGCAGAACGCCCGGATCGCCGACTGGCTGGT
CCGCCTGACGACCGCCCATCGCGACTGGGGCTTTGGCCTGTGCTACCTGTACCTGCG
CAACGTGAAGGGCTTTGGCTGGAATCACAAGCGGGTCTACCGGATCTACCGCGAGC
TGGAGTTGAACCTGCGGATCAAGCCGAGGAAGCGGCTGGTGCGTGAGCGGCCCGAG
CCTCTGGCGGTGCCGGAGGCTATCAACCAGGTCTGGTCGATGGACTTCATGCACGAC
CAGTTGGCCGACGGCCGCAGCTTCCGGCTGTTCAATGTGCTCGACGACTTCAATCGC
GAGGGGCTGGGGATCGAGGTGGATCTGTCGCTGCCGTCAGCCCGTGTGATCCGATC
GCTGGAGCAGATCATCGAGTGGCGCGGCAAGCCCGCCGTGATCCGCTGCGACAACG
GCCCTGAATACATCAGTGGCGCGTTGCTGTCCTGGGCGCAGCGGCATGGTATCCGGG
TCGAGCACATCCAGCCGGGCAAGCCACAGCAGAACGCCTACGTTGAACGCTACAAC
CGCACCATCCGCTACGCCTGGCTCGCCCAAACCCTGTTCGACACCATCGACCAGGTG
CAGGACAAAGCCACCCGCTGGCTATGGACGTACAACCACGAGCGTCCGAATATGGC
GCTCGGCGGCATCACGCCTGCCATGAAGTTGGCGATGGCCGCTTAGCTCCACTTCTG
GCGACCGCTAAAAGTGGGGGGATTACCCTGTTCAGCATCCAGGCGGCACGACAGGA
GGCCGGGGCAGCGAGTCCAGTTGGTTGACCGCCTCCAGAACTGATTCCACTGCGTCT
TGGTGCGCTTGCGATTGAGGATTCATTCCATAGGTCTCTCTAATCATCCGGGATGCG
ATATCCCTCTGTTGGCTTTCGGGAACATGGACCAGTACGTTGCGTACCTTATCGGCG
AAATTCGATGAAATTACTTGTGGTTCTGCATATTTCTCCCACGCGGTAGGTTTCTTTT
GGTTATATCTAAGTGTTTTTGCAGTATGTTGATCTTCGTCCGGATGTACTAACGGAAT
TAACCCTTCAAGACGGTTTTTCAGAGATTCGGCGCTGCTTTTTTCCATCGACCAACGA
CCTTCCTCGGCTGGCTTTCCGCTCCATGCGCTGTCTTTCAATCGAACAGCCGGACCGT
TGGCCCAGGCATCCATGACTACCGTTGACCTCGGGTCGCGCCTTCCGCGCCTCAGTT
CGTTCCAGATATGCCCCGCTTTGACTGGGATCTCATTCCCGTCGTCGGTTATCTCTTT
CATTGGTACGGAAGCTTCAGCAGTCACACTCCTCTCGTTGACCGCCATATGGGGAGC

ATGAAGGATGGCGCAAAGGGGAGCGGACTGATCGCAGACCGCTGCTCCCATGACGA
CGCCCACAGCAAGCTTGCTGTGGGCGTCCGCTGCCGATTTTGCAACCGACGATCCGA
TGCCATTATGGCCTTGGGAGGGAATGTCGTCCGCTTTAACGTTTCCTCGGCCATGCTT
GAGGATCAGGCGGACGTCGTTAACAGTTTGCCCGGCTGTCGCGAGTCTTGTCGGACT
TGTGACTTCGTCGCCGACGACATATGCTCCCATGTGGTAAGCCACCTGCTGGATCTG
ATGCTGTTCTAGAGATTCATCTCCGCACCTTCGCAAACGGTCGGGACGCGTCCCTAG
GGAGACTCTCAGCGCTGATAGCTCCGGCGATTGCGGCGACATGACGCTTTCAAAGCT
GGAAACAGTGGAGGCTCTCGGCAACTCTTCCCCGTCATATGCGGCTGGTTCAGGTGA
CCCTGCAACGCTCTGCTTCGAAACGCATAGACCCATTTTGATCTCCGACAGTGACCT
GAGAATCAAGTCTGGCGAGGAACTCTCTGGATGTCGTTGGTTCACCGAAGTCATGTT
TGGGACCACGAAAGATTGGCGCTGGCTCCCCACTTACAATGCCATTGCGGGCGTACT
ACCCAGTCCACAGGGGAAGTGTGCAGCACATCAACCTGTATGGCCAATTCGACGTC
GCCATGAATACCCGGCTCGATCTGCCGTGATCCGAAATGGCGGATGGTGACACGAC
TGTTGGGTCTAACCCCTACCAGGCGTTTCGTTCCAGCGGCAATCTCCAAGTTTTCCGG
CCTCAGCTCGACGAGCTGAGCCGCCTGCATCGGCAATGGCTGTCCGACCTTGGTGCC
TTCAAAGAATCGCTGCGTACACATGGTGCTGAACCCAAAGTGCTGGAATACGTGAA
CGAGGCTTTCGGCCGTCTGGCCGAGCGTAATCAAGCAGCTTGCTGGGTGACCGGCA
AAGTAGCCGACGGAAATTTCGGCGGTTTGTGGTTCAACCCCTTGATTAGCCCTGATC
GAGGGTCGGCAGGGATTCGTGTAAAAAACAGCCAAAAGTGAGCTAACTTGCTGTCA
GAAAAGAAATCTCTCACAACTTCTTGCCGGCCCTCCAGACGGGTTCAGGACCGCCTT
CATGTCTGCGCCTCGCCCCCGGTCGTCGAGGGTTGCCAGGATCCCCCGAAGTTGTGT
GCTGCCGTGCGGATTCGGCAAGGAGTGGAGGGACCAGAGATCTTTTAGTCTTGGGC
GACCATCTTTTCTATAAAGAGGTATGCCTGATGGATCCCATTCGTCCGCGCACGTCA
AGTCCTGCCGACGAACTTTTGGCCGGACCCCAGCCGGATAGGGTTCAGCCGCAGCC
GACTGCAGATCGTGGGGGGGCTCCGCCTGCTGGCAGCCCCCTGGATGGCTTGCCCGC
TCGACGGACGATGTCCCGAACCCGTCTCCCGTCTCCCCCTGCCCCCTTGCCTGCGTTC
TCAGCGGGCAGTTTCAGCGATCTGCTCCGTCAGTTCGATCCGTCGCTTCTTGATACAT
CGCTTTTTGATTCGATGTCTGCCTTCGGCGCGCCTCATACAGAGGCTGCCTCAGGAG
AGGGGGATGAGGTGCAATCGGGTCTGCGTGCAGCCGATGACCCGCACCCCACCGTG
CAGGTCGCTGTGACGGCCGCGCGACCGCCGCGCGCCAAGCCGGCGCCGCGACGGCG
TGCTGCGCACACCTCTGACGCTTCGCCGGCCGGGCAGGTCGATCTATGCACGCTCGG
CTACAGCCAGCAGCAGCAAGAGAAGATCAAACTGAAGGCGCGTTCGACAGTAGCAC
AGCACCACGAGGCACTGATCGGCCATGGGTTTACACGTGCGCACATCGTTGCGCTCA
GCCAACACCCGGCAGCCTTAGGGACCGTCGCTGTCAAGTACCAGGCCATGATCGCG
GCGTTGCCGGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGG
CGCACGCGCCCTGGAAGCATTGCTCACGGTGTCGGGAGAGTTGAGAGGTCCACCGT
TACAGTTGGACACAGGCCAACTTCTCAAGATTGCAAAACGTGGCGGCGTGACCGCG
GTAGAGGCAGTGCATGCATGGCGCAATGCACTGACGGGCGCTCCCCTGAACCTGAC
CCCGGACCAGGTGGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGA
CGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAG
GTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCG
GCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCA
TCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCG
GTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAA
TGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCG
AGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGC

AAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGG
CCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGG
AGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGAC
CAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCA
GCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGG
CCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTG
CCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAG
CCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCACGATGGC
GGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACA
TGGCCTGATCCCGGCGCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGG
CGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACC
CCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGAC
GGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGATCCCGGCGCAGG
TGGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTG
TTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGC
CAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGC
TGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCACGAT
GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCA
ACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGC
AGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTG
ACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGA
GACGGTGCAGCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACC
AGGTGGTGGCCATCGCCAGCAATGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCA
GCGGCTGTTGCCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGG
CCATCGCCAGCCACGATGGCGGCAAGCCGGCGCTGGAGACGGTGCAGCGGCTGTTG
CCGGTGCTGTGCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAG
CAATGGCGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGT
GCGAGCAACATGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATGGCGGC
GGCAAGCAGGCGCTGGAGAGCATTTTTGCCCAGTTATCTCGCCCTGATCAGGCGTTG
GCCGCGTTGACCAACGACCACCTCGTCGCCTTGGCCTGCCTCGGCGGGCGTCCTGCG
CTGGAGGCAGTGAAAAAGGGATTGCCGCACGCGCCGACCTTGATCAAAAGAACCAA
TCGCCATCTTCCCGAACGCACGTCCCATCGCGTTGCCGACCACGCGCAAGTGGCTCG
CGTGTTGGGTTTTTTCCAGTGCCACTCCCACCCAGCGCAAGCATTTGATGAAGCCAT
GACGCAGTTCGGGATGAGCAGGCACGGGTTGTTACAGCTATTTCGCAGAGCCGGCG
TCACCGAACTCGAGGCCCACAGTGGAACGCTCCCCCCAGCCTCGCAGCGTTGGCAC
CGTATCCTCCAGGCATCAGGGATGAAAAGGGCCGAACCGTCCGGTGCTTCCGCTCA
AACGCCGGACCAGGCGTCTTTGCATGCATTCGCCGATGCGCTGGAGCGTGAGCTGG
ATGCGCCCAGCCCAATAGACCGGGCGGGCCAGGCGCTGGCAAGCAGCAGCCGTAAA
CGGTCCCGATCGGAGAGTTCTGTCACCGGCTCCTTCGCACAGCAAGCTGTCGAGGTG
CGCGTTCCCGAACAGCGCGATGCGCTGCATTTCCTCCCCCTCAGCTGGGGTGTAAAA
CGCCCGCGTACCAGGATCGGGGGCGGCCTCCCGGATCCTGGTACGCCCATGGACGC
CGACCTGGCACCGTCCAGCACCGTGATGTGGGAACAAGATGCTGACCCCTTCGCAG
GGGCAGCGGATGATTTTCCGGCATTCAACGAAGAGGAGATGGCATGGTTGATGGAG
CTATTTCCTCAGTGAGGCTTAATCGGAGATTACCTGAGCGTTGGCAGTGATGGGGGT
AATAAACCTCTCACTCAATGAGTTAGCTCACTTTTGGCTGTTTTTTACACGAATCCCT

GACGACCCTCTGATCTCTCAGCTATGAAAATTGCGCGATCCGCAGAGCTGGCTCCGT
GGCGGCACACTGCCTGCCAGCAGCTTCTTGGTCGCCTTGACCTTGTTATCGGCTCTTT
TGGTGCGCGTAGCGCACGGCGACCGCTCTTAAACGCAGACCACCCGAAGCCGGCAG
CGGGTCCAGGGTCATGCCCTGCCTCGAAGTCCGGTTAATCACTCCCTGCTGACCCGC
TGTTGCTCTTTGCGGGAGGGAACCGCCGCCACTGCGCTCTCATCGGGTTTGGTTCCC
CGATCTGTCTGATGCGCAGTGGCGGTGGGAGGGGCGCAGCCCCTCTATCGAGCTTCT
AACGTTGTCGGCGTCGCGTAGCGGCGGCGACGCTGATTTGCTGTTGAGCTCGTACTG
ACTGTGCTTTCAACCAAAGGTTGACTGGATGGCCTGCCAAATAAACAGGACAAATA
ATGCTATTTCTCCACAGGCTTATCCCCCGCCACGGTGGATAACTTGGCGCGCGGCAT
GGCGTTGAGGGGTTCGGCTCGGTGATCGAGGTCGCCGCAGGCGGTGCGGCTCGCCTT
TTCGCGGCAGATTTGCCGGCGTGCATGTCTGTTTTTGCGGCATCGGGCAGCGGCGAG
CCTGCCGGCGATGGCTAGTTGAACAGGGCGAGGATGTCGGAAATGTGCCGCACGTC
GCTCTGTGCGTCGCCTGGCGGGTCGCCCGTGGCGCGGTGCGCCGCAGGACGGGCAT
CGCCCTTGCGCGGCGTGGTGCGCGGCTCCTGGGCCGGCTGACGGCGTGCTTGCGCCT
GCGTGAAGGACGCAACGCCTGTGCCGCGTGCCTGGTCGGCGTGCTTGCGATCGCGCT
TGAGCAACCTGCTGGTCCCCACCATCTTGCACAGCCGCTGGAAGTTGAGCCGGCGTA
CGCCCGCCTCGCTGCACCACAGGTCACCGCCATCGGCGTCTTTGGTGTGCCGCCGCA
CCTGTTTGGTGGGGTAATCCCCCCGCCCATTAGCAGACGCCAGAAGTGGAATTTTCT
CGTATCCTTTCTCGAGGAGGTTCCATGAAGAAGTCCCGCTTTACCGACAGCCAGATC
ATCGCCGTGCTCAAGCAGGCCCAGGCCGGTGCGCCCGTGCCGGAGCTGTGCCGCGA
GCACGGCATCAGCTCGGCCACGTTCTACAAGTGGCGCAGCAAGTTCGGCGGCATGG
ACGTGTCCATGGTCGCGCGCATGAAGGAGCTGGAGGAGGAGAACCGCCGGCTCAAG
AAGATGTACGCCGAGGCGCAGCTCAGTACCGACCTGCTGAAGGAAGCGCTCGCAAA
AAAATGGTGAGGCCATCTCAGCGACGCGAGATGGCCCAATCGGCAGTCACGAGCGG
GCGTACGAACATCCGCCACGCCTGCCAGGCCTTCGCGGTGAGCGAGACCTGCTTCCG
CTACCAGGCCAAGGCCAGCGAGCAGAACGCCCGGATCGCCGACTGGCTGGTCCGCC
TGACGACCGCCCATCACGACTGGGGCTTTGGCCTGTGCTACCTGTACCTGCGCAACG
TGAAGGGCTTTGGCTGGAATCACAAGCGGGTCTACCGGATCTACCGCGAGCTGGAG
TTGAACCTGCGGATCAAGCCGAGGAAGCGGCTGGTGCGTGAGCGGCCCGAGCCTCT
GGCGGTGCCGGAGGCTATCAACCAGGTCTGGTCGATGGACTTCATGCACGACCAGTT
GGCCGACGGCCGCAGCTTCCGGCTGTTCAATGTGCTCGACGACTTCAATCGCGAGGG
GCTGGGGATCGAGGTGGATCTGTCGCTGCCGTCAGCCCGTGTGATCCGATCGCTGGA
GCAGATCATCGAGTGGCGCGGCAAGCCCGCCGTGATCCGCTGCGACAACGGCCCTG
AATACATCAGTGGCGCGTTGCTGTCCTGGGCGCAGCGGCATGGTATCCGGGTCGAGC
ACATCCAGCCGGGCAAGCCACAGCAGAACGCCTACGTTGAACGCTACAACCGCACC
ATCCGCTACGCCTGGCTCGCCCAAACCCTGTTCGACACCATCGACCAGGTGCAGGAC
AAAGCCACCCGCTGGCTATGGACGTACAACCACGAGCGTCCGAATATGGCGCTCGG
CGGCATCACGCCTGCCATGAAGTTGGCGATGGCCGCTTAGCTCCACTTCTGGCGACC
GCTAAAAGTGGGGGGATTACCCTTCCAGGGCGCGCACCAGTCCATGGCGAGTAGCG
TCCAGTACGCCCGGCCGGCGCTTCTTGCCGGCCAGCGGCTTGGGCTGACGCCAGTGC
ACACGCTTGTGCGGCCGACCGCTGGGAGCGGTCGGTTCGGGGCCTTGAGCGTGTGG
GCCTGTGTCTACTGCTGGCATGTCCCGCGCCGATGAACGGCAGCGGGGCAAGGTGG
TCGGCGTTTTGGCCTTTGCATCCTTTTCCCTCTTGCACGGCTTCGCCGGCGGGATTAG
GATGGAGCTATCTGACTAGCCCATTGGTCTGGTTTTTCCCGCCAGCCCAGTGATGAA
AAAGCCGCCCTGGCAGGGGCGGTTTTTTCGTTTTCGGCATCTGCTCATCACACCAGC
GGGCTGCCTCGTTGTTTTGGTACGTCTGGTGCTGAATCTACCTTGTTGCAGGCGCGA

GCGTACTCGCAGGAAAGGTGGTGGACGGTACGCGGTCATCAAAACCGGTGGGCAGA
AGCGGGAACCGGCAGGCCTTGCGGCCTCCGCGCACCGCCCACCATAGAACTGGCGG
GCCAAATGAAAGCGCCAGCATCGGACGATGGGCGCTTGCGCGCTCCTGCCACCTTT
ATCGGTTTGATGTCCCGAATCGCTGATTGTGCAGCGATAGCTCAATCCTGTCGCCAA
CCGGCGAACGTGTCAACCGCCTCTTAAACCGTCGCCTCCTCGACTGCCGCCGCTGCT
TTGGCTTGCAGGTTGAGGAGCGTGGCGATGACGGCGTGCAAGTCGACGTACTGGGT
GTGACGTTGGAAGGCAGCTGCCGGCGATTCCGGATCACTTGGATCGGTCTGCAAATG
CTCTTGGAACTGCTCGGCCAGCAAACCAGCCATGAAAGCGACGTGCGCGTGGTTGG
CCTGCAACACGCTGCGGCCAATGGCCTCGAAGTTGTGCAGGCTGTCAGATAGCCAC
ATGACATCGCGCGCCACCGTGCGACTGGCAAGGGTGGTATCGGCACGGTAGGGACT
GTATTGCCGGATGGCCAACGCGGTGTTGCCAGTGAAAAGTACCAAGCTGGCCAACT
CATCGGTGTAATTGATGCCCATGGTTTTTGGCTTCCTCGGGAAAGCAGGGAGGGTTT
ATGAAAGTTCTGTGCTGACAGCGAGTTGGGGTTCAGCCGACATACGTGTAAAAATC
ACTCACAGAATCCAATTTCTCTCGTAAAAACATGACTTTGCCAGATAAAAACGGGTT
CCACCGGCTCACCCACTATCTGGTAATCCCCCCAGCTACGAGTGGTTCTATTAGCAA
ACCACATTACCGTTCGCTCTGCGTCACGGGATGCGTTGGAATTCACACGTGGCCACC
TGCAGGTGGCCACGAAGCGCAACAGCTCCTGCAGCTGTGCATTGGCGCGCGACTTTC
CTGAAGCTGGTCGCGGCGCCGCAAGCCAGCAACGTGACCGTTATCCAATGTTGACA
ACTGGGCTCGTATCGGCTTATAATTACCTCAACACGGCCCCTCTCGCCATGCGCAAG
TAAGATCAGCGCATGGTGCTGAGGGGTTCTTCATTTTTCCTACCCCCAAACTGCTTGC
CAATTCGATGGCGCCCCCATGTCTTCGACTGTCGGGACTCTCTGGACAAGATTAGGA
ATGAAGTTCGGAAAGGTATTTAGATGTGCCTTCATTCTATCCGACCAAGTTGATCCA
GGATTAATAACGCTCATCAATTGTTGAACCATCTTCAATAACATGAAGCAGCGAGCA
GTAGCTGTGGAATTTGCAGGGAAATGAACGGCCCAAGCAGCCTCATCTGATCCAAG
AAGCTTGGGCTGGTCAACAATATTTCTGTTCCACAACCTACTGTGATGAGCGCAAAC
GTTTCTTAGATAATTAAGGCTACGCAGCCAACTGGCGAAAGTTCGACCATTTGAAAC
GCCATACTTTATAGAGATTGCATCTTGCTCAGCCTCACGCATCCCATCGTATAGCGT
AGACATCGCACCGAAATCCCAAACTTCGCAGACGACCCAAATTGCTAATGGCAGGC
CATATTTTTCCTTATTATGCCGAACAAATTCCTCCTTGGAACGAGCAACAAGTTGGC
CATTTTTTCCAAGCCAAGCATGATGTGGAGTAACACCCTTCTGAGCATCATATGACT
GGGTGAACTTCTCGTGCAAGAAATTGGGGTTAGCATATGCAAAGGTGTCGAGCTTTC
CCAGGGTATGAGAAATATCAACCCGTAGTGCAATTTCTATACGCTCCAACGCATCTC
CGACAAGATGACGCAGTCGCTTATCAAAAACGTAAAGGTCTACCACTCCGCCGAAC
GAGGTATCTGGCACAAACCAGTCAGTAGCTAGGATTTTGACCTTAGATTTCTTGGGG
CGCTCAGCCTTGTCGTTCAACTCCATATATTCCGCACGGAGCCGAAATGGATACCAG
TATCCGCTGAGGCGGTAGTAGCCAATTTTCTCCAAGTAAAGAAGCGCTTTGGGCTCA
TCAGTTACAGCCATATTTCGATCTTTGAGCTGATCGAGCTGCTCTTGGTAACTTCTCC
ATGCCTTCTGATAGCCCATTACCGCCTCCTTAATCAATCACACAGCACTGTCTGCAA
GCTTAAATTTACTACTACCGTGAAACTCGATCAAATTCTCGGGCTCAGAGGCGAACC
ATATGTAGGAGCCCCATGCTATTTCCGACACACCTTTCCTGAAAACACTTTTCCCAC
GATCTTGATAAGCGGTTAAAAAGCAAGACTTTCCTCATTGAAGCCAGCTTCCAGAGC
CAGCGACATAAGTGCAACTTTACGCTCCCTGTGTCATTCTCAGAAGACGACTGCACA
GAATGTCAGGAGTCGGCTGCACTAGTGGATTGCAGCGCGGCGTATAGAGCCGTCTT
GCTCACCTTGAGCCGTCTAGCAGCCTCTCGGACATTAAGCCCGTTGGCGATGTGCTC
CCGCGCTCGCTGCAACTTGTCGGCGGTGACAACCGGCTTTCGCCCGCCCTTCCGCCC
GCGTGCGGCGGCGGCGGTCAGCCCGGCCTTGGTTCGCTCGCGGATCAAGTCGCGCTC

GAACTGGCCCAGTGCGCCGAAGACGTGGAAGATGAGCCGTCCGCCCGGCGTGGTGG
TGTCGATGCTTTCGGTCAGCGACCGGAAGCCGACGCCACGCGCTTCCAGCGCGGCG
ATTGTCTCGATGAGGTGTGGCATCGAGCGCCCGAGCCGATCCAGACGCCAGACGGC
CAGCACGTCGCCGGCGCGCAGGTAGGCCAGCGCATCAGCCAAGCCGGGCCTGTCGG
CCTTGGCCCCAGAAGCCGTGTCCTCAAAAACGCGCTCGCAGCCTGCCTTGCGTAGCG
CATCCGTCTGCAAAGCGGTGTCCTGTTCCGTCGTCGATACCCGCGCATAGCCGATCA
GTACCATTACCGCCCCTCTTGTCCGTCTTTCCGTCCGCCTATCATAATGTCCGAAAAG
CCATTGAACAAGCACATTTCCGGACAATGTCCGGCATGGCCGCCTAATGATCGTTTG
ACGGACATCCTTGATTAGGCAGCTGGGTCTTGTATGCCATCAGTTCCTCACCCTCAA
TGTCGAAGCGCGGCTGAGCCATTTGGCCAGCACCTCATTCGATGAGGATATCCACGC
CGATCATGTCGATCAGGCGCATTGCTGTTCAGCATCCAGGCGGCACGACAGGAGGC
CGGGGCAGCGAGTCCAGTTGGTTGACCGCCTCCAGAACTGATTCCACTGCGTCTTGG
TGCGCTTGCGATTGAGGATTCATTCCATAGGTCTCTCTAATCATCCGGGATGCGATA
TCCCTCTGTTGGCTTTCGGGAACATGGACCAGTACGTTGCGTACCTTATCGGCGAAA
TTCGATGAAATTACTTGTGGTTCTGCATATTTCTCCCACGCGGTAGGTTTCTTTTGGT
TATATCTAAGTGTTTTTGCAGTATGTTGATCTTCGTCCGGATGTACTAACGGAATTAA
CCCTTCAAGACGGTTTTTCAGAGATTCGGCGCTGCTTTTTTCCATCGACCAACGACCT
TCCTCGGCTGGCTTTCCGCTCCATGCGCTGTCTTTCAATCGAACAGCCGGACCGTTG
GCCCAGGCATCCATGACTACCGTTGACCTCGGGTCGCGCCTTCCGCGCCTCAGTTCG
TTCCAGATATGCCCCGCTTTGACTGGGATCTCATTCCCGTCGTCGGTTATCTCTTTCA
TTGGTACGGAAGCTTCAGCAGTCACACTCCTCTCGTTGACCGCCATATGGGGAGCAT
GAAGGATGGCGCAAAGGGGAGCGGACTGATCGCAGACCGCTGCTCCCATGACGACG
CCCACAGCAAGCTTGCTGTGGGCGTCCGCTGCCGATTTTGCAACCGACGATCCGATG
CCATTATGGCCTTGGGAGGGAATGTCGTCCGCTTTAACGTTTCCTCGGCCATGCTTG
AGGATCAGGCGGACGTCGTTAACAGTTTGCCCGGCTGTCGCGAGTCTTGTCGGACTT
GTGACTTCGTCGCCGACGACATATGCTCCCATGTGGTAAGCCACCTGCTGGATCTGA
TGCTGTTCTAGAGATTCATCTCCGCACCTTCGCAAACGGTCGGGACGCGTCCCTAGG
GAGACTCTCAGCGCTGATAGCTCCGGCGATTGCGGCGACATGACGCTTTCAAAGCTG
GAAACAGTGGAGGCTCTCGGCAACTCTTCCCCGTCATATGCGGCTGGTTCAGGTGAC
CCTGCAACGCTCTGCTTCGAAACGCATAGACCCATTTTGATCTCCGACAGTGACCTG
AGAATCAAGTCTGGCGAGGAACTCTCTGGATGTCGTTGGTTCACCGAAGTCATGTTT
GGGACCACGAAAGATTGGCGCTGGCTCCCCACTTACAATGCCATTGCGGGCGTACT
ACCCAGTCCACAGGGGAAGTGTGCAGCACATCAACCTGTATGGCCAATTCGACGTC
GCCATGAATACCCGGCTCGATCTGCCGTGATCCGAAATGGCGGATGGTGACACGAC
TGTTGGGTCTAACCCCTACCAGGCGTTTCGTTCCAGCGGCAATCTCCAAGTTTTCCGG
CCTCAGCTCGACGAGCTGAGCCGCCTGCATCGGCAATGGCTGTCCGACCTTGGTGCC
TTCAAAGAATCGCTGCGTACACATGGTGCTGAACCCAAAGTGCTGGAATACGTGAA
CGAGGCTTTCGGCCGTCTGGCCGAGCGTAATCAAGCAGCTTGCTGGGTGACCGGCA
AAGTAGCCGACGGAAATTTCGGCGGTTTGTGGTTCAACCCCTAAAACATCCAAATCC
GACATTAACGAAAAATGCAGGTGTCGGTTCTATTGTTTAGCTGCTACCTAGCAGCAA
CCTAGCTACTAGATTGTTCGTAATGAGCAAGGGCCTTTTCCAATATCTCGACCAGCA
GCACACCTTCACGCTGCGCCAGGGCGCGGATGCGAGCATCCCATTCCGGCGTCACCC
GCGTGGCGAACTGCACCGTGCGGTTCGTCTTGCGGGCGCTCCGGCCGTCGATCTTCT
TGGCCGTCCGCGCTTGCGGTGGTGGGTTGGCGGCGGCGACCGGCCTGGTTACCGGC
GCAACCTCCGGCGCGGTCAGATTCGTACTGGCCTCGTCCAGCGTCGGCGGTGCGCCC
AGCTCGTCGCGCCGGCCTTTCAGTTTCTTGAGCATGGATACGTCAGCCATTTGTCACC

CGGTCGATGAGCGATTGAATGAGGGTGTCGGCCCGCTCGTTCAGACCCGCAAAGCG
GGTTTCTGTGATGCTGTGACCGGTATTTTGCGCCTGGCGGTAGGCCGGCCGGTCGCG
TAGGCTCCCGTCCAGGACGGTGTAGCCGGCCTGCTGCAAATAGACCCGCGCCGCAA
CCTCTTCCGCCGGCGTGCCGATGCGGTTGAGGGCGAAGGCCAGGCGGTCGTTCGGG
ACACCGGCCTTCGTCAGTGCGTGAAACTCGCGCACGGCCGGAATGAGGTCATCCCTC
GATGCGCCAGACGGCTGCACGATCAGATCAGCAGCGCGGGCTATGTCGAGCGTCGC
CGAGCTGGTGCGGGCCGGCCCGTCGATGATGAGAAGGTCGTAACCCTTGGCCGCTT
CCAGCGCCTGCGCGGCCGTGGCGAACGCCTCCACGGACACCACGGGTTCAATTCCCT
GGGAGAGCCGCAGGCGATGCCAATCGACAGCCGTTCCCTGCTGCGTATCGAGGTCT
GCAACCTTCACGCGCAATCCGCCGGCGGCCGCTTCCCGTGCAAGGGCACGCGCCAG
GGTGGATTTCCCGACGCCGCCTTTCTGTGAAACGAATGCGACAATAGCCGCCATCGC
TTAACCTCTATCTAGCTGCTAGGTAGCAGCAATTATAATCATTGAGCTAGATGAAAG
GTAGTGGCTAGATAGCTTTTACCCGTTGAAATTGCGCCGCCGCTCGCGCTCTTTAGG
GCTTGGCAGTAATGATGGCTGTGGCAGTCTGGTGGCTATCCAGCCACGCGGCTGTAA
CCGCGCAGCGGTACGTCACCAGCACCAATATCGTCTGGTAGTCATCTGCCAACGCTC
GGGCAAAGATCGGTTCGTGGTGCGCGATGCGGTTCCGCAGCAGGCGCACGCGCTCC
AAATCGCCATAGATAGCATTCCGCAGAGCGGCAACAGGCTTGGCCGGGTCAATGTT
CGGCATGACGCGCCGCAGATTCTGATCCCAAATGCGGGTGTCGTAACGGCCCGTAA
ACATCTTCTGCCAGAAAACGAACTTTAGTTCCGGGATCACCTTACCCGTGGTCTGCG
CTGACCGCCTGGCGCTTTGCAGATCACGGCGCGGGCTGTAGCCTTGCTTGGGATCGG
GCAAGCTGCGCTCGAATGTTGCCGACCACGGCCAGCGCGGCCCATAAAGGGCTTCC
AGCGCATCAGACACAGCGTTTCGGATGACTACCTCACACACATGCAGCGGGGCCAG
GAACGCGCCGGACACTTCAGCGTTCCATGCGTAGAGCAACAGGGCGGCAGGGTCAT
CATCCGATTGCACCCCAGCTGCGGTATCGAAGGTTCCCATGCGAGCTGCTGACAACG
CGGCCCGTACTGCATCAATGGTTGTTTTCTGGCTGTTGTTGATATACCCGCCCGGAA
GCCTTATATTTGAGTCACTAGCCACGGGACTAGCGGGCGCAAGCCTCCCCCCGAGG
ACATACGATAGTACGAAAAACCCGCCAACCTTGGCGGGTTTTTCCTTTCTGGTCATC
TCCGTTGCCCCTGTGCCCTGTGGAACCGCAGCCCCAACATGCGTGTATAGGGTGGTT
CTGGCCTTCTGTGAGACGAATGCGACACTCTAGTAGCTAGTAGAGGATAGAAGCTA
ACTATATTTTAGCCTCTAAATTGCGACGCCGCGTCCGCTCGGCGTCGGCCAAGTCGC
TGGGTCACGTCGCCAGCTGGAACTGGGGCGACGGAATTTTCTGGGCTTCCGGCTTAC
ATCCCCTTCTTGGCCGCCTTGTTATCGGCTCTTTTGGTGCGCGTAGCGCACGACGACC
GCTTTTCACCGCAGACCACCCGGAGCCGGCAGCCGTTCCAGGGTCATGCCCTGCCTC
GAAGTCCGGTTAATCACTCCCTGCTGACCCGCTGTTGCTTTTTGCGGGAGGGAACCG
CCGCCACTGCGTGCTCATCGGGTTTGGTTTCCCGATCTGGCTGATGCGCAGTGGCGG
TGGGAGGGGCGCAGCCCCTCTATCGAGCTTCTAACGTTGTCGGCGTCGCGTAGCGGC
GGCGACACCGATTTGCTGTTGATTTTCTACTGACTGTGCTTTCAACCAAAGGTTGACT
GGATGGCCTGCCAAATAAACAGGACAAATAACGCTATTTCTCCACAGGCTTATCCCC
CGCCACGGTGGATAACTTGGCGCGCGGCATGGCGTTGAGGGGGACGGCTCGGTGGT
CGAGGTCGCCGCAGGCGGTGCGGCTCGCTTTTTCGCGACAGATTTGCCGGCGTGGAC
GCCTGTTTTTGCGGCGTCGGGCAGGGGCTAGCCTGCCGGCGACGGCTAGTTGAACAG
GGCGAGGATGTCGGAAATGTGCCGCACGGCGCTCTGTGCGCCGCCTGGCGGGTCGC
CCGTTGCTCGGTGCGCCGCAGGACGGGCAGCGCCCTTGCGCGGCGCGGTGCGCGGC
TCCTGGGCAGGCTGACGGCGTGCTTGCGCCTGCGTGAAGGACGCAACGCCTGTGCC
GCGTGCCTGGTCGGCGTGCTTGCGATCGCGCTTGAGCAATCTGCTCGTCCCCACCAT
GTCGCACAGCCGCTGGAAGTTGAGCCGGCGCACGCCCGCCTCGCTGCACCACAGGT

CGCCGCCCTCGGCGTCTTTGGTGTGCCGCCGCACCTGTTTGGTCGGAAATGCCCAGC
CAAGGCTCACCATCATGTCGAGCGAGCGCCGTGCGCGCCGCTGCTCGGCGTCGCTCT
GTCCGCCATAGGCGAGCTGGGCAAGGTCGGGCAGCGTGTAGCGCTGCCAGGGGCCG
CCGCCCTGACGTGGCTTGCCCACAAAGCCGGTGCGCACGTCAGCCAGGTAGAGCAG
CGCTACGGCCATGGCCAACAACGAGTCGGCCCCATCCTTGCGCAGTCCGCGCACCC
GCGTCAGCGCCTTGTCCAGGGCATAACCGGCCACGATGACGTTGTGGCGAATCCGCT
TCCACGCGCCTTTCCGTTCTTCCAGGGCGCGCACCAGTCCATGGCGAGTAGCGTCCA
GTACGCCCGGCCGGCGCTTCTTGCCGGCCAGCGGCTTGGGCTGACGCCAGTGCACAC
GCTTGTGCGGCCGACCGCTGGGAGCGGTCGGTTCGGGGCCTTGAGCGTGTGGGCCTG
TGTCTACTGCTGGCATGTCCCGCGCCGATGAACGGCAGCGGGGCAAGGTGGTCGGC
GTTTCGGCCTTTGCATCCTTTTCCCTCTTGCACGGCTTCGCCGGCGGGATTAGGATGG
AGCTATGTGACTAGCCCATTGACTGACTGGTTTTTCCCGCCAGCTGACTGACTGGTTT
TTCCCGCCAGCCCAGTGATGAAAAAGCCGCCCTGCCAGGGGCGGTTTTTTCGTTTTC
GGCCTCCGCTCATCACACCAGCAGGCTGCCTCGTTGTTTTGGTACGTCTGGTGCTGA
ATCTACCTTCATGCAGGCGCGAGCGTACTCGCAGGAAAGGTGGTGGACGGTACGCG
GCTATCAAAACCGGATAGGAAAGAGCGATAACCGGCAGGCCTTGCGGCCTCCGCGC
ACCGCCCACCATAGAACTGGCGAGCTAAATAAAAGCGCCAGCATCGGACGATGGGC
GCTTGCGCGCTCTTTCCACCTTTATCGGTTTGATAGGCCGAATCGCTGATTGTGCAGC
GATAGCTCAATCCTCTCGCCAACCGGCGAACGTGTCAACCACTCTTAAACCGTCGCC
TCCTCGACCGCCGCCGCCGCTTTGGCTTGCAGGTTGAGGAGCGTGGCGATGACGGCG
TGCAAGTCGACGTACTGGGTGTGACGTTGGAAAGCAGCTGCCGGCGATTCCGGGTC
ATTTGGATCGGTCTGCAAATGCTCTTGGAACTGCTCGGCCAGCAAACCAGCCATGAA
AGCAACGTGCGCGTGGTTGGCCTGCAACACGCTGCGGCCAATGGCCTCGAAGTTGT
GCAGGCTGTCAGATAGCCACATGACATCGCGCGCCACCGTGCGACTGGCAAGGGTG
GTATCGGCACGGTAGGGACTGTATTGCCGGATGGCCAACGCGGTGTTGCCTGTAAAC
CGTACTAGGTTGGCAAGTTCATCGGTGTAGTTGATGCCCATGGTTTTTGGCTTCCTCG
GGAAAGCAGGGAGGGTTCTTGAAAGTTCTTTGCTGACAGCGAGTTAGCTCACTTTTG
ACTTTTTTACACAAATCCCTGCCGACCCTCAAAATCGAAAAGCTGCAAACACAGAA
ATACGATTTTACGATTTCCGCGACCATTCCCGCAGCAGCTCGCGCGCAACGTCGGCG
ATCTTCACGCCCTCGTCGGCGCACTTGCGCTTGATGTCACGGTGCAGGCTTTCGGCA
ACGTCGATGGTCAGCCGCTTCATCTTCTCGGCCGATTCGGCAGGTGCGGTCGCCGAG
GCGGCACGGTTAGTAACCCATTCGTCGGCGTTGCCCGCCGCAGCCCTGCTGGCGGTC
GGTTTGCCGCCGATCTGGATTTTCTTTCCGCTCATTGTGCGAACTCCATCAGCGCCTT
GCGCACGGCCTCGATCTCGGCGGCCGCCGGTCCCTCGGCGTCGACTTCATGGACGGC
CTGCCCTCGTGCAACAGCTTCGGCGAATAGGACGCGCTGCGTGACCGATACCGTCA
GCGCGGGGACAGGGTAAGCGGCGAGGGCTTCACCCACGTCCCGACCAATCGCCGTA
TTTACGATTTTACGATTCACCAGAAAAGCGGCTTTCAGTGCGCTCTTGTAGACGCGG
GCTTCCTCAATCAGCTTCATGACTTCATCGGCCGCCCATATGTCGTAGGGCGAGGGC
TGGACAGGAATCAGCACAACGTCCGCTGCCATGATCGCGCTGCGGGCTAGGTCGGT
AACGCGAGGCGGGCCGTCGATCACGATGTGGTCGTAGCCCTGCCCGAGTTGGGCAA
TGTCCTTGTGGACGGTCGGCCGGGGAAACCCGACCACCGAGAACAGCGGTTCGCCC
TCACGGGCGGCGGCCCAATCCAAGGCGCTGCCTTGGGGGTCGGCGTCGATCAGCAG
GACGCGGGAACCGTGCCGCGAGAAGCTGGCAGCGAGGTTCACGGCCAGCGTCGTCT
TCCCGACGCCGCCTTTCTGGTTCAACAGACCAACGATCACGATCGCCTCTTTTTCCGT
ATTTACGGAATTACGATTTTACGATTAAACGGAAAAAATGCAAGCGGCGTGCGCCG
CTCGTCCGCTCGGGATAGGGGACAAGGAACCGCACGCCGTTGAAGTGGATCAGGTG

CGGGGGCGCATCGGCCACCCACACCTCGGTTTCCCCATGGCGATTTCGCCAAGGTAT
CGGCCCAGGATGGCCCGGCTCGGGCAAGCTCGGCGTGCCGCTTGCTATCGCCCCTCC
CGTGACTGGTGACGGATTCGCCCAGCAGCAGCCAATTCCAAATAAACACCAAGACT
TCTCATTTTTTATATTTGTATGCGACGTTTGTCACTCTTTCGCAACCACGGGCTTACA
TCCGGCACGTATATATGATTTCATAAAAAATGAATCGCATCAAACGCGAAACGTAA
GACTTGAATCGCAAAAAACTCAACCCCAAGCGTGGAAAGCTGAACGTGACCGTGGC
ATTACTTCCCTCTTTGTTTTTCAACGATGAGGAAGCAATGCTCACCCTGACCAAGAA
AGAGCTGGCAGTGCTCGATGGGTAATCCCCCCGCCCATTAGCAGACGCCAGAAGTG
GAATTTTCTCGTATCCTTTCTCGAGGAGGTTCCATGAAGAAGTCCCGCTTTACCGAC
AGCCAGATCATCGCCGTGCTCAAGCAGGCCCAGGCCGGTGCGCCCGTGCCGGAGCT
GTGCCGCGAGCACGGCATCAGCTCGGCCACGTTCTACAAGTGGCGCAGCAAGTTCG
GCGGCATGGACGTGTCCATGGTCGCGCGCATGAAGGAGCTGGAGGAGGAGAACCGC
CGGCTCAAGAAGATGTACGCCGAGGCGCAGCTCAGTACCGACCTGCTGAAGGAAGC
GCTCGCAAAAAAATGGTGAGGCCATCTCAGCGACGCGAGATGGCCCAATCGGCAGT
CACGAGCGGGCGTACGAACATCCGCCACGCCTGCCAGGCCTTCGCGGTGAGCGAGA
CCTGCTTCCGCTACCAGGCCAAGGCCAGCGAGCAGAACGCCCGGATCGCCGACTGG
CTGGTCCGCCTGACGACCGCCCATCGCGACTGGGGCTTTGGCCTGTGCTACCTGTAC
CTGCGCAACGTGAAGGGCTTTGGCTGGAATCACAAGCGGGTCTACCGGATCTACCG
CGAGCTGGAGTTGAACCTGCGGATCAAGCCGAGGAAGCGGCTGGTGCGTGAGCGGC
CCGAGCCTCTGGCGGTGCCGGAGGCTATCAACCAGGTCTGGTCGATGGACTTCATGC
ACGACCAGTTGGCCGACGGCCGCAGCTTCCGGCTGTTCAATGTGCTCGACGACTTCA
ATCGCGAGGGGCTGGGGATCGAGGTGGATCTGTCGCTGCCGTCAGCCCGTGTGATCC
GATCGCTGGAGCAGATCATCGAGTGGCGCGGCAAGCCCGCCGTGATCCGCTGCGAC
AACGGCCCTGAATACATCAGTGGCGCGTTGCTGTCCTGGGCGCAGCGGCATGGTATC
CGGGTCGAGCACATCCAGCCGGGCAAGCCACAGCAGAACGCCTACGTTGAACGCTA
CAACCGCACCATCCGCTACGCCTGGCTCGCCCAAACCCTGTTCGACACCATCGACCA
GGTGCAGGACAAAGCCACCCGCTGGCTATGGACGTACAACCACGAGCGTCCGAATA
TGGCGCTCGGCGGCATCACGCCTGCCATGAAGTTGGCGATGGCCGCTTAGCTCCACT
TCTGGCGACCACTAAAAGTGGGGGGATTACCGTACTACACCGCAACAGGCTCGCGA
GAAGAACCCATCGTTTTCAAGTTGGCCGGATGGGCTACGCTGGAAGACGAGGGTAC
TGTCGTAGGTATGATCTCCGCGTCGGACGTGACGACGGACGATAACGTGGCTCGGCT
TGCAACGCCGCCGCCGATCAAGGGGCGGTATCTGCCTGAAGAACTACTGACTGATG
AACAGAAAGCGGCTGCGAAACGCGGGTAACGAGGTAGGGAGAACGGGCGTACAGC
CCTGACGCAAAAGGACAAAGGCTTCTCGCAGACCATCGTTTGCGAGAAGCCTTGCT
GAGAAGGTGGCGCGGTCTAGCGCCGATGTGGCGGTGAGGTTCTCAAAAATCGTTCT
CAGTTTAAACGACCAGGAACGGCGAGTTTCAAGAACGCTGGACAGTTGCCAGCGGC
AAGAACAAGTTTAGTGGCGAGTCTGGCAAAGCAATGAAGCCGTGATACCGGTAGAA
CTCCGCTGCTGCCTCGTCCTTGGCGTCCACCATCAGAGCGAAGGCGGCAATATCCGA
ACGGGCGGCACGGTCGAGCGCATCGGCCAGCAGCGCACCGCCGAGACCTTGCCCCT
TGAATGCCTGATCGACGGCCAGACGACCCATGCGGACGGCAGGCACGGTCGGATAA
CGCGGCAGCTTCTTGCCGGTGCTGGCCGGAAGATCGGCCAGCAGCAGGCTTGCCGA
CGCCAAGGTGTAGTAGCCCGCAATGCGTTGCCCGTCCGCCAAGGCCACGAAGCAGG
CGGCCACGCGGCGACGAACGTCCTGTGTAACCTGCTCGCGCAGGTAGCGATTCAGC
GGCTCCGAATCACTGCGGAAGGCGGCGCGATCATGTGCCGCATCGAGCTGCACAAG
CTGGAACGGTACGCGGCTCATTCAGAGCGCAGGAGCTTGCGACGGCGAGCGAATGC
GCGTTTCAGAGCAGGGACCGGTGTCGGCGGCGACAGCAGGGCTTGCGCGAAACATT

CCTGATCGGCCAGAGACAGGCGGATGACTTCGGCTTGCTCGATGGCGCGCTGTGCG
GCGTCCTGCACGGCAGACACCACGAAATCGGTCATGGTGCGCCCCTGAAGTTCAGC
GGCACGCTTGAGCATCGAATGCAGATCGGTGCTGATTCGGGCTTCGAGACGGGCGG
TAGAGGTGGCGACGGGCATGGTTTTGTCCTCCTTTATGCAAAAATACGGCAAATTGC
CGTACAAATCAAGGCAGAATTTCCGGCGGTTCCGGCTTATCGGTCTCTTTCGGTACA
TACAGCCAACAGTGTGGCAAGTCGGAACCACCGAAAGCTCTACCTGTGTTGCATGTA
ACAAAAAGAGGCGTAAACTTGTTGCATGTAACGGTGAGGCCAACATGGCACACGTC
AATTCACGAGTTCAAAAGCACCGTGACGCCCTGCGCATGGCAGGGCTGCGTCCGGT
GCAAATTTGGGTGCCGGACACACGGCGGCCTGACTTCGCCGAGGAATGCCGCCGTC
AGTGTCGTCTTGCTGCACAAGCGGACATGGAGGACACCGACATGCAGCGCTTCATG
GATGAGGCGCTAGCAGACATGGATGGCTGGACGGAATGATGCGGGGTGACTTTGTG
ACCATCGCCATGCAAGGCGACTTCGGAAAACCACGGCCCGCTCTGGTCATCCAGGC
CAACCAGTTTGACGAACACACCACCGTGACGGTGCTGCCGGTGACAAGCACGCTTG
TTGCTGCGCCGCTGCTGCGCATTACGGTTCATCCAAGTACCGACAATGGTTTGCAGA
AGCCGTCTCAGGTGATGGTGGACAAGGCCATGACCGTGAAGCGTGACAAGGTAGGT
CGAGCCTTTGGGCGCGTAGACGCTGATGCGTTGGTGGAAATTGAGCGCTGCCTGGCC
GTGTTCTTGGGCATCGCTAAGTGACGAAAAACGACGGTTTGTCCGAACCGCTTAACT
GAGTTGGCTGTTCCGTTGCTCCCCGGCCTCCACTCCCTAGCCAAAGCGTCCGTACTC
GGACGGTTTCTTGCCGCTCGCCGGCTGCTTGCGCTGCTGTTGGAACTGCTGACGCTG
CAACGCCTGCTTGCCGTCGGCTGCGCCTTTGGGTTGCATACCGGCGGCGCGGCGCTG
CAATTGGCGTTCGCGCTGGATCGCTAGGGCCGTCGTCTTGTCGAAGCGCCGCGCGAT
GGCGGCCGGCAGCTCCTTGACCGAGTTCGTGTAGACGCGCGCCTCGTGGCAGGCGC
GGCTGATGGCGACGTAATACAGATTCATCGACGTGGTGCGGCTCTTGGTATCGAGCG
CGATGAGGGCGCGGTCGTTGGTCAAGCCCTGCGCGCTGTGGACGGTGGACGCATAG
GCGTGCTCAAGGTGGAGCGGTCGATTCGTCGGCAGCTCAAGCGCGCGTGCCGGTCG
GCCGTCGCGCTGCTCGACCGATTCGAGCTTGACGACGCCGCCGATCACGCCGGCCAC
GCGCATGCGCTCGCCGTTCGTGAGGTCGCGGCCGGGATCGTTGCGGTTGATCCGCAC
GGTATCCCCGACCGACAGCTCGGCGCGTTCAAGGCGGTAGACGCTGAGCTGCGTCG
CCTTGCGCGGGTTGATCGTGGTGGTAGTGCCGTCTTGCCGCTGCAAGACCAGCGCAT
TGCCGGGCAATGCGTCCTTGACGCGGTAGGTCTCGCCTCGCGTCAGGCCGGCTTTCT
GGTCGTCCTTCTCGGGCTGGATCACCATGCCCGGCTGATAGCTCGGCGCAAAGCGGC
GCTGTGCCTGCGTCATGTCCACGCGGGTCAGCGTGTCGAACTCCCGCCCCTTGCCGG
TCAGGTCCAGCGACTCGCGCACCATGCGGTTGATCTCGCGGCGGGCCTGGTTGGTGC
CGGCAACGATCAGCGTTTCGCGGCGTTCGGGTTCGGTGAGCTGGATGGGTAATCCCC
CCACTTTTAGCGGTCGCCAGAAGTGGAGCTAAGCGGCCATCGCCAACTTCATGGCA
GGCGTGATGCCGCCGAGCGCCATATTCGGACGCTCGTGGTTGTACGTCCATAGCCAG
CGGGTGGCTTTGTCCTGCACCTGGTCGATGGTGTCGAACAGGGTTTGGGCGAGCCAG
GCGTAGCGGATGGTGCGGTTGTAGCGTTCAACGTAGGCGTTCTGCTGTGGCTTGCCC
GGCTGGATGTGCTCGACCCGGATACCATGCCGCTGCGCCCAGGACAGCAACGCGCC
ACTGATGTATTCAGGGCCGTTGTCGCAGCGGATCACGGCGGGCTTGCCGCGCCACTC
GATGATCTGCTCCAGCGATCGGATCACACGGGCTGACGGCAGCGACAGATCCACCT
CGATCCCCAGCCCCTCGCGATTGAAGTCGTCGAGCACATTGAACAGCCGGAAGCTG
CGGCCGTCGGCCAACTGGTCGTGCATGAAGTCCATCGACCAGACCTGGTTGATAGCC
TCCGGCACCGCCAGAGGCTCGGGCCGCTCACGCACCAGCCGCTTCCTCGGCTTGATC
CGCAGGTTCAACTCCAGCTCGCGGTAGATCCGGTAGACCCGCTTGTGATTCCAGCCA
AAGCCCTTCACGTTGCGCAGGTACAGGTAGCACAGGCCAAAGCCCCAGTCGTGATG

GGCGGTCGTCAGGCGGACCAGCCAGTCGGCGATCCGGGCGTTCTGCTCGCTGGCCTT
GGCCTGGTAGCGGAAGCAGGTCTCGCTCACCGCGAAGGCCTGGCAGGCGTGGCGGA
TGTTCGTACGCCCGCTCGTGACTATCTCGCGTCGCTGAGATGGCCTCACCATTTTTTT
GCGAGCGCTTCCTTCAGCAGGTCGGTACTGAGCTGCGCCTCGGCGTACATCTTCTTG
AGCCGGCGGTTCTCCTCCTCCAGCTCCTTCATGCGCGCGACCATGGACACGTCCATG
CCGCCGAACTTGCTGCGCCACTTGTAGAACGTGGCCGAGCTGATGCCGTGCTCGCGG
CACAGTTCCGGCACGGGCGCACCGGCCTGGGCCTGCTTGAGCACGGCGATGATCTG
GCTGTCGGTAAAGCGGGACTTCTTCATGGAACCTCCTCGAGAAAGGATACGAGAAA
ATTCCACTTCTGGCGTCTGCTAATGGGCGGGGGGATTACCGATGTAGTCGTTCACGA
TGGCCCGGTGCCGCTCGATCGGCTCTTTCAGCTCCTCAACGTGCTTGATGTGCGCAA
GCGATGGCGTGACACGGCCCTCGGCCGCGTGCTCGACAGCGGTTTTCAGCTCGTGGT
CTTTCTGGCGCTGAATCTCGCTGATTCGCGCGGTCTGCATGCCGTCCTGCTGGAGCTG
CGCGAACGGCTTGCCGGCCTCGATGGCCTCGGTCTGCTTGGTGTCGCCAAGCAAGAC
CATGCGAGCGCCGGCTTTCTCGACAACCCGCATGACCTGTTCCATCTGGCGAGCGCC
GACGACGCCCGATTCGTCCAGAACGATGATCGTCTTGCCGTCGATGGGCTTGTCCTT
GGTGCGCAGAAAGCTCGCCAGCGTGTGCGCTTCCAGCCCTTCATTCTTGAGCGCCTT
GACCTGATTGCCGTAGGGCGCGAGCGCAACGGTGCGGTAGCCCTCGCTCACGGCCG
ACGCCTGCTTGATGAGCGCGACCGCCTGATTCACGGTATAGGTTTTCCCGGTGCCGG
CGTCGCCCTGGATGCCGACGAACCGATTCTTGGTGCTGACGATGGTTTCCACGGCGA
AACGCTGGCCGGCGCTCAACGCGGTGCCTTCCAACGCCGTTTTGACGGCGGCCGCCG
TCATGATCGGCTCGATGGCTCCGCGCCCGGTGCGCTCGATGGCGAGAATCGCTTTCT
CGCGGGCGAGCGCCTTCTGTGTCGTGTAACGCTTCTCGGCCGGCACGAGCGAACCCT
GCTTGATCGCGTTGCCGACGTACTGCTGCGCCTGTTTGTCGGACCAGCCTTTCAGCTC
TTGCAGATAGCTTTTCCAGCCGGCCGACGAGAGCGCCGGCCCGTCCTTGCGGTCGGC
CATGCGATACGCCGGCACGGCTTCGATCAACGCGCCTTGGCCGGCGAGCCGCTTGAT
TTCGGCGCGAACCTCGTCCGGTCCTGCCAGCCCGACCGCGCGACGCAACGCCGTGG
CCGTGAGCGCGGTTTCGCGAACAACAGCCTCACGCTCGGTCAGGTGGTTGATGGCGT
ACTGCACGACGGCTTGTGCCGGCGTGAGGCTCGCCGGCAGGCTCGTGGCTGCGATG
CGGTCGCCGGCGTCGCCCCGCTCACGGCCACGGGTGCCGAAGCTGTCGCCTGCCTCG
TAGGTGCGGCCATCGAGCTGCGAGCGCAGGCCGTAGTCGATACCAAGCTCGCGGCT
TTTTCGACCCAATACTGCTTGACCAGCTCGCGGTCGGACTCGTCTTTGCGCGGCCG
CGTGGCAAGCGAAATGATCTGCTTTTCAAGCGTTGTCGCCGTCGCTCGCGTCTTGCC
CTCGTTGGCAAGCGCTTCCTCGATGACGCGACTACGCGCCGAAAACGCCTCGATCTG
GTCGCGGCTGATGTGGTCCAGCTCGAAATTGCCCTGGTCATCGACCAGGCGTATCGC
GTAGCCAAGCGCTTGCAAGCCCCTCGCCAGCTCGGCCTTGTAGAGCGCGTCTACCTC
GTGCTGCACGCGGAATATGTCCTCGTTCGACAGGGCGCGCCAAGCGCCGTCTGCGC
GCTGCGTCATGTTGAGCACGACCGCATGCGTGTGTAGCTGCGGGTCTTTGGCCCGGC
TCATCTCGTGCCGAAACTTGCCGATGACCATGTTTGCGGTGCGCTCGCGGTAGCTCT
TTCCCTTGACCTTCTTCCGGGCCTCGGCCAGCCGCTCGACCTGCTCCAGTGCGCGCGT
CACGGCGCGGTCGTGCGCAGCGGTGACGTCCTTGTCGCCAGCAACAAGCGCCTGCA
TGGAAACCGACTTCGGGGCGGAAAACGTCAGGTCCAGCCCCATGCGCTTCTTGTTGT
CGGTCGGGTCAAAGGTCGTCTGGATGCGCTCGCCGTTGGGGAGTTGGCCGTCCAGCA
GCCGCGAGAGCTGCGCCTGTTCCACCGGCCCCGTCAGGCCAAGCGCCTGTGCGCCTT
GGCCTTGCCACTCGCCGGGGTGCTCCTTGGCGTAGTAGTCGTCGGCCGCCGAGAAGT
AATGCGCGGCCGCGTACTGGTTGTTGCCGCGAATCGGCGTCACATTGAGCATGGCCC
GGCCCCCGTCACTTGATGTCGATGGCGGCCGCGCGATGCGGATACTTCACGTAGGGC

AGCTTGATCTTCGCGATGGGGAAGTCCTCGGCAAACATGACGTAGCCCTCCAGGTCC
TTGAGGTTTGAGATTTCCGAGTCCAATACCACCGGCTCCGGTGGACTGGCGACAACA
TGGCGGCTGCGGCCGCCTCCACCGCTTCCCCCTGCGTTGGTCGTAATCTGGATGCGC
TCGACGTGCTGCTTGCCAAGAATCTCGCTCGCCTTGTCGGCATTGAGCGCGTTCGAC
GCGCCAAAGATGAGGTAATTACGGAAGCAGCCTAGCAAGGTCTTGGCCGCATCCTT
GCCGTAGGTCTCATCGAGCTGCGCCCAATCCTGTATCGAGGCCGCCATGCGCAGCCC
GTGCTTACGGCCCTTCGTGGCGGCCGGCACGAACGATTCGAGCTTGCCGAGCGATTC
CAGCTCATCGAGGAACAGCCACAGACGCCGGGAGAGGTGTTGCGCGTCGTCGGAGA
TACTGAGAATCGTCGCGCAAATCGTGTCGATCCATGTCGCCACAAGCGGGCGCTGC
GCGGCCCGCATGTCCTCGCGCCACGTGATGAACAGATTGCCGGCGTTGGGGTCATGC
ACCCACTTATGCAGCGAGAAATCGCCCTTGGTCATGAAGCGCAGCGGCCGCACGTA
CTTGTTCATCATGAACTGGATCGACGCGATGGCCTTCTCGGCGTTCTCGCGGAAGTA
GCCCTGCGAATCGGTGTTGGCGAGAAACGCGCGGATGACTTCACCATCCTCGCGCA
CCAACAGGTTGACCAGCGTGTCTTGATCCGGGTTGTTCGTCTCGACCAGCTTGCGCA
TGGTGTCGGCCAGCACGTCACGCGCATAGGCGCACCACTGTTCATCGCTTGGATCGA
TCTGCGGCGGGATGACACTTTTCGCCATCCGGTCGAAGTCATGCACGCCCTTGATCT
CGTTGAACAGCGTCCAGCCCGACGAGCGACGGTCAAACGGGTTGAGAATCGTGTCG
CCGGGAAAGCTGAATTTCGAATAGAACGTCCCGTTCGGGTCGATCACGGCCATCTTG
TCGCGCCGTTTCACGGCCGACGAGATCATGCTTTCCAACGCAACCGATTTGCCCGCG
CCGATGGACGCGCAAATCAGCGTGTTACGGTTCTCAAGGTGCAGCGGCATCGGCAT
GGGGCCGATCATGATGGGGGGCAAATCCTTCGCGCCCCTCTGTTCGGCCCGTCGCCG
GCGATTCTCCCGGCGATTGGCCGCGTTGACCTTGCTCTTTACGCTGTGCCAGTTCGCC
ATCTGCGAGCCGCGCATCCAGCGGCGATACGGTGCGCCCCGGAACCCATCGTCGCC
GTACTCGTGGAACAGATACCCGGTGGTCGCGGCCAGTCCCAACCCGGCCACAACCC
CGCCAACCATGAACGGATTGTTCGGCGTCGCGTGTAGCGCAGCGGTGAGCGCTTGC
GGGAACGGTGACAGCGGCACATTCGCTGCATACGTGCCGGCGAGCCAGCCAGCAAT
CGGCAAACTAACGCCAAAGAACAGCGTGGCATCCTCGGCTTCGTGCTTCTGCATCGC
TTACACCCTCCCGAGAATGGAATCGCAGCGGCGACCCCATAGGTCGGCCAGCAGCT
CGGCATCGTGGGCCGCATCAACGCAATCGCGGTGCAGCGCATCGGCGAGCGCGGCC
TTAAGCCAGCGGCTCGTCGTCGGTGCCGCGATCACGGCATTGATTTCCGCCTCGACC
TGGCCGGTGGTCTGTAACCCCGATGTCACCCCGGAGCAAACGGCCTCGTACGTCAGC
GTATAGAAATAAGACCGCTGCCCTCGAATTTCATGCTCGTCGTCTATTCAATTCCGA
GCTTGTTCTTCGTGATCTTTGCGGGCGTGATGAACACACGGCTCATGCAATCGAGCT
TGCTGGCAGGAAACGAGTACGCGACGATCGCATCGCCGATCTGCGCGCCCAAGGTG
TAGACGGCATGCTCATTGACGCCGCGCGCGTAGTCCATCACGATGAGGATCGGATA
CTGTTTTCGCAGCGATGCCATCGAGACGCAGCTGCCCGTCACCTCGAAGGATGCCGA
ATCTGCGCGATTGTCCGCGCCCATGACCACCACGGCGTTCGACGCGAACAGCGACG
GAGCGATTTGCTTCGACGTCTTGCCAGGCTCCTGCCCATTGGTCATGGCGACCGAGT
CGGGGCGGGGGGAGACGAGCGTCAGTACGTCTTTTCCATTCCCTGGAAGAGCCAAT
GCAATTTTTTCCGAGAAAGCCCAGAGCTGCTTTTCAGCGGCGGCGGCGGCGCGCGGT
ACGGCTGCGGCAGGCACGAGCGGAGCAGTCGCGAGGACGACCGCGCCAAGGGCGG
CGACGAGAGTCTTCATGGTTTATTGCTCCTCTTGCGAAGGGGGATTGAGTTTGTCGT
ACTCGTTGCCGTAGTACACCTTGTAGTTCTCACCGGTAGTGGACGTGACATTGGCCT
CGCAGAATGCATCACCCACACGTTGCGCCAGATCGGCTCCCCCGGCCGCGATCGTG
GACAGCAGCGCCGGAGCGTTCGCCGCGATCAGGCTGATGTCGGCGCCTCCCTGGGA
GGTATCCAGGATCTCGTTGCGCGCCTGTCAACAACAACTTCAAACTGACCCACCTTG

AACGCTGATCAGCAACTTAGAGTTGACCCACCTCGATGGGTTCGGCAGTTTGGCGGG
ATGGAATGAGGCCGGCCTTGCGTTTGTCTTTGAGTCGGAAGCTTTCTCCTTTGATCTG
GATGACGTGAGCGTGGTGCAACAATCGATCCAGCAAGGCGGCCGTGAGGGTGGTAT
CGCCGGCGAAGGCTTGATCCCATTGTCCGAACGGCAAGTTGCTGGTCAGCACCATG
GCCCCCTTCTCGTAGCGCTTGGCCACGACCTGGAAGAACAGATTGGCCTGGTCGCGA
TTCATCGGCAGGTAACCGATCTCGTCGATGATGAGCAGCCGTGGCCCCATCACGTTG
CGGCGGATGAAATCGGCCAAGCGATCCTGGCGTTGGGCCGCAATCATGGCCACCAT
CAGGTCGGCGGCACTGACGAAGCGGGTTTTCATGCCGGCCTGGGTGGCCAAATAGC
CCAAGCTGATCGCCAGGTGGGTTTTGCCGACGCCGGAGGGACCCAACAGCACCAGA
TTCTCGGCCCGCTCGACGAACGCAAGTCCGGACAGTTCCTGCATTGCCTTCTTGCTG
GTGCCATGGGCCGCGGTGAAGTCGAACTGCTCCAAGGTCTTGATCACGGGGAAGCC
GGCCACCCGGGCCAGCATGTGTCGCGAACGGCTTTGCCGTGTGGTCACTTCATTTTT
GAGCAATCCTTCCAGGAAATCGGCGAAGCTCAGTTGGTTGCTCACCGCTTTGCTAGC
CAGTTCCGGGTAGTGCGTGGCCATCGCCTCCAGCTTCAGTTGGCCACACAGTTCATG
GATGCGCTGATGTTGCAATTGCATCAGTGCACCTCCGGGAACAATTGCTCATAGACC
GACAAGGCGTGCTGCAAGGGCTGTCGGGGAAAGCGTTCCCAGTCCGGTGCTGCCGC
TTGCGCGGGCGGGGCCGCTTTGGGCAACTGACCGACGTAGGGCGATGGCACGCCTT
GCAGCGCCGCGCATTCTTGCTGCAATTGCTGTTGCGGCACCTTCCCGGTCGTGCCAT
GCACGCGGGCATTGGCCGTGGTGCGCAACCACGTCAGCACCGCCGTATTGGCGGTA
TCGGCGTCCAGGCGTAGACCGTCCTGGGCCAGCTTGGACGCCAACGGGTTGTAGAA
GCTGCGCCGCAGGTAACCGTTGAAGCGTTCGACCTTGCCCTTGGTCTTGGCCCGATA
GGGTCGACACAGCTTGGGGGCGAAGCCGTAGTGTCCGGCAAAGTCCAGGAAGGCCG
GCTGGAAGCGATGTTGGCCCGGACCGTAAGCATCGCGTTCGAGCACAACGGTTTTCA
TGTTGTCGTACAGCACCTGGCGCGACACGCCGCCGAAGAACGCAAAGGCATGCTCA
TGGCACGCTAGCAGCGTCTCCAGCCGCTCATCGGTGACAAACTCCACGTAGCTGGCC
CGTGACCAGCCTAGCGTAGCCACAAACGCCGACAGTGGATGTTTGCCGCGTCGGAA
AACAATCCAGTCGCACTGCAACTGCTGTCCGGGCGGGGTCTCAAAACGCACCAGCG
GCTCTTGCGGACGGACCACGCGCAGGCTGCGCATGAACTGGCGCAGCCGCGACATG
CCGCCCCGATAGCCCAACAGCGCGATCTCCTGCTTGATCACCGTGGCTGGCAACCAG
GCCGGATGCGCTGCGCGAACGCGCTGCTCAATGTAGGTCCGGTACGGGTCCAGCTT
GTGTGGACGGACCCCGCGTGGTTGGCGCTGGGCCGCTGCCACATCGCGCAGATAAC
GCCGCACTGTATTGCGCGATACCTGCAATTGCCGGGATATCTCCCGGATGCTCATGC
CCTGTCGGGCCAATATCTTGATCTCCACTTGTTCCTCTACAGTCAGCATGGCCCTGCA
AAAGGGCCAAAGGCTACGGAGGCGGTCAATTTGAAGTTGCCGCGGTGGGTCAGATT
AAAAGTGCTGGTGACACGCCGTGGAGTTGTTGAGCACGGCCCGTCCCTCATTTGTGC
AGGCCTTGGCGATGTAGTCCTCCCGTGTGCCGAGCACCCAGCCCTGATCCAGGTATA
GGTGGCCCATCTCGTGGCTGACCATGGACACTCTGGAGTACAACCCAGGGATGTCGT
TGGAGGCACACGAATAGTTCTGGATGTAGACAGTGCTGCCCGGCTTGTCCAGATAG
GTGGCTGGGCCATACACGAAGGGGAACGTGGTGCCGTTGACCTGACCGGTGAGCGT
GCTGGAGGCAAAGATGACGAAGTCCACGCATTCGCCAAGTCCCGTGTTCACGGGCT
GCTGCTGTGCATGGGCCTGAGTGACAAATAAGGCCGACATGGCGATGACGGCCGCG
CTGAACAACTTCATGTTTCCTCTTTCCTTGGGGTGTGAGCGTTTTCTAAAGCAGGCAG
AAGCTTGAATGCAAGCACGGTGCGCCGCAAATTTGACGTGCCTGGCCAGCTTCGTTT
TGTCACCATCGTTAGTCGTCCCGTGCCGGGGGCCACTGGCGCGCTGTGTGGATAAGG
GTCATCACCCATACCGTGTCGGCAGCTTCATCCACTTCGTAGACGAGCCGGTAGCTT
TCATGCGGGATAAGCTCGCGAGTACCAAGAACTTTCCCGGTGTGGCCGAGCTTGGG

GAACTCTCCCAGCCGCTTGGCCGCCTCGCTGAAAAGATCATCCATCTTGAGCGCTGC
GCGGGGGGTCTCTACCGCGATGTAATCGATGATATCGGCACGGTCCTGACGTGCGGC
CGCGATCCAGCGGACTTTCACAATCCAGCCTCTCCGGCGTGGCGTAGCAGGTCCACG
CGGCGCGCGGCCGCATCGGCTTCCACTTCCTCATTGGAGAAGTGCAGGCCGGCGTCA
CGCTGCTTGCGGGCGACCTCCACTTTGCGGCGCACGAACTCGTCGTATTCCCGTGCC
TCGCGACGGCGGGTGATGTAGTCGCGCATCAGCTCGCGCACGACCTGCGCCGCCGG
GCGGTCATCTGCGGCGGCTTCGGCCATGAACGTGTCGCGCAGCTCCGGTTCCAGCTT
CATCGTGAAAACGGCTTTCGTTGACATAACGGGCCTCCTGCTTGTTTGCTATTTCCAG
TATATACCAGGGCATTACCTTTGCCGCCTCCTTTTGGTGACATTACGAATCCTCCGTA
AAGCGGGTTAGCAGTGCAGCCAGCTCGGGCAGGGAGGCGGCCAGCACGTCGGCCG
GCACGCCGGATTTCTCGCATGCGAGAACCAGCTCGTTGACCTCGCCCGTGGTCGCCT
TCTTGCGCACCAGAAAAAGCAGCATCGCCAGAAGCGCATGCGCGCCTTTCTGGCTG
CGCTCGATCTCGGCCAAACGGCCGACAAGTTCCTGTCGGTCGGCCCATGCTTCCATG
CTGTCGCGGCCCGACTCGCGATGGCTGCCACGGTCGAGAGACTTGTCGCGGATGTAC
TTCGACAGGTGCTTGTAGCCGGCCAGAGCGGCCGCTTGTTCCACCCTGATGCGCTCG
TCCTCCGTGTACCAAACGGACACAGATTTTGCGGTACTCATCGGCGTTTTTACTCTAT
TGATTTTTCAATAGAGCTTAGCCTAGCTTTGATGAGGCATTCCGGCAAATTCCTAAT
GAAAAAATATTGACTTCAATGCGGTGAGAATCTGGCCTGAAGCCATATCTGCCAAG
CATTTTGTACCCCCAGAGGGGGTGCGTAATGTGTATTGAGATTTCGGGCCTGTTAAG
GACCGGAATCTCTAGGGGTTAAGGGCGTACAGGTCGCCAGTACGCGTAGCGTGCTA
GGGAAAAGACACGGATCGGACAAAGTGAGGACGGCTGCAGGCTTGTTCTGGTTGAC
AGTGGGAAGTAAATAGGACATTTTTAGGACAATTTTCAAAGTCAACAACGGAGGGG
GATGGATGTGGATGTCAATAAGCTGAAGCAGTCGATCACTGAACTCACTGGCGGCA
CAAAAGCGGCAAAGCTGCGCCGTGTGATGCCAGAGATCGAGGAGAAGCTGGACCA
GGGCGTCCGACACGAGGAAATCGTCGCGGCGCTGCACGCGGGTGGGCTAGACGTCA
CGTTGGAAACCTTCCGCAAGACCCTCTACAGGTATCGCGCGCAACGTGCGGGCGGA
ACCAAGAAGACCCCTCCAAAGGTGGTCAGGGCCACCACGGCGGACGGCAATCCGCA
GGTCCGGGCAGATGCGCCTACGGAGGATTTAACGGCCGACGACGCGGCAGGCGATG
CCCCGACATCAACGAGCTTTGACGACGCGCTCGACCGTCGCAAGCGCGACCAGGTG
GGGGAGTCATACCTGACGCGCCGGAAGCCAATCATCGGCAAGAAACGCACTACCTA
GGAGTGATAATGAAGATTGCAGTGATTAATTTTTCCGGCAACACCGGCAAGTCCACA
GTGTCAAAGCACCTGCTGTATCCCCGTCTCAAGGACGCGGAATACATCGCGGTCGA
GTCGATCAACGCTGACGAAGGCGAAGGCGATAGCGTGCGCGGCAAGCAATTCGGTG
CCCTGCAGGAGCAGCTGCTGGTGATCGACTCGGCCGTCATCGATGTTGGCTCGTCCA
ACGTCGAGGATTTCGTGAAGCTCATGCGCCAGTACCGTGGGTCGCATGAAGATATG
GACCTGTTTGTGATTCCGGCCGTCAAGGAAGCAAAGCAGATCAAGGACACGATCGC
CACGATCCAAGCCCTGGCCGCGATGGGCGTGCCGGCGAAGAAAATCCGCGTGGTAT
TCAACAAGCTTGAGGCGGACGATACGGTCGAAGACGCCTTCTATCCCCTGATCGCGT
TCCATCAGGACACGAAGTCCTTCACCCTGCGCCCTGCGGCGGCAATCCAGTATTCCG
AGCTGTTCCAGCGTCTGCGGGCCTTGAACGTGACAATCGGCGACCTTGTTTCTGACA
CAACGGACTACAAGGCACTCCTGCGCGAGGCCAAGACGCCGGAGGACAAGCAGGC
GATCGCGGCCAAGATTTCTGCCAAGCGCCTGGCGGCGTCGGCACAGGAAAATCTGG
ATAGCGTGTTCGCTGTTTTGACCAAATAACCAAGGGGGCATCAATGGCCGGACTGA
GTACACATCGGGAAGCGCTCATGGCCGAGCTGCTGCATGATGTCGGCGCCATGGTG
CGTCGCTTCGAGCAAGCAGATGAGGCGCTATCCGGAAAGATCGAGAAGGCAACAAC
GGACGCGGCCGGCAAGGCGTTCCTGGCCACCAAGCTCAATTTCGAGTCGGTCATCG

ACAAGAACGCCGAAAAGCTGACCGAGGCTGGCCGACACGCGGCCGCGCAGATCGG
GAATCAGCTCAACAGCGGCGCGGCGCAGCTCGTCGCCGCGAATGCGGCCCTTGAGA
GCAAGGCGCGGCGGTTCGTGCTGCTGCTGGCTGGGTTCGCATTTGTCGCCGGCGTGG
TGGGCGGCTTCGTTGGCGCCAAGCTGGCCGGGATGTGAGACACCGGGCGCTCGCGA
GAGCCCTCTTTTTCGCCCGTCTCCTGTCGTCACCAGCACTTCTTAATCTGACCCAGTT
CAGCAACTTAAAATGACCCACCTCTCGTCGTCTTTGGCCCTTTTGCAGGGCCATGCT
GACCGTAAAGGGGTTGTAAGCCGGAACCCCAGAAAAATCCGTCTCTCATTTCCAGC
CGCCATCGCATTCGTCCCAGAACGAACCGGTTGCCTCGTGCTTGACACTCATGTCAT
CGACCACGCAGTTGCGCAGGTCGGCCAAATGCAAGATTTCGTGCAGTAGATCGTCG
ATCTGCTGCTTCAAATCCGCATCGTCGGCGTAACGAATCACCAGTCGGTATTCGGTA
TCACTGAGCTTTGTGCTCTTGTGTGAGGTGGCGATCAGGTCGTCGATGTCCTCCCGA
GCCCGCTTTTTGCCTCGGGTGAACTTGCTCCCGTTTTCGACATGGAGCCAGAGAATC
ACGGCAACCGATGCCGCCGCCATGAGTTCGGTGCTCGGTGCCGGTGCGCTGAGCTTT
GCGTGTTTGACGGCCTGGCTGATGGTGCTGCGGCTGCATCCCACCGTCTTTTGCACC
ACGCTCCACGAGCTGCCGCTGGCCAGCAGCCGATTGATCGCACCGTAGCGCGCCCG
GTCGACTTGGCGGCCTCGATATTTGCCTGCGACCTTGGCTTTCGCGATGCCTTGGGCC
TGCCGCTGCCGACGCTGTTCATAGTCGCGCCGAGCGATGGCGGCCAGCATGTCCAG
GAGCATGTCATTGATCGCACCGAACATGCGAGCGTCGAACTCGTTTTGCAGGGGGG
CCAAGTGCTGCCACGTCGTCGGTACATTGACCGCGACAACGCGCACTGCGCGCTCGC
GGATCAGCGTTTTCAGCTTCTCCCACTCGGCCCTGCCAAGACGTGACAGACGATCCA
CGTCCTCGACCAGCAACACGTCATGCGGCCGGCTGTCGGCCAGCAACCGAAACAGC
TCGGCTCGGTCCAAGCGTGCGCCGGATGCGTTCTCAGCGTAGTAGTTGCAGATGCGA
ACGCCCCGATCGCGCGCAAACTGATCCAAGCCGCTGCGAGCACGCTGGGCATCCTG
TTCGTCTGTGGAGGCGCGCAGGTAGGCGCGGACAAAGATGGTCATCGGTACGGTTA
GGGTCGATTGGGTGAACCGGTTTGTTTAGGGTCGCACGGTTGGCAACAGAGTACCG
GCCTTTGATGGTTCGACAAGGGGATACCCAAACAAAACCAGCATCCTGGCAGGGCA
GGATGCTCAACGTGGATGCACGCGTGGCCGGACTGGAGCTGGGGTGACGGAAATTT
CGGCGGTTCCGGCTTACCCCCCCTTTATCGCCGTCGATAGCCGCATCGCTCCGATAT
CGGTGCCGCAGTCCGGACCCAGCGGGTCGCAGATTGCACTGCATCGCGGCACGCTT
GTGATGGGAGCCGACTTGTACATACCGCCTTATCGGCGCCTTCGCCAACGCGCAGAT
CGCTTCCGGCTTGGCCGTCGTGCTGCGTCGCTGACGAGCATGATCCAGCGCTCCACC
CATCACCAGGGACGCAAATGATGAATATCGGAGCCAAGCTCTTGCTTGCCACTTGCA
TTATCGCAAGCACGGCCAATGCAGCGGAAACCGGGCAAACTGCGGCGACCGCATCG
GCGGGCCGCACTGCGGCGATCTCGTCCTGGGAAACGACGATCCGGCAGGAAGCTCC
CACTATGGAGGGATGCTTCCGTTCGTCGTTCCCCAACAATGGCTGGCAGGCAACCGC
CTGCGTGCCCTCGCAAAAGCTGCTTTCGAGGCCGCCGGCCATCACTGTCCACAAGGA
TGGCGCCATCACCGCGCGCACCGATGCGGCGCTGACCGCCGGTAACGGTGTCGACT
ACGTTGCAGGCACACGCCAGCTTATCCAATCGACGTTGGGAACATTCCCGAGCGTG
ACTGGCGTGGCCACTGGCGTGGCCGACTATTCGCTACAGATCAACACCAATCTCGAT
AGGAACGCTGTGACCTGCGCACAGTTCGGATATTCCTCGTGCTGGACCTGGCAGCAG
TTCTTCTACTCCACCGACTACGACCGCGACCCGATCAACGGCCGCACGCCGGTCATC
TTTATCGAGAACTGGTTCTATGCCGGCAATGCCGAGGAATTTGAGGCAAAGGGCTGC
CCGTCTGGTTG