Ph.D. Thesis in Agronomy and Bio-engineering

Université Catholique de Louvain – Institut de Recherche pour le Développement

Céline Pesce

Functional analysis of type III secretion system and transcription activator-like effectors of *Xanthomonas translucens*, the causal agent of bacterial leaf streak on cereals

Earth and Life Institute – Applied Microbiology (UCL, Belgium)

Interaction Plantes-Microorganismes-Environnement (IRD, France)

November 20, 2015

PhD defense committee

Promoters:	Claude Bragard	UCL (ELIM, Belgium)
	Ralf Koebnik	IRD (IPME, France)
Reviewers:	Jacques Mahillon	UCL (ELIM, Belgium)
	Jens Boch	Leibniz Universität Hannover (Germany)
Secretary:	Anne Legrève	UCL (ELIM, Belgium)
President:	Bruno Delvaux	UCL (ELIE, Belgium)

"**Ce que tu vis au sommet te change profondément et te devient indispensable...**"

André Gide

Une thèse, c'est comme l'ascension d'une montagne.

On commence souvent trop vite, on s'essouffle puis on s'égare. Mais il y a toujours de surprenantes découvertes après chaque virage. Et en chemin, on fait de belles rencontres qui nous aident à avancer.

Au final le sommet en vaut vraiment la peine.

I would like to thank all my jury members: Prof. Bruno Delvaux, Prof. Anne Legrève, Prof. Jacques Mahillon and Prof. Jens Boch for their participation on my work.

Je souhaite remercier Michel Nicole et Valérie Verdier pour m'avoir accueillie et permis de travailler dans de très bonnes conditions à l'IRD.

Je voudrais remercier très fortement mes deux directeurs Claude Bragard et Ralf Koebnik. Merci à vous deux de m'avoir fait confiance pendant ces 4 ans, de m'avoir permis d'assister à de nombreux congrès et échanges scientifiques. J'ai énormément appris à vos côtés. Merci à Claude de m'avoir épaulée aussi bien au niveau scientifique qu'administratif. Enfin un grand merci à Ralf pour toutes les connaissances que tu m'as apportées. Tu m'as toujours expliqué mes erreurs sans me juger, tu n'as jamais perdu patience malgré mon anglais approximatif et tout cela dans la bonne humeur.

Merci Boris qui m'a permis de faire mes premiers pas dans cette équipe. Merci à toi de m'avoir guidée, formée et surtout supportée pendant ces 4 ans. Merci Sébastien pour tes conseils et ta pédagogie. Grâce à toi, le clonage et les stats me paraissent (légèrement) plus simples!

Merci à tous les membres du labo pour avoir supporté mes ronchonnements et mes « Célinades». Merci à Servane, Jon et Séverine pour vos conseils très précieux et nos efforts sportifs partagés. Merci également à Sandrine et Flo les reines de l'organisation et de la bonne humeur.

Merci à Jean Paul et Martine pour le partage de leurs connaissances.

Merci à Sylvie et Isabelle pour leur disponibilité et leur immense patience.

Un grand merci à Ouioui pour toutes les fois où tu m'as aidée au boulot comme en dehors pour ma voiture, mes skis, mon vélo, ...

Merci à Jamel pour ton aide et des seaux d'eau.

Merci à Harold pour son aide précieuse en serre.

Bonne route à Edwige, ma super stagiaire. Merci pour ton attention, ton travail et tes gâteaux!

I would like to thanks all the collaborators who were involved in my thesis. Thanks to Jens Boch and Jana Streubel for the work that we did together and for you welcome in Halle. Thanks to Camilo Lopez and Adrianna Bernal for the opportunity to work with your teams in Bogota. That was a great moment. Thanks to Daniela Büttner for your work on the translocon part. Et Merci à Geneviève Conerejo pour ton aide précieuse avec le confocal.

Un immense MERCI à mes deux fidèles compagnons de route qui ont toujours été là pour me faire remonter en selle. Merci à vous mes deux « alcoolites » pour tous ces beaux moments riches en partage et en émotions. Faisons que nos routes se croisent encore et encore...

Un merci groupé (alors que vous méritez tellement plus…) aux joyeux lurons qui savent se serrer les coudes: Chloé pour nos grandes discussions sans queue ni tête, Maxime pour les moments de décompression que l'on a partagés, Célia et son picon, Lucile et ses toc-toc podaux et cérébraux, Emilie ma messagère ardennaise, Rémi pour nos incompréhensions partagées, Mathild2 pour ses monologues, Cécile et ses petits cris, Hélène avec sa banane et ses jeux de sociétés, Jérémie et ses mains magiques, Fabien, Sixtine, Julien, Vianney, FX... et de nouveau Mathilde et Poulinette. Merci pour les concerts (avec ou sans trompettes) et les sorties que vous avez organisés avec brio. Bonne route à tous et à très vite...!

I had the opportunity to meet several fantastic persons from many nationalities. All of them help me to evolve during these years. Thank a lot to them: Hinda our wise girl, our only guy: Tu, Taca, Bear (and your red cheekbones), Issa, Kossi, Ganna, Daniel, Jilian, Kim, Mustapha...

Thanks to all the south Americans for your zest of life full of music: Omar, Gabriel, Andres(s), Nathalia(s), Maira, Carolina, Louana, Guilherm, Elizabeth, Alvaro, César(s), Paula, Vivian, Carlos...

Grand Merci à mes deux plus « vieilles » connaissances montpelliéraines : Erika et Carole. Vous m'avez guidée lorsque je perdais mon chemin grâce à vos expériences respectives. MERCI.

Je n'oublie pas les copains rencontrés en voyage, à l'escalade, au monocycle ou les compagnons de course à pieds. Parmi eux, merci beaucoup à Max pour tous les bouts de chemin que l'on a fait ensemble, j'espère qu'on en aura plein d'autres...

Merci à tous les Isérois résidants ou de passage (Céline), vous m'avez apporté des grands bols d'air « pur» et beaucoup de fraicheur.

Muchas gracias Ale, Hector y Yolanda por su cálida bienvenida en Colombia. Estos pequeños momentos fuera de lo común, me han hecho tanto bien. Hasta pronto...

Merci Claudia et Lisa pour la cohabitation que l'on partagée. C'était tellement agréable de rentrer le soir et de s'assoir autour d'un petit repas et un verre de vin pour refaire le monde. On remet ça quand et où vous voulez... Hasta Pronto !

Enfin un grand MERCI à mes parents pour la confiance qu'ils m'ont toujours accordée, malgré la distance géographique et mon sujet de thèse (non je ne travaille pas sur les bonbons et le cannabis !). Vous m'avez toujours épaulée dans mes choix. Vous m'avez aidé à monter les pentes raides Je sais que je n'ai pas fait beaucoup d'aller-retour à Houldizy, mais qu'ils furent agréables et ressourçants. Merci Damien pour les moments que l'on a partagés entre escalade, ski et kayak itinérant. Merci aussi pour tes infos voyages, musiques, bouquins, ...

Merci à ma grande famille qui malgré la distance et surtout le flou autour de mon travail m'a beaucoup aidée à garder les pieds sur terre. Merci aux deux vadrouilleurs, votre belle maison cévenole m'a été d'une grande inspiration durant ma rédaction.

Et une pensée pour ma grand-mère qui n'aura pas connu l'aboutissement de toutes mes années d'études mais je suis sure qu'elle aurait été très heureuse de me voir finir ce grand voyage.

MERCI à tous…

Preamble

This PhD thesis was supported by FRIA (Fond pour la Recherche dans l'Industrie et dans l'Agriculture) and was carried out at the "Earth and Life Institute" (ELIM). Most experiments were performed in the laboratories of the "Institut de Recherche pour le Développement" (IRD) in Montpellier (France).

The initial objective of this project was a better understanding of the host-pathogen interaction between cereals and *Xanthomonas translucens*, the causal agent of bacterial leaf streak. At the start of this project, virtually nothing was known about pathogenicity mechanisms of *X. translucens*. First, inoculation methods were established and used for phenotyping of barley-pathogenic xanthomonads under laboratory condition (chapter III). Then, candidate virulence factors were studied, such as the type III secretion system (T3SS) and its translocated effector proteins. Interestingly, the genetic organization of the type III secretion gene cluster of *X. translucens* differs from those of others xanthomonads. Functional analyses of this secretion system, based on a mutagenesis approach targeting one conserved gene, demonstrated that the T3SS is required for disease development and bacterial colonization of barley leaves (chapter IV). In the same chapter, a new gene is described to encode a new translocon component. Finally, one type III effector belonging to the TALE (Transcription Activator-Like Effector) family was found to play an important role in virulence during the interaction of *X. translucens* and barley (chapter V).

This thesis contains six manuscripts (four published and two in preparation) done in collaboration with other scientists. The results are displayed as three chapters, as illustrated in following pages. In addition, several sections of this work have been presented at different scientific meetings as oral communications or posters (see Appendix B).

Included manuscripts

Published papers:

1. Streubel J.*, Pesce C.*, Hutin M., Koebnik K., Boch J. and Szurek B. 2013. Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* **200**, 808-819 [doi: 10.1111/nph.12411]
   (*These authors contributed equally to this work)
2. Pesce C., Bolot S., Cunnac S., Portier P., Fischer-Le Saux M., Jacques M.A, Gagnevin L., Arlat M., Noël L.D., Carrère S., Bragard C., and Koebnik R. 2015. High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc.* **12**, e01574-14 [doi:10.1128/genomeA.01574-14]
3. Jacobs J.M., Pesce C., Lefeuvre P., and Koebnik R. 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front. Plant Sci.* **6**, 431 [doi: 10.3389/fpls.2015.00431]
4. Pesce C., Bolot C., Berthelot E., Bragard C., Cunnac S., Fischer-Le Saux M., Portier P., Arlat M., Gagnevin L., Jacques M.A, Noël L.D., Carrère S., and Koebnik R. 2015. Draft genome sequence of *Xanthomonas translucens* pv. *graminis* pathotype strain CFBP 2053. *Genome Announc.* **3**, e01174-15 [doi: 10.1128/genomeA.01174-15]

Manuscripts in preparation

- Pesce C., Büttner D., Bragard C., and Koebnik R. The noncanonical Hrp type III secretion system of *Xanthomonas translucens* contains a new translocon component and is required to cause a hypersensitive response on non-host plants and to colonize the host plant barley. *Manuscript in preparation.*
- Pesce C., Szurek B., Bragard C., and Koebnik R. Distinct transcription activator-like effectors contribute to pathogenicity of barley-pathogenic xanthomonads. *Manuscript in preparation.*

Abstract

Xanthomonas translucens is the causal agent of bacterial leaf streak, the most common bacterial disease of small grain cereals.

To cause disease, most xanthomonads depend on a conserved type III secretion system (Hrp system), which translocates type III effectors (T3E) into plant host cells. Comparative genomics of several *X. translucens* strains revealed that they contain a *hrp* gene cluster the genomic organization which is different to that found in other *Xanthomonas* species. Inoculation methods of barley plants were optimized in order to permit quantitative pathogenicity assays. Mutations were generated in two conserved genes of the *X. translucens hrp* cluster, affecting *hrcT* and *hgiA* (HrpG-induced gene). Both mutations caused drastic reductions in symptom development on the host plant barley and a loss of hypersensitive response in the non-host plant pepper. Evidence is provided that the HgiA protein is an analog of the putative translocon protein HrpF.

A specific family of T3Es, called Transcription Activator-Like Effectors (TALEs), modulates plant gene expression. Artificial TALEs were established as a test system for the contribution of plant target genes to disease. TALEs from *X. translucens* were found to have distinct characteristics compared to other *Xanthomonas* TALEs. This study identified a conserved TALE in barley-pathogenic xanthomonads, which strongly contributes to disease development.

List of figures

List of tables

List of supplemental data

List of acronyms and abbreviations

a.k.a.: also known as
BLS: Bacterial Leaf Streak
bp: base pair
CFBP: Collection Française de Bacteries Phytopathogénes
cv.: cultivar
DAMP: damage-associated molecular pattern
DNA: deoxyribonucleic acid
dpi: days post inoculation
EBE: effector-binding element
EPPO: European and Mediterranean Plant Protection Organization
EPS: extracellular polysaccharide
ETI: effector-trigged immunity
GFP: green fluorescence protein
Hgi: hrpG-induced
Hpa: hrp associated
Hrp: hypersensitive response and pathogenicity
Hrc: HR conserved
HR: hypersensitive response
i.e.: *id est* (it is)
MAMP: microbial-associated molecular pattern
MLSA: multilocus sequencing analysis
NB-LRR: nucleotide binding–leucin rich repeat
PAMP: pathogen-associated molecular pattern
PIP box: plant-inducible promoter box
PTI: PAMP-triggered immunity
pv.: pathovar
R gene: resistance gene
RVD: repeat variable diresidue
S gene: susceptibility gene
spp.: species
syn.: synonymous
T3E: type III effector
T3SS: type III secretion system
TALE: transcription activator-like effector
TM: transmembrane
UPB: Unité de Phytopathologie et Bactériologie
var.: variety

Contents

Chapter I

Introduction

Biotic stresses on plants may result from many pathogens such as fungi, bacteria, oomycetes, nematodes, insects, viruses, phytoplasmas and parasitic plants. Worldwide, plant diseases and pests pose a serious threat to food security. Estimation of yield reduction caused by diseases are at 35% for pre-harvest crop (Popp et al., 2013). Controlling plant diseases is therefore crucial to meet the challenge of reaching a sustainable production of fodder. Bacterial pathogens cause serious diseases on plants around the world. *Xanthomonas*, one of the most important genus of plant pathogens, is able to attack at least 400 different plants species. Bacterial leaf streak (BLS) of small grain cereals caused by *Xanthomonas translucens* is one of diseases that has rapidly emerged over the last years and which has a worldwile distribution.

For this reason, the objective of this thesis is to better understand the host–pathogen interaction between cereals and *X. translucens.* Functional analyses have been performed on the conserved type three secretion system, which is known to be involved in virulence of many mammal and plant pathogens. This bacterial syringe translocates effector proteins into the host plant cells. One specific family of these translocated proteins, called Transcription Activator-Like Effectors (TALEs) is known to be of the uttermost importance for virulence of some xanthomonads. Therefore, the role of *X. translucens* TALEs during BLS development on cereals will be studied.

1. Plant defenses against pathogens

Bacterial diseases have generally been disregarded as compared to fungal and viral diseases. Yet, the economic impact of bacterial diseases under warm and humid climate is very high (Gnanamanickam, 2007). For instance, yield losses on cassava due to *Xanthomonas axonopodis* pv. *manihotis* vary between 12 to 100%, depending on the area (Verdier et al., 2004). In 2012, plant pathologists have voted for their top-10 bacterial plant pathogens, representing the scientifically and economically most important pathogens (Table I.1).

Due to the scarcity of effective control measures, breeding for resistant cultivars is often the first line of management of bacterial diseases. Despite being fixed to the ground, most plants are resistant to the majority of plant pathogens. The first layer of disease prevention consists of a physical barrier (waxy cuticular skin layers), which is a passive protection without visible symptoms. The second resistance layer (or host resistance) is an active mechanism called hypersensitive response (HR). It corresponds to a rapid cell death in the local region surrounding the infection. Both these responses belong to what is called an “incompatible reaction”. In such situation, the pathogen is called “avirulent” and the plant host is defined as “resistant”. At the opposite, a “compatible reaction” results from the interaction between a “virulent pathogen” on a “susceptible plant host”, where it causes disease.

Table I.1 Top-10 bacterial plant pathogens

Rank	Bacterial species	Hosts plants	Remarks
1	*Pseudomonas syringae* (29 pathovars)	Tomato, horse chestnut, cereals, beans, *Prunus* species, olive trees, …	Increasing importance (resurgence of old disease or new infection)
2	*Ralstonia solanacearum*	200 plant species in over 50 families (tomato, tabacco, eggplant, banana, …)	US$1 billion in losses each year worldwide on potato
3	*Agrobacterium tumefaciens*	93 families of plants	Ability for genetic transformation
4	*Xanthomonas oryzae* pv. *oryzae*	Rice	Yield losses of 10%-50% (Ou, 1972)
5	*Xanthomonas campestris* (pathovars)	Brassicas, pepper, tomato, cotton, …	
6	*Xanthomonas axonopodis* pv. *manihotis*	Cassava	Yield losses of 12%-100% (Lozano, 1986; Verdier et al., 2004)
7	*Erwinia amylovora*	Apple, pear, quince, blackberry, raspberry, …	Sporadically and occasionally destructive
8	*Xylella fastidiosa*	Grapevine, citrus, almond, elm, oak, maple, coffee, peach, pear, …	Emergent in Europe since 2013 (Italia)
9	*Dickeya dadantii* and *Dickeya solani*	10 monocot and 16 dicot families (*Saintpaulia*, potato, banana, …)	
10	*Pectobacterium carotovorum* and *Pectobacterium atrosepticum*	Crop plants (potato, beet, …)	Yield losses of 5-70% in Montana area on beet (Zidack and Jacobsen, 2001)

The table represents the ranked list of bacteria as voted by plant bacteriologists associated with the journal *Molecular Plant Pathology* in 2012. Table redraft based on information presented in the paper of Mansfield et al., 2012.

Four phases have been proposed to explain the plant-pathogen interaction, in a model called “zig-zag” model (Jones & Dangl, 2006). Each of these steps is described in the figure I.1.

The first one, PAMP-triggered immunity, corresponds to a basal disease defense line. It recognizes common molecules shared by many classes of microbes. Typically, trans-membrane pattern recognition receptors (PRRs) recognize microbial or pathogen-associated

molecular patterns (MAMPS or PAMPs). The best known elicitor of PAMP-triggered immunity (PTI) is the bacterial flagellin, which is the major component of the flagellum used by the bacteria for their motility. A 22-amino acids peptide (flg22) from a conserved flagellin domain is sufficient to bind and activate the PRR receptor kinase FLS2 from many plants. The recognition of PAMPs (or MAMPs) by PRRs results in PTI, which ultimately halts the pathogen colonization.

There is then an intermediate step, called ETS for effector-triggered susceptibility. This step happens when the pathogen deploys effectors in the host that suppress the PTI and thus contribute to the pathogen's virulence. Plant-pathogenic bacteria are able to deliver ~10 to ~30 effectors into the host plant cell (Jones & Dangl, 2006).

If one of the effectors is recognized by the plant's surveillance system, another level of plant defense is triggered, leading to the so-called ETI (effector-trigged immunity). This defense step corresponds to the specific recognition of a bacterial effector gene, also known as an avirulence gene (*avr*), by a corresponding resistance gene (*R* gene), typically inducing an HR. This gene-for-gene interaction was proposed as a new concept in plant pathology by Flor (Flor, 1942). It requires that *R* gene products recognize *avr*-dependent signals, which induce activation of defense mechanisms. The process is able to totally block the pathogen's growth. During a compatible interaction one of these factors is missing and the disease appear. ETI takes place inside the plant cell. It is known that most *R* genes encodes NB-LRR proteins (nucleotide binding – leucin-rich repeat domain) (Rairdan & Moffett, 2007). Approximately 600 NB-LRR proteins have been identified in the rice genome (Goff et al., 2002). On the other hand,

many effector families have been reported so far. Yet, very little is known about the mechanism of activation of NB-LRR proteins by effectors (Rafiqi et al., 2009).

The last step of the zig-zag-model is reached when the pathogen suppresses the ETI, which results again in a compatible interaction and disease development. This bypass of ETI by virulent pathogens is generally explained by the lost of an Avr effector or the gain of a new ETI-suppressing effector, which is not recognized by the plant defense system.

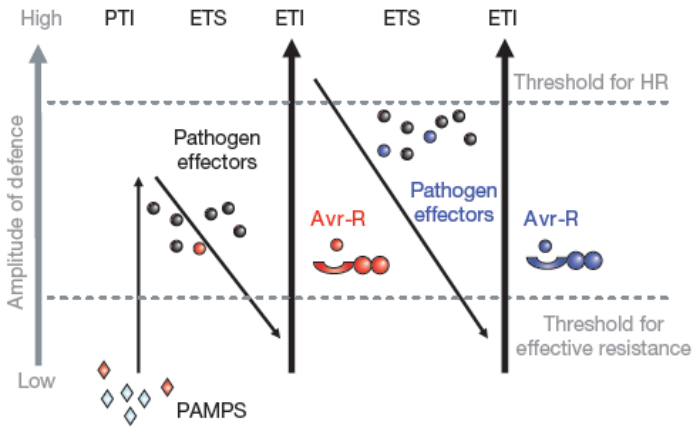


Figure I.1 Zig-zag model proposed to schematize the plant immune system in response to pathogen attack. (Jones & Dangl, 2006)
PAMPs: pathogen-associated molecular patterns, PTI: PAMP-triggered immunity, ETS: effector-triggered susceptibility, ETI: effector-trigged immunity, Avr: avirulence gene, R: resistance gene, HR: hypersensitive response.

In 2014, Pritchard and Birch proposed to modify the initial zig-zag model. They argued that such a model should not be a dogma for our plant–pathogen interaction understanding. Host-pathogen interactions are often more complex and dynamic. Some parameters are not taken into consideration in the zig-zag model, which is based

on growth and reproduction of biotrophic pathogens. But necrotrophic pathogens do not really fit into this model, because plant cell death could also be a source of nutrients for such pathogens. Also during PTI, 'damage-associated molecular patterns' (DAMPs) are not accounted in the original zig-zag model. The environmental context is also absent in the proposed model, even if it could induce either a negative (abiotic stresses) or a positive impact (factors promoting systemic acquired resistance) on the host–pathogen interaction. For these reasons, microbe–plants interactions have to be appraised as being more complex.

Beside the interest of understanding the mechanisms underlying the bacteria–plant interaction with the view of controlling the disease they cause, bacteria are also interesting models to study. Their capacity to bypass rapidly the host plant resistance genes is providing the researcher with a toolbox to unravel the complexity of such domain.

2. Plant-pathogenic bacteria

Among the 7100 classified bacterial species, only roughly 150 species cause diseases on plants (Buonaurio, 2008). Bacterial diseases on plants are characterized by symptoms as leaf or fruit spots, roots tumors, cankers or vascular wilts. The most common phytopathogenic bacteria are Gram-negative and belong to the genera *Acidovorax*, *Agrobacterium, Dickeya*, *Erwinia*, *Pantoea*, *Pectobacterium*, *Pseudomonas*, *Ralstonia*, *Xanthomonas* and *Xylella*. But also Gram-positive bacteria, such as *Clavibacter*, represent important plant pathogens.

Plant-pathogenic bacteria use effective secretions systems to deliver proteins into the host cells or into the intercellular spaces. Protein secretion plays a central role in modulating interactions between plants and bacteria. Six types of secretion systems are known in Gram-negative plant-associated bacteria (Tseng et al., 2009). The six secretion systems are classified into two classes based on where the proteins are secreted. The first class includes the type I secretion system (T1SS), the type II pathway (T2SS) and the type V secretion system (T5SS) all of which secrete proteins into the host intercellular spaces. In contrast, the type III secretion system (T3SS), the type IV secretion system (T4SS) and the type VI secretion system (T6SS) are able to deliver proteins and/or nucleic acids directly into the host cell.

The type I secretion system is the simplest structure and it is found in almost all phytopathogenic bacteria. Type I-secreted proteins are toxins or extracellular enzymes some of which play a role as virulence factor (Delepelaire, 2004).

The type II secretion system is common to Gram-negative bacteria and delivers various proteins, such as toxins, enzymes and other virulence factors. Most *Xanthomonas* species have two T2SS per cell and use them to secrete virulence factors involved in host cell wall degradation, such as pectate lyases, polygalacturonases and cellulases (Tseng et al., 2009).

The type III secretion system is well studied in animal pathogens, but also in plant pathogens, such as *Pseudomonas syringae* (Grant et al., 2006). Pathogenicity of several Gram-negative bacteria, including *Xanthomonas*, *Ralstonia*, *Pseudomonas* and *Erwinia* depends on effector proteins secreted via the T3SS. The structure and the

machinery of T3SS looks like a needle. Sequences similarities exists between components of the T3SS and the flagellum machinery (Blocker et al., 2003).

The type IV secretion system is the only pathway which is able to contribute to the secretion of both proteins and nucleic acids. The model system to study T4SS is the Gram-negative bacterium *Agrobacterium tumefaciens,* because it is the pathway used to inject the tumor-inducing *Ti* plasmid into the host cell where it causes the formation of crown gall tumors (Cascales, 2008).

The type V secretion system, also called auto-transporter, is present in almost all plant-pathogenic bacteria, including *Xylella* and *Xanthomonas* (Hahn et al., 1996). Auto-transporters are surface proteins that cross the outer membrane autonomously without cytosolic energy, i.e. the translocation of the polypeptide chain across the outer membrane does not use adenosine triphosphate (ATP) or a proton gradient (Leo et al., 2012). There are three sub-classes of T5SS and a large number of proteins including adhesions are secreted via such system (Tseng et al., 2009).

The type VI secretion system is comparatively poorly understood. More than a quarter of sequenced bacterial genomes possess a T6SS (Bingle et al., 2008). Gram-negative bacteria have usually one or two copies of the T6SS cluster. This system contains sequences with similarities to components of the tail spike complex of bacteriophage T4. That is why the T6SS is also called a phage-tail-spike-like injectisome. It has the potential to introduce effector proteins directly into the cytoplasm of host cells.

3. Type III secretion system

The type III secretion system is well conserved and present in most of the Gram-negative plant pathogens. T3SS is known to be essential for the host interaction of many pathogenic Gram-negative bacteria including plant pathogens and symbionts, with the exception of *A. tumefaciens* and *Xylella fastidiosa* (Simpson et al., 2000; Wood et al., 2001). The conservation of this system in pathogenic and symbiotic bacteria indicates that it seems to be a major factor of bacterial adaptation to eukaryotic hosts. Contrary to the other secretion systems, this one is clearly involved in virulence of both animal and plant pathogens.

Lindgren et al. (1986) were the first to link pathogenicity and the ability to elicit the HR in plants. With these works, based on a transposon mutagenesis strategy, the DNA region was identified on *Pseudomonas* spp. to be important for virulence and hypersensitivity (Niepold et al., 1985; Lindgren et al., 1986). Some mutants lost their pathogenicity on host plants and the ability to induce an HR on non-host plants. Such mutants were called *hrp* for hypersensitive response and pathogenicity (Arlat, 1989).

In 1992, Gough and colleagues revealed a striking homology of *hrp* genes from *Ralstonia solanacearum* (at that time called *Pseudomonas solanacearum*) with the key virulence determinant of animal-pathogenic bacteria of the genus *Yersinia*.

Shortly before, Rosqvist and collegues had shown in 1991 that the mammal pathogen *Yersinia* spp. carries a plasmid of about 70 kb in size, which encodes several virulence determinants called *Yop* for *Yersinia* outer protein. They demonstrated that some Yop proteins are

secreted into the culture supernatant and can cause cytotoxic effects. In particular, YopE which has cytotoxic effects is dependent on the critical role of YopD, a host cell membrane translocator (Rosqvist et al., 1991). Surprisingly, the Yop proteins did not have a classic signal sequence for secretion but their secretion was found to dependent on the products of several genes clustering in the *virC* locus, which comprises 13 genes (Lory, 1992).

The core type III secretion apparatus is conserved among plant- and animal-pathogenic bacteria (Alfano & Collmer, 1997). In 1996, Bogdanove and colleagues established a new nomenclature with *hrc* genes (*hrp* conserved genes) for those elements that are conserved in both plant and animal pathogens.

This needle-like structure or injectisome shows many similarities with the bacterial flagellum (Groisman & Ochman, 1993). The three-dimensional structure of the T3SS from *Salmonella enterica* was revealed in 1998 by Kubori and colleagues. The architecture of Gram-negative bacteria dictates that proteins that are secreted outside of the cell must cross two barriers, the inner and the outer membrane. Coincidentally, the T3SS is built from two ring-like structures that interact with the inner membrane and the outer membrane of the bacterium. In animal-pathogenic bacteria, a needle-like structure is linked to this complex. In contrast, plant-pathogenic bacteria harbor a filamentous extension. This extracellular pilus, also called “Hrp pilus”, is supposedly a hollow structure. It serves as a transport channel for secreted proteins. Another structure of the T3SS, embedded in the host plant membrane, is called translocon. It forms a pore in the plant plasma membrane, which permits the entry of the

secreted protein into the plant cell (Fig. I.2). This pore-forming translocon is not conserved between animal and plant pathogens. HrpF has been identified as the protein that fulfills the role in *Xanthomonas euvesicatoria* (Rossier et al., 2000; Büttner et al., 2002), a role that is fulfilled by it is homolog PopF in *R. solanacearum* (Meyer et al., 2006).

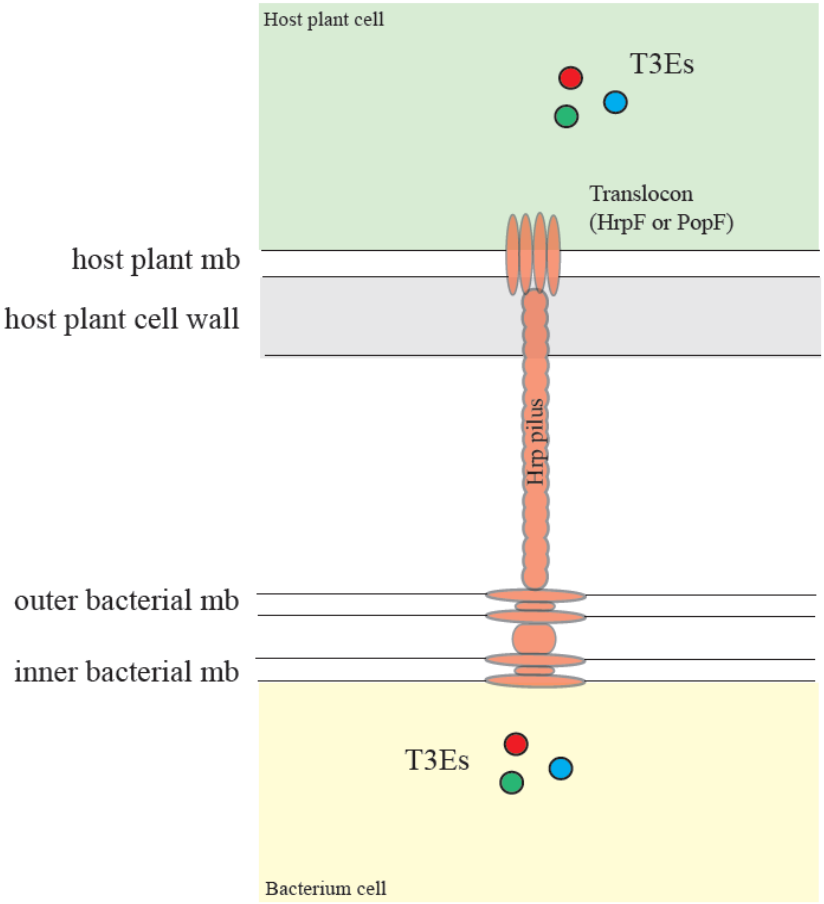


Figure I.2 Schematic representation of the T3SS from plant pathogenic bacteria.
Yellow and green rectangles represent the bacterial cell with the two membranes (inner mb and outer mb) and the host plant cell, respectively. The grey rectangle represents the host plant cell wall. Redraft from Büttner and He (2009). T3Es stands for type III effectors.

The T3SS machinery is composed of one to several subunits of approximately 20 bacterial proteins, many of which form oligomers and are embedded in the inner or outer membrane (Burkinshaw & Strynadka, 2014). The corresponding structural genes are localized in the chromosomal *hrp* gene cluster with a size of about 25 kb (Alfano & Collmer, 1997).

hrp genes are not constitutively expressed, but they are activated when the bacteria enter the plant or when they are cultivated in certain *hrp*-inducing minimal media (Tang et al., 2006). A conserved regulatory cascade induces the expression of the *hrp* cluster genes, which correspond to T3SS structural genes.

In *Xanthomonas* species, gene expression in the *hrp* cluster is regulated by two key components (Wengelnik et al., 1999). The first key regulator, HrpG, is probably activated by phosphorylation when the bacteria enter into the plant apoplast. This response regulator is thought to perceive an environmental signal via the cognate sensor kinase HpaS (Li et al., 2014). HrpG is able to activate the expression of the second key regulatory factor, *hrpX*. Then, HrpX binds to a conserved sequence motif called plant-inducible promoter (PIP) box, which is present in the promoter region of most of *hrpX*-dependent genes. The PIP box consensus is TTCGB–N_{15}–TTCGB (B=any nucleotide except A) and is followed by a –10 promoter motif (YANNNT; Y=pyrimidine) in a distance of 30 to 32 bases pairs (Koebnik et al., 2006). PIP boxes are present in front of *hrp* cluster operons, genes encoding cell wall-degrading enzymes, and also in front of genes for translocated effector proteins (Fig. I.3).

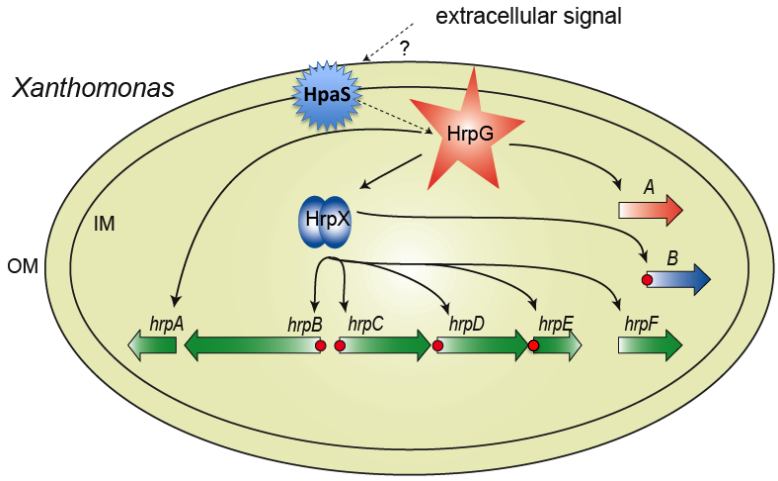


Figure I.3 Key regulators controlling the *hrp* cluster in *Xanthomonas*.
Extracellular signals are perceived by HpaS, a kinase sensor, which then phosphorylates HrpG. This regulator activates the second key regulatory gene *hrpX*, the *hrpA* gene, and other unknown genes. HrpX binds to the plant-inducible promoter box (PIP box, red circles) in front of *hrp* operons and induces their expression. Schema adapted from Berger et al. (2006).

4. Type III effectors

Type III effectors (T3Es) are secreted protein delivered from the bacterial cell via the T3SS to host cell where they support the colonization of the plant tissue or serve in the dissemination of the bacteria. Each bacterial strain injects a pool of specific T3Es. T3Es were first identified via the capacity to induce defense responses. At the moment, a faster way to identify such T3Es uses *in silico* analyzes, for instance via the search for the presence of a PIP box or sequence motifs that are associated with eukaryotes. The reduced cost for genome sequencing helped to identify T3E repertoires via sequence homology. Presently, many T3Es are identified, but for many of them the function is still unknown. It is not an easy task to characterize T3E functions. In 2009, White and colleagues have proposed to classify

candidate and known T3Es of *Xanthomonas* into 40 groups based on sequence similarity. It is important to highlight the possibility of one effector to have simultaneous functions. Often, functional redundancies among effector proteins are found, and consequently, inactivation of only one T3Es does not affect the virulence of the bacteria (Vivian and Arnold, 2000; Noël et al., 2003). T3Es target several host cell mechanisms, such as callose deposition, which reinforces the plant cell wall (Hauck et al., 2003). More recent research focuses on the role of the T3Es in determining the host plant specificity (Ryan et al., 2011).

5. The genus *Xanthomonas*

5.1. General characteristics and taxonomy of xanthomonads

Xanthomonas is well-known as a large genus of plant-pathogenic bacteria. The genus *Xanthomonas* (from the Greek *xanthos*, meaning 'yellow', and *monas*, meaning 'entity') was proposed by Dowson in 1939. The *Xanthomonadaceae* family belongs to the gamma subdivision of the Proteobacteria. These Gram-negative bacteria are straight rods, motile by a single polar flagellum and strictly aerobic. Growth characteristics on artificial medium are the formation of shiny, convex, and yellow circular colonies. This yellow color of most xanthomonads is due to the synthesis of a pigment, xanthomonadin (carotenoid-like pigments) (Jenskins et al., 1985). Another characteristic to the genus is the large production of extracellular polysaccharides (EPS), which are responsible for the mucoid aspect of the colonies on solid medium. This compound, also called xanthan

gum and known as additive E415, is used as a food thickener and stabilizing agent since 1968.

Xanthomonas causes important diseases on a wide variety of plant species. The 27 species (Parkinson et al., 2009) of the genus are pathogenic to at least 120 monocotyledons species and 270 dicotyledons species, including fruit trees, nut trees, solanaceous plants, brassicaceous plants and cereals (Hayward, 1993). Symptoms are variable according to the species and host. Xanthomonads are able to cause necrosis, cankers, spots or blight. They can affect many parts of the plants, like leaves, stems and fruits (Leyns et al., 1984). In their life cycle, these bacteria have generally an epiphytic phase. They enter into the host via wounds, stomata or hydathodes. Individual species can comprise multiple pathovars. *Xanthomonas* pathovars show host plant specificity and tissues specificity. Depending of the pathovar, bacteria are able to colonize the vascular system (e.g. *Xanthomonas oryzae* pv. *oryzae* on rice) or the mesophyll tissue (e.g. *X. oryzae* pv. *oryzicola* on rice). Xanthomonads are well adapted to exploit the great diversity of host plants and their different host tissues.

The genome of *Xanthomonas* bacteria consists of a single circular chromosome of about five megabases (except for the *X. albilineans* genus: 3.7 Mb) (Pieretti et al., 2009). In addition, strains may contain one to several plasmids. For instance, *X. euvesicatoria* strain 85-10 has four plasmids with a size of 2, 19, 38, and 183 kb (Thieme et al., 2005). Presently, around 400 assembled genomes of *Xanthomonas* are available in the NCBI database.

Until 1995, the taxonomy of the genus *Xanthomonas* was based on host specificity. Since the first use of the term *Xanthomonas* in

1939 by Dowson, it was common practice to define a plant-pathogenic *Xanthomonas* isolated from a new host plant as a new *Xanthomonas* species (Vautrein et al., 1995). Vauterin and colleagues published in 1995 a new *Xanthomonas* classification based on DNA-DNA hybridization and correlated their data with pathogenic specialization. Shortly after, 16S rDNA sequences were compared to establish a new taxonomic classification of *Xanthomonas* which consisted of 20 species, as proposed by Hauben and colleagues in 1997. The high level of 16S rDNA identity also led to the creation of phylogenetic groups (Fig. I.4) (Hauben et al., 1997).

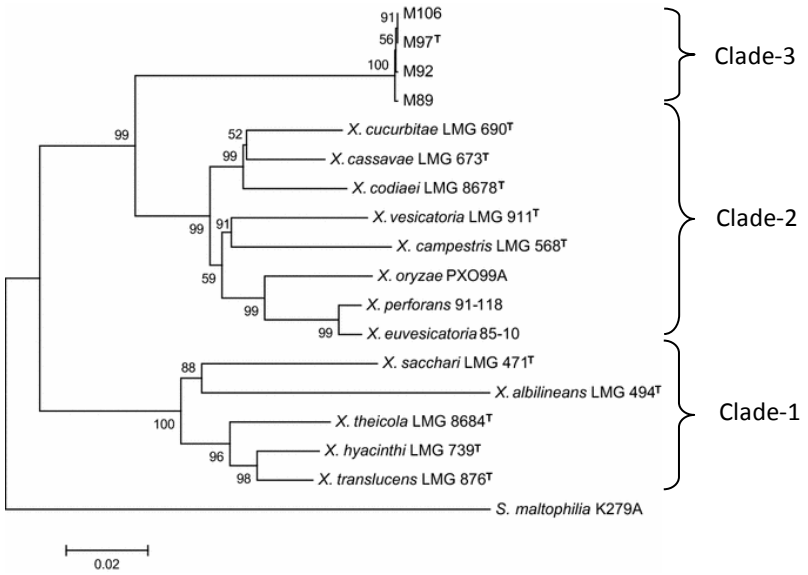


Figure I.4 Phylogenetic tree of *Xanthomonas* spp. based on seven concatenated partial housekeeping genes (*atpD*, *dnaK*, *efp*, *glnA*, *gyrB*, *lepA*, and *rpoD*).
Phylogeny was inferred using the neighbor-joining method with 1,000 bootstrap replicates. Evolutionary distances were computed using the Maximum Composite Likelihood method. Adapted from Triplett et al. (2015).

For the clade-2, which is the largest and best-described group, 16S rDNA sequences have a high level of identity; species are distinguished often by only one or two SNPs (Hauben et al., 1997). In 2008, based on multilocus sequence analysis (MLSA), the taxonomic status of the genus *Xanthomonas* was improved (Young et al., 2008). In this study, clade-1 contained five highly diverse species, *X. albilineans*, *X. sacchari,* *X. theicola*, *X. translucens* and *X. hyacinthi*. The largest group, clade-2, included 15 species such as *X. axonopodis*, *X. oryzae*, *X. campestris*, and an additional group of strains isolated from diseased *Dysoxylum* sp. that could represent a new species. Recently, based on MLSA, 16S rRNA and whole genome sequence analysis, a novel species corresponding to the non-pathogenic rice isolate *Xanthomonas maliensis* was proposed (Triplett et al., 2015). This new species is the founder of a new clade, clade-3 (Fig. I.4).

Xanthomonas classification is still evolving with the identification of new species or reclassification of other taxons (new pathovars or renaming of species) based on new genome comparison techniques. Indeed, during writing of this thesis, a new species, *Xanthomonas pseudoalbilineans*, was proposed to be a member of clade-1 (Pieretti et al., 2015).

Interestingly, most of the *Xanthomonas* species contain a set of *hrp* genes encoding a functional T3SS, but four species, *X. albilineans*, *X. pseudoalbilineans*, *X. sacchari* (all from clade-1) and *X. maliensis* (clade-3), do not possess these genes. Two of the four exceptions, *X. albilineans* and *X. sacchari*, are pathogens of sugarcane (Pieretti et al., 2009; Studholme et al., 2011; Triplett et al.,

2015). In *Xanthomonas* and other plant pathogens, the genes for the T3SS are highly conserved and clustered in neighboring operons in the genome. The genetic organization of this canonical 20-kb region is conserved in clade-2 xanthomonads (Weber et al., 2007).

Some specific factors from *Xanthomonas* are known to contribute to virulence on host plants, including cell wall-degrading enzymes, extracellular polysaccharides, adhesins and secreted proteins. But the key pathogenicity factor is the T3SS and the secreted proteins: T3Es. They are translocated into the host cell where they manipulate the plant cell. The T3SS from *Xanthomonas spp*. translocates a cocktail of different effectors into the plant cell. 37 effectors or effector candidates have been identified up to now in *X. oryzae* pv. *oryzae* ($PXO99^{A}$ strain) (http://www.xanthomonas.org). Data collection of all available *Xanthomonas* genomes has helped in the identification of more than 60 T3Es families based on sequence relatedness (Ryan et al., 2011) (http://www.xanthomonas.org/t3e.html). In *Xanthomonas* spp. one specific family is known to be of the uttermost importance for virulence: the transcriptional activator-like effector or TALE. This largest effector family is widely conserved within *Xanthomonas* spp. (Scholze & Boch, 2010).

5.2. Transcription activator-life effectors (TALEs)

5.2.1 Structure and function of TALEs

Most of the effectors play a specific role at the protein level within the host cell cytoplasm, while TALEs act directly on the host gene expression. They show specific structures and function. TALEs mimic eukaryotic transcriptions factors and are thus able to reprogram the

transcriptome of their target cells (Fig. I.5). TALEs have been shown to activate host resistance genes or susceptibility genes (*S* genes) (Boch & Bonas, 2010b). Induction of an *S* gene is beneficial for the bacteria. It is associated with the disease development and with bacterial multiplication and colonization within the plant tissue. Proteins related to TALEs have been found in the plant-pathogenic genus *Xanthomonas* and in the plant-pathogenic species *R. solanacearum*, where they are called RipTALs (de Lange et al., 2013; Li et al., 2013).

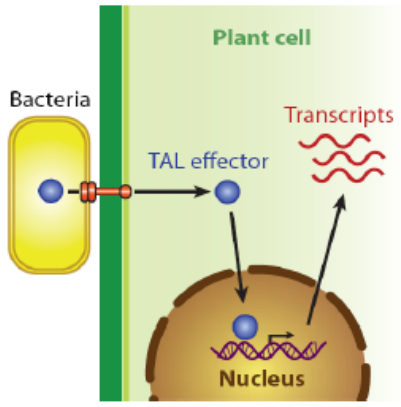


Figure I.5 Mode of action of transcription activator-like effectors.
Xanthomonas translocates a cocktail of effector proteins via the T3SS (red syringe) into plant cells, among them transcription activator-like effectors (TALEs, blue dot). TALEs are imported into the plant cell nucleus, where they induce expression of specific target genes (Boch et al., 2009).

More recently, TALE-like proteins, called Bats, have been identified in *Burkholderia rhizoxinica*, an endosymbiotic bacterium of the fungal plant pathogen *Rhizopus microspores* (de Lange et al., 2014; Juillerat et al., 2014). Surprisingly, TALE-related DNA-binding proteins (MOrTLs) were also found to be encoded in marine bacterial metagenomics sequences (de Lange et al., 2015). TALEs have been

mostly studied in *Xanthomonas* species and a role in virulence was demonstrated for many of them (Muñoz Bodnar et al., 2013).

The first member of the TALE family, AvrBs3 from *X. euvesicatoria,* was identified in 1989 by Bonas et al. AvrBs3 was identified by the elucidation of the incompatibility of *X. euvesicatoria* on pepper (cultivar ECW-30R) by activation of the *R* gene *Bs3*. The second member of the TALE family, PthA from *Xanthomonas citri*, was identified by Yang *et al*. (1995). They were looking for virulence factors of *X. citri* on citrus. These two effectors turned out to share the same structure and the same ability of DNA binding. They were classified as the AvrBs3/PthA family. After different works to elucidate the specificity of DNA binding, this family is now called TALE family and constitutes the best-studied family of T3Es from *Xanthomonas*. Today, more than 200 proteins are associated to the TALE family. The number of TALEs per strain is variable and comprised between 0 for *X. campestris* pv. *campestris* strain 8004 to 28 for *X. oryzae* pv. *oryzicola* strain BLS256 (White et al., 2009).

The TALE protein consists of three regions, the N- and the C-terminal regions, which are highly conserved, combined with a unique central domain of variable length (Fig. I.6). The N-terminal region contains a secretion / translocation signal recognized by the T3SS (Boch & Bonas, 2010a). The C-terminal region harbors one or two nuclear localization signals (NLS) and an acidic transcriptional activation domain for eukaryotic organisms (AAD) (Gürlebeck et al., 2006). NLS motif(s) are essential for the transport of TALEs into the host plant nucleus via an importin (Szurek et al., 2001). Both the NLS motifs as well as the AAD were shown to be essential for virulence.

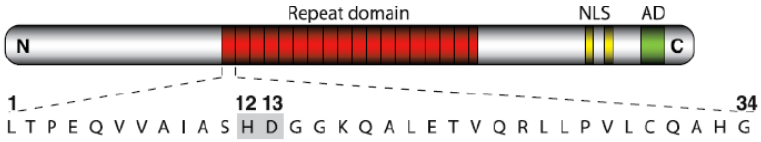


Figure I.6 Schematical structure of a transcription activator-like effector.
TALE contain a central tandem repeats domain, nuclear localization signals (NLSs), and an acidic activation domain (AD). Shown is the amino acid sequence of the first repeat of AvrBs3. Repeat variable diresidues (RVDs) at positions 12 and 13 are shaded in grey (Boch & Bonas, 2010b).

The central protein region is characteristic for TALEs. It consists of an array of tandemly repeated (between 1.5 to 33.5) nearly identical modules of 30-42 amino acids (Lahaye & Bonas, 2001) (Fig. I.6). Each repeat sequence is well conserved, expect for two hypervariable amino acids at position 12 and 13 (Kay et al., 2007). These di-peptides are called Repeat Variable Diresidues (RVDs). This central repeat array mediates DNA binding in a sequence-specific manner at an effector-binding element (EBE) in the plant genome. The binding specificity is governed by the RVDs of each module, following a code where one dipeptide corresponds to one base pair (Boch et al., 2009, Moscou & Bogdanove, 2009) (Fig. I.7). The order of distinct RVDs determines which EBEs are targeted.

Different RVDs have different specificity, as shown in figure I.7. HD is very specific for cytosine while NN is able to bind to guanine and adenine. Furthermore, a minimum of 6.5 repeats is needed for detectable induction of a target gene in a reporter system (Boch et al., 2009). It was also shown that a thymidine or a cytosine at the 5' position (position 0) in front of the targeted EBE is more efficient than other nucleotides (Doyle et al., 2013).

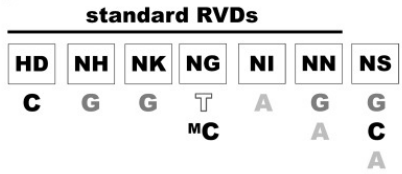


Figure I.7 Over-represented Repeat Variable Diresidues (RVDs) are specificity for one or several nucleotides.
Nuance of grey/black correspond to the strengths of DNA binding. Amino acids correspond to H for histidine, D for aspartate, N for asparagine, K for lysine, G for glycine, I for isoleucine, and S for serine. Nucleobases of the DNA target boxes are G for guanine, T for thymine, A for adenine and C for cytosine. ^{M}C indicates methylated cytosine (Jankele & Svoboda, 2014).

Three years ago, the first crystal structure of a TALE-DNA co-complex was published by Mak *et al*. (2012) and by Deng *et al.* (2012). These high-resolution structures helped to understand the specific interaction between the TALE protein and the DNA molecule (Fig. I.8).

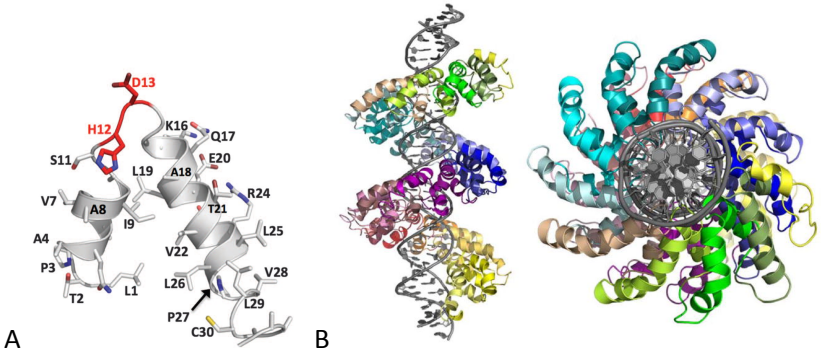


Figure I.8 Three-dimensional structure of the TALE PthXo1 from *Xanthomonas oryzae* pv. *oryzae* in complex with its bound DNA.
A. Structure of a single TALE repeat. The sequence and structure of a representative repeat (number 14) is shown; RVD residues (HD) that recognize cytosine are shown in red. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; H, His; I, Ile; K, Lys; L, Leu; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; and V, Val. B. Structure of the PthXo1 DNA-binding region in complex with its target site. DNA is in grey and each repeat has an individual color (Mak et al., 2012).

Each repeat is formed by two helices separated by the RVD (Fig. I.8A). The three-dimensional structure shows that the repeats self-associate to form a right-handed superhelix wrapped around the major groove of the target DNA (Fig. I.8B). This conformation permits the RVDs to directly contact the corresponding DNA base pair. The first RVD residue stabilizes the conformation of the RVD, while the second makes the specific contact to the DNA sense strand (Mak et al., 2012). The affinity of binding between one RVD and one nucleobase is variable. So called "strong" RVDs, such as HD and NN, make van der Waal contacts and hydrogen bonds, while "weaker" NI and NG repeats have a less affine interaction that creates an unusual contact pattern with a nonpolar van der Waals interaction (Mak et al., 2012) (Fig. I.9). A more recent study revealed a surprising observation: mismatches at the 5' end of the target DNA sequence had more drastic effects on the affinity than mismatches at the 3' end (Meckler et al., 2013).

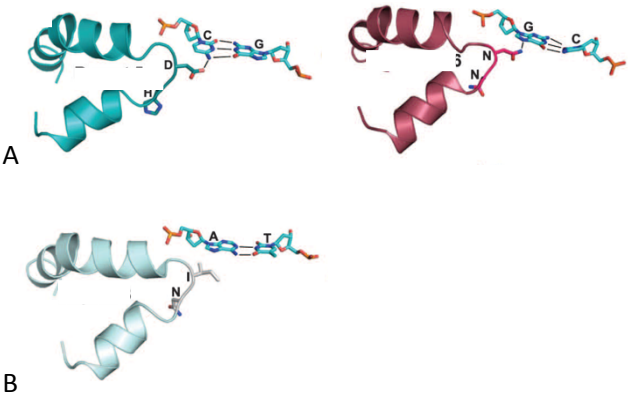


Figure I.9 Binding between TALE repeats and their cognate nucleobases.
A. Interaction of strong RVDs (HD and NN) and nucleobases (C for cytosine and G for guanine). B. Interaction of weak RVD (NI) and nucleobase (A for adenine). Adapted from Mak *et al*., 2012.

5.2.2 Target genes of TALEs causing susceptibility

TALEs are able to modify the gene expression in the plant host, which contributes to the fitness of the pathogen in the absence of the corresponding *R* gene (Bonas et al., 1989). For most of the target genes, their role is still unclear due to either redundancy or the subtlety of their actions. Nevertheless, more and more target genes and their associated TALEs have been identified, thanks to the TALE code. Up to now nine targets, paying a role in susceptibility, were confirmed (Table I.2), but their role during the establishment of the disease is not fully understood yet. Among TALEs, two major functional families are emerging, transporters and transcription factors. The first, most studied family includes the *SWEET*-inducing TALEs. SWEET genes code for sugar and/or copper transporters (Talbot et al., 2010). Two members of this family were identified to be specific targets of *X. oryzae* pv. *oryzae* (Table I.2). These genes were defined as susceptibility (*S*) genes (White et al., 2009). Induction of these genes by the pathogen is important for the disease development. It is thought that *Xanthomonas* bacteria induce the expression of a *SWEET* gene in order to trigger sugar efflux, which serves as a carbon source and allows the bacteria to multiply. In the case of *X. oryzae* pv. *oryzae*, induction of *OsSWEET* genes causes an glucose efflux into the xylem and/or apoplasm where bacteria multiply (Chen et al., 2010, Talbot, 2010). Their induction is thus beneficial for the bacterial multiplication in the xylem vessels. This hijacking of plant nutrients by the pathogen is also found during bacterial blight of cassava caused by *X. axonopodis* pv. *manihotis* (Table I.2) (Cohn et al., 2014).

TALEs are also able to induce different transcriptional factors. This has been first described for AvrBs3 from *X. euvesicatoria,* which induces the expression of *upa20*. UPA20 activates subsequently more than 20 secondary targets, including *upa7*. Activation of *upa7* is then sufficient to cause an hypertrophy of the mesophyll cells (Kay et al., 2007). PthXo6 and PthXo7 from *X. oryzae* are other examples of TALEs that are able to induce transcription factors (Sugio et al., 2007) (Table I.2).

Table I.2 Validated TALE targets, paying a role in susceptibility, for *Xanthomonas* spp.

Bacterial species	TALE	Target gene	Role of the induced protein	Reference
X. euvesicatoria	AvrBs3	*upa20*/*upa16*	Transcription factor, induces hypertrophy of mesophyll cells	Kay et al., 2007 Kay et al., 2009
X. axonopodis pv. *manihotis*	$TAL20_{Xam668}$	*MeSWEET10a*	Sugar transport	Cohn et al., 2014
X. oryzae pv. *oryzae*	PthXo6	*OsTFX1*	Transcription factor	Sugio et al., 2007
X. oryzae pv. *oryzae*	PthXo7	*OsTFIIA1*	Transcription factor	Sugio et al., 2007
X. oryzae pv. *oryzae*	AvrXa7	*OsSWEET14*	Sugar transport	Antony et al., 2010
X. oryzae pv. *oryzae*	TalC	*OsSWEET14*	Sugar transport	Yu et al., 2011
X. oryzae pv. *oryzae*	PthXo1	*OsSWEET11*	Sugar and/or copper transport	Yang et al., 2006
X. oryzae pv. *oryzae*	Tal9a	*OsHen1*	sRNA synthesis	Moscou and Bogdanove, 2009
X. oryzae pv. *oryzicola*	Tal1c	*OsHen1*	sRNA synthesis	Moscou and Bogdanove, 2009
X. citri pv. *citri*	Pth4	*CsLOB1*	Transcription factor, induces cell expansion	Hu et al., 2014

Ten different TALEs from five *Xanthomonas* species have a known target gene. The details of their role are not fully understood yet. TF, transcription factor.

More than 200 available TALE sequences belong to *Xanthomonas* spp. Thanks to the knowledge of the TALE DNA-recognition code and supported by structural and transcriptomic studies, several bioinformatic tools were developed for the prediction of target genes: TALE-NT 2.0 suite (Doyle et al., 2012), TALgetter (Grau et al., 2013), Storyteller (Pérez-Quintero et al., 2013), and TALVEZ (Pérez-Quintero et al., 2013). However, due to prediction of collateral target genes, it is necessary to perform additional functional analyses in order to identify the *bona fide* target gene(s).

5.2.3 Plant responses to the presence of TALEs

As for the other T3Es, plants are able to bypass TALEs as virulence factors. Some plants display polymorphisms in the promoter region of *S* genes. Because of these variations in the DNA sequences, TALEs may not be able to bind anymore. This was shown in the case of *xa13,* which contains polymorphism in the promoter region of *OsSWEET13* (Fig. I.10B) (Römer et al., 2010).

One interesting but not fully understood resistance pathway is *xa5,* which codes for a subunit of transcription factor IIA. TFAII is a general eukaryotic transcription factor with no known role in disease resistance. Substitution of two nucleotides in its gene allows the rice *xa5* gene to interfere with several functions of multiple TALEs (Fig. I.10C) (Iyer & McCouch, 2004).

The second class of resistance linked to TALEs is called "executor *R* genes" (Bogdanove et al., 2010; Zhang et al., 2015). Executor *R* genes exploit TALE activity, i.e. TALE-mediated gene expression, for resistance. In contrast to many other *R* genes,

resistance thus does not depend on protein-protein interaction but relies on the deregulation or induction of silent host genes. This type of resistance is observed for *Bs3* from pepper and *Xa27* and *Xa23* from rice, which recognize AvrBs3 from *X. euvesicatoria* and AvrXa27 and *AvrXa23* from *X. oryzae* pv. *oryzae*, respectively (Gu et al., 2005; Römer et al., 2007; Wang et al., 2015). All three *R* genes contain a corresponding EBE in the promoter region that is targeted by TALEs. Interestingly, *Bs3* and *Xa27* share no sequence homology with any other plant *R* gene. The same holds true for *Xa10,* which confers resistance to strains carrying AvrXa10 (Tian et al., 2014) (Fig. I.10D).

TALEs can also be recognized in the host cell cytoplasm. This is the case in tomato plants that contain the NB-LRR gene *Bs4*. The product of this *R* gene binds to N-terminal part of the AvrBs4-TALE and prevents the translocation of the TALE into the host cell nucleus (Fig. I.10E).

5.2.4 TALEs as biotechnological tools

The specificity of TALE/DNA binding has been elucidated and is currently used for genome editing in a variety of cell types and different organisms (Wright et al., 2014). The possibility to realize precise modifications of different genomes is a major issue in biotechnology. TALEs are also used to specifically activate a certain gene. It is possible to design and synthesize TALEs to target a given gene. One example of a biotechnological tool based on TALEs is the TALEN technology for transcription activator-like effector nuclease (Boch, 2011). Nuclease domains are fused to TALE proteins and act

as a double strand DNA scissor, permitting to induce genomic modifications in a site-specific manner. Recently, TALENs were used to create fragant rice or to engineer resistance of rice plants against *X. oryzae* pv. *oryzae* (Li et al., 2012).

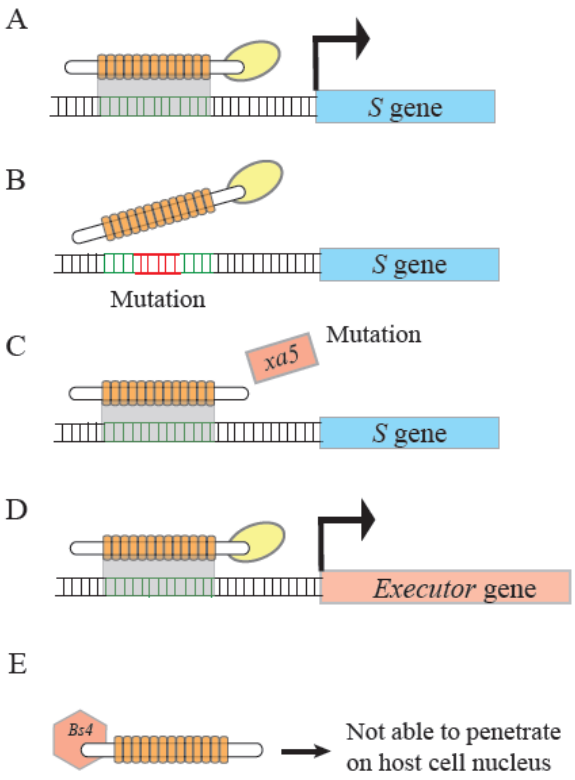


Figure I.10 Resistance pathways against TALEs in diverse pathosystems.
A. Pathway during compatible interaction. The TALE is able to induce the cognate *S* gene, thus causing disease. Activation of *S* genes can be suppressed either by mutations in the individual EBE boxes (B) or by a variation in the RNA II polymerase complex, such as xa5 (C). Plants have also evolved activation traps for resistance in which executor *R* genes are under transcriptional control of EBE boxes that would normally be present in *S* gene promoters (D). TALEs can also be detected in the host cytoplasm. The NB-LRR protein Bs4 recognizes the TALE (AvrBs4) and blocks the its import into the nucleus. Redrawn from Bogdanove et al., 2010.

5.3 Model pathogen of this study: *Xanthomonas translucens*

Before 1995, *Xanthomonas translucens* was classified as *Xanthomonas campestris* based on DNA-DNA hybridization and 16S DNA sequence identity (Vauterin et al., 1995). *X. translucens* is now considered as a distinct species. Ten pathovars have been described within *X. translucens*, using a classification based on the host range. *X. translucens sensu lato* possesses a broad host range including several grasses and small grain cereals (barley, wheat, rye, triticale, some grasses and other *Poaceae*) (Bull et al., 2010). Recently, Giblot-Ducray and colleagues (2009) have defined a new pathovar within the *X. translucens* species. This eleventh pathovar corresponds to *X. translucens* pv. *pistaciae,* which is pathogenic on pistachio trees. This new pathovar increases considerably the host range, which was previously restricted to monocotyledonous plants and now includes dicotyledonous plants. DNA-DNA hybridization and sequence analysis of housekeeping genes have confirmed the affiliation of this bacterium to the *X. translucens* species (Marefat et al., 2006a,b).

The *X. translucens* classification into pathovars is rather confusing and uncertain. It was elaborated first based on the initial host from which the strains were isolated instead of performing an extensive differential host range study. This would have allowed for a clear attribution of the pathovar status, which should have included the use of standard pathogenicity assays. In 1997, the classification was re-evaluated based on the pathogenicity and including pheno- and genotyping tools, such as restriction fragment length polymorphisms (RFLP), sodium dodecyl sulfate-polyacrylamide gel electrophoresis

(SDS-PAGE) of proteins and amplified fragment length polymorphisms (AFLP) (Bragard et al., 1997).

Whereas a few *X. translucens* pathovars attack cereals (Bragard et al., 1997), others are more specific to grasses or are even able to attack woody plants (Rademaker et al., 2006). For these reasons, the *X. translucens* species could be divided in four sub-groups. First, one subgroup encompasses the grasses pathogens, including pathovars *arrhenatheri, graminis, phlei, phleipratensis,* and *poae*. The second sub-group includes all the causal agents of BLS on wheat and other small grain cereals corresponding to the pathovars *hordei, secalis, translucens* and *undulosa* (Vauterin et al., 1991) (Table I.3) and the third subgroup including pathovar *cerealis* are called "translucens group". The pathovar *pistaciae* remains unallocated by now, or constitutes a fourth sub-group.

Table I.3 Association between host plant families and *X. translucens* pathovars

Grasses pathogens	*X. translucens* pv. *phleipratensis* *X. translucens* pv. *poae* *X. translucens* pv. *arrhenatheri* *X. translucens* pv. *phlei* *X. translucens* pv. *graminis*
Small grain cereals pathogens	*X. translucens* pv. *translucens* *X. translucens* pv. *hordei* *X. translucens* pv. *undulosa* *X. translucens* pv. *secalis*
Grasses and small grain cereals pathogens	*X. translucens* pv. *cerealis*
Anacardiaceae pathogen	*X. translucens* pv. *pistaciae*

Four groups are distinguished, pathogens of grasses, pathogens of small grain cereals (such as wheat, barley), pathogens of both, and pathogens of *Anacardiaceae* plants (such as pistachio trees).

The *X. translucens* species belongs to clade-1 of *Xanthomonas* spp., also called "early-branching group" (Fig. I.4) (Parkinson et al., 2007; Triplett et al., 2015) . Currently, this clade is smaller than clade-2 of xanthomonads, but nevertheless it is much more diverse than clade-2.

The focus of my work is the "*translucens* group", especially strains that are pathogenic on barley. Two pathovars of *X. translucens* are concerned and qualified as barley pathogens, which do not attack wheat. They are named *X. translucens* pv. *translucens* and *X. translucens* pv. *hordei* (Jones et al., 1917). Based on taxonomical rules, such pathovars should be considered as strict synonyms (Young et al., 1996). Nevertheless, due do inconsistencies in the host range of the pathotype strains, it was decided for this thesis to keep up with the names that were initially attributed to the strains by the different authors. Additionally, I will also discuss *X. translucens* pv. *undulosa*, which attacks wheat, but also barley, triticale and rye (Smith et al., 1919).

Only a few recent studies about *X. translucens* highlight the peculiarity of this plant-pathogenic bacterium. Nevertheless, the need for a better understanding of the *X. translucens* pathosystem is emphasized by its seed transmission, by the economic importance cereal crops, by its adaptation to warm and humid climates where more and more small grain cereals are cultivated as well as by recent epidemics caused by the bacterium (Adhikari et al., 2012). *X. translucens* is the causal agent of BLS, also known as bacterial leaf stripe on small grain cereals (Jones et al., 1917). The disease is sometimes called black chaff when associated with the glumes. Most

of the host plants belong to the *Poaceae* family. The major hosts of this family are common wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*). Amongst other cultivated species, rye (*Secale cereale*), triticale (*Triticosecale* sp) and *Bromus* sp., which is a forage grass, may be considered as minor hosts of *X. translucens*.

BLS is the most common bacterial disease on cereals. The disease has been reported since the end of the 19th century. The first identification was made on barley in the USA (Jones *et al.*, 1917) and then on wheat (Smith et al. 1919), rye (Reddy et al., 1924), grasses (Wallin, 1946) and finally on triticale (Zillinsky et al., 1971).

X. translucens is able to multiply epiphytically (Duveiller et al., 1991). The bacteria are able to colonize the leaf surface and to enter the host plant via wounds or natural opening like stomata. Symptoms on cereals plants include first translucent stripes, which then elongate and turn into light brown lesions. At a later stage, water-soaked lesions with honey-like exudates are visible on leaves. Disease symptoms on glumes are usually dark brown discolorations of the peduncle with black longitudinal stripes (Fig. I.11).

The colonization process of *X. translucens* during the infection is still not fully understood. Briefly, the bacteria enter through the stomata and multiply to large masses in the parenchyma. This causes elongated streaks limited by the veins, which act as barriers (Duveiller et al., 1997).

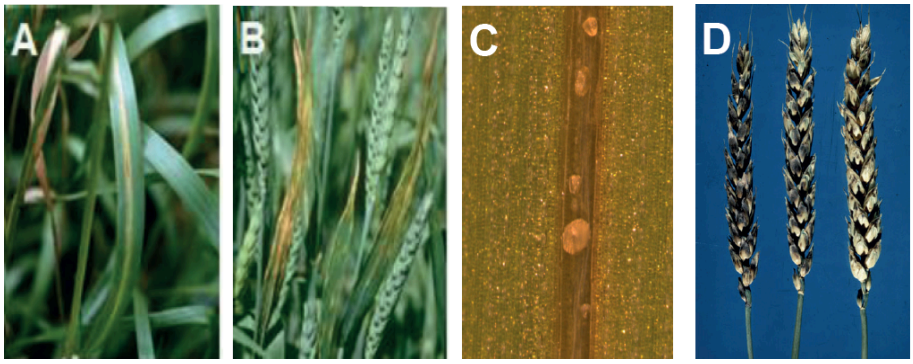


Figure I.11 Bacterial leaf streak and black chaff symptoms on host plants due to *X. translucens*.
A. *X. translucens* pv. *graminis* on annual ryegrass (*Lolium multiflorum*), B. *X. translucens* pv. *translucens* on wheat leaves (*Triticum* spp.), C. bacterial exudates due to *X. translucens* pv. *hordei* on barley leaves (*Hordum vulgare*), D. black chaff on wheat glume (*Triticum* spp.). Pictures are taken from http://www.pflanzenkrankheiten.ch (A), E.A. Milus, University of Arkansas, Bugwood.org (B), and http://www.entofito.com (D).

X. translucens is a seed-borne pathogen, and thus an important constraint for international germplasm exchange (Duveiller et al., 1997). Several countries include *X. translucens* on their quarantine list. For instance, *X. translucens* pv. *translucens* is listed on the A2 list (pests which are locally present in the EPPO region) from the European and Mediterranean Plant Protection Organization (EPPO).

BLS is a sporadic disease with a widespread repartition and it has been reported from very diverse locations worldwide until the end of last century (Fig. I.12). Since 1980s, the disease is more and more frequent and obtains increased attention (Duveiller, 1989; Adhikari et al., 2012). The disease develops very fast under different climatic conditions (sprinkler irrigated fields in temperate climate or in rainfall subtropical environments, warm temperature with cool nights, etc.).

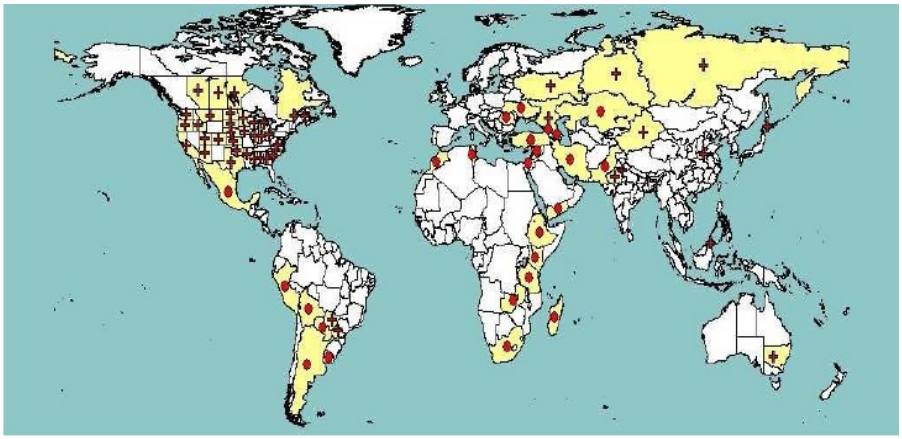


Figure I.12 Worldwide distribution of *X. translucens* pv. *translucens*.
Red circles present national record; red crosses present subnational record. Extract from EPPO database on quarantine pest (2013) (http://www.eppo.int).

Since the eighties of the last century, the number of reported epidemics increased in warm and often humid areas such as in Brazil, Mexico, Iran, Ethiopia (Pereira et al. 1983; Korobko et al., 1985; Alizadeh et al., 1989; Mehta, 1990). More recently, BLS epidemics have been reported repeatedly in the United States. In 2012, BLS on wheat was observed in the Upper Midwest of the United States, particularly in Minnesota, North Dakota and South Dakota (Adhikari et al., 2012). In summer 2015, epidemics on barley have been reported in the Colorado State (Jan Leach, Colorado State University, personal communication). More frequent occurrence of epidemics could be explained by several factors. First, the increase of wheat production (especially winter wheat) favors an increase of the bacterial inoculums. Second, a climate change with milder and more humid weather conditions during the small grain cereals growing season may have supported a general increase of the disease. The permissive range of temperatures for *X. translucens* growth is 15-30°C (Duveiller et al.,

1991), but the optimal temperature is around 26°C under moist condition (>30% of humidity). Quantitative information on losses caused by BLS is limited (Duveiller, 1994). Yield losses as high as 20-40% have been reported (Duveiller et al., 1997), such as in Idaho with yield losses of 40% (Forster 1982). Generally, yield losses are estimated at around 10% (Forster et al., 1986).

Large-scale dissemination and the primary inoculums of *X. translucens* are linked to the seeds. At the field level, bacteria are transmitted by splashes, contact between plants and maybe by insects. Bacteria are able to survive on seeds for up to 63-81 months (Forster et al., 1990). The precise location of *X. translucens* in seeds of cereals seems to be mostly in the external seed coat (Duveiller et al., 1992). Since *X. translucens* is considered to have a very low transmission rate, a zero tolerance of the bacteria may not be required for effective disease management (Schaad et al., 2000). Collectively, *X. translucens* has a broad host range, but alternative hosts are probably not important for the cereals-restricted bacteria to survive during winter (e.g. on grasses) (Fig. I.13).

The most efficient control of BLS on cereals is to avoid sowing of infected seeds, which requires robust and sensitive protocols for detection of *X. translucens* in seed batches in order to certify pathogen-free seeds (Duveiller and Bragard, 1992, Duveiller et al., 1991). There is no direct control measure available. The efficacy of copper-based compounds for controlling the bacteria is limited (Duveiller, 1994). Breeding of plants with resistance against the bacteria is certainly the most durable method to reduce the disease on cultivated cereals.

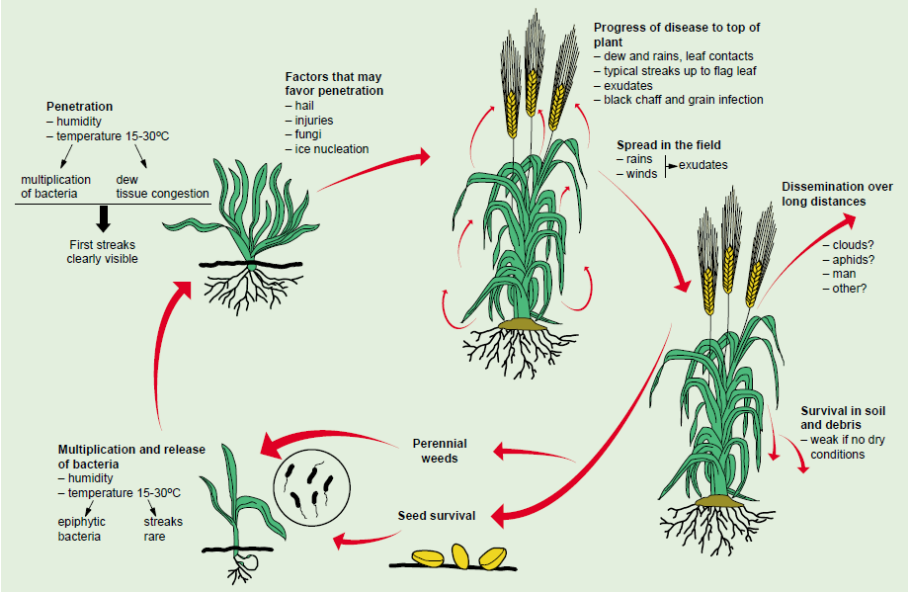


Figure I.13 Bacterial infection cycle of *X. translucens* and possible ways how the disease may spread (Duveiller et al., 1997).

5.4 Hosts of *Xanthomonas translucens*: cereals

Cereals, including maize, rice, wheat and barley, are among the first crops for human consumption. Cereal grains provide more food energy worldwide than any other type of crop. Wheat (700 million tons/year) is the third most-produced cereal after maize (1 billon tons/year) and rice (740 million tons/year) worldwide, before barley with more than 140 million tons per year on average (FAO, 2013; http://faostat.fao.org). Wheat and other small grain cereals (i.e. not including rice and maize) are major crops worldwide. They are considered as important 4F plants for Food, Feed, Fiber and Fuel (http://www.4fcrops.eu/).

Wheat (*Triticum* spp.) is a well-adapted cereal to a wide range of agrosystems and climatic conditions. This grain is grown on more land area than any other commercial food. Moreover, wheat is the

primary source of dietary protein for humans. Wheat genetics is among the most complicated in all cultivated plants. Most wheat varieties are polyploids (tetraploid or hexaploid) (Hancock, 2004). Last year, the International Wheat Genome Sequencing Consortium (IWGSC) published a draft genome of the bread wheat (*T. aestivum*, variety "Chinese spring"). This tetraploid genome contains 21 pairs of chromosomes. The draft genome sequence includes 17 gigabases with 124,201 gene loci (IWGSC, 2014).

In comparison to wheat, the barley (*H. vulgare*) genome is easier to study due to a smaller genome size. The barley genome is diploid, contains only seven chromosomes and has a size of 5.1 gigabases. It is one of the largest diploid genomes sequenced up to now, which has been published in 2012 by the International Barley Genome Sequencing Consortium (IBGSC). An annotation database and BLAST tools are available (http://webblast.ipk-gatersleben.de/barley/). Barley belongs to the *Poaceae* family, making it a model plant for the genetics and genomics within the family, which also includes the polyploid wheat and rye species. Consequently, this well-known crop is now used as a model plant for genetics, phytopathology and biotechnology studies (Hockett et al., 1985; Gurushidze et al., 2014).

Barley is also an important model for understanding ecological adaptation. It is has been cultivated in highly diverse regions worldwide, from the Arctic Circle to the tropics and from -330 m below sea level up to 4200 m above sea level. Barley is widely adapted to various climatic and soil conditions because of its tolerance towards cold, drought, alkali and salinity. This cereal serves mostly

for animal feed (two-thirds of global crop), but also for malting and brewing (one third). To a lesser extent, barley is used for human consumption. In dry areas like Ethiopia, barley is the major source of calories (Mulatu, et al., 2011). In Europe and the USA, barley attracts more and more attention nowadays because it offers possible health benefits like control of cholesterol levels or regulation of blood pressure (Ames & Rhymer, 2008).

Chapter II

Objectives

Xanthomonas species cause important diseases on several major plants including fruit trees, *Brassicaceae*, solanaceous plants, and cereals. Our study focuses on bacterial leaf streak (BLS), caused by *X. translucens*, which is the most common bacterial disease of small grain cereals, such as wheat, barley and triticale. This disease has been reported at diverse locations worldwide until the end of last century and received increased attention in recent years. BLS on barley and wheat is a seed-borne disease and thus a constraint for international germplasm exchange. Several countries list *X. translucens* as a quarantine organism (EPPO A2 list n°183). Yield losses as high as 40 percent have occurred in the most severely diseased fields in the United States, although losses are generally 10 percent or less. BLS is often sporadic, but occurs over a range of very different conditions. Despite its discovery in 1919 - almost 100 years ago! - virtually nothing is known about pathogenicity factors of *X. translucens* on cereals.

Most of the *Xanthomonas* species possess a type III secretion system (T3SS). This secretion system resembles a needle which translocates proteins into the plant cells. The T3SS is well conserved among Gram-negative mammal and plant pathogens. The secretion machinery is encoding by approximately 20 gene localized on a chromosomal cluster called *hrp* cluster. This system is essential for many pathogenic bacteria and well-studied in some *Xanthomonas* species such as the tomato- and pepper-pathogenic bacterium

X. euvesicatoria. Most mutations in the *hrp* gene cluster cause drastic reduction of pathogenicity. Several of the proteins translocated by the T3SS, called type III effectors, are known to play an important role in the disease development. In the *Xanthomonas* genus, one specific T3E family is known to be of particular importance for disease development. This is the family of TALEs for transcription activator-like effectors. These conserved effectors mimic eukaryotic transcription factors and reprogram the target cells gene expression. More than 200 TALEs sequences are available for *Xanthomonas* species but only few target genes have been identified in the host plants. There are basically no data available on the T3SS and TALEs for the *X. translucens*–cereals pathosystem.

My main objective is therefore to unravel pathogenicity mechanisms of *X. translucens* by taking advantage of recent technological advances in high throughput sequencing technologies and by learning from the comparatively well-studied *X. oryzae*–rice pathosystem.

My contribution will be organized around three major topics. First, standardized methods of inoculation had to be established. The work will focus on the barley species which is genetically easier to study. Fives methods will be tested in order to propose reproducible and easy-to-use standard methods. Moreover, fluorescence microscopy will be applied to follow the infection process upon different modes of inoculation. For this purpose, GFP-labeled *X. translucens* strains will be constructed and two modes of inoculations will be examined on barley.

The second objective is to study the peculiarities of the *X. translucens* T3SS. To this end, genomes of three selected *X. translucens* strains will be sequenced and their *hrp* gene clusters will be compared with those from other bacteria which have similar gene clusters. Functional analyses will be performed on different components of the T3SS, using the standard inoculation methods optimized in the previous part.

Once a major role of the T3SS in pathogenicity is established, studies will focus on TALEs from barley-pathogenic *X. translucens*. Based on previous studies, the presence of TALEs was suspected for strains of *X. translucens* and a contribution to pathogenicity could be suspected but was not at all established. Mutational screens combined with functional analyses will be performed in order to identify major virulence TALEs during the interaction of *X. translucens* with barley.

Chapter III

Comparison of inoculation methods of *X. translucens* and development of GFP-labeled bacteria for the study of bacterial leaf streak on barley

The *Xanthomonas translucens* species is poorly studied and characterized. Eleven pathovars have been described, corresponding to a broad host range including grasses, major cereals crops and the dicotyledonous pistachio tree. Classification into pathovars is based on pathogenicity assay, but remains uncertain due to the absence of standardized inoculation methods for *X. translucens*. Moreover, the way strains of *X. translucens* infect the host plant remains unclear. Depending on the pathovar and the host plant, bacteria could behave either as vascular or parenchyme-restricted agents. In this chapter, five methods of inoculation are tested on barley plants. Among them, leaf clipping and leaf infiltration will be retained as standard methods to assess the pathogenicity of wild-type strains and mutants in follow-up studies, in particular for the assessment of the contribution of components of the type III secretion system and of transcription activator-like effectors to the disease. In addition, a microscopic tool has been established with preliminary results that aim to clarify the localization of the bacteria *in planta*.

This work is written in the format of a short communication.

Pesce C, Conejero G, Bragard C, Koebnik R.

Introduction

Xanthomonas translucens was discovered at the beginning of the 20th century as a bacterial pathogen of barley and other small grain cereals. This bacterium is the causal agent of a disease named bacterial leaf streak (BLS). At first evidenced by on barley and named « *Bacterium translucens* » (Jones et al., 1917), the disease was soon reported on wheat as caused by *Bacterium translucens* var *undulosum* (Smith et al., 1919), then later on rye as *Bacterium translucens* var *secalis* (Reddy et al., 1924). Following the new designation of the genus *Xanthomonas* by Dowson (1939), the disease was then extensively studied by Hagborg in Canada, with the proposal of two additional names: *X. translucens* f. sp. *hordei* and *X. translucens* f. sp. *cerealis*. Related bacteria were also discovered on several grasses in the USA and the name *X. translucens* var. *phleipratensis* were proposed (Wallin 1946). Egli and co-workers proposed later the name *X. translucens* pv. *graminis* (Egli et al., 1975) and introduced the additional pathovars *arrhenatheri, phlei* and *poae* (Egli et al., 1982).

In an effort to rationalize the taxonomy of plant-pathogenic bacteria, Young and colleagues proposed in 1977 a new nomenclature. Before, different species within the genus *Xanthomonas* were all grouped under the single species name *Xanthomonas campestris* and were differentiated as pathovars. A pathovar is defined as a group of strains distinguished from other organisms from the same species or subspecies, mainly by characteristics of pathogenicity, such as host range and/or disease symptoms (Bradbury, 1983). The bacteria of the species *Xanthomonas translucens* were then called *X. campestris* and further described as pathovars (i.e. *X. campestris* pv. *translucens*,

X. campestris pv. *undulosa*, *X. campestris* pv. *cerealis*, etc.). Such approach introduced confusion because some authors named all the bacteria they found as *X. campestris* pv. *translucens*, referring to the previous species name, while others kept trying to identify the bacteria at the pathovar level.

Later, Vauterin and co-workers proposed to distinguish the five pathovars affecting small grain cereals (pvs. *cerealis*, *hordei*, *secalis*, *translucens*, *undulosa*) from those recorded on grasses (pvs. *arrhenatheri*, *graminis*, *phlei*, *phleipratensis*, *poae*) (Vauterin et al., 1991). Such approach was encouraged by comparative studies using restriction fragment length polymorphism, amplified fragment length polymorphism, together with additional phenotypical characteristics (Bragard et al., 1995, Bragard et al., 1997). They finally kept the different pathovars under the common species name *X. translucens* (Vauterin et al., 1995).

More recently, additional isolates have been identified as *X. translucens*, but not on the traditionally reported hosts. Rademaker and colleagues identified *X. translucens* on *Asparagus* spp. in the USA (Rademaker et al., 2006). *X. translucens* has been also isolated from a pistachio tree (Facelli et al., 2005).

To unravel the complex situation of the different pathovars within the "*translucens* group", there is a need for i) a standardized and widely recognized inoculation methodology and ii) a better knowledge of the host-bacteria interaction, especially to understand the long distance spread of the bacteria and its seed transmission.

In this chapter, five different inoculation methods will be proposed for *X. translucens*. Additionally, GFP-labeled fluorescent

X. translucens strains will be evaluated as a tool to monitor and assess the spread and *in vivo* localization of the bacteria in barley and other host and non-host plants. These experiments have been set up in order to support the work in the following chapters on the pathogenicity determinants of the *X. translucens* on barley plants.

Materials and methods

Bacterial strains

A selection of fifteen *Xanthomonas translucens* strains, belonging to nine different pathovars, was used for this study. Twelve strains came from the UPB collection (Unité de Phytopathologie et Bactériologie, UCL, Louvain-la-Neuve, Belgium) and three from the French national collection CFBP (Collection Française de Bactéries Phytopathogène, Angers, France). They were selected to be representative of the *X. translucens* diversity, based on previous work of Bragard *et al.* (1997). Plant from which the strains were isolated, their pathovar classification and sampling sites are listed in Table III.1. *Xanthomonas* strains were cultivated at 28 °C in PSA medium (10 g peptone, 10 g sucrose, 1 g glutamic acid, 16 g agar, l^{-1} H2O).

An *Escherichia coli* strain, containing the vector pNEO::*gfp*, was provided by Lui et al. (2008). This replicative vector, compatible with *Xanthomonas* spp., expressed the green fluorescence protein (GFP) under the constitutive glutamate dehydrogenase promoter and contains also a kanamycin resistance gene. The plasmid was introduced into two rifampicin resistant strains of *X. translucens,* UPB886^{R} and UPB787^{R} (Table III.2), by conjugation using pRK2013 as a helper plasmid in a triparental mating (Figurski and Helinski, 1979).

Rifampicin and kanamycin resistant clones were selected upon plating on PSA medium and one isolate was chosen for further experiments. Antibiotics were added to the medium at the following final concentrations: rifampicin, 100 µg/ml; kanamycin, 50 µg/ml; spectinomycin, 100 µg/ml.

Table III.1 ***X. translucens*** **strains from strain collections**

Collection n°	Pathovar	Isolated from	Origin
UBP455*	*arrhenatheri*	*Arrhenatherum* spp.	Switzerland
CFBP 2541*	*cerealis*	*Bromus inermis*	USA
CFBP 2053*	*graminis*	*Dactylis glomerata*	Switzerland
UPB458*	*hordei*	*Hordeum vulgare*	India
UPB684	*hordei*	*Hordeum vulgare*	Iran
UPB820	*hordei*	*Hordeum vulgare*	Iran
UPB886	*hordei*	*Hordeum vulgare*	Iran
UPB906	*hordei*	*Hordeum vulgare*	Iran
UPB441*	*phleipratensis*	*Phleum pratense*	USA
UPB454*	*poae*	*Poa trivialis*	Switzerland
UPB469*	*secalis*	*Secale cereale*	Canada
CFBP 2054*	*translucens*	*Hordeum vulgare*	USA
UPB545	*translucens*	*Hordeum vulgare*	Mexico
UPB787	*translucens*	*Hordeum vulgare*	Paraguay
UPB513	*undulosa*	*Triticosecale*	Mexico

Strains with * are pathotype or neopathotype strains.

All strains and transformants of *X. translucens* were stored at -80°C in glycerol suspensions and retrieved before the plant inoculation, by plating on PSA agar medium supplemented with appropriate antibiotics.

Plant material and plant inoculations

All experiments were done with two barley (*Hordeum vulgare*) cultivars which are commonly used as genetic models, the two-row spring barley cultivar Betzes and the six-row spring barley cultivar Morex the genome of

which was sequenced (IBGSC, 2012). In order to validate the growth conditions and pathogenicity assays, the barley variety Corona, which was used in previous experiments with *X. translucens* was used for comparison (Bragard et al., 1997). Altogether, this choice covers a wide spectrum of barley genotypes.

In addition to barley, other host and non-host plants have been used under specific combinations. These include spring wheat (*Triticum aestivum* cv. Alondra), rice (*Oryza sativa* cv. Azucena) and the dicotyledonous bell pepper (*Capsicum annuum* cv. Early California Wonder line ECW-10R) that contain the resistance gene *Bs1* which recognizes the effector AvrBs1 from *X. euvesicatoria*.

Table III.2 ***Xanthomonas*** **strains and plasmids used in this study**

Plasmids/ *Xanthomonas* strains	**Description**	**Reference**
pRK2013	Helper plasmid for *Xanthomonas* spp. conjugation; Sp^R	Figurski & Helinski, 1979
pNEO::*gfp*	Compatible plasmid for *Xanthomonas* spp. containing *gfp*; Km^R	Lui et al., 2008
$UPB886^R$	Rifampicin-resistant clone	This study
$UPB787^R$	Rifampicin-resistant clone	This study
$UPB820^R$	Rifampicin-resistant clone	This study
CFBP 2054^R	Rifampicin-resistant clone	This study
$UP886^R$ [pNEO::*gfp*]	Fluorescent strain	This study
$UPB787^R$ [pNEO::*gfp*]	Fluorescent strain	This study
$UPB820^R$ [pNEO::*gfp*]	Fluorescent strain	This study
CFBP 2054^R [pNEO::*gfp*]	Fluorescent strain	This study

Resistances to the antibiotics kanamycin, spectinomycin and rifampicin are annotated as Km^R, Sp^R and Rf^R, respectively.

Inoculation methods

A comparison of five different inoculation methods was performed, with a limited set of strains and based on previously published methodologies:

- Pseudo-stem inoculation proposed by Bragard et al. (1997) was used on the different bacterial strains to assess their pathogenicity on the barley cvs. Morex and Betzes as well as on bread wheat cv. Alondra. Pseudo-stem inoculation consisted first of water infiltration into the pseudo-stem with the help of a syringe, followed by application of a bacterial suspension originating from agar plates around the hydrated lesion. Lesions were scored ten days after inoculation.
- Leaf clipping was performed with scissors that had been briefly soaked in a bacterial suspension (Masuch et al., 1989; Kölliker et al., 2006; Studer et al., 2006; Chaves et al., 2013; Wichmann et al., 2013). Three-week old barley leaves were clipped with a bacterial suspension at optical density at 600 nm (OD_{600}) of 0.5 corresponding to 3 x 10^8 CFU/ml. Sizes of lesions were measured two and three weeks post-inoculation.
- Leaf infiltration was performed with a bacterial suspension at OD_{600} of 0.5, which was injected into the leaf tissue using a needle-less plastic syringe (Bragard et al., 1997). Water soaking and translucent lesions were observed five days after inoculation.
- Seed inoculation involved soaking of seeds in a bacterial suspension at OD_{600} of 1 during 4 hours at 22 °C. Immediate after the seeds are sowed in soil. Twelve days after inoculation, water-soaked lesions were observed on the top of the barley leaf.

- Sponge inoculation was proposed as a method with the idea of mimicking the natural field infection of the bacteria. Briefly, a bacterial suspension (OD_{600} = 0.5) applied by rubbing three-week old leaves with a sponge, thus allowing do deliver a continous liquid layer over the inoculated leaf, but without causing any visible damage to the leaf tissue. Disease symptoms were observed ten days after inoculation.

For inoculation experiments, plants were grown with 16 hours of light per day at 22 °C and 50% relative humidity. After inoculations, plants were placed at 100% relative humidity in an incubation chamber during the first 24 hours post inoculation. Visual observations and microscopic analyses were done in the following days after inoculation.

Binocular stereomicroscope imaging of *X. translucens in planta*.

Inoculated barley cv. Morex leaves (leaf clipping and pseuso-stem inoculation) were examined with a binocular Nikon stereomicroscope (model SMZ1500) without sample preparation and magnification.

Localization of *X. translucens* in planta by confocal microscopy.

Emerging symptomatic and inoculated leaves were examined by confocal microscopy. For pseudo-stem inoculations, three centimeters long leaf sections were collected around the lesion area, where symptoms were easily visible. For leaf clipping, sections of three cm were collected two cm away from the inoculation cut. Leaf sections were immobilized in a 3% agarose gel and transversal sections were done with the vibratom (Thermo Scientific Microm HM650V). Sizes

of the slices were between 80 to 100 micrometers. Samples were washed in phosphate-buffered saline (PBS; 0.1 M NaCl, pH 7.2) and then put with PBS on a microscopy slide, covered with a cover slip and fixed with silicone grease. Samples were analysed with a confocal microscope (Carl Zeiss LSM 700; Jena, Germany). Pictures were processed using the Carl Zeiss Confocal Software (Zen, http://microscopy.zeiss.com). The GFP was excited at 488 nm. Four independent leaf clipping experiments and three independent sponge inoculations were done, involving at least two plants per experiment.

Results

All five small grain cereals pathovars (pvs. *cerealis*, *hordei*, *secalis*, *translucens*, *undulosa*) and the pathovar *arrhenatheri* are able to have compatible interactions with barley.

Five methods of inoculation were tested on barley plants. All methods resulted in similar symptoms (Fig. III.1) but the onset of symptom formation differed. First, water-soaked lesions occured around the inoculation area, followed by chlorosis and/or appearance of translucent areas or streaks along the leaf blade. Later, the lesions showed a greasy appearance linked to exopolysaccharide production and at late stage of infection, exudates were observed.

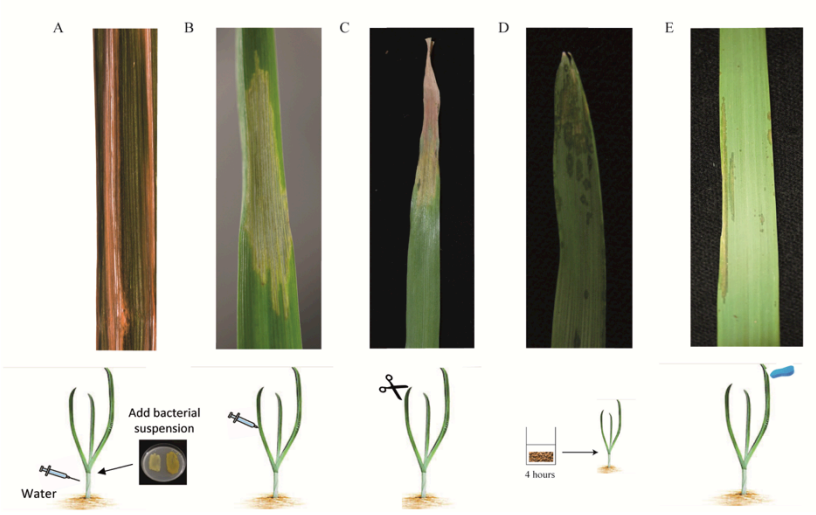


Figure III.1 Inoculation methods of *X. translucens* on barley plants
A. Pseudo-stem inoculation. Symptoms were scored ten days after inoculation. B. Infiltration on barley leaves. Symptoms were observed five days after inoculation. C. Leaf clipping method. Lesion lengths of disease were scored two and three weeks after inoculation. C. Seed inoculation. Seeds were dipped into a bacterial suspension for four hours and symptom development appeared 12 days after inoculation. D. Sponge inoculation. 3-week old leaves are rubbed with a sponge. Symptoms appeared ten days after inoculation.

Pseudo-stem inoculation and infiltration were tested on two cultivars of barley (cvs. Morex and Betzes). For all *X. translucens* strains tested, symptoms observed on Morex were similar or stronger than symptoms on Betzes plants (Figs. III.2 and III.3). Hence, the Morex cultivar appears to be more susceptible to strains belonging to different pathovars, at least when using the pseudo-stem inoculation or infiltration methods.

Figure III.2 Comparison of susceptibility of two varieties of barley (cvs. Morex and Betzes) to infection by *X. translucens* strains
Strain UPB469 (*X. translucens* pv. *secalis*) and CFBP 2541 (*X. translucens* pv. *cerealis*) were inoculated into the barley cvs. Morex and Betzes. Pictures were taken ten days after pseudo-stem inoculation. Symptoms caused by UPB469 on cv. Morex involved extensive lesions (annotated as +), while symptoms caused by CFBP 2541 on cv. Morex corresponded to small restricted lesions (annonoted as +/-). Note that no symptoms were observed for both bacterial strains on the barley cv. Betzes.

Pseudo-stem inoculation was used to compare 15 *X. translucens* strains. In order to obtain semi-quantitative results, three levels of symptom severity were defined (Fig. III.1 and Table III.3). Firstly, after ten days, a "+" corresponds to translucent lesions covering more than 50% of the leaf around the inoculation area. Secondly, the category annotated as "+/–" corresponds to translucent lesions which cover less than 50% of the leaf. Finally, the category "–" indicates that the bacterial strain did not cause any macroscopically visible symptoms 10 days after inoculation (Fig. III.3).

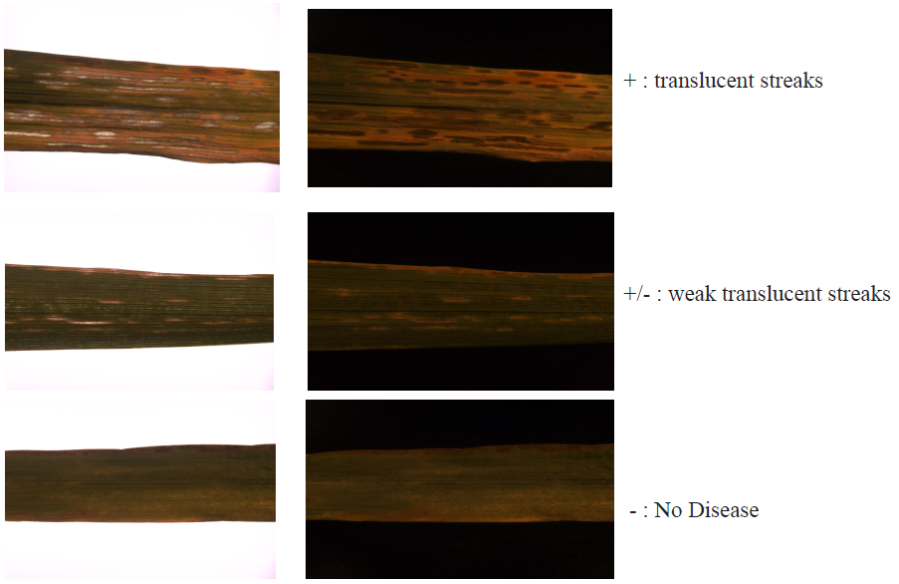


Figure III.3 Comparison of two varieties of barley (cv. Morex) for susceptibility to *X. translucens* strains
Strains UPB6844 (*X. translucens* pv. *hordei*) corresponds to a "+", CFBP 2541 (*X. translucens* pv. *cerealis*) corresponds to the "+/-" and UPB458 corresponds to "-". These three strains were inoculated into Morex barley plants. Pictures were taken ten days after pseudo-stem inoculation.

As expected, pseudo-stem inoculation revealed that almost all "*translucens* group" members caused symptoms on barley cv. Morex (including *translucens*, *hordei*, *undulosa*, *secalis* and *cerealis*) (Table III.3). The pathovar *hordei* strain UPB458 was an exception. For comparison, this strain was also tested on the barley cv. Corona, which was used in previous studies when the pathovar classification was proposed (Bragard et al., 1997). However, under our conditions, neither pseudo-stem inoculation nor infiltration of UPB458 caused symptoms on Corona (data not shown). Excluding the trivial explanation that the strain got mislabeled at some point, this finding

could indicate that pathogenicity got lost during passage in the laboratory.

Table III.3 Pseudo-stem inoculations scored with three levels of disease severity on barley cvs. Morex and Betzes and on the bread wheat cv. Alondra.

Strain	Species / pathovar	Barley		Wheat
		Morex	Betzes	Alondra
CFBP 2541	*X. translucens* pv. *cerealis*	+/–	–	+
UPB458	*X. translucens* pv. *hordei*	–	–	–
UPB684	*X. translucens* pv. *hordei*	+	+/–	–
UPB820	*X. translucens* pv. *hordei*	+	+	–
UPB886	*X. translucens* pv. *hordei*	+	+	–
UPB906	*X. translucens* pv. *hordei*	+	+/–	–
UPB469	*X. translucens* pv. *secalis*	+/–	–	–
CFBP 2054	*X. translucens* pv. *translucens*	+	+	–
UPB545	*X. translucens* pv. *translucens*	+	+/–	–
UPB787	*X. translucens* pv. *translucens*	+	+	–
UPB513	*X. translucens* pv. *undulosa*	+	+	+
UBP455	*X. translucens* pv. *arrhenatheri*	+/–	–	–
CFBP 2053	*X. translucens* pv. *graminis*	–	–	–
UPB441	*X. translucens* pv. *phleipratensis*	–	–	ND
UPB454	*X. translucens* pv. *poae*	–	–	ND

+, strong translucent streaks with water soaking;, +/–, reduced and restricted translucent lesions; –, no symptoms. ND, no determined. Results were scored ten days after inoculation of barley cvs Morex and Betzes and of wheat cv. Alondra.

The second method used was a bacterial infiltration of barley leaves (data not shown). Symptoms were similar to pseudo-stem inoculation with all 15 tested *X. translucens* strains (Table III.3).

In addition to the pseudo-stem inoculation and infiltration, leaf clipping was evaluated as a method that could lead to more quantitative results, thus allowing to compare aggressivity of bacterial strains. Lesion length and disease symptoms for two strains of *X. translucens* pv. *translucens* and *X. translucens* pv. *hordei* and a

reference *X. translucens* pv. *undulosa* strain are shown in Figure III.4. Differences of virulence were noticeable between different strains by leaf clipping, but no lesions were observed for the *X. translucens* pv. *undulosa* strain during the experiment (i.e. until 20 dpi), even if the same strain caused clear symptoms on barley leaves upon infiltration or pseudo-stem inoculation (Table III.3).

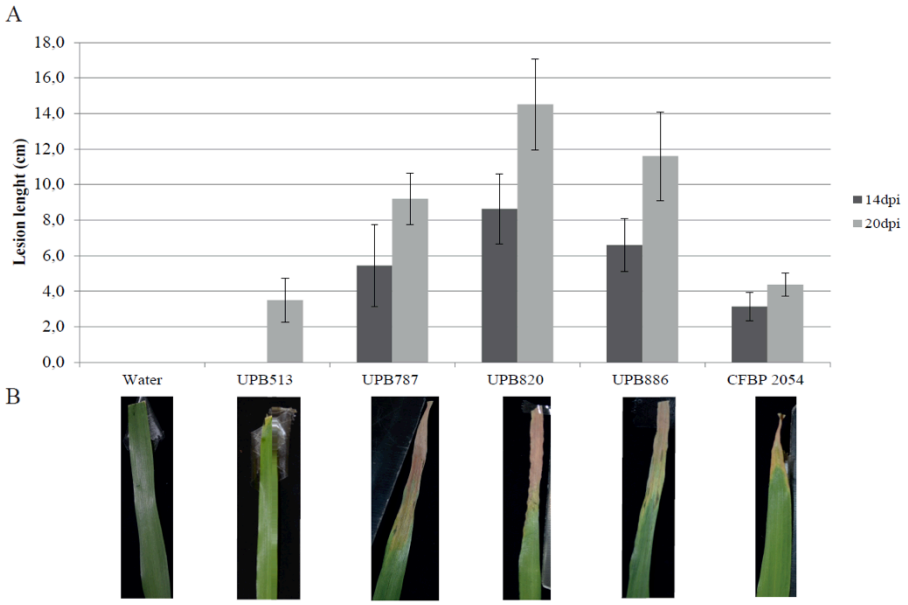


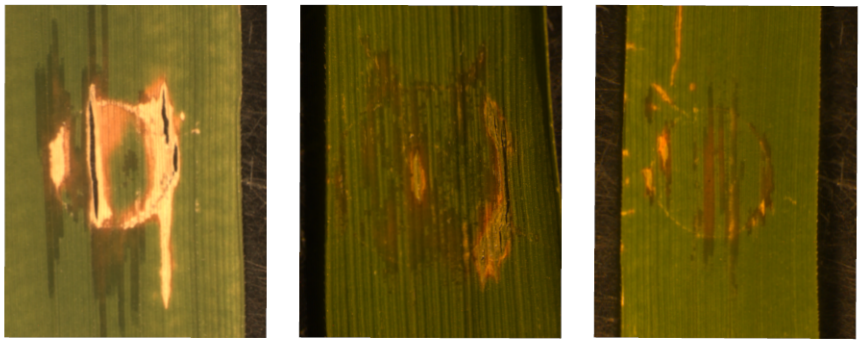
Figure III.4 Disease symptoms of Morex plants upon leaf clip inoculation with *X. translucens*.
A. Lesion lengths were measured 14 and 20 days after inoculation. B. Representative pictures of yellow streaks obtained 14 days after leaf clipping. *X. translucens* pv. *undulosa* (UPB513), *X. translucens* pv. *translucens* (UPB787 and CFPB 2054) and *X. translucens* pv. *hordei* (UPB820 and UPB886) were used.

Strains of the pathovars *translucens* and *hordei* are able to cause symptoms on rice and trigger a hypersensitive reaction on pepper.

X. translucens pvs. *undulosa*, *translucens* and *hordei* (UPB513, UPB787 and UPB820) were also infiltrated into leaves of rice (cv.

Azucena), another monocotyledonous plant of the *Poaceae* family. Seven days after infiltration, typical symptoms corresponding to water soaking in the infiltrated area were present, as observed in barley (Fig. III.5A). Three strains of *X. translucens* representative of three pathovars were also infiltrated into leaves of the dicotyledonous pepper cultivar ECW-10R. This non-host plant contains the resistance gene *Bs1* which recognizes the effector AvrBs1 from *X. euvesicatoria*. 72 hours after infiltration, a hypersensitive response was observed for *X. translucens* pv. *translucens* strain CFBP 2054 and *X. translucens* pv. *cerealis* strain CPBP 2541, but no reaction was visible for the *X. translucens* pv. *graminis* strain CFBP 2053 (Fig. III.5B). This finding correlates with the presence of an *avrBs1* homolog in strain CPBP 2541 and absence of a homolog in strain CFBP 2053 (Pesce et al., 2015a,b).

A



UPB513
X. translucens pv.
undulosa

UPB787
X. translucens pv.
translucens

UPB820
X. translucens pv.
hordei

B

CFBP 2054
X. translucens pv.
translucens

CFBP 2541
X. translucens pv.
graminis

CFBP 2053
X. translucens pv.
cerealis

Figure III.5 Strains of *X. translucens* cause symptoms on rice and can trigger a hypersensitive response on pepper.
A. UBP513 (*X. translucens* pv. *undulosa*), UPB787 (*X. translucens* pv. *translucens*) and UPB820 (*X. translucens* pv. *hordei*) were infiltrated into leaves of Azucena rice plants. Six days after infiltration, water soaking symptoms were observed for all three strains. B. Leaves of ECW-10R pepper plants were infiltrated with CFBP 2054 (*X. translucens* pv. *translucens*), CFBP 2541 (*X. translucens* pv. *cerealis*) and CFBP 2053 (*X. translucens* pv. *graminis*). 72 h after infiltration, hypersensitive responses are observed for the CFBP 2541 and CFBP 2054. No response was visible for CFBP 2053.

GFP-labeled fluorescent *X. translucens* strains serve as a tool to follow long distance spread of the bacteria in barley plants.

The behavior of GFP-labeled fluorescent *X. translucens* strains in barley was analyzed using a binocular stereomicroscope.

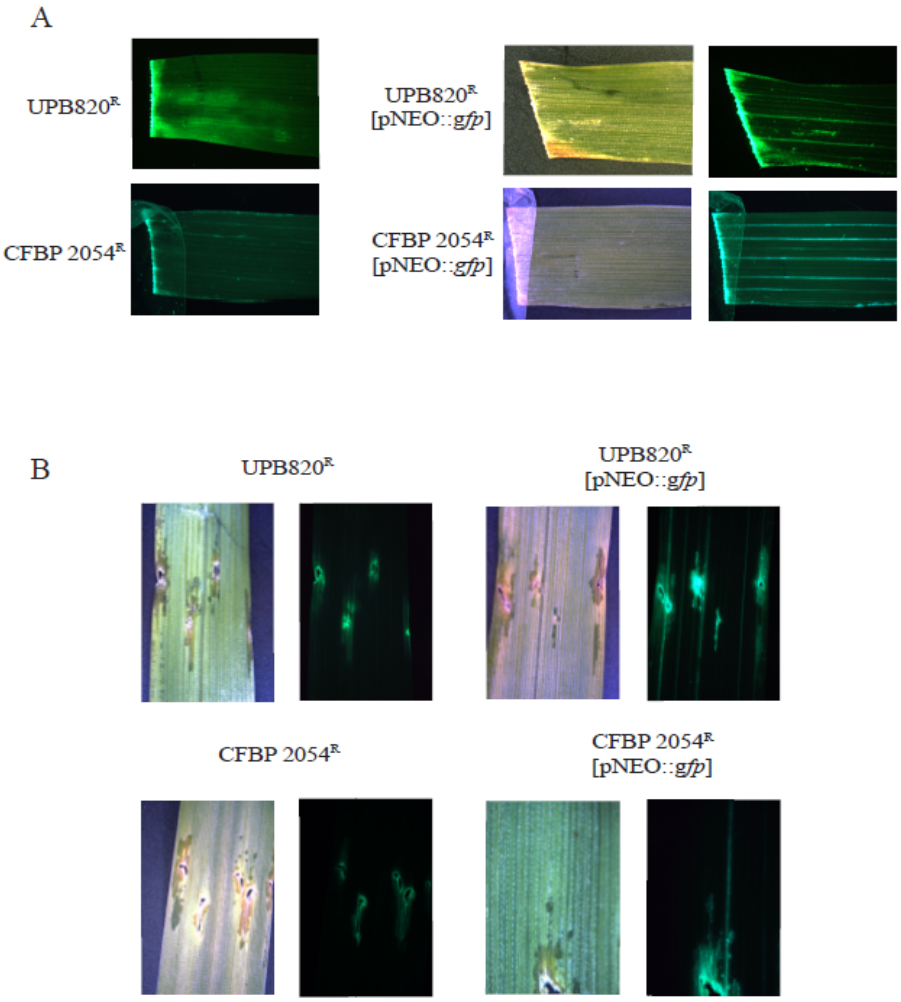


Figure III.6 Localization of GFP-labeld *X. translucens* in barley leaves
A. Leaf clipping inoculation of wild type strains UPB820^R (*X. translucens* pv. *hordei*) and CFBP 2054^R (*X. translucens* pv. *translucens*) on the left side and the corresponding GFP-labeled strains on the right side. B. Same strains were inoculated in the pseudo-stem. Pictures without fluorescence (the left of the paired pictures) are shown for symptom evaluation. For both inoculation methods, pictures were taken 7 days after inoculation on Morex plants.

Upon inoculation of the strains UPB820^{R} (*X. translucens* pv. *hordei)* and CFBP 2054^{R} (*X. translucens* pv. *translucens*), fine green fluorescent lines, co-localizing with the veins, were observed, indicative of a colonization of the xylem vessels by the bacteria (Fig. III.6). This pattern of bacterial localization was obtained using two different methods of inoculation, like leaf clipping (Fig. III.6A) or pseudo-stem inoculation (Fig. III.6B). The same localization was also oberved for the strains *X. translucens* pv. *translucens* UPB787^{R}, *X. translucens* pv. *hordei* UPB886^{R} and *X. translucens* pv. *undulosa* UPB513^{R} (data not shown). However, with the resolution provided by the binocular stereomicroscope, a bacterial colonization within the parenchyma next to the veins could not be excluded.

To improve resolution, confocal imaging of transversal sections was then performed seven days after leaf clipping or sponge inoculation of barley leaves (Fig. III.7). These very first confocal images of *X. translucens* inoculated to barley leaves did not lead to a clear conclusion with respect to tissue specificity. Using strain *X. translucens* pv. *translucens* UPB886^{R} on barley cv. Morex, localization of the bacteria appeared to depend on the inoculation method. Leaf clipping lead to colonization of the vascular tissue, as shown in Figure III.7. Interestingly however, upon sponge inoculation of the same strain, bacteria were observed in symptom-free tissue of the parenchyma (data not shown).

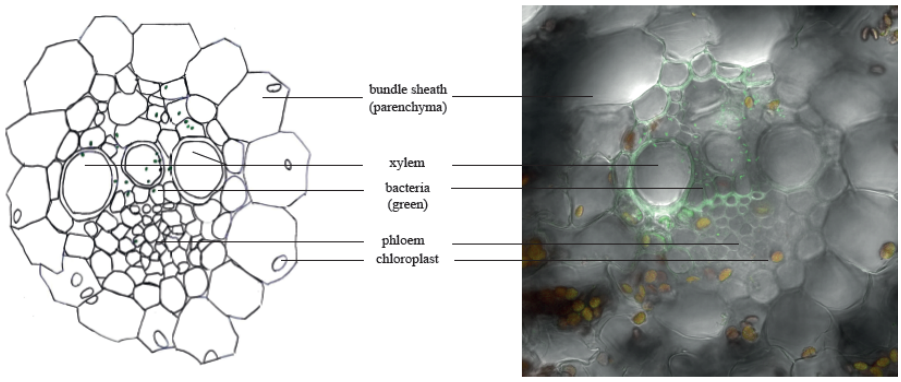


Figure III.7 Confocal imaging of a barley leaf after leaf clipping inoculation with GFP-labeled *X. translucens* pv. *hordei* strain UPB886^{R}.
A. Image of a transversal cut, showing the vascular vessel of a barley leaf inoculated with strain UPB886^{R} (right side) and its schematic representation (left side). The picture was taken seven days after inoculation.

Discussion

Plant-pathogenic bacteria have been commonly named based on the disease they cause. For instance, *X. translucens* was named based on the symptoms observed on cereals and grasses, such as long translucent streaks on leaves. Different pathovars of *X. translucens* were then named mostly based on their host range and sometimes also on differences at the symptom level, thus generating a complex nomenclature. Nowadays, there is a need for rapid and easy-to-use diagnostic methods to identify pathovars. The availability of more and more genome data (e.g. Pesce et al., 2015a,b) as well as advances in molecular biology allows to revisit the current nomenclature and to develop new diagnostic tools.

Despite previous efforts to clarify the nomenclature, some confusion about the different, currently used pathovar designations[1] remains. Such lack of clarity is due to the fast that different definitions have been proposed, most of the time, based on a case-by-case study. Hence, pathovar designations are at risk of wrong assignment when using different panels of cultivars and different inoculation methods for pathogenicity tests.

The host range is difficult to assess.

In this study, we have shown that the host range is difficult to be doubtlessly appraised. Some strains from the same pathovar (e.g. pv. *hordei*) cause different symptoms level depending of the host cultivars of the same species, as shown for the two cultivars Morex and Betzes. Similarly, strains of the pathovars *secalis* or *cerealis* can even cause symptoms or not on different cultivars of barley. Moreover, strains of *X. translucens* from different pathovars were also able to cause symptoms on rice (at least on the cultivar Azucena), a generally not considered host for *X. translucens*. This finding highlights the usefulness of the concept of "natural host range" as compared to the "experimental host range" (Bradbury and Sadler, 2002, Bull and Koike, 2015). Finally, when experimentally determining the host range of strains one needs to make sure to include a representative and

[1] infra-sub-species taxon, defined as a group of strains distinguished from other organisms from the same species or subspecies, mainly by characteristics of pathogenicity. These differences in pathogenicity can be in symptoms or host range. Pathovars may also differ in other characteristics, but such differences must be of minor importance (Bradbury, 1983).

identical panel of host species and cultivars in the tests in order to avoid misnomers.

Standardized inoculation methods are required for different types of studies.

Another reason for the uncertainty of the current pathovar nomenclature is certainly the lack of standardization for the pathogenicity tests. Most of the different studies on *X. translucens* have been performed using different host cultivars, different inoculation methods and even different inocula (bacterial concentrations). Here we have compared different inoculation methods with strains of different pathovars. Each inoculation method has been proposed with a specific purpose. From a conceptual point of view, one should certainly use the method used in the initial pathovar description for achieving a correct designation. However, this is very difficult to achieve taking into account the changes in varieties and cultivars that are available nowadays compared to those that were once distributed and used.

Nevertheless, our study shows that the tested strains of the pathovars *hordei* and *translucens* are pathogenic on barley, and not on the wheat cultivar Alondra based on pseudo-stem inoculation. Similar results have been observed by infiltration. For this qualitative analysis (being pathogenic or not), leaf clipping appeared to be more precise because this inoculation method resulted in major differences between strains of the same pathovar. For instance, *X. translucens* pv. *translucens* is traditionally described as a pathogen of barley but not of wheat, while *X. translucens* pv. *undulosa* is described as a pathogen

of wheat and barley. Strikingly, *X. translucens* pvs. *hordei* and *translucens* cause lesion after leaf clipping between 2.5 cm (CFBP 2054) and 8.2 cm (UPB820) after 14 days post inoculation. Interestingly after 14 days no lesion are observed for the *X. translucens* pv. *undulosa* (UPB513), while strong symptoms are observed after stem inoculation of the strain on the same host variety.

Hence, choice of parameters (host cultivar, inoculum, time point for readout) could influence the pathovar designation, especially when using leaf clipping inoculations. Despite these pitfalls, leaf clipping provides a unique opportunity to follow the colonization of the plant tissue in a quantitative manner, at least for pathogens that cause a systemic infection. This method is therefore ideal to characterize mutants in candidate pathogenicity factors.

Finally, sponge inoculation and seed inoculation are certainly more close to the natural infection process. In this case, however, even using a highly concentrated bacterial suspension as inoculum, the level of colonization remained rather low. Whereas this problem certainly impedes systematic studies using these techniques, they may still find their application in microscopic studies that aim at clarifying the natural colonization process and tissue specificity of the pathogen.

GFP-labeled fluorescent xanthomonads enable to follow the bacterial colonization of the leaf tissues.

The use of GFP-labeled strains helped to study the colonization process of the bacteria following infection. All the strains of the pathovars *translucens* and *hordei* tested infected the mesophyll of barley leaves. This result is in agreement with a previous study by

Shekhawat et al. (1976). Nevertheless, when bacteria were inoculated by the leaf clipping method, they were observed in the xylem vessels of the inoculated plant, thus indicating that *X. translucens* pv. *hordei* may behave like a vascular pathogen, at least when the inoculation method allows the direct entry into the xylem vessels. Within the species *X. translucens*, such vascular behavior was already evidenced for the pathovar *graminis* (Masuch et al., 1989). It is of the uttermost importance to understand, from a practical point of view, if seed transmission of *X. translucens* occurs internal or external of the seeds (Mehta et al., 1990; Mehta et al., 2001). It is conceivable that different pathovars of *X. translucens* use different tissues for colonization, a situation that is reminiscent of the rice-pathogenic xanthomonads which either colonize the vascular bundle (*X. oryzae* pv. *oryzae*) or remain restricted to the apoplasm (*X. oryzae* pv. *oryzicola*) (Ryan et al., 2011). Strikingly, these two rice pathogens are also genetically distinct and can be easily differentiated using a simple multiplex PCR scheme (Lang et al., 2010). Future work using comparative genomics, careful pathogenicity tests and microscopic analyses will therefore clarify the taxonomy of the species *X. translucens* and may allow to develop similar easy-to-use and robust diagnostic tools.

References

Bradbury JF. 1983.The new bacterial nomenclature, what to do. *Phytopathology* **73**, 1349-1350.

Bradbury JF, saddler GS. 2002. A guide to plant pathogenic bacteria. Wallingford, UK:Oxford Univ. press. 2nd ed.

Bragard C, Singer E, Alizadeh A, Vauterin L, Maraite H, Swings J. 1997. *Xanthomonas translucens* from small grains: diversity and phytopathological relevance. *Phytopathology* **87**, 1111-1117.

Bragard C, Verdier V, Maraite H. 1995. Genetic diversity among *Xanthomonas campestris* strains pathogenic for small grains. *Appl Environ Microbiol* **61**, 1020-1026.

Bull CT, Koike ST. 2015**.** Practical benefits of knowing the enemy: modern molecular tools for diagnosing the etiology of bacterial diseases and understanding the taxonomy and diversity of plant-pathogenic bacteria. *Annu Rev Phytopathol* **53**, 157-180.

Chaves A, Mitkowski N. 2013. Virulence of *Xanthomonas translucens* pv. *poae* isolated from *Poa annua*. *Plant Pathol J* **29**, 93-98.

Dowson WJ. 1939. On the systematic position and generic names of the Gram-negative bacterial plant pathogens. *Zentralbl Bakteriol Parasitenk Infektionskr Hyg Abt II* **100**, 177-193.

Egli T, Goto M, Schmidt D. 1975. Bacterial wilt, a new forage grass disease. *J Phytopathol* **82**, 111-121.

Egli T, Schmidt D. 1982. Pathogenic variation among the causal agents of bacterial wilt of forage grasses. *J Phytopathol* **104**, 138-150.

Facelli E, Taylor C, Scott E, Fegan M, Huys G, Noble RD, Swing J, Sedgley M. 2005. Identification of the causal agent of pistachio dieback in Australia. *Eur J Plant Pathol* **112**, 155-165.

Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**, 1648-1652.

Hagborg WA. 1974. Notes on bacterial diseases of cereals and some other crop plants. *Can Plant Dis Surv* **54**, 129-151.

International Barley Genome Sequencing Consortium (IBGSC). 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711-716.

Jones LR, Johnson AG, Reddy CS. 1917. Bacterial blight of barley. *J Agric Res* **11**, 625-643.

Kölliker R, Kraehenbuehl R, Boller B, Widmer F. 2006. Genetic diversity and pathogenicity of the grass pathogen *Xanthomonas translucens* pv. *graminis*. *Syst Appl Microbiol* **29**, 109-119.

Lang JM, Hamilton JP, Diaz MG, Van Sluys MA, Burgos MR, Vera Cruz CM, Buell CR, Tisserat NA, Leach JE. 2010. Genomics-based diagnostic marker development for *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*. *Plant Dis.* **94**, 311-319.

Liu C, Li J, Zhang X, Liu Q, Liu H, Gong P, Zhang G, Yao L. 2008. Stable expression of green fluorescent protein mediated by GCV in *Giardia canis*. *Parasitol Int* **57**, 320-324.

Masuch G, Schoene K, Paul VH. 1989. Histological investigations on the pathogenesis of *Xanthomonas campestris* pv. *graminis* to *Lolium multiflorum*. *EPPO Bulletin* **19**, 73-80.

Mehta YR. 1990. Management of *Xanthomonas translucens* pv. *undulosa* and *hordei* through cereal seed testing. *Seed Sci Technol* **18**, 467.

Mehta YR, Arias CAA. 2001. Inheritance of resistance to *Stemphylium solani* and to its phytotoxin in cotton cultivars. *Fitopathologia brasileira* **26**, 761-765.

Pesce C, Bolot S, Cunnac S, Portier P, Fischer-Le Saux M, Jacques MA, Gagnevin L, Arlat M, Noël LD, Carrère S, Bragard C, Koebnik R. 2015a. High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc* **3**, e01574-14.

Pesce C, Bolot C, Berthelot E, Bragard C, Cunnac S, Fischer-Le Saux M, Portier P, Arlat M, Gagnevin L, Jacques MA, Noël LD, Carrère S, Koebnik R. 2015b. Draft genome sequence of *Xanthomonas translucens* pv. *graminis* pathotype strain CFBP 2053. *Genome Announc* **5**, e01174-15.

Rademaker JL, Norman DJ, Forster RL, Louws FJ, Schultz MH, De Bruijn FJ. 2006. Classification and identification of *Xanthomonas translucens* isolates, including those pathogenic to ornamental asparagus. *Phytopathol* **96**, 876-884.

Reddy CS, Godkin J, Johnson AG. 1924. Bacterial blight of rye. *J Agric Res* **28**, 1039-1040.

Ryan RP, Vorhölter FJ, Potnis N, Jones JB, Van Sluys MA, Bogdanove AJ, Dow JM. 2011. Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat Rev Microbiol* **9**, 344-355.

Shekhawat GS, Patel PN. 1978. Histology of barley plant and rice leaf infected with *Xanthomonas translucens* f. sp. hordei. *J Phytopathol* **93**, 105-112.

Smith EF, Jones LR, Reddy CS. 1919. The black chaff of wheat. *Science* **50**, 48.

Studer B, Boller B, Herrmann D, Bauer E, Posselt U, Widmer F, Kölliker R. 2006. Genetic mapping reveals a single major QTL for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). *Theor Appl Genet* **113**, 661-671.

Vauterin L, Hoste B, Kersters K, Swings J. 1995. Reclassification of *Xanthomonas*. *Int J Syst Bacteriol* **45**, 472-489.

Vauterin L, Swings J, Kersters K. 1991. Grouping of *Xanthomonas campestris* pathovars by SDS-PAGE of proteins. *J Gen Microbiol* **137**, 1677-1687.

Wallin JR. 1946. Parasitism of *Xanthomonas translucens* on grasses and cereals. *Iowa St Coll J Sci* **20**, 171-193.

Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R. 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol Plant Pathol* **14**, 576-588.

Young JM, Dye DW, Bradbury JF, Panagopoulos CF, Robbs CF. 1977). A proposed nomenclature and classification for plant pathogenic bacteria. *N Z J Agric Res* **21,** *153*-177.

Chapter IV

Genomics of *Xanthomonas translucens* to facilitate the discovery of new pathogenicity factors and to develop diagnostic tools

Collectively, pathovars of *Xanthomonas translucens* cause various bacterial diseases on small grain cereals and grasses. Different pathovars can be distinguished from each other by means of DNA-based molecular typing, such as DNA fingerprints or sequencing of one or a few housekeeping genes. Yet, a robust phylogenetic framework is still missing for the *X. translucens* species, and virtually nothing is known about the emergence of epidemic clones and the repertoire of pathogenicity factors in the different species. Because no genome sequence was available for any strain of *X. translucens* at the beginning of this project, we decided to sequence the genomes of three pathotype strains, *X. translucens* pv. *graminis* strain CFBP 2053, which was isolated from *Dactylis glomerata* in Switzerland in 1973, *X. translucens* pv. *cerealis* strain CFBP 2541, which was isolated from *Bromus inermis* in the United States in 1941, and *X. translucens* pv. *translucens* strain CFBP 2054, which was isolated from *Hordeum vulgare* in the United States in 1933 (unpublished data).

Among the different *X. translucens* pathovars, *X. translucens* pv. *cerealis* is unique because of its broad host range. Strains of *X. translucens* pv. *cerealis* cause disease on both small grain cereals (wheat, barley, oat, rye) as well as on grasses (e.g. species of *Agropyron*, *Bromus*, *Dactylis*, *Leymus*, *Lolium*). In contrast, *X. translucens* pv. *graminis* is restricted to grasses, while *X.*

translucens pv. *translucens* causes disease on small grain cereals. All three genome sequences were found to encode a noncanonical hypersensitive response and pathogenicity (Hrp) type III protein secretion system, the genetic organization of which differs from that of clade-2 xanthomonads, as also reported for *X. translucens* pv. *graminis* strain ART-Xtg29 (Wichmann et al., 2013). Two of the genome sequences have been deposited at GenBank. With only 31 contigs (CFBP 2541) and 48 contigs (CFBP 2053), these two genome sequences of pathotype strains are expected to serve as reference sequences. Thanks to the chosen strategy and the very high coverage, we succeeded to assemble the *tal* genes, which are notoriously difficult to be assembled from short reads due to their repetitive nature. While this work was in progress, three other strains of *X. translucens* have been sequenced, which for comparison contain 788 contigs (strain ART-Xtg29), 404 contigs (strain DAR61454), and 551 contigs (strain DSM 18974). Very recently, the first complete genome sequence of *X. translucens* was generated using PacBio technology (pv. *undulosa* strain Xtu 4699).

These new genomic resources will help to study the role of type III effectors, including TAL effectors, in the pathogenicity of *X. translucens*. Moreover, a robust taxonomic framework is expected to emerge from these data, and pathovar-specific diagnostic tools can be developed more easily.

This work was published in two manuscripts.

Pesce C, Bolot S, Cunnac S, Portier P, Fischer-Le Saux M, Jacques MA, Gagnevin L, Arlat M, Noël LD, Carrère S, Bragard C, Koebnik R. 2015. High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc.* 3 (1): e01574-14. doi:10.1128/genomeA.01574-14.

Pesce C, Bolot S, Berthelot E, Bragard C, Cunnac S, Fischer-Le Saux M, Portier P, Arlat M, Gagnevin L, Jacques MA, Noël LD, Carrère S, Koebnik R. 2015. Draft genome sequence of *Xanthomonas translucens* pv. *graminis* pathotype strain CFBP 2053. *Genome Announc.* 3 (5): e01174-15. doi:10.1128/genomeA.01174-15.

High-Quality Draft Genome Sequence of the *Xanthomonas translucens* pv. *cerealis* Pathotype Strain CFBP 2541

Céline Pesce,[a,b] Stéphanie Bolot,[c,d] Sébastien Cunnac,[a] Perrine Portier,[e,f,g,h] Marion Fischer-Le Saux,[e,f,g,h] Marie-Agnès Jacques,[e,f,g] Lionel Gagnevin,[i] Matthieu Arlat,[c,d,j] Laurent D. Noël,[c,d] Sébastien Carrère,[c,d] Claude Bragard,[b] Ralf Koebnik[a]

[a] UMR 186 IRD-Cirad-Université Montpellier 2 Résistance des Plantes aux Bioagresseurs, Montpellier, France
[b] Earth and Life Institute, Applied Microbiology Phytopathology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[c] INRA, Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR 441, Castanet-Tolosan, France
[d] CNRS, LIPM, UMR 2594, Castanet-Tolosan, France
[e] INRA, UMR1345 Institut de Recherche en Horticulture et Semences (IRHS), Angers, France
[f] Agrocampus Ouest, UMR 1345 IRHS, Angers, France
[g] Université d'Angers, UMR 1345 IRHS, SFR4207 QUASAV, PRES L'UNAM, Angers, France
[h] INRA, CIRM-CFBP Collection Française de Bactéries associées aux Plantes, Angers France
[i] UMR Peuplements Végétaux et Bioagresseurs en Milieu Tropical (PVBMT), CIRAD, Saint-Pierre, La Réunion, France
[j] Université Paul Sabatier, Toulouse, France

Running title: Genome Sequence of Pathotype Strain CFBP 2541

E-mail: koebnik@gmx.de

Abstract

Xanthomonas translucens pv. *cerealis* is the causal agent of bacterial leaf streak on true grasses. The genome of the pathotype strain CFBP 2541 was sequenced in order to decipher mechanisms that provoke disease and to elucidate the role of transcription activator-like (TAL) type III effectors in pathogenicity.

Results

Wheat and other small grain cereals are major crops worldwide and are considered important 4F (food, feed, fiber, and fuel) plants. In human consumption, wheat ranks as the second most-produced crop plant after rice, and wheat is grown on more land area than any other commercial crop (see http://faostat3.fao.org/home/E).

Xanthomonas translucens pv. *cerealis* has been found on crops, like wheat (*Triticum* spp.), barley (*Hordeum* spp.), and rye (*Secale cereale*) (1–3), and it also naturally occurs on smooth bromegrass and quack grass (4). Bacterial leaf streak caused by strains of *X. translucens* (5) is the most common bacterial disease of wheat. As a seed-borne disease, it is a constraint for international germplasm exchange (6). The symptoms include translucent stripes at the leaf blade at the early infection state, which later develop into elongated water-soaked lesions, as well as the production of exudates at late infection state (7). While most plant-pathogenic xanthomonads studied thus far belong to the group II clade, the strains of *X. translucens* belong to the group I clade, which also includes the

species *Xanthomonas albilineans*, *Xanthomonas hyacinthi*, *Xanthomonas sacchari*, and *Xanthomonas theicola* (8).

Pathotype strain CFBP 2541 (LMG 679, NCPPB 1944) was isolated from *Bromus inermis* in the United States in 1941. We tested this strain on barley (*Hordeum vulgare* L. Morex and Betzes) and wheat (*Triticum aestivum* L. Alondra) under laboratory conditions. Strong symptoms were obtained with the "Morex" and "Alondra" plants, while "Betzes" remained symptomless.

To obtain new insights into the molecular determinants provoking disease or resistance, we sequenced strain CFBP 2541 using the Illumina HiSeq 2000 platform (GATC Biotech, Germany). The shotgun sequencing yielded 59,447,151 read pairs (26,337,209 100-bp paired-end reads, with an insert size of 250 bp, and 33,109,942 50-bp mate-pair reads, with an insert size of 3 kb). A combination of Velvet (9), SOAP*denovo*, and SOAPGapCloser (10) yielded 31 contigs >500 bp (N_{50}, 1,399,657 bp), with the largest contig being 1,809 kb, for a total assembly size of 4,515,938 bp, corresponding to 1,926 x coverage.

The genome was found to encode a noncanonical hypersensitive response and pathogenicity (Hrp) type III protein secretion system, the genetic organization of which differs from that of clade II xanthomonads, as previously reported for *X. translucens* pv. *graminis* strain Xtg29 (11). In contrast to strain Xtg29, however, the genome assembly of strain CFBP 2541 indicated the presence of two type III transcription activator-like (TAL) effector genes (12, 13), which was supported by Southern blot hybridization. Since *tal* genes are notoriously difficult to be assembled from short reads due to their

repetitive nature, we sequenced the *tal* genes upon PCR amplification. Surprisingly, the two genome-assembled *tal* genes turned out to be correctly assembled, probably due to the very high coverage and a significant number of single-nucleotide polymorphisms (on average, 1 per 10 bp) that distinguish all individual repeats from each other. This information opens the way for studying the role of *tal* genes in the pathogenicity of *X. translucens*.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. JWHD00000000. The version described in this paper is the first version, JWHD01000000.

Acknowledgments

This work was supported by grant ANR-2010-GENM-013 from the French Agence Nationale de la Recherche. C.P. thanks the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (093604) for support. The LIPM is part of the LABEX TULIP (ANR-10-LABX-41). Strain CFBP 2541 is available at the CIRM-CFBP, French Collection for Plant-Associated bacteria (http://www6.inra.fr/cirm_eng/CFBP-Plant-Associated-Bacteria).

References

1. **Rademaker JL, Norman DJ, Forster RL, Louws FJ, Schultz MH, de Bruijn FJ.** 2006. Classification and identification of *Xanthomonas translucens* isolates, including those pathogenic to ornamental asparagus. Phytopathology **96**:876–884. http://dx.doi.org/10.1094/PHYTO-96-0876.
2. **Wagoner P, Schauer A.** 1990. Intermediate wheatgrass as a perennial grain crop, p 143–145. *In* Janick J, Simon JE (eds), Advances in new crops. Timber Press, Portland, OR.
3. **Bragard C, Singer E, Alizadeh A, Vauterin L, Maraite H, Swings J.** 1997. *Xanthomonas translucens* from small grains: diversity and

phytopathological relevance. Phytopathology **87**:1111–1117. http://dx.doi.org/10.1094/PHYTO.1997.87.11.1111.

4. **Fang CT, Allen ON, Riker AJ, Dickson JG.** 1950. The pathogenic, physiological and serological reactions of the form species of *Xanthomonas translucens*. Phytopathology **40**:44–64.
5. **Vauterin L, Hoste B, Kersters K, Swings J.** 1995. Reclassification of *Xanthomonas*. Int J Syst Bacteriol **45**:472–489. http://dx.doi.org/10.1099/00207713-45-3-472.
6. **Duveiller E, Bragard C, Rudolph K, Fucikovsky L.** 1997. The bacterial diseases of wheat: concepts and methods of disease management, p 78. International Maize and Wheat Improvement Center, Estado de México, Mexico.
7. **Smith EF.** 1917. A new disease of wheat. J Agric Res **10**:51–54.
8. **Young JM, Park DC, Shearman HM, Fargier E.** 2008. A multilocus sequence analysis of the genus *Xanthomonas*. Syst Appl Microbiol **31**:366–377. http://dx.doi.org/10.1016/ j.syapm.2008.06.004.
9. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18**:821– 829. http://dx.doi.org/10.1101/gr.074492.107.
10. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J.** 2012. SOAP*denovo*2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **1**:18. http://dx.doi.org/10.1186/2047-217X-1-18.
11. **Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R.** 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. Mol Plant Pathol **14**:576–588. http://dx.doi.org/10.1111/mpp.12030.
12. **Bragard C, Verdier V, Maraite H.** 1995. Genetic diversity among *Xanthomonas campestris* strains pathogenic for small grains. Appl Environ Microbiol **61**:1020–1026.
13. **Boch J, Bonas U.** 2010. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol **48**:419–436. http://dx.doi.org/10.1146/annurev-phyto-080508-081936.

Draft Genome Sequence of *Xanthomonas translucens* pv. graminis Pathotype Strain CFBP 2053

Céline Pesce[a,b] , Stéphanie Bolot[c,d], Edwige Berthelot[a], Claude Bragard[b], Sébastien Cunnac[a], Marion Fischer-Le Saux[e,f,g,h], Perrine Portier[e,f,g,h], Matthieu Arlat[c,d,i], Lionel Gagnevin[a,j], Marie-Agnès Jacques[e,f,g], Laurent D. Noël[c,d], Sébastien Carrère[c,d], Ralf Koebnik[a]

[a] UMR Interactions Plantes Micro-organismes Environnement (IPME), IRD-Cirad-Université Montpellier, Montpellier, France
[b] Earth and Life Institute, Applied Microbiology Phytopathology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[c] INRA, Laboratoire des Interactions Plantes Micro-organismes (LIPM), UMR 441, Castanet-Tolosan, France
[d] CNRS, LIPM, UMR 2594, Castanet-Tolosan, France
[e] INRA, UMR1345 Institut de Recherche en Horticulture et Semences (IRHS), Angers, France
[f] Agrocampus Ouest, UMR 1345 IRHS, Angers, France
[g] Université d'Angers, UMR 1345 IRHS, SFR4207 QUASAV, PRES L'UNAM, Angers, France
[h] INRA, CIRM-CFBP Collection Française de Bactéries associées aux Plantes, Angers France
[i] Université Paul Sabatier, Toulouse, France
[j] UMR Peuplements Végétaux et Bioagresseurs en Milieu Tropical (PVBMT), CIRAD, Saint-Pierre, La Réunion, France

Running title: *Xanthomonas translucens* pv. graminis CFBP 2053 Genome

E-mail: koebnik@gmx.de

Abstract

Strains of *Xanthomonas translucens* pv. *graminis* cause bacterial wilt on several forage grasses. A draft genome sequence of pathotype strain CFBP 2053 was generated to facilitate the discovery of new pathogenicity factors and to develop diagnostic tools for the species *X. translucens*.

Results

Forage grasses, such as bluegrass (*Poa* spp.), bromegrass (*Bromus* spp.), fescue (*Festuca* spp.), oat grass (*Arrhenatherum elatius*), orchardgrass (*Dactylis glomerata*), quack grass (*Agropyron repens*), ryegrass (*Lolium* spp.), and timothy (*Phleum pratense*), are major crops that serve to feed livestock throughout the world. Some grasses, such as fescue, smooth broome and crested wheatgrass, are also commonly used as turf in landscape gardening and for sports grounds. In addition, grasses are of economic interest for revegetation of dumped or burned areas. Last, but not least, different grasses serve the survival of wildlife, for example, as nesting sites of birds, as cover for small animals, as a habitat for foraging raptors, or simply to feed all kinds of animals that depend on grass leaves, shoots, roots or seeds as a source of calories.

All of the above-mentioned grass species can be infected by different pathovars (pvs. *arrhenatheri*, *cerealis*, *graminis*, *phlei*, *phleipratensis*, and *poae*) of *Xanthomonas translucens*. Recently, we sequenced the genome of *X. translucens* pv. *cerealis* pathotype strain CFBP 2541, which was isolated from *Bromus inermis* (1). To gain

further insight into the pathogenicity of grass-pathogenic xanthomonads, we then sequenced another genome of an *X. translucens* pv. *graminis* strain (2). *X. translucens* pv. *graminis* can invade the plant via wounds (e.g., after feeding or mowing), followed by colonization of the protoxylem lacunae and of the adjacent xylem parenchyma (3). Upon entry into the xylem vessels, the pathogen can spread throughout the leaf, leading to symptoms such as wilting of leaves and necrosis of the entire plant (3).

Pathotype strain CFBP 2053 (ATCC 29091, LMG 726, NCPPB 2700), which was isolated from *D. glomerata* in Switzerland in 1973, was sequenced using the Illumina HiSeq 2000 platform (GATC Biotech, Germany). The shotgun sequencing yielded 18,569,445 read pairs (14,855,556 100-bp paired-end reads, with an insert size of 250 bp, and 3,713,889 50-bp mate-pair reads, with an insert size of 3 kb). A combination of Velvet (4), SOAPdenovo, and SOAPGapCloser (5) yielded 48 contigs ≥500 bp (N_{50}, 126,255 bp), with the largest contig being 386,541 bp, for a total assembly size of 4,335,530 bp, corresponding to 691 x coverage. Contigs were scaffolded into two pseudomolecules using *Xanthomonas euvesicatoria* strain 85-10 as a reference sequence (6).

With these characteristics, our new genome sequence is of much better quality than the first *X. translucens* pv. *graminis* genome sequence, which consists of 788 contigs (N_{50}, 8,376 bp) (2). The genome was found to encode a noncanonical hypersensitive response and pathogenicity (Hrp) type III protein secretion system, the genetic organization of which corresponds to that of other *X. translucens* strains (1, 2, 7). Based on the catalog of *Xanthomonas* type III

effectors (http://www.xanthomonas.org), strain CFBP 2053 has the same set of type III effectors as the other sequenced strain of this pathovar, except for *avrBs2*, which is present in ART-Xtg29 but not in CFBP 2053.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number LHSI00000000. The version described in this paper is the first version, LHSI01000000.

Acknowledgments

This work was supported by grant ANR-2010-GENM-013 from the French Agence Nationale de la Recherche. C.P. thanks the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (093604) for support. The LIPM is grateful to the LABEX TULIP (ANR-10-LABX-41). This work benefited from interactions promoted by COST Action FA 1208 (https://www.cost-sustain.org). Strain CFBP 2053 is available at the CIRM-CFBP, French Collection for Plant-Associated bacteria (http://www6.inra.fr/cirm_eng/CFBP-Plant-Associated-Bacteria).

References

1. **Pesce C, Bolot S, Cunnac S, Portier P, Fischer-Le Saux M, Jacques MA, Gagnevin L, Arlat M, Noël LD, Carrère S, Bragard C, Koebnik R.** 2015. High-quality draft genome sequence of the *Xanthomonas translucens* pv. cerealis pathotype strain CFBP 2541. Genome Announc **3**:e01574-14. http://dx.doi.org/10.1128/genomeA.01574-14.
2. **Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R.** 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. Mol Plant Pathol **14**:576–588. http://dx.doi.org/10.1111/mpp.12030.

3. **Masuch G, Schoene K, Paul VH.** 1989. Histological investigations on the pathogenesis of *Xanthomonas campestris* pv. *graminis* to *Lolium multiflorum*. EPPO Bull **19**:73–80. http://dx.doi.org/10.1111/j.1365-2338.1989.tb00131.x.
4. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18**:821–829. http://dx.doi.org/10.1101/gr.074492.107.
5. **Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J.** 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **1**:18. http://dx.doi.org/10.1186/2047-217X-1-18.
6. **Thieme F, Koebnik R, Bekel T, Berger C, Boch J, Büttner D, Caldana C, Gaigalat L, Goesmann A, Kay S, Kirchner O, Lanz C, Linke B, McHardy AC, Meyer F, Mittenhuber G, Nies DH, Niesbach-Klösgen U, Patschkowski T, Rückert C, Rupp O, Schneiker S, Schuster SC, Vorhölter FJ, Weber E, Pühler A, Bonas U, Bartels D, Kaiser O.** 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. vesicatoria revealed by the complete genome sequence. J Bacteriol **187**:7254–7266. http://dx.doi.org/10.1128/JB.187.21.7254-7266.2005.
7. **Gardiner DM, Upadhyaya NM, Stiller J, Ellis JG, Dodds PN, Kazan K, Manners JM.** 2014. Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. PLoS One **9**:e84995. http://dx.doi.org/10.1371/journal.pone.0084995.

Chapter V

The noncanonical Hrp type III secretion system of *Xanthomonas translucens* contains a new putative translocon component and is required to cause a hypersensitive response on non-host plants and to colonize the host plant barley

Most *Xanthomonas* species rely on a highly conserved and complex type III secretion (T3SS) system, which delivers Type III Effectors (T3Es) into the plant host cell. Recent work on *Xanthomonas translucens* pv. *graminis* had demonstrated that the T3SS is important for symptom development but is not required for survival of bacteria *in planta*. Comparative genomics revealed that the noncanonical *hrp* T3SS gene cluster of *X. translucens* does not contain the putative translocon gene *hrpF*, which is present in all clade-2 *Xanthomonas hrp* clusters. Instead, a new gene was discovered in *X. translucens*, called *hgiA*, which could fulfill the function of *hrpF*. Mutagenesis experiments and plant inoculations revealed that components of the T3SS, such as *hrcT* and *hgiA*, strongly contribute to pathogenicity on the host plant barley and to the development of a hypersensitive response on non-host pepper plants.

This work is in preparation for publication.

Pesce C, Büttner D, Bragard C, Koebnik R.
The noncanonical Hrp type III secretion system of *Xanthomonas translucens* contains a new putative translocon component and is required to cause a hypersensitive response on non-host plants and to colonize the host plant barley.

The noncanonical Hrp type III secretion system of *Xanthomonas translucens* contains a new putative translocon component and is required to cause a hypersensitive response on non-host plants and to colonize the host plant barley

Céline Pesce[a,b] , Daniela Büttner[c], Claude Bragard[b], Ralf Koebnik[a]

[a] UMR 186 IRD-Cirad-Université Montpellier IPME, Montpellier, France
[b] Earth and Life Institute, Applied Microbiology Phytopathology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[c] Institute of Biology, Department of Genetics, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany

E-mail: koebnik@gmx.de

Abstract

Xanthomonas translucens is the causal agent of bacterial leaf streak, the most common bacterial disease on cereals such as wheat or barley. To cause disease, most xanthomonads depend on a highly conserved type III secretion system, the hypersensitive response and pathogenicity (Hrp) system, which translocates type III effectors into host plant cells. Type III effectors are delivered to the host cell by a surface appendage, the hollow Hrp pilus, and a translocon that is inserted into the plant cell plasma membrane. Homologs of the HrpF protein, including PopF from *Ralstonia solanacearum* and NolX from rhizobia, are thought to act as a translocon protein. Comparative genomics revealed that strains of *X. translucens* harbor a noncanonical *hrp* cluster, which shares features with type III secretion systems from *Ralstonia solanacearum*, *Burkholderia andropogonis*, *Collimonas fungivorans* and *Uliginosobacterium gangwonense*. Surprisingly, most of these bacteria, including, *X. translucens*, do not encode an *hrpF* homolog. Instead, they possess another gene next to the *hrp* cluster, which is co-regulated with the *hrp* cluster, called *hgiA* (HrpG-induced). Mutagenesis of the conserved type III secretion gene *hrcT* indicates that the *X. translucens* type III secretion system is required to cause disease on the host plant barley and to trigger a non-host hypersensitive response in pepper leaves. Similarly, an insertional mutant in the *hgiA* gene, which is the first gene of a two-gene operon, displayed an Hrp phenotype. The *hgiA* mutant could be partially complemented by either *hgiA* or the downstream gene, *hpaH*. These findings suggest that both HgiA and HpaH contribute to the injection of type III effectors into plant cells.

Introduction

To colonize their host organisms, most *Xanthomonas* species rely on a highly conserved type III protein secretion system (T3SS), which is common to many animal and plants pathogens. The T3SS delivers so-called Type III Effectors (T3Es) into the target host cell without cleavage of classical signal peptide where they act for the benefit of the bacterium (Büttner & Bonas, 2010). Genes encoding the T3SS are typically clustered in a ~20 kb region of the genome. In the case of plant pathogens, the T3SS genes have been collectively called *hrp* (hypersensitive response and pathogenicity) genes since mutants in these genes were found to be impaired in their ability to cause an hypersensitive response (HR) on non-host plants and to have lost pathogenicity on host plants (Büttner & He, 2009). Later, *hrp* genes have been further categorized according to their broad conservation and to subtle phenotypes of some of the corresponding mutants. Eleven *hrp* genes that are conserved among plant and animal pathogens and which are critical for causing disease have been renamed into *hrc* (HR conserved) genes (Bogdanove et al., 1996). For example, Tn*5* transposon insertion mutants in the *hrcT* (syn. *hrpB8*) gene of *X. euvesicatoria* strain 75-3 were unable to cause symptoms in susceptible host plants or to trigger an HR in resistant cultivars (Bonas et al. 1991, Fenselau & Bonas, 1995). Similarly, it was shown that an *hrcT* mutant of *Xanthomonas oryzae* pv. *oryzicola* was unable to induce an HR on tobacco or to cause disease symptoms on rice (Li et al., 2011a). And those genes that contribute to but are not essential for the plant-bacteria interaction are now called *hpa* (*hrp* associated) genes.

Two different types of *hrp* gene clusters have been described for plant pathogens, based on their gene repertoires, operon structures and mode of regulation (Büttner & Bonas, 2002a). Group I comprises *Pseudomonas syringae* and plant-pathogenic enterobacteria while group II includes the *Ralstonia solanacearum* complex and species of *Acidovorax* and *Xanthomonas*. Group-II *hrp* genes and their co-regulated type III effector genes are under control of two key regulatory genes, *hrpG* and *hrpX*. HrpG shares characteristics with two-component response regulators of the OmpR family and is able to induce the expression of *hrpX*. HrpX is an AraC-type transcriptional activator that binds to the plant-inducible promoter (PIP) box (TTCGB-N_{15}-TTCGB), which is present in front of most of the *hrp* operons and many T3E genes (Koebnik et al., 2006).

The first *hrp* gene clusters from *Xanthomonas* have been described in 1991 for *Xanthomonas euvesicatoria* (a.k.a. *Xanthomonas campestris* pv. *vesicatoria*) and for *Xanthomonas campestris* pv. *campestris* (Arlat et al. 1991, Bonas et al. 1991). The core *hrp* cluster consists of six *hrp* operons, *hrpA* to *hrpF*, encompassing a total of eleven *hrc*, seven *hrp*, and three *hpa* genes with identical organization in other *Xanthomonas* strains (Weber et al., 2007).

Expression of the *hrp* cluster results in the formation of the contiguous syringe that spans the bacterial cell envelope. It is predicted to consist of a multi-ring structure that is embedded in both inner and outer membrane and linked to the ATPase complex in the cytosol (Lorenz et al., 2012). In contrast to animal pathogens with their needle complex, plant pathogens have evolved a pilus-like

structure for cross-kingdom protein delivery into the plant's cells, which in the case of *Xanthomonas* is composed of the HrpE protein (Weber et al., 2005). Type III-secreted proteins were found to be in the vicinity of the Hrp pilus during their secretion, suggesting that Hrp pili serve as conduits for the translocation of type III effector proteins (Romantschuk et al., 2001). Entry into the host cells is a final step in protein delivery; for this purpose, animal pathogens have evolved a multi-protein pore-forming translocon complex consisting of YopB and YopD, or homologs thereof (Chatterjee et al., 2013). However, no homologs of YopB or YopD were found in plant pathogens. Based on genetics and biophysical experiments, HrpF was identified to fulfill the role of a translocon protein in *Xanthomonas* (Rossier et al., 2000, Büttner & Bonas, 2002b); similarly, its homolog PopF is required for protein delivery from *R. solanacearum* (Meyer et al., 2006).

Essentially all functional studies of the *Xanthomonas hrp* cluster were done on members of the clade-2, which is the largest and the most conserved clade of xanthomonads and contains the well-studied model species *X. euvesicatoria*, *X. campestris*, and *Xanthomonas oryzae* (Parkinson et al., 2007). By far most of the clade-2 xanthomonads contain a canonical *hrp* cluster with the same genetic organization as in the model strains 85-10 and 8004 (Arlat et al. 1991, Bonas et al. 1991). Only recently, clade-2 strains originating from barley or cannabis have been described that are pathogenic despite the absence of a Hrp T3SS (Ignatov et al., 2015; Jacobs et al., 2015).

Much less work has been performed on members of the clade-1, which comprises the five highly diverse species *Xanthomonas albilineans*, *Xanthomonas sacchari, Xanthomonas theicola*,

Xanthomonas translucens and *Xanthomonas hyacinthi.* For two of them, *X. albilineans* and *X. sacchari,* genomic analyses demonstrated that they do not contain an Hrp T3SS (Pieretti et al., 2009; Studholme et al., 2011). In contrast, the five available genome sequences from four different pathovars of *X. translucens* revealed that all of them contain a Hrp T3SS the genetic organization of which is at variance to those from the clade-2 xanthomonads (Wichmann et al., 2013; Gardiner et al., 2014; Pesce et al., 2015).

X. translucens forms a diverse group of bacteria most of which are pathogenic to monocotyledonous plants, such as various grasses and small-grain cereals (*eg* wheat, barley, rye). Strains have been assigned to pathovars based on symptoms on different host plants, totaling to at least ten pathovars. Mutants in *hrcR*, *hrpE*, and *hrpG* of the forage grass pathogen *X. translucens* pv. *graminis* were impaired in their ability to cause symptoms when compared with the wild-type strain *Xtg*29 (Wichmann et al., 2013). Notably, the *hrpG* mutant, which was the only mutant that was complemented in this study, still caused clearly detectable symptoms, and complementation was only very partially achieved. Moreover, bacterial colonization of the mutants within the plant tissue was hardly impaired, if at all, over a period of two weeks post inoculation, and all mutants were still quantified in substantial numbers (Wichmann et al., 2013). These observations, which are in sharp contrast to those with clade-2 xanthomonads, cast doubts on the importance of the Hrp T3SS for the pathogenicity of *X. translucens*.

In this study, we compared the core *hrp* gene clusters and their flanking sequences of several *X. translucens* strains with those from

other bacterial pathogens, and scrutinized them for novel genes that might be important for pathogenicity. Knockout mutagenesis of a conserved *hrc* gene demonstrates the importance of the Hrp T3SS for pathogenicity of barley-pathogenic *X. translucens* strains. Another gene, which is found next to the core *hrp* cluster and which is conserved among *X. translucens*, was found to be equally important for pathogenicity. Given the absence of *hrpF* homologs in strains of *X. translucens*, we provide evidence that this gene might encode the hitherto undiscovered translocon component of the *X. translucens* T3SS.

Materials and Methods

Bacterial strains and growth conditions

Xanthomonas strains used in this study are listed in table V.1. Strains were cultivated at 28°C in PSA medium (10 g peptone, 10 g sucrose, 1 g glutamic acid, 16 g agar, l^{-1} H_2O). *Escherichia coli* DH10b (Durfee et al., 2008) bacteria, which were used for molecular cloning, were cultivated at 37°C in lysogenic broth (LB).

Rifampicin-resistant *Xanthomonas* mutants were selected upon plating on rifampicin-containing PSA medium at high cell density and one clone was chosen for further experiments. Plasmids were introduced into *E. coli* by thermo-transformation and into *X. translucens* by conjugation using pRK2013 as a helper plasmid in tri-parental mating (Figurski & Helinski, 1979). Antibiotics were added to the medium at the following final concentrations: rifampicin, 100 µg/ml; gentamicin, 20 µg/ml, kanamycin 50 µg/ml.

Plant material and plant inoculations

All plants (barley and pepper) were grown in growth chambers with cycles of 16 hours of light per day at 22 °C and 50% relative humidity.

Plants of the barley (*Hordeum vulgare* L.) cultivar Morex (six-rowed spring barley) were used for virulence assays and to follow the bacterial colonization *in planta*. For inoculation of barley leaves, three-week old plants were cut at about two centimeter below the leaf tip with sterile scissors that have been soaked in a bacterial suspension at optical density at 600 nm (OD_{600}) of 0.5, corresponding to 3 x 10^8 CFU/mL, and symptom development was followed over time (until 15 days post-inoculation). Immediately after inoculation, plants were transferred for 24 hours into a home-built chamber that allowed to reach nearly saturated relative humidity. Disease symptoms were assessed using five replications per condition. Statistical significance of the results was assessed using the Tukey honest significant difference test for post-ANOVA pairwise comparisons, set at 5% ($P < 0.05$). Alternatively, bacterial suspension was applied on the hydrated lesion made by water infiltration into the pseudo-stem with the help of a needle. Symptom development was assayed 10 days post-inoculation. Water was used as negative control for all inoculation and infiltration experiments.

To monitor the ability to trigger a non-host HR, pepper varieties ECW-10R and ECW-20R were used (Kousik & Ritchie, 1999). Bacterial suspensions at an OD_{600} of 0.2 or 0.3 were infiltrated into the leaves of three-week old pepper plants using a needleless syringe and

leaves were scored for an HR at two to eight days after inoculation, depending on the assayed avirulence activity.

Molecular cloning techniques and construction of mutants

Plasmid DNA was isolated using the Wizard® Plus SV Minipreps DNA Purification System (Promega, USA). Restriction enzymes were used according to the manufacturer's recommendations (New England Biolabs, USA). Cloning reactions were performed using the Fermentas ligation kit. Polymerase chain reactions (PCR) were conducted in 20 µL volumes using GoTaq® G2 Polymerase (Promega, USA). Cells of *E. coli* were transformed with plasmid DNA following a thermal shock and the resulting clones were validated by PCR and DNA sequencing. Primer sequences are provided as supplemental material (Suppl. data V.S1).

Chromosomal knockout mutants in *X. translucens* were obtained upon introduction of the suicide vector pVO155 (Oke & Long, 1999) that contained an internal fragment of the target gene. Consequently, single crossing-over events via homologous recombination at the target gene led to gene disruptions. Mutations were confirmed by PCR and DNA sequencing. For primer design, the draft genome sequence of *X. translucens* pv. *translucens* strain DSM 18974 (GenBank accession number CAPJ01000000) was used because of its phylogenetic proximity to other strains of the same pathovar and to strains of the pathovar *hordei*. To knockout the *hrcT* gene, a conserved and essential component of the T3SS, a 419-bp DNA fragment was PCR amplified from genomic DNA of strain CFBP 2054 (a sibling of strain DSM 18974) with oligonucleotide primers

containing unique restriction sites for *Xba*I and *Bam*HI at their 5' ends, thus facilitating subsequent cloning into pVO155. Similarly, 410-bp and 430-bp DNA fragments were amplified from genomic DNA of strain CFBP 2054 and subsequently cloned as *Xba*I-*Bam*HI fragments into pVO155 to create knockouts in *hgiA* and *hgiB*, respectively.

Mutants were complemented using the medium-copy plasmid pBBR1MCS-5 (Kovach et al., 1994) as a vector into which the corresponding DNA fragments were cloned upon PCR amplification from the sequenced *X. translucens* pv. *cerealis* pathotype strain CFBP 2541 (Pesce et al., 2015). All plasmid constructs were checked by PCR, sequenced and introduced into *X. translucens* strains by conjugation. For complementation of *hrcT*, an 825-bp DNA fragment was amplified and cloned using the restriction enzymes *Sal*I and *EcoR*I. To complement the *hgiA* mutation, which was expected to have a polar effect on the downstream *hpaH* gene, a 1347-bp *hgiA* fragment (start to stop codon) was cloned using *Hin*dIII and *Xba*I. Similarly, a 684-bp *hpaH* fragment (start to stop codon) was cloned into pBBR1MCS-5 using *Hin*dIII and *Xba*I.

In planta growth assays

Leaves of four-week old barley plants were infiltrated with a bacterial suspension of *X. translucens* at an OD_{600} of 0.2. One square centimeter leaf segments were collected six hours, two days and five days after infiltration, and ground into a fine powder using the Qiagen TissueLyser system (30 rps for 30 s). Ground material was resuspended in 500 µl of 10 mM $MgCl_2$, and 5-µl drops of a tenfold

dilution series were spotted as triplicates onto selective PSA plates containing rifampicin. Three technical replicates were done for each of the three biological replicates.

Bioinformatic analyses

Database searches were performed using BLAST or PSI-BLAST (Altschul et al. 1990, 1997) For PSI-BLAST searches at NCBI (http://blast.ncbi.nlm.nih.gov), hits with e-values smaller than 10^{-10} were used for iterative cycles. Multiple sequence alignments were generated using MUSCLE (Edgar, 2004) at the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/muscle/). The Artemis genome browser at http://www.sanger.ac.uk/resources/software/artemis/ was used to (re)annotate genomic regions of interest, such as the *hrp* clusters (Rutherford et al., 2000). The consensus sequence logo was generated at the website http://weblogo.berkeley.edu/logo.cgi (Crooks et al., 2004).

Table V.1 ***Xanthomonas*** **strains and plasmids used in this study**

Bacterial strains and plasmids	Description	Reference
pVO155	pUC119-derived suicid vector (Km^R)	Oke and Long, 1999
pVO155::*hrcT*	pVO155 containing a 419-bp internal *hrcT* fragment from CFBP 2054	This study
pVO155::*hgiA*	pVO155 containing a 410-bp internal *hgiA* fragment from CFBP 2054	This study
pVO155::*hgiB*	pVO155 containing a 430-bp internal *hgiB* fragment from CFBP 2054	This study
pBBR1MCS-5	Broad-host-range vector (Gm^R)	Kovach et al., 1995
pBBR1MCS-5::*hrcT*	pBBR1MCS-5 containing *hrcT* from CFBP 2541	This study
pBBR1MCS-5::*hgiA*	pBBR1MCS-5 containing *hgiA* from CFBP 2541	This study
pBBR1MCS-5::*hpaH*	pBBR1MCS-5 containing *hpaH* from CFBP 2541	This study
pBBR1MCS-5::*hrpG**	pBBR1MCS-5 containing the *hrpG** allele which renders *hrp* gene expression constitutive in rich medium	Wengelnik et al., 1999
CFBP 2053^R	*X. translucens* pv. *graminis* (Rf^R)	This study
CFBP 2054^R	*X. translucens* pv. *translucens* (Rf^R)	This study
UPB 787^R	*X. translucens* pv. *translucens* (Rf^R)	This study
UPB 820^R	*X. translucens* pv. *hordei* (Rf^R)	This study
UPB 886^R	*X. translucens* pv. *Hordei* (Rf^R)	This study
CFBP 2054^RΔ*hrcT*	CFBP 2054^R with pVO155 inserted into *hrcT*	This study
UPB 787^RΔ*hrcT*	UPB 787^R with pVO155 inserted into *hrcT*	This study
UPB 820^RΔ*hrcT*	UPB 820^R with pVO155 inserted into *hrcT*	This study
UPB 886^RRΔ*hrcT*	UPB 886^R with pVO155 inserted into *hrcT*	This study
UPB 787^RΔ*hgiB*	UPB 787^R with pVO155 inserted into *hgiB*	This study
UPB 820^RΔ*hgiB*	UPB 820^R with pVO155 inserted into *hgiB*	This study
UPB 886^RΔ*hgiB*	UPB 886^R with pVO155 inserted into *hgiB*	This study
CFBP 2054^RΔ*hgiA*	CFBP 2054^R with pVO155 inserted into *hgiA*	This study
UPB 787^RΔ*hgiA*	UPB 787^R with pVO155 inserted into *hgiA*	This study
UPB 820^RΔ*hgiA*	UPB 820^R with pVO155 inserted into *hgiA*	This study
UPB 886^RΔ*hgiA*	UPB 886^R with pVO155 inserted into *hgiA*	This study

Δ: insertion mutation. Resistances to the antibiotics kanamycin, gentamycin and rifampicin are annotated as Km^R, Gm^R and Rf^R, respectively.

Results

Clade-1 strains of *Xanthomonas* harbor a noncanonical *hrp* cluster, which shares features with type III secretion systems from *Ralstonia solanacearum*, *Burkholderia andropogonis*, *Collimonas fungivorans* and *Uliginosobacterium gangwonense*

Wichmann and colleagues have shown that the genetic organization of the *hrp* cluster of *X. translucens* pv. *graminis* strain ART-*Xtg*29 is distinct to that of all other clade-2 xanthomonads (Wichmann et al., 2013). This finding was later confirmed for other pathovars of *X. translucens* (Gardiner et al., 2014; Pesce et al., 2015). Notably, it was found that the two key regulators *hrpX* and *hrpG* are located within the *hrp* cluster whereas all clade-2 xanthomonads have *hrpG* and *hrpX* at a distinct genomic location between *radA* and *hsp70* (Jacobs et al., 2015). Interestingly, no orthologs of *hrpG* or *hrpX* were found in strains of two other clade-1 species, *X. albilineans* or *X. sacchari* (Pieretti et al., 2009; Studholme et al., 2011).

To better understand the structure and molecular function of the T3SS from different *X. translucens* pathovars, we compared their *hrp* clusters with each other. In *X. translucens*, the *hrpX* gene is always present downstream of *hrcT* and upstream of *hrcC*, which probably forms an operon with *hrpX*. In contrast, clade-2 xanthomonads have their *hrcC* gene immediately downstream of the last gene of the *hrpB* operon, *hrcT*. The second Hrp regulator of *X. translucens*, HrpG, is encoded at the other side of the *hrp* cluster, downstream of *hpaB*. This genetic organization of the *hrp* cluster is reminiscent of *R. solanacearum*, where *hrpX* is called *hrpB* (van Gijsegem et al., 1995; Brito et al., 1999). Most *hrp* operons were found to be preceded

by a PIP box and a properly spaced -10 promoter motif (Fig. V.1) (Koebnik et al., 2006).

A comparison of all five available genome sequences from different pathovars of *X. translucens* revealed that three genes, most likely belonging to two HrpX-controlled operons, are present downstream of *hrcC* (Fig. V.2). First, *xopA* (syn. *hpa1*) is found next to *hrcC*, but transcribed in opposite direction, and contains all elements of a plant-inducible promoter (PIP box and -10 motif), reminiscent of the situation in clade-2 xanthomonads. Further downstream, a divergent

Plant-inducible promoter is predicted to control the expression of an unknown gene, which we will call *hgiA* for *hrpG*-induced gene A, followed by *hpaH* (syn. *hpa2*). In contrast, clade-2 xanthomonads contain *hpaH* and *xopA* next to each other. On the other edge of the core *hrp* cluster, *hrpG* is encoded downstream of *hpaB*, followed by the helper gene *hpaD* and its associated effector gene *xopF*. Further downstream, another unknown gene is found that contains all elements of a plant-inducible promoter (except for strain DAR61454, which has one mismatch in the -10 motif), which we therefore call *hgiB*. *X. translucens* pv. *translucens* strain DSM 18974 contains an additional gene between *xopF* and *hgiB*, which may encode another type III effector. Interestingly, this gene is also present in *X. translucens* pv. *graminis* strain ART-Xtg29 where it, however, appears to be inactivated due to a frameshift mutation.

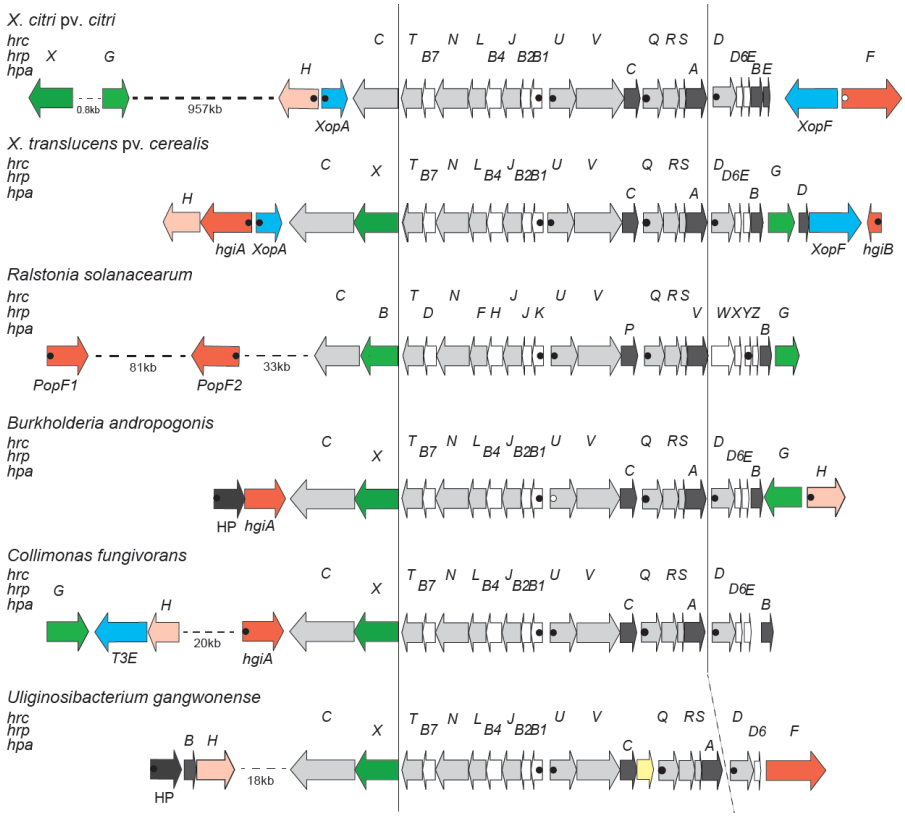


Figure V.1 Schematic overview of the *hrp* gene cluster and flanking regions from *Xanthomonas citri* pv. *citri* (strain 306), *Xanthomonas translucens* pv. *cerealis* (strain CFBP 2541), *Ralstonia solanacearum* (strain GMI1000), *Burkholderia andropogonis* (strain ICMP 2807), *Collimonas fungivorans* (strain Ter331) and *Uliginosibacterium gangwonense* (strain DSM 18521).

Conserved *hrc, hrp* and *hpa* genes are represented by light grey arrows, white arrows and dark grey arrows, respectively. Genes for secreted proteins, key regulators, putative translocons, the *hpaH* gene and mobile genetic elements are represented by blue arrows, green arrows, dark orange arrows, light orange arrow and yellow arrow, respectively. Canonical plant-inducible promoters are indicated by black circles. White circles correspond to putative noncanonical plant-inducible promoters with minor deviations from the consensus sequence or in the spacer lengths between the conserved sequence motifs. Distances between operons are not drawn to scale. T3E, candidate type III effector; HP, hypothetical protein.

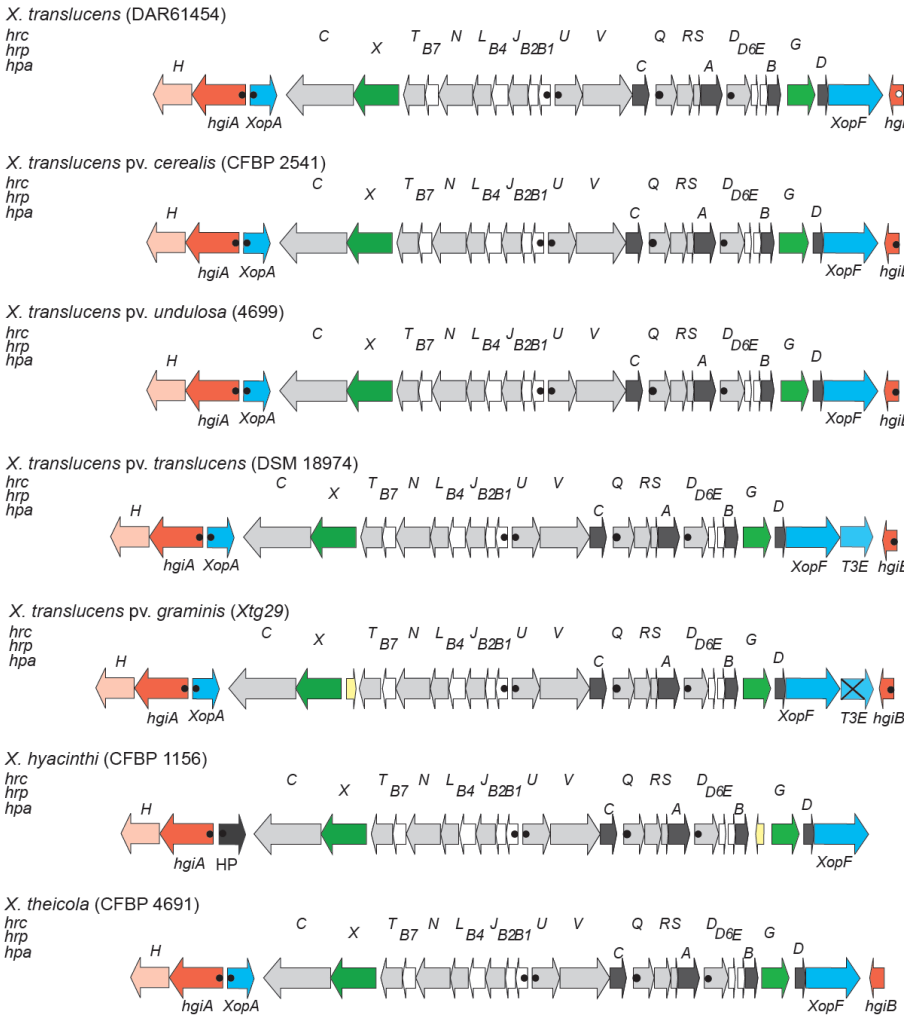
In contrast to all other xanthomonads, strain ART-Xtg29 contains IS element remnants between *hrcT* and *hrpX*, which, however, are unlikely to affect the functionality of the T3SS. From our

collaborators, we got access to the draft genome sequences of two more clade-1 species, *Xanthomonas hyacinthi* strain CFBP 1156 and *Xanthomonas theicola* strain CFBP 4691 (Marion Fischer-Le Saux, personal communication). The genetic organization of their *hrp* clusters appears to be identical to that of the *X. translucens hrp* clusters, except that some remnants of IS elements appear to separate *hpaB* from *hrpG* in strain CFBP 1156 (Fig. V.2).

Surprisingly, we did not find any homolog of the translocon protein HrpF (called PopF in *R. solanacearum*), which is encoded next to the *hrp* cluster in clade-2 xanthomonads and somewhere else in the genome of *R. solanacearum* (Fig. V.1). We also did not find any homologs of several proteins from *Pseudomonas syringae* that have been shown to promote translocation of type III effectors, namely Hrpk1, HrpW1, HrpZ1, and HopAK1 (Kvitko et al., 2007). Neither did we find any homolog of the translocon components from human pathogens (LcrV-YopB-YopD from *Yersinia*, PcrV-PopB-PopD from *Pseudomonas*, SipB-SipC from *Salmonella*, IpaB-IpaC from *Shigella*) (Chatterjee et al., 2013). We therefore speculate that one of the unknown proteins that are encoded next to the *X. translucens* core *hrp* clusters, HgiA or HgiB, could play the role of a translocon. Multiple sequence alignments of both proteins revealed a surprisingly high level of divergence, a hallmark that these proteins share with other translocon proteins (Fig. V.3) (Büttner & Bonas, 2002a; Büttner et al., 2007).

Further insight in the structure and molecular function of the T3SS was expected from a comparison with other, taxonomically distant bacteria. BLAST and PSI-BLAST database searches revealed

similar T3SS exist in the genera *Burkholderia*, *Collimonas*, and *Uliginosibacterium* (Fig. V.1). Strains of these bacteria have the same order of genes from *hrcC* to *hrcD*. Downstream of *hrcD*, *Uliginosibacterium gangwonense* strain DSM 18521 has an *hrpD6* ortholog encoding an 81-amino acid protein, while the other two bacteria encode polypeptides of similar length, which are likely to be orthologs of *hrpD6*. Short open reading frames between *hrpD6* and *hpaB* may encode pilin-like proteins in *Collimonas fungivorans* strain Ter331 and *Burkholderia andropogonis* strain ICMP 2807. Hrp pilin proteins are intrinsically hard to identify due to their small size and because they are under diversifying selection (Guttman et al., 2006; Weber et al., 2007) BLAST searches did not detect an *hrpG* ortholog in the genome of *Uliginosibacterium gangwonense* strain DSM 18521, whereas *B. andropogonis* contains *hrpG* next to *hpaB*, as in *X. translucens*, but in opposite direction. *C. fungivorans* has an *hrpG* ortholog ~22 kb away from the *hrp* cluster in a region that appears to encode at least three type III effectors, in addition to HpaH. Most interestingly, while *U. gangwonense* has an HrpF ortholog, but no homologs of HgiA, the other two bacteria contain HgiA orthologs but do not encode HrpF. This mutual exclusivity suggests that HrpF and HgiA could functionally substitute for each other.



<u>Figure V.2</u> Genetic organization of *hrp* gene clusters in clade-1 *Xanthomonas* strains.

Conserved *hrc, hrp* and *hpa* genes are represented by light grey arrows, white arrows and dark grey arrows, respectively. Genes for secreted proteins, key regulators, putative translocons, the *hpaH* gene and mobile genetic elements are represented by blue arrows, green arrows, dark orange arrows, light orange and yellow arrows, respectively. Canonical plant-inducible promoters are indicated by black circles. White circles correspond to putative noncanonical plant-inducible promoters with minor deviations from the consensus sequence or in the spacer lengths between the conserved sequence motifs. Distances between operons are not drawn to scale. The crossed arrow for *X. translucens* pv. *graminis* Xtg29 corresponds to a frame-shifted variant of a candidate type III effector gene from *X. translucens* pv. *cerealis* DSM 18974.

T3E, candidate type III effector.

A

```
XTG_ART-Xtg29      MSFKSFIKSIGHGITHAAEDVAHGIEQGVKDVAKGVEGVAEHGAQALGDVMQGDFRGGLK
XT_DAR61454        MSFKSFFKSVGHGIAHAAEDVAHGVEQGAKDVAKGVEGVVEHGVNAVGDVMQGDFKGSLH
XTU_Xtu4699        MSFKSFFKSVGHGIAHAAEDVAHGVEQGAKDVAKGVEGVVEHGVNAVGDVMQGDFKGSLH
XTT_DSM18974       MSLKSFFKSIGHGITNAVEDVAHGVEQGAKDIAKGVEGAVEHGAHALGDVMQGDFRGGLH
XTC_CFBP 2541      MSLKSFFKSIGHGITNAVEDVAHGVEQGAKDIAKGVEGAVEHGVHAFGDVLQGDFRGGLH
                   **:***:**:****::*.******:***.**:******..***.:*.***:****.*.*:

XTG_ART-Xtg29      ELGQAGKSIGDTAKAEFKMGLTVGGAMTDLALGTVADAHLNKGLTKLAQSAKGVMDKTRD
XT_DAR61454        ELGQAGKAVGDTAKAEFKLGLAVGGAMTDLALGTVADAHLSKGLTKLAQSAKRGMDKARD
XTU_Xtu4699        ELGQAGKAVGDTAKAEFKLGLAVGGAMTDLALGTVADAHLSKGLTKLAQSAKRGMDKARD
XTT_DSM18974       ELGQAGKAVGDTAKAEFKLGLTVGGAMTDLALGTVADAHLSKGLTKLAETAKGGLDKARD
XTC_CFBP 2541      ELGQAGKAVGDTAKAEFKLGLTVGGAMTDLALGTVADAHLSKGLTKLAETAKGGLDKARD
                   *******::*********:**:******************.*******::**  :**:**

XTG_ART-Xtg29      AVDKSVNQFVDSAEGAVAGGVRCAEDVAHGRLDRLGKDSLSVASNAFNVCTSLTPEGMVA
XT_DAR61454        AVDKSVDQVVDSAEGVVAGGARFADDAVHGRFDRLGKDSLSVASNAFNVCTSLTPEGMAA
XTU_Xtu4699        AVDKSVDQVVDSAEGVVAGGARFADDAVHGRFDRLGKDSLSVASNAFNVCTSLTPEGMAA
XTT_DSM18974       AVDKSLNQVVDSAEGAVAGGVRCVDDVAHGRFDRLGKDSLSVAGNAFNLCTSLTPEGMAA
XTC_CFBP 2541      AVDKSLNQVVDSAEGAVAGGMRCVDDVAHGRFDRLGKDSLSVAGNAFNLCTSLTPEGMVA
                   *****::*.******.**** * .:*..***:***********.****:*********.*

XTG_ART-Xtg29      SVGANLTRTFLSDTPVGKYANVVGDTMARKPLWMLRDSAKDVAEPLLEPLKKDLASADAS
XT_DAR61454        NVGANLTRSFMDGTSLSKYADLVGDTMARKPLWMLRDGAKDVAEPLLDPLKKDLASADAA
XTU_Xtu4699        NVGANLTRSFMDGTSLSKYADVVGDTMARKPLWMLRDGAKDVAEPLLDPLKKDLASADAA
XTT_DSM18974       NVGANLTRSFMDGTSLSKYADVVGDTMARKPLWMLRDGAKDVAEPLLDPLKKDLASADAA
XTC_CFBP 2541      NVGANLTRSFMDGTPLSKYANVIGDTMARKPLWMLRDGAKDVAEPFLDPLKKDLASADAA
                   .*******:*:..*.:.***:::***************.*******:*:***********:

XTG_ART-Xtg29      VDRTLDKAEAGVSRTLDKAEAGVSRTLAEAAEQAQTASLSGAIEYAIARELEKSLQPDDT
XT_DAR61454        VDR-----------TLDKAENGISRTLAEAA-----------------------------
XTU_Xtu4699        VDR-----------TLDKAENGISRTLAEAA-----------------------------
XTT_DSM18974       VDR-----------TLDKAEDGISRTSTEAA-----------------------------
XTC_CFBP 2541      VNRTLDKAENGISGTLDKAENGISRTLAEAA-----------------------------
                   *:*           ****** *:*** :***

XTG_ART-Xtg29      QSRGLGDLSIQIGLGDTGSLSGAMEDAIARELEQSLQPDDTQSRGLGDLSIQIGLGDTGS
XT_DAR61454        ----------------------------------------------------VASGLAVQGD
XTU_Xtu4699        ----------------------------------------------------VASGLAVQGD
XTT_DSM18974       ----------------------------------------------------VASGLAVQGD
XTC_CFBP 2541      ----------------------------------------------------VASGLAVQGG
                                                                       :  **.  *.

XTG_ART-Xtg29      LSGAMEDAIARELEQSLQPDDTQSRGLGDLSIQIGLGDTPSLSGATEDAIARQLADALEG
XT_DAR61454        MTAQ--------------------------------SQTPPLSEAQEDAIARQLADAMER
XTU_Xtu4699        MTAQ--------------------------------SQTPPLSEAQEDAIARQLADAMER
XTT_DSM18974       MTAQ--------------------------------SQTPLLSEAQEDAIAGQLADALQG
XTC_CFBP 2541      MTAQ--------------------------------SQTPQLSEAQEDAIARQLADAMER
                   ::.                                .:** ** * ***** *****::

XTG_ART-Xtg29      KDKQ-------STQDGRGDTPLLSDAQADSIAGQLYDALQDNLSTQDGSGDTPPLSEAQE
XT_DAR61454        KDEQ-----------------------------------------------------
XTU_Xtu4699        KDKQ-------STQDGRGDTPLLSEAQEDSIAGQLADALQDNLSTQDGRGDTPLLSEAQE
XTT_DSM18974       ----------------------------------------NLSTRDGRGDTSLLSDAQE
XTC_CFBP 2541      KDTQSGGQDNLSTQDGRGDTPLLSEAQEDAIAGQLADALNDNLSTQDGRGDTPQLSEAQE


XTG_ART-Xtg29      DSIADQLADDLQDNLATQGGS---------------------------------------
XT_DAR61454        --SGGQV------NLSTQNGS---------------------------------------
XTU_Xtu4699        DSIAGQLADALQGNLSTQDGSGDTPPLSEAQEDAIARQLADAMERKDEQSGGQVNLSTQN
XTT_DSM18974       DAIAGQLADALQGNLSTQDGR---------------------------------------
XTC_CFBP 2541      DAIAGQLADALNDNLSTQDGR---------------------------------------
                      ..*:      **:**.*

XTG_ART-Xtg29      --GDMPPLSEAQQEAIQRQLDDAMERKARQSGRDEGL---------LSASADVSFQFGAV
XT_DAR61454        --GDTPLLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSASADVSFQLRAV
XTU_Xtu4699        GSGDTPLLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSAADVSFQLRAV
XTT_DSM18974       --GDTPQLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSASADVSFQLHTV
XTC_CFBP 2541      --GDTPQLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSASADVSFQLHTV
                     ** * *****::** *** **** * .****:***         ***:******: :*
```

B

```
XTG_ART-Xtg29      --MGIGRCFGSGG-PSARYQQDEHAQGSATPRSPQERDPYDPIALNDPAFAGLTTRRVSQ
XT_DAR61454        MMGNINRCFGSGGSSSARYQQDEHAQGSATPGAPHERDPYQPIVSNDQALAGL-RPRVSN
XTU_Xtu4699        MMGNINRCFGSGGSSSARYQQDEHAQGSATPGAPHERDPYQPIVSNDQALAGL-RPRVSN
XTC_CFBP 2541      MMGNINRCFRSGE-SSARYQQDEHAHGSATPGSPHERDPYQPVASNNPALAGL-RPRVSN
XTT_DSM18974       --MGIGRCFGSGG-SSARYQQDEHAQGSAAPHSPHERDPYQPIALNDPAFAGL-PRRVSQ
                      .*.*** **  .**********:***:* :*:*****:*:. *: *:***   ***:

XTG_ART-Xtg29      VPAHQRGLATNVAGLQREHSRALQLRNQIAQQVANGYPNLQLLQNKVDGILGNIDQKLAG
XT_DAR61454        APVHQRGQAINVAGLQGALRQAEELHSQIDEQVQQGEPGLQSTLESLETTIRKIRGKLAG
XTU_Xtu4699        APVHQRGQAINVAGLQGALRQAEELHSQIDEQVQQGEPGLQSTLESLETTIRKIRGKLAG
XTC_CFBP 2541      APVHQRGVATNEAGLQAVLREAQQLSREIAEQAEHGDPGLQPILESLNKTIRKINGKLNG
XTT_DSM18974       VPAHQRGQVANQAALQEVRREGEELRSQIAEQFQQ-EPALKPILESLDTIIRKINGKLAG
                   .*.**** . * *.**     . :*  :* :*  :  * *:   :.::  : :*  ** *

XTG_ART-Xtg29      TRSGGSGPDEYRHQIEEVRQDFNACNSVGID-PSRHFEHSGQLDNYNPYNRD
XT_DAR61454        SL-GSRGPDDYLAEIEAVREKFNEFNRAAME-PGTNFGYSGSLHSGSF-ALR
XTU_Xtu4699        SL-GSRGPDDYLAEIEAVREKFNEFNRAAME-PGTNFGYSGSLHSGSF-ALR
XTC_CFBP 2541      SL-GSRGPDEYRSEIQQVRRDFNAFNEAAGGYQGANFGYSGVLHSGSFGGGR
XTT_DSM18974       SR-GSRGPDEYRSEIAQVRRDFKAFNEAASGYQGANFGYSGVLHSGSF-GPR
                   :  *. ***:*  :*  ** .*:  * ..    . :* :** * . .
```

Figure V.3 Multiple alignment of the HgiA and HgiB amino acids sequences from five *X. translucens* strains.
A. Multiple alignment of HgiA sequences. B. Multiple alignment of HgiB sequences. The following *Xanthomonas* strains were analyzed: XTG_ART-Xtg29, *X. translucens* pv. *graminis* strain ART-Xtg29; XT_DAR61454, *X. translucens* strain DAR61454, XTU_Xtu4699, *X. translucens* pv. *undulosa* strain Xtu 4699, XTC_CFBP 2541, *X. translucens* pv. *cerealis* strain CFBP 2541, XTT_DSM18974, *X. translucens* pv. *translucens* strain DSM 18974.

The *Xanthomonas translucens hrp* cluster is required to cause disease on barley and to trigger a non-host hypersensitive response in pepper leaves

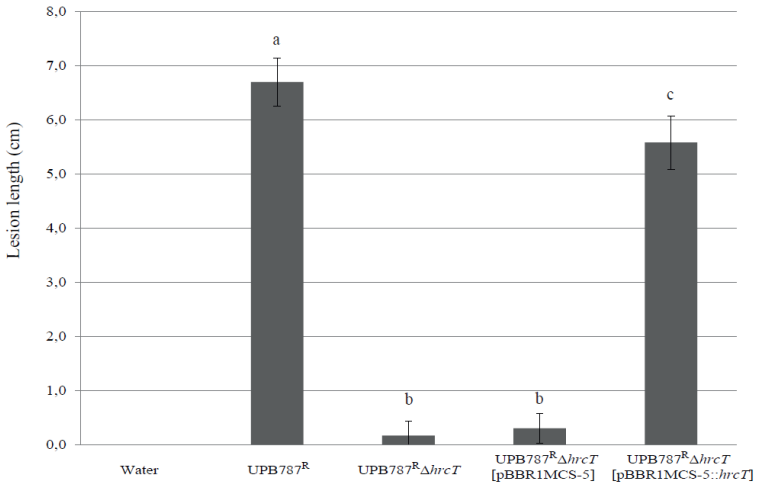
To test the functional role of the *hrp* cluster, knock-out mutants were constructed in one of the conserved *hrc* genes, *hrcT*. This gene was chosen since it is a relatively small gene and it is located at the end of an operon (Fig. V.1), thus facilitating complementation assays. Mutations in *hrcT* of other xanthomonads resulted in a null phenotytpe, i.e. the mutants were unable to cause disease on host plants or to elicit an HR on non-host plants (Li et al., 2011a) or resistant host plants (Bonas et al. 1991, Fenselau & Bonas, 1995). Consequently, *hrcT* mutants were generated in four *X. translucens* strains that are pathogenic to barley (strains CFBP 2054^R, $UPB787^R$,

UPB820^{R} and UPB886^{R}). Data are shown only for the strains UPB787^{R} and 886^{R}.

To evaluate the role of the Hrp T3SS in symptom formation on barley, two different assays were performed. When bacterial suspension was applied on the hydrated lesion localized on stem, *hrcT* mutants of all four strains caused significantly reduced symptoms in comparison to the wild type (Suppl. data V.S2B). To quantify the effect of the mutation on the progression of the disease, bacteria were inoculated by leaf clipping. Two weeks after inoculation, lesion lengths decreased by more than 90% for the mutant strains UPB787^{R}Δ*hrcT* (Fig. V.4A) and UPB886^{R}Δ*hrcT* (Suppl. data V.S2A). When complemented with a plasmid-borne *hrcT* gene from *X. translucens* pv. *translucens* strain CFBP 2054, both mutants fully regained the ability to cause symptoms. (Fig. V.4A and Suppl. data V.S2). These data clearly demonstrate that *hrcT* gene important for T3SS of *X. translucens* is required for symptom formation on barley.

To assess the colonization of the leaf blade during development of the disease, we infiltrated the *hrcT* mutant of strain UPB787^{R} and its complemented derivative into barley leaves. Colonization of the leaf blade was monitored over a period of five days. Later time points were not considered because at that time the tissue had started to collapse.

A



B

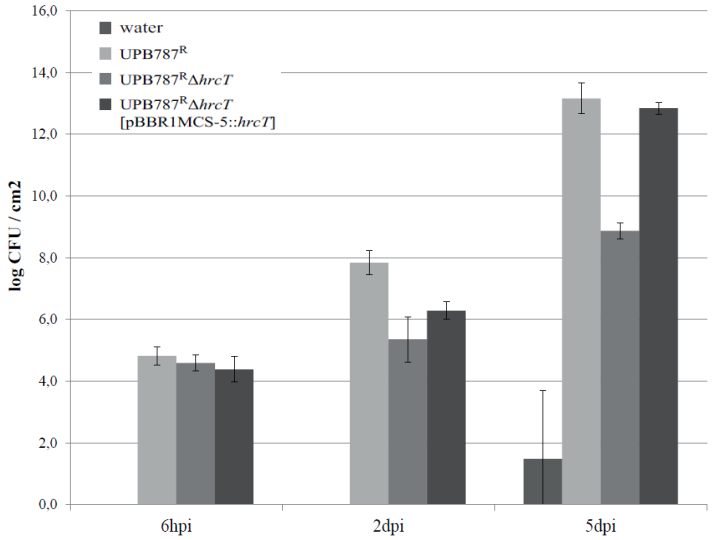


Figure V.4 The T3SS is required for symptom development and bacterial colonization of barley leaves.

A. Barley leaves were inoculated with the wild-type strain UPB787^R, with the *hrcT* mutant (UPB787^{R}Δ*hrcT*), with the *hrcT* mutant containing an empty vector (UPB787^{R}RΔ*hrcT* [pBBR1MCS-5]) and with the *hrcT* mutant complemented with a plasmid-borne *hrcT* gene (UPB787R^{R}Δ*hrcT* [pBBR1MCS-5::*hrcT*]). Lesion length was measured at 15 dpi. Data are the mean of five assays. Bars with the same letter are not statistically different based on a Tukey's honest significant difference test (a = 0.05). B. Bacterial counts in barley leaf segments infiltrated with the wild-type strain UPB787^R, with the *hrcT* mutant (UPB787^{R}Δ*hrcT*) and with the *hrcT* mutant complemented with the *hrcT* gene (UPB787^{R}Δ*hrcT* [pBBR1MCS-5::*hrcT*]) were analyzed at 6 hpi, 2 dpi and 5 dpi. Data are the mean of three assays. Error bars represent ± SD.

Strain UPB787RΔ*hrcT* showed a significant reduction of bacterial growth at 2 and 5 days after inoculation when compared to the wild-type strain (Fig. V.4B). The complemented derivative recovered the same population densities as the wild type at five days after inoculation (Fig. V.4B), thus indicating that *hrcT* from T3SS of *X. translucens* pv. *translucens* is important for plant colonization.

X. translucens pv. *translucens* strain UPB787^{R} was found to cause a non-host HR when inoculated into pepper leaves of the cultivar ECW-10R (Fig. V.5A). In contrast, the *X. translucens* pv. *hordei* strain UPB886^{R} did not trigger an HR within 8 days post inoculation. Since HR elicitation is often associated with a functional Hrp T3SS, we inoculated the UPB787 Δ*hrcT* mutant and its complemented derivative into pepper leaves. As expected, the mutant failed to trigger an HR while the complemented strain regained to ability to cause an HR comparable to that caused by the wild type. In conclusion, strains of cereal-pathogenic *X. translucens* require a functional Hrp system to cause an HR on non-host plants and to colonize the host plant barley.

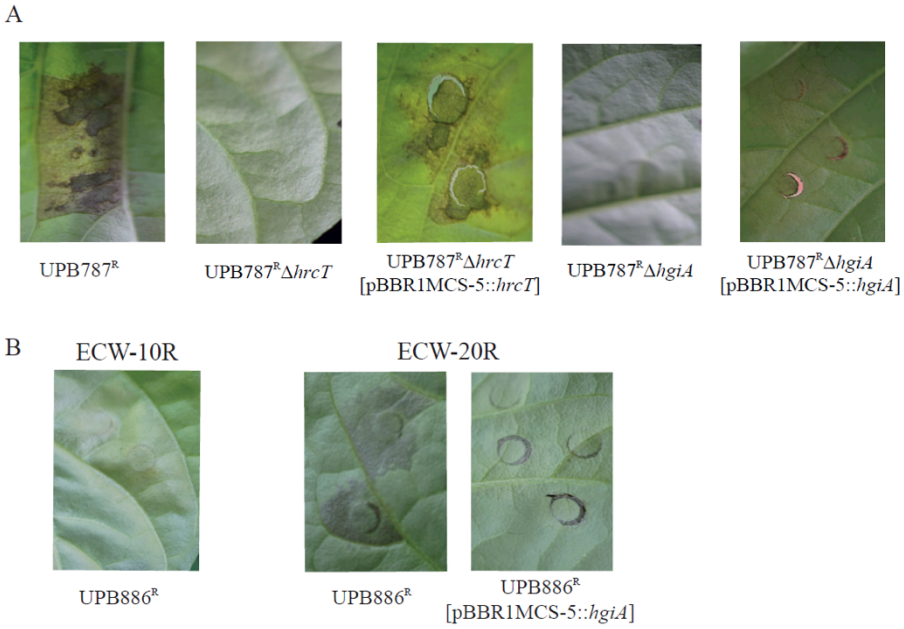


Figure V.5 The type III secretions system and the *hgiA* gene are required for HR elicitation on non-host pepper plants.
A. Pepper leaves of the variety ECW-10R were infiltrated with bacterial solution (OD_{600nm} = 0.3) of the wild-type strain UPB787^R, the mutants in *hrcT* (UPB787^{R}Δ*hrcT*) and in *hgiA* (UPB787^{R}Δ*hgiA*), and both mutated strains complemented with a mutated gene (UPB787^{R}Δ*hrcT* [pBBR1MCS-5::*hrcT*] and UPB787^{R}Δ*hgiA* [pBBR1MCS-5::*hgiA*]). Pictures were taken at 7 dpi. B. The wild-type strain UPB886^R was infiltrated into pepper leaves of the varieties ECW-10R and in ECW20R. Pictures of ECW-10R were taken at 7 dpi and pictures of ECW-20R were taken at 72 hpi.

The *hrp* cluster-associated, conserved *hgiA* gene contributes to symptom formation in barley

Comparison of five *X. translucens* strains had revealed the presence of two conserved, presumably HrpX-regulated genes, *hgiA* and *hgiB*, at the borders of their *hrp* clusters. Indeed, RNA-seq experiments had shown that both genes are strongly induced in a *X. translucens* pv. *translucens* strain CFBP 2054^R that ectopically expresses a constitutively active form of the master regulator HrpG, called HrpG*

(Wengelnik et al., 1999) from *X. euvesicatoria* 85-10 (unpublished data). In order to decipher a possible contribution of these two genes to disease, we created pVO155 insertion mutants in *hgiA* and *hgiB*. The genomic context of both genes indicated that *hgiB* is a single gene while *hgiA* could be the first gene of a two-gene operon (Fig. V.2). Consequently, the *hgiA* insertion mutant was expected to have a polar effect on the downstream gene, *hpaH*.

Insertion mutations in *hgiB* were constructed for three *X. translucens* strains (UPB787^{R}, UPB820^{R} and UPB886^{R}). When the mutants were infiltrated into barley leaves, they behaved similar to the wild-type strains (data not shown). Neither did leaf clip inoculations reveal any differences between UPB787^{R} wild-type strains and its *hgiB* mutants (Fig. V.6A). These experiments indicate that *hgiB* does not critically contribute to disease development on barley plants.

Next, the *hgiA* gene was knocked out in the three *X. translucens* strains UPB787^{R}, UPB820^{R} and UPB886^{R}. Leaf clip inoculations revealed that all three *hgiA* mutants were strongly impaired in their ability to cause symptoms, resulting in phenotypes similar to that of the *hrcT* mutants (Fig. V.6B, Suppl. data V.S3). For complementation assays, we individually cloned *hgiA* or *hpaH* into the broad-host range vector pBBR1MCS-5 and conjugated them into the mutant strains. Strains were inoculated by leaf clipping and plants were scored 15 days after inoculation. Introducing *hgiA* into UPB787^{R}Δ*hgiA*::pVO155 caused a partial complementation, leading to lesions that reached ca. 60% of the lesion length caused by the wild type (Fig. V.6B). Surprisingly, we observed a slightly lower, but still significant level of complementation when only the *hpaH* gene was

introduced into the UPB787^{R}Δ*hgiA*::pVO155 mutant (Fig. V.6B). Similar results were obtained with the *X. translucens* pv. *hordei* strains UPB886^{R} and UPB820^{R} (Suppl. data V.S3). These data suggest that both genes, *hgiA* and *hpaH*, contribute to virulence but none of them is absolutely required for causing symptoms.

The HgiA protein is a likely component of the *Xanthomonas translucens* translocon, which is required for non-host HR elicitation

The strong phenotype of the *hgiA* mutant suggested that the gene is either a key component for the delivery of type III effectors into host cells, or it encodes itself a major virulence effector that contributes critically to disease. In the latter case we expected that a mutation in *hgiA* would not have a drastic effect on the ability to trigger a non-host HR, unless it is the HgiA protein itself that triggers an HR.

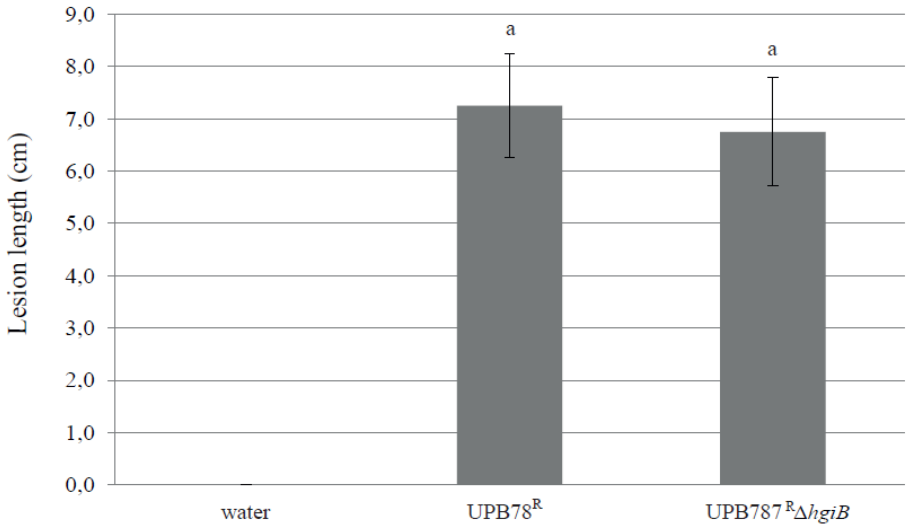
First, we tested the UPB787^{R}Δ*hgiA*::pVO155 mutant for HR elicitation on pepper plants of the cultivar ECW-10R (Fig. V.5A). While the wild type resulted in a strong HR, the mutant did not lead to any reaction, thus mirroring the phenotype of the *hrcT* mutant, which is unable to secrete effectors. When a plasmid-borne copy of *hgiA* was introduced into the UPB787^{R}Δ*hgiA*::pVO155 mutant, a weak HR was observed, thus demonstrating a partial complementation of the mutant phenotype (Fig. V.5A).

Since UPB886^{R} did not cause an HR on ECW-10R pepper plants, we wondered whether this strain would cause an HR on ECW-20R plants. PCR with two pairs of *avrBs2*-specific primers suggested the existence of a functional *avrBs2* gene. We therefore expected that

UPB886^{R} would cause an HR on ECW-20R. When we inoculated the wild-type strain and mutants in *hrcT* and *hgiA* into ECW-20R plants, we only observed an HR with the wild type (Fig. V.5B). These results suggest that AvrBs2 from *X. translucens* is recognized by the *Bs2* gene in pepper, and that delivery of AvrBs2 into the plant's cells depends on *hrcT* and *hgiA*, as one would expect it these two genes are required for type III protein translocation.

In conclusion, HR elicitation by an own *avr* gene (most likely *avrBs2*) depends on a functional T3SS, as shown by the phenotypes of the *hrcT* mutant. Moreover, an insertion mutation in *hgiA* mirrors the effect of a knock-out mutation in *hrcT*. This finding is compatible with the idea that HgiA is a component of the translocon that delivers effector proteins into the plant cell.

A



B

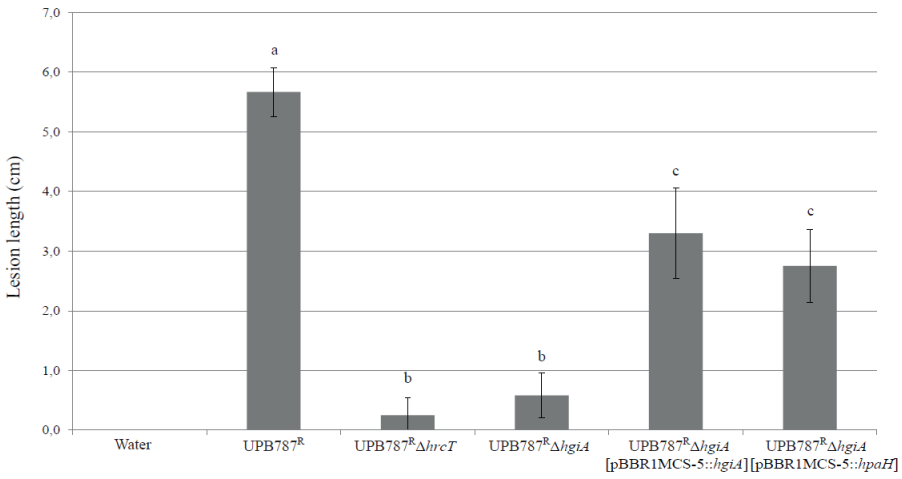


Figure V.6 Roles of *hgiA* and *hgiB* genes in disease development on barley.

A. Barley leaves were inoculated with the wild-type strain UPB787^R and the *hgiB* mutant (UPB787^{R}Δ*hgiB*). Lesion length was measured at 15 dpi. Data are the mean of five assays. B. Barley leaves were inoculated with the wild-type strain UPB787^R, the *hrcT* mutant (UPB787^{R}Δ*hrcT*), the *hgiA* mutant (UPB787^{R}Δ*hgiA*), the *hgiA* mutant complemented with the *hgiA* gene (UPB787^{R}Δ*hgiA* [pBBR1MCS-5::*hgiA*]) or with the *hpaH* gene (UPB787^{R}Δ*hgiA* [pBBR1MCS-5::*hpaH*]). Lesion length was measured at 15 dpi. Data are the mean of five assays. Bars with same letters are not statistically different based on a Tukey's honest significant difference test (a = 0.05). Error bars represent ± SD.

Discussion

The first *hrp* gene clusters from *Xanthomonas* were discovered in the clade-2 xanthomonads *X. euvesicatoria* and *X. campestris*, which rapidly became model organisms to study the pathogenicity of *Xanthomonas*. Other species, such as the rice pathogen *X. oryzae* and the citrus pathogen *X. citri* pv. *citri* were studied in detail as well, last not least due to their high economic interest and concerns about associated diseases, namely bacterial leaf blight and leaf streak of rice and citrus cancer. Collectively, tremendous amounts of data have been accumulated for these species over the last 25 years. In contrast, much less is known for species of the clade-1. *X. albilineans* GPE PC73 was the first clade-1 strain whose genome was sequenced (Pieretti et al., 2009), followed by strains of *X. sacchari* (Studholme et al., 2011). It came as a surprise that these clade-1 strains did not have an *hrp* gene cluster because it was known from previous work that at least pathovars of *X. translucens* contain *hrp* genes and type III effectors of the TAL family (Bragard et al., 1997). Indeed, when the first *X. translucens* genome was sequenced, a complete, but noncanonical *hrp* gene cluster was found (Wichmann et al., 2011). In contrast to clade-2 strains, however, a knockout in a conserved *hrp* gene of the grass pathogen *X. translucens* pv. *graminis* did not lead to the typically observed drastic loss of virulence (Wichmann et al., 2013).

In this study, we could demonstrate that a mutation of the conserved *hrcT* gene of the barley pathogens *X. translucens* pv. *hordei* and *X. translucens* pv. *translucens* resulted in a loss of pathogenicity. Mutants were strongly attenuated in symptom formation and could hardly colonize the leaf blade upon leaf clip inoculation, a typical

phenotype that is shared with clade-2 strains. Perhaps type III effectors as a whole are more import for the colonization of small-grain cereals *versus* the colonization of grasses. It will therefore be interesting to compare *hrp* mutants of other pathovars of *X. translucens*, such as the small-grain pathovars *undulosa* and the grass pathovars *arrhenateri*, *poae*, *phlei*, and *phleipratensis*. It will also be worthwhile to analyze the repertoires of type III effectors of all these *X. translucens* pathovars.

A comparison of the five available *X. translucens hrp* gene clusters revealed that they all have the same genetic organization, which to some extend resembles that of the *hrp* cluster from *R. solanacearum*. In both species, the two key regulatory genes, *hrpG* and *hrpX*/*hrpB*, are encoded on the left and right side of the *hrp* cluster. In general, both flanking regions are largely conserved among the five *X. translucens* strains (Fig. V.2). One of the few differences was the presence of a large open reading frame beyond *xopF* in the *X. translucens* pv. *graminis* strain DSM 18974. A homologous sequence was only found in *X. translucens* pv. *graminis* strain ART-Xtg29 which, however, contains a frameshift mutation. Inactivation of genes at the boundaries of *hrp* clusters has been observed for clade-2 xanthomonads and is often linked to type III effector genes. For instance, *X. oryzae* pv. *oryzae* encodes the LRR effector XopAE (syn. HpaF) beyond *hrpF* while the corresponding gene in *X. euvesicatoria* 85-10 suffers from a frameshift mutation (Sugio et al., 2005) Likewise, the candidate chaperone gene *hpa3* in front of *hpa4* (syn. *xopF*) from *X. oryzae* pv. *oryzae* contains a frameshift in the *X. euvesicatoria* homolog (Sugio et al., 2005). Whether the

DSM 18974 gene beyond *xopF* indeed encodes a type III effector is unknown although its conservation in many plant pathogens of the genera *Acidovorax*, *Pseudomonas* and *Xanthomonas* may suggest it. Moreover, homologs exist in *hrp* cluster-containing strains of the species *X. arboricola* and *X. cannabis* but not in related strains without an *hrp* cluster (Ignatov et al., 2015, Jacobs et al., 2015). Interestingly, a homolog is also present in the genome of *C. fungivorans*, not far from the *hrp* gene cluster and in close vicinity of other candidate type III effector genes. Reporter fusion assays may reveal the presence of a type III secretion signal and encourage further studies on this candidate type III effector.

In addition to the atypical location of the *hrpG* and *hrpX* genes, we detected two other genes in the *hrp* regions that are conserved among strains of *X. translucens* and in *X. hyacinthi*. Both genes are preceded by a canonical plant-inducible promoter sequence (Koebnik et al., 2006). Indeed, RNAseq experiments with strain CFBP 2054, which is a sibling of DSM 18974, revealed that both genes are indeed strongly activated by HrpG (data not shown). Therefore, these genes have been dubbed *hgiA* and *hgiB*, *hgi* for HrpG-induced. Both genes show considerable sequence diversity among the analyzed *X. translucens* strains (Fig. V.3), a finding that let us speculate that the genes may be under diversifying selection and that the gene products may be exposed to the plant surveillance system. Yet, the small number of available sequences does not allow to test this hypothesis. Homologs of *hgiB* have only been detected in *X. translucens* while remote homologs of *hgiA* were also found in *X. hyacinthi*, *B. andropogonis*, *C. fungivorans*, in two strains of *Ralstonia* and in

the *Burkholderiaceae* bacterium 26 where they are always encoded in the vicinity of the *hrp* cluster.

Insertional knockout mutants in *hgiB* did not show any defect in virulence when tested in three different strains. Apparently, this gene does not contribute to the type III secretion and effector translocation. Further work will address the question if *hgiB* encodes another type III effector. In contrast, when *hgiA* was knocked-out in three different strains by insertion of a plasmid which contains a partial *hgiA* sequence by single homologous recombination, the mutant bacteria turned out to be non-pathogenic on the barley variety Morex. Moreover, we found two strains of *X. translucens* that caused a non-host HR in pepper, likely due to the recognition of *avrBs2* from strain UPB886^{R} in the *Bs2*-containing pepper cultivar ECW-20R, and an unidentified type III effector from strain UPB787^{R} in the pepper cultivar ECW-10R. Both hypersensitive responses were dependent on an intact *hgiA* gene. This feature corresponds to the typical phenotype of a *hrp* mutant which is non-pathogenic on susceptible host plants and cannot cause a non-host HR, and strongly suggests that *hgiA* directly contributes to the delivery of type III effector proteins into plant cells.

Strains of *X. translucens* do not possess the translocon protein HrpF, which is ubiquitous in all clade-2 xanthomonads that have an Hrp system. Wichmann and colleagues speculated that cells of *X. translucens* pv. *graminis* may use breaches to gain direct access to the xylem cells in which they reside, and as these cells do not contain a cell membrane, the *hrpF* gene may be dispensable for effector translocation and, consequently, may have been lost during evolution

(Wichmann et al., 2013). However, using GFP-labeled bacteria of *X. translucens* pv. *translucens* we could detect the bacteria in the vascular bundle upon leaf clipping, clearly demonstrating that the bacteria colonize the intercellular space of barley leaves (own unpublished results). We therefore favor the second alternative that was put forward by Wichmann et al., namely that the *hrpF* gene might be present, but cannot be recognized on the basis of sequence identity.

Comparison of type III secretion gene clusters with a similar genetic organization and mode of regulation (via HrpX) from various bacteria revealed that all the strains that do not have a detectable *hrpF* gene do instead possess a homolog of *hgiA*. This mutual exclusivity can be taken as circumstantial evidence that *hgiA* encodes a functional analog of HrpF. No transmembrane segment was predicted for HgiA by any of the following algorithms: TMHMM v.2.0 (http://www.cbs.dtu.dk/ services/ TMHMM/), Phobius (http://phobius.sbc.su.se/), TopCons (http://topcons.net), a feature that is shared with other putative translocon proteins, HrpF from *X. euvesicatoria* and PopF1 and PopF2 from *R. solanacearum* (data not shown). Moreover, both protein families, the HrpF family and the HgiA family, share the presence of repeated sequence motifs as a common structural feature. HrpF from *X. campestris* pv. *campestris* has three imperfect repeats of ~110 amino acids, whereas the prototype HrpF protein from *X. euvesicatoria* has only two repeats, and the corresponding homolog in *Sinorhizobium fredii*, NolX, has only one copy of the repeated sequences (Sugio et al., 2005). Another polymorphic region is found in the C-terminal region of HrpF proteins where 5 to 15 copies of an imperfect tetrapeptide repeat are present in

different species of *Xanthomonas* (unpublished data). Similarly, sequence comparison of HgiA proteins from *X. translucens* revealed the presence of repeated sequence motifs (Suppl. data V.S4). For instance, strain Xtu4699 has five copies of a 20-amino acid motif (D/Q-TP-L/P-LSEAQED-A/S-IA-R/G-QLADA), whereas a related sequence motif is present six times in strain ART-Xtg29, four times in strain CFBP 2541, three times in strain DSM 18974 and only two times in strain DAR61454. It is tempting to speculate whether or not these variations function in the adaptation to specific host plants and/or evolved to escape from detection by the plant immune system. We also tested whether or not an *hgiA* mutant can be complemented by *hrpF* from *X. euvesicatoria*, and *vice versa*, whether an *hrpF* mutant can be complemented by *hgiA* from *X. translucens* strain CFBP 2541 (Büttner et al., 2002). None of these combinations restored symptom formation on the corresponding host plant, barley or pepper (data not shown). This result was not really surprising since even the HrpF homolog NolX, which is 48% identical to HrpF, was not able to complement an *hrpF* mutant (Huguet & Bonas, 1997). Hence, HgiA and HrpF may be analogous translocon proteins, which are probably not functionally interchangeable.

Surprisingly, the polar insertion mutant in *hgiA* could be partially complemented by *hgiA*, but also by the downstream gene *hpaH*. It therefore appears that both proteins contribute to the function of the type III secretion system. The results show that HgiA, although a possible facilitator of transport, is not absolutely required for the translocation of effector proteins into plant cells. Similar observations have been made for the *hrpF* gene from *X. oryzae*. Mutations in *hrpF*

led to drastic loss of virulence but did not completely eliminate pathogenicity (Sugio et al., 2005; Cho et al., 2008; Li et al., 2011b). This finding is also reminiscent of the phenotype of a mutant in *hrpK*, which is a possible homolog of *hrpF* in *P. syringae* (Petnicki-Ocwieja et al., 2005). Interestingly, Li and co-workers have shown that the Hpa2 (syn. HpaH) protein from *X. oryzae* pv. *oryzicola* is HpaB-dependently secreted via the Hrp system and has the ability to bind to the host cell membrane (Li et al., 2011b). Moreover, it was shown by yeast two-hybrid and protein pull-down experiments that Hpa2 interacts with HrpF, and bimolecular fluorescence complementation assays using split YFP demonstrated that both proteins can interact in the plant cell membrane (Li et al., 2011b). From these results, along with data from seveal avirulence reporter experiments, it was concluded that both proteins, Hpa2 and HrpF, act together to translocate type III effectors into plant cells, even if each protein alone has sufficient activity to allow effector translocation at much reduced rate. Our complementation data suggest a similar mechanism for *X. translucens* where both HpaH and HgiA contribute to effector translocation, yet the presence of only one of the two protein is still sufficient to allow some effector delivery to the plant cell. A similar redundancy of translocon components was also found in *P. syringae* where a quintuple mutant in the candidate translocon components *hrpK1*, *hrpW1*, *hrpZ1*, *hopP1*, and *hopAK1* could be complemented by all these genes alone except for *hopP1*, which suggested that a consortium of semiredundant translocators cooperate in the translocon formation (Kvitko et al., 2007). A similar multi-component translocon complex could operate in *Xanthomonas*, that may involve glycine-rich

proteins (e.g. XopA, syn. Hpa1) as a third component (Sugio et al., 2005). These results challenge the view of a simple, single-protein translocon pore; more work is needed to understand the molecular details at the "port of entry" (Büttner & Bonas, 2002b).

References

Adhikari TB, Gurung S, Hansen JM, Bonman JM. 2012. Pathogenic and genetic diversity of *Xanthomonas translucens* pv. *undulosa* in North Dakota. *Phytopathology* **102**, 390-402.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Bogdanove AJ, Beer SV, Bonas U, Boucher CA, Collmer A, Coplin DL, Cornelis GR, Huang HC, Hutcheson SW, Panopoulos NJ, Van Gijsegem F. 1996. Unified nomenclature for broadly conserved *hrp* genes of phytopathogenic bacteria. *Mol Microbiol* **20**, 681-683.

Bragard C, Singer E, Alizadeh A, Vauterin L, Maraite H, Swings J. 1997. *Xanthomonas translucens* from small grains: diversity and phytopathological relevance. *Phytopathology* **87**, 1111-1117.

Brito B, Marenda M, Barberis P, Boucher C, Genin S. 1999. *prhJ* and *hrpG*, two new components of the plant signal-dependent regulatory cascade controlled by PrhA in *Ralstonia solanacearum*. *Mol Microbiol* **31**, 237-251.

Büttner D, Bonas U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol Rev* **34**, 107-133.

Büttner D, Bonas U. 2002a. Getting across – bacterial type III effector proteins on their way to the plant cell. *EMBO J* **21**, 5313-5322.

Büttner D, Bonas U. 2002b. Port of entry – the type III secretion translocon. *Trends Microbiol* **10**, 186-192.

Büttner D, He SY. 2009. Type III protein secretion in plant pathogenic bacteria. *Plant Physiol* **150**, 1656-1664.

Büttner D, Nennstiel D, Klüsener B, Bonas U. 2002. Functional analysis of HrpF, a putative type III translocon protein from *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* **184**, 2389-2398.

Büttner D, Noël L, Stuttmann J, Bonas U. 2007. Characterization of the nonconserved *hpaB-hrpF* region in the *hrp* pathogenicity island from *Xanthomonas campestris* pv. *vesicatoria*. *Mol Plant Microbe Interact* **20**, 1063-1074.

Chatterjee S, Chaudhury S, Mcshan AC, Kaur K, De Guzman RN. 2013. Structure and biophysics of type III secretion in bacteria. *Biochemistry* **52**, 2508-2517.

Cho HJ, Park YJ, Noh TH, Kim YT, Kim JG, Song ES, Lee DH, Lee BM. 2008. Molecular analysis of the *hrp* gene cluster in *Xanthomonas oryzae* pathovar *oryzae* KACC10859. *Microb Pathog* **44**, 473-483.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190.

Durfee T, Nelson R, Baldwin S, Plunkett G 3rd, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M, Gibbs RA, Csörgo B, Pósfai G, Weinstock GM, Blattner FR. 2008. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* **190**, 2597-2606.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**, 1648-1652.

Gardiner DM, Upadhyaya NM, Stiller J, Ellis JG, Dodds PN, Kazan K, Manners JM. 2014. Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS One* **9**, e84995.

Guttman DS, Gropp SJ, Morgan RL, Wang PW. 2006. Diversifying selection drives the evolution of the type III secretion system pilus of *Pseudomonas syringae*. *Mol Biol Evol* **23**, 2342-2354.

Huguet E, Bonas U. 1997. *hrpF* of *Xanthomonas campestris* pv. *vesicatoria* encodes an 87-kDa protein with homology to NolX of *Rhizobium fredii*. *Mol Plant Microbe Interact* **10**, 488-498.

Ignatov AN, Kyrova EI, Vinogradova SV, Kamionskaya AM, Schaad NW, Luster DG. 2015. Draft genome sequence of *Xanthomonas arboricola* strain 3004, a causal agent of bacterial disease on barley. *Genome Announc* **3**, e01572-14.

Jacobs JM, Pesce C, Lefeuvre P, Koebnik R. 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front Plant Sci* **6**, 431.

Koebnik R, Krüger A, Thieme F, Urban A, Bonas U. 2006. Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J Bacteriol* **188**, 7652-7660.

Kousik CS, Ritchie DF. 1999. Development of bacterial spot on near-isogenic lines of bell pepper carrying gene pyramids composed of defeated major resistance genes. *Phytopathology* **89**, 1066-1072.

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM 2nd, Peterson KM. 1995. Four new derivatives of the broad-host-range cloning vector

pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175-176.

Kvitko BH, Ramos AR, Morello JE, Oh HS, Collmer A. 2007. Identification of harpins in *Pseudomonas syringae* pv. *tomato* DC3000, which are functionally similar to HrpK1 in promoting translocation of type III secretion system effectors. *J Bacteriol* **189**, 8059-8072.

Li YR, Zou HS, Che YZ, Cui YP, Guo W, Zou LF, Chatterjee S, Biddle EM, Yang CH, Chen GY. 2011a. A novel regulatory role of HrpD6 in regulating *hrp-hrc-hpa* genes in *Xanthomonas oryzae* pv. *oryzicola*. *Mol Plant Microbe Interact* **24**, 1086-1101.

Li YR, Che YZ, Zou HS, Cui YP, Guo W, Zou LF, Biddle EM, Yang CH, Chen GY. 2011b. Hpa2 required by HrpF to translocate *Xanthomonas oryzae* transcriptional activator-like effectors into rice for pathogenicity. *Appl Environ Microbiol* **77**, 3809-3818.

Lorenz C, Hausner J, Büttner D. 2012. HrcQ provides a docking site for early and late type III secretion substrates from *Xanthomonas*. *PLoS One* **7**, e51063.

Meyer D, Cunnac S, Guéneron M, Declercq C, Van Gijsegem F, Lauber E, Boucher C, Arlat M. 2006. PopF1 and PopF2, two proteins secreted by the type III protein secretion system of *Ralstonia solanacearum*, are translocators belonging to the HrpF/NopX family. *J Bacteriol* **188**, 4903-4917.

Oke V, Long SR. 1999. Bacterial genes induced within the nodule during the *Rhizobium*–legume symbiosis. *Mol Microbiol* **32**, 837-849.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D. 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int J Syst Evol Microbiol* **57**, 2881-2887.

Pesce C, Bolot S, Cunnac S, Portier P, Fischer-Le Saux M, Jacques MA, Gagnevin L, Arlat M, Noël LD, Carrère S, Bragard C, Koebnik R. 2015. High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc* **3**, e01574-14.

Petnicki-Ocwieja T, Van Dijk K, Alfano JR. 2005. The *hrpK* operon of *Pseudomonas syringae* pv. *tomato* DC3000 encodes two proteins secreted by the type III (Hrp) protein secretion system: HopB1 and HrpK, a putative type III translocator. *J Bacteriol* **187**, 649-663.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques MA, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M, Rott P. 2009. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* **10**, 616.

Romantschuk M, Roine E, Taira S. 2001. Hrp pilus – eeaching through the plant cell wall. *Eur J Plant Pathol* **107**, 153-160.

Rossier O, Van Den Ackerveken G, Bonas U. 2000. HrpB2 and HrpF from *Xanthomonas* are type III-secreted proteins and essential for pathogenicity and recognition by the host plant. *Mol Microbiol* **38**, 828-838.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, Grant M. 2011. Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel)* **2**, 1050-1065.

Sugio A, Yang B, White FF. 2005. Characterization of the *hrpF* pathogenicity peninsula of *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Microbe Interact* **18**, 546-554.

Van Gijsegem F, Gough C, Zischek C, Niqueux E, Arlat M, Genin S, Barberis P, German S, Castello P, Boucher C. 1995. The *hrp* gene locus of *Pseudomonas solanacearum*, which controls the production of a type III secretion system, encodes eight proteins related to components of the bacterial flagellar biogenesis complex. *Mol Microbiol* **15**, 1095-1114.

Weber E, Berger C, Bonas U, Koebnik R. 2007. Refinement of the *Xanthomonas campestris* pv. *vesicatoria hrpD* and *hrpE* operon structure. *Mol Plant Microbe Interact* **20**, 559-567.

Weber E, Ojanen-Reuhs T, Huguet E, Hause G, Romantschuk M, Korhonen TK, Bonas U, Koebnik R. 2005. The type III-dependent Hrp pilus is required for productive interaction of *Xanthomonas campestris* pv. *vesicatoria* with pepper host plants. *J Bacteriol* **187**, 2458-2468.

Wengelnik K, Rossier O, Bonas U. 1999. Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. *vesicatoria* result in constitutive expression of all *hrp* genes. *J Bacteriol* **181**, 6828-6831.

Wichmann F, Asp T, Widmer F, Kölliker R. 2011. Transcriptional responses of Italian ryegrass during interaction with *Xanthomonas translucens* pv. *graminis* reveal novel candidate genes for bacterial wilt resistance. *Theor Appl Genet* **122**, 567-579.

Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R. 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol Plant Pathol* **14**, 576-588.

Supplemental data V.S1: Oligonucleotides used in this study.

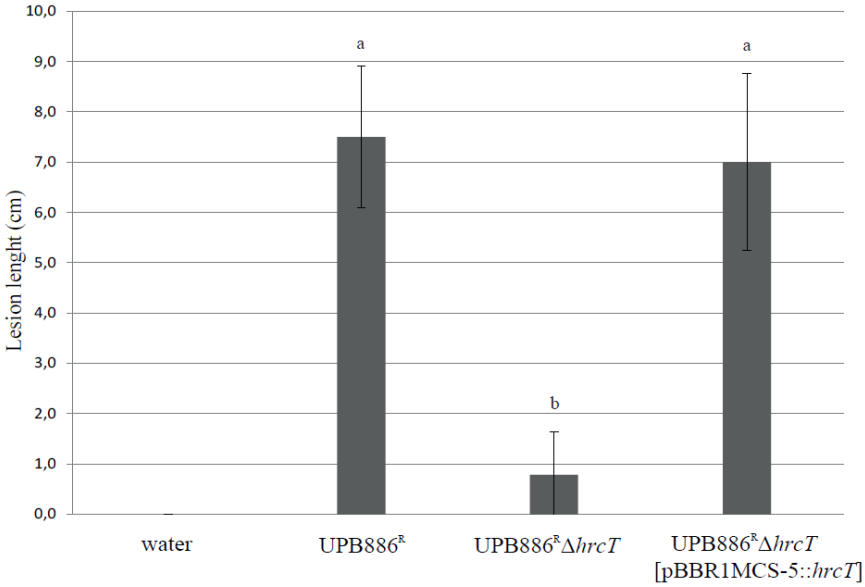
Name	Sequence	Purpose
pVO155-Fw	5′-GAGATCCCCAGCCCGCCTAATG	Verification of constructs in pVO155
pVO155-Rev2	5′-GCGCGTTACAAGAAAGCCGG	Verification of constructs in pVO155
pBBR-Fw	5'-AGAAGCACACGGTCACACTG	Verification of constructs in pBBR1MCS-5
pBBR-Rev3	5'-CCGGCTCGTATGTTGTGTGG	Verification of constructs in pBBR1MCS-5
Xt-hrcT-Fw	5'-GCAGGATCCGACCTGGCCGGTTTCAAC	Cloning of an internal *hrcT* fragment for mutagenesis
Xt-hrcT-Rev	5'-CGCTCTAGAAGCAACTGCGAAACGA	Cloning of an internal *hrcT* fragment for mutagenesis
Xt-hrcT-cpt-Fw	5'-CTAGGTCGACGAGAGAATCGTTATGAATGTGTCCGATGCCG	Cloning of *hrcT* in pBBR1MCS-5 for complementation
Xt-hrcT-cpt-Rev	5'-CTAGGAATTCTTAGTGGGTAGCAGGAGACC	Cloning of *hrcT* in pBBR1MCS-5 for complementation
Xt-hgiA-Fw	5'-GCAGGATCCACCGCCAAGGCCGAGTTC	Cloning of an internal *hgiA* fragment for mutagenesis
Xt-hgiA-rev	5'-CGCTCTAGAGCATCCACAGCGGCTTGC	Cloning of an internal *hgiA* fragment for mutagenesis
Xt-hgiA-cpt-Fw1	5'-CTAGAAGCTTGAGAGAATCGTTATGAGCCTTAAAAGTTTCT	Cloning of *hgiA* in pBBR1MCS-5 for complementation
Xt-hgiA-cpt- Rev1	5'-CTAGTCTAGACTACACCGTATGGAGCTGGAA	Cloning of *hgiA* in pBBR1MCS-5 for complementation
Xt-hpaH-cpt-fw	5'-CTAGAAGCTTGAGAGAATCGTTATGCGATGGACAGGATGGC	Cloning of *hpaH* in pBBR1MCS-5 for complementation
Xt-hpaH-cpt-rev	5'-CTAGTCTAGATCAACACGCAGCCAGATCCT	Cloning of *hpaH* in pBBR1MCS-5 for complementation
Xt-hgiB-Fw	5'-GCAGGATCCAGCCTCGCTGGCGCAGATGA	Cloning of an internal *hgiB* fragment in pVO155
Xt-hgiB-Rev	5'-CGCTCTAGAGCACCGCCTGCAGACCGGCC	Cloning of an internal *hgiB* fragment in pVO155

Restriction sites of the enzymes used for cloning are underlined.

<u>Supplemental data V.S2</u> The T3SS is required for symptom development.

Barley leaves were inoculated with the wild-type strain UPB886^{R}, the *hrcT* mutant (UPB886^{R}Δ*hrcT*) and the *hrcT* mutant complemented with the *hrcT* gene (UPB886^{R}Δ*hrcT* [pBBR1MCS-5::*hrcT*]). A. Leaf clipping inoculations were scored at 15 dpi. Data are the mean of five assays. Error bars represent +/- SD. Bars with the same letters are not statistically different based on a Tukey's honest significant difference test (a = 0.05). B. Stem inoculations of barley leaves were scored at 7 dpi.

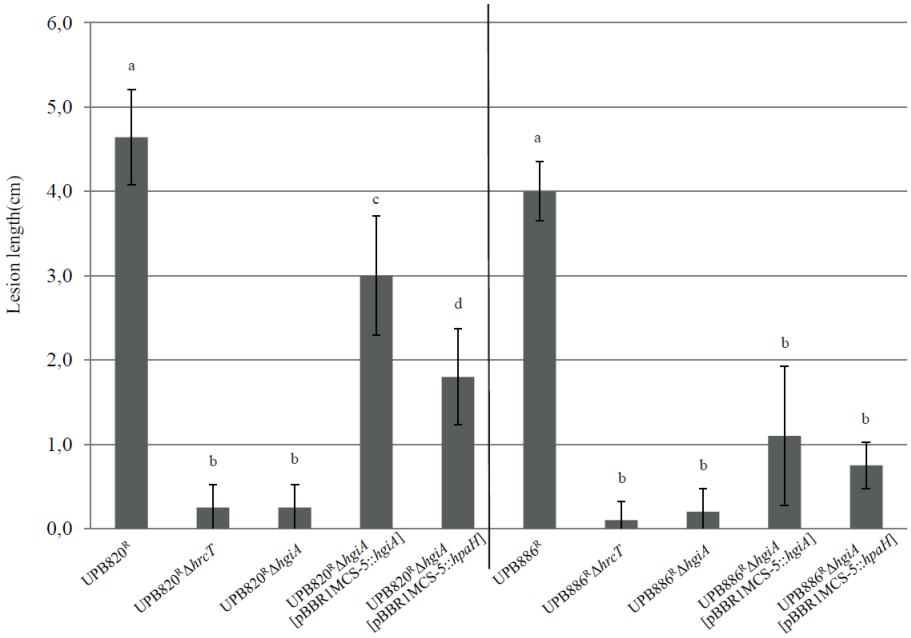
A



B

<u>Supplemental data V.S3</u> Role of the *X. translucens hgiA* operon for disease development on barley leaves.

Barley leaves of the cultivar Morex were inoculated with the wild-type strains UPB820^{R} and UPB886^{R}, their *hrcT* mutants (UPB820RΔ*hrcT,* UPB886RΔ*hrcT*), *hgiA* mutants (UPB820^{R}Δ*hgiA* and UPB886^{R}Δ*hgiA*), and the *hgiA* mutants complemented with either the *hgiA* gene (UPB820^{R}Δ*hgiA* [pBBR1MCS-5::*hgiA*], UPB886^{R}Δ*hgiA* [pBBR1MCS-5::*hgiA*]) or the *hpaH* gene (UPB820^{R}Δ*hgiA* [pBBR1MCS-5::*hpaH*], UPB886^{R}RΔ*hgiA* [pBBR1MCS-5::*hpaH*]). Lesion length was measured at 15 dpi. Data are the mean of five assays. Error bars represent ± SD. Bars with same letters are not statistically different based on a Tukey's honest significant difference test (a = 0.05).



Supplemental data V.S4 Presence of a repeated sequence motif in the HgiA proteins from five *X. translucens* strains.

A. Amino acids sequences of HgiA from five *Xanthomonas* strains were analyzed: XTG_ART-Xtg29 (*X. translucens* pv. *graminis*), XT_DAR61454 (*X. translucens*), XTU_Xtu4699 (*X. translucens* pv. *undulosa*), XTC_CFBP 2541 (*X. translucens* pv. *cerealis*), XTT_DSM18974 (*X. translucens* pv. *translucens*). Red characters represent the conserved 20-amino acid sequence motif. B. Consensus sequence logo of the repeated sequence motif.

A

>XT_DAR61454
MSFKSFFKSVGHGIAHAAEDVAHGVEQGAKDVAKGVEGVVEHGVNAVGDVMQGDFKGSLHELGQAGKAVG
DTAKAEFKLGLAVGGAMTDLALGTVADAHLSKGLTKLAQSAKRGMDKARDAVDKSVDQVVDSAEGVVAGG
ARFADDAVHGRFDRLGKDSLSVASNAFNVCTSLTPEGMAANVGANLTRSFMDGTSLSKYADLVGDTMARK
PLWMLRDGAKDVAEPLLDPLKKDLASADAAVDRTLDKAENGISRTLAEAAVASGLAVQGDMTAQS
QTPPLSEAQEDAIARQLADAMERKDEQSGGQVNLSTQNGSG
DTPLLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSASADVSFQLRAV

>XTT_DSM18974
MSLKSFFKSIGHGITNAVEDVAHGVEQGAKDIAKGVEGAVEHGAHALGDVMQGDFRGGLHELGQAGKAVG
DTAKAEFKLGLTVGGAMTDLALGTVADAHLSKGLTKLAETAKGGLDKARDAVDKSLNQVVDSAEGAVAGG
VRCVDDVAHGRFDRLGKDSLSVAGNAFNLCTSLTPEGMAANVGANLTRSFMDGTSLSKYADVVGDTMARK
PLWMLRDGAKDVAEPLLDPLKKDLASADAAVDRTLDKAEDGISRTSTEAAVASGLAVQGDMTAQS
QTPLLSEAQEDAIAGQLADALQGNLSTRDGRG
DTSLLSDAQEDAIAGQLADALQGNLSTQDGRG
DTPQLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSASADVSFQLHTV

>XTC_CFBP 2541
MSLKSFFKSIGHGITNAVEDVAHGVEQGAKDIAKGVEGAVEHGVHAFGDVLQGDFRGGLHELGQAGKAVG
DTAKAEFKLGLTVGGAMTDLALGTVADAHLSKGLTKLAETAKGGLDKARDAVDKSLNQVVDSAEGAVAGG
MRCVDDVAHGRFDRLGKDSLSVAGNAFNLCTSLTPEGMVANVGANLTRSFMDGTPLSKYANVIGDTMARK
PLWMLRDGAKDVAEPFLDPLKKDLASADAAVNRTLDKAENGISGTLDKAENGISRTLAEAAVASGLAVQG
GMTAQS
QTPQLSEAQEDAIARQLADAMERKDTQSGGQDNLSTQDGRG
DTPLLSEAQEDAIAGQLADALNDNLSTQDGRG
DTPQLSEAQEDAIAGQLADALNDNLSTQDGRG
DTPQLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSASADVSFQLHTV

>XTG_ART-Xtg29
MSFKSFIKSIGHGITHAAEDVAHGIEQGVKDVAKGVEGVAEHGAQALGDVMQGDFRGGLKELGQAGKSIG
DTAKAEFKMGLTVGGAMTDLALGTVADAHLNKGLTKLAQSAKGVMDKTRDAVDKSVNQFVDSAEGAVAGG
VRCAEDVAHGRLDRLGKDSLSVASNAFNVCTSLTPEGMVASVGANLTRTFLSDTPVGKYANVVGDTMARK
PLWMLRDSAKDVAEPLLEPLKKDLASADASVDRTLDKAEAGVSRTLDKAEAGVSRTLAEAAEQA
QTASLSGAIEYAIARELEKSLQPDDTQSRGLGDLSIQIGLG
DTGSLSGAMEDAIARELEQSLQPDDTQSRGLGDLSIQIGLG
DTGSLSGAMEDAIARELEQSLQPDDTQSRGLGDLSIQIGLG
DTPSLSGATEDAIARQLADALEGKDKQSTQDGRG
DTPLLSDAQADSIAGQLYDALQDNLSTQDGSG
DTPPLSEAQEDSIADQLADDLQDNLATQGGSG
DMPPLSEAQQEAIQRQLDDAMERKARQSGRDEGLLSASADVSFQFGAV

>XTU_Xtu4699
MSFKSFFKSVGHGIAHAAEDVAHGVEQGAKDVAKGVEGVVEHGVNAVGDVMQGDFKGSLHELGQAGKAVG
DTAKAEFKLGLAVGGAMTDLALGTVADAHLSKGLTKLAQSAKRGMDKARDAVDKSVDQVVDSAEGVVAGG
ARFADDAVHGRFDRLGKDSLSVASNAFNVCTSLTPEGMAANVGANLTRSFMDGTSLSKYADVVGDTMARK
PLWMLRDGAKDVAEPLLDPLKKDLASADAAVDRTLDKAENGISRTLAEAAVASGLAVQGDMTAQS
QTPPLSEAQEDAIARQLADAMERKDKQSTQDGRG
DTPLLSEAQEDSIAGQLADALQDNLSTQDGRG
DTPLLSEAQEDSIAGQLADALQGNLSTQDGSG
DTPPLSEAQEDAIARQLADAMERKDEQSGGQVNLSTQNGSG
DTPLLSEAQEDAIARQLADAMEGKDKQSGREEGLLASLFGEPALSAAADVSFQLRAV

B

Chapter VI

A conserved transcription activator-like effector contributes to pathogenicity of barley-pathogenic xanthomonads

Transcription-Activator Like Effectors (TALEs) form a distinct family of type III effectors, which are injected into the plant host cell where they act as transcription activators. TALEs are present in many species and pathovars of *Xanthomonas*. In some *Xanthomonas* pathosystems, certain TALEs have been isolated as major virulence factors and their target genes have been identified by a combination of transcriptomic and bioinformatic approaches (Hutin et al, 2015). TALEs proteins from *Xanthomonas* are composed of three well-defined domains, an N-terminal domain containing the secretion/translocation signal, the central, DNA-binding domain consisting of highly similar repetitive sequence motifs, and the C-terminal domain containing the nuclear localization signals and an acidic transcription activation domain (Boch & Bonas, 2010). Proteins with a sequence-related DNA-binding domain have also been found in strains of *Ralstonia solanacearum* and *Burkholderia rhizoxinica* (Boch et al., 2014).

Currently, nothing is known about the conservation, diversity structure and function of TALEs in *X. translucens*. Here, we therefore analyzed the repertoire of TALEs from several *X. translucens* strains, sequenced some of them and performed a *tale*-directed mutagenesis. Certain *tale* mutants were compromised in pathogenicity, suggesting

an important role of some TALEs in the pathogenicity of *X. translucens* on barley plants.

This work is in preparation for publication.
Pesce C, Szurek B, Bragard C, Koebnik R.
A conserved transcription activator-like effector contributes to pathogenicity of barley-pathogenic xanthomonads.

A conserved transcription activator-like effector contributes to pathogenicity of barley-pathogenic xanthomonads

Céline Pesce[a,b], Boris Szurek[a], Claude Bragard[b], Ralf Koebnik[a]

[a] UMR 186 Interactions Plantes Microorganismes Environnement (IPME), IRD-Cirad-Université Montpellier, Montpellier, France
[b] Earth and Life Institute, Applied Microbiology Phytopathology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

E-mail: koebnik@gmx.de

Abstract

Transcription activator-like effectors (TALEs) are widespread in strains of the plant pathogens *Xanthomonas* and *Ralstonia solanacearum*. For several so-called major-virulence TALEs, an important contribution to pathogenicity has been demonstrated. TALEs are composed of three conserved protein domains. The central domain, which binds to double-stranded DNA in a sequence-specific manner, consists of a repetitive array of typically 34 or 35 amino acid residues, as found in most TALEs from *Xanthomonas* or *R. solanacearum*, respectively. Analysis of *tale* gene repertoires from representative strains of nine pathovars of *Xanthomonas translucens* revealed that they contain between zero to eight *tale* genes per strain. Repeat arrays from eight *X. translucens tale* genes were cloned, including both *tale* genes from *X. translucens* pv. *cerealis* strain CFBP 2541^{R} and all four *tale* genes from *X. translucens* pv. *translucens* strain UPB787^{R}. Sequence analyses revealed that all eight *X. translucens tale* genes encode hybrid TALEs consisting of both 34- and 35-aa repeats. Moreover, two distinct subfamilies of TALEs were found to co-exist in individual strains of *X. translucens*, as revealed by a substantial number of sequence polymorphisms in the N- and C-terminal domains. Three strains of *X. translucens* were subjected to *tale*-specific knockout mutagenesis. Mutants with defects in pathogenicity on barley plants were found to be affected in a conserved *tale* gene from barley-pathogenic xanthomonads belonging to the pathovars *hordei* and *translucens*. This is the first report on a conserved major virulence TALE in *X. translucens*, which opens the way to define its target sequences in cereals.

Introduction

Collectively, bacteria of the genus *Xanthomonas* infect a plethora of host plants many of which are of tremendous economic interest as crops or ornamentals. Both dicots and monocots are attacked by different species and pathovars of *Xanthomonas*. For instance, distinct pathovars of *Xanthomonas citri* infect citrus (pv. *citri*), cotton (pv. *malvacearum*), mango (pv. *mangiferaeindicae*), or soybean (pv. *glycines*). Strains of *Xanthomonas oryzae* cause bacterial leaf blight (pv. *oryzae*) or bacterial leaf streak (pv. *oryzicola*) on rice plants, and strains of *Xanthomonas translucens* can infect small-grain cereals and several grasses. In many cases, a key role in pathogenicity was attributed to the Hypersensitive Response and Pathogenicity (Hrp) gene cluster, which encodes a type III protein secretion system (T3SS) (Tampakaki et al., 2010). Type III effectors (T3Es), which are injected into plant cells via the type III secretion systems, are among the most important and best studied pathogenicity factors of xanthomonads and other pathogens of plants and animals (He et al., 2004).

In *Xanthomonas* spp., T3Es of the AvrBs3/PthA family are known to be frequently involved in pathogenicity. Since these effector proteins act as transcription activators within the host cells, they have been called Transcription Activator-Like Effectors (TALEs) (Yang et al., 2006). Different species and pathovars of *Xanthomonas* have between 0 and ~30 TALEs per strain (Scholze & Boch, 2010, 2011). The architecture of TALEs is largely conserved within the genus *Xanthomonas* spp. and also among strains of *Ralstonia solanacearum*. Three distinct regions have been defined for the polypeptide chain of TALEs from *Xanthomonas* spp. and *R. solanacearum*. The first region

of ~290 amino acids at the N terminus of *Xanthomonas* TALEs includes the secretion / translocation signal that is recognized by the T3SS, and sequence motifs further downstream that are involved in unspecific protein-DNA contacts (Guo et al., 2011). The C-terminal region of *Xanthomonas* TALEs (~280 amino acids) contains nuclear localization signals (NLS), which are essential for the import of the protein into the plant cell nucleus, and an acidic activation domain (AAD) that promotes transcription initiation in the plant cell (Gürlebeck et al., 2006). These two regions are well conserved among *Xanthomonas* spp. A comparison of the N-terminal region of three Hax proteins (TALEs from *Xanthomonas campestris*) with AvrBs3 (from *Xanthomonas euvesicatoria*) revealed only two amino acid exchanges (Kay et al., 2005). In contrast, the C-terminal region appears to be a little less conserved because 28, 12 and 0 amino acid exchanges were found when Hax2, Hax3 and Hax4 were compared with AvrBs3, respectively (Kay et al., 2005).

The hallmark of TALEs, however, is their central region which typically consists of 1.5 to 33.5 tandemly repeated nearly identical modules of 33 to 35 amino acids and an additional truncated repeat with 20 amino acids (Boch & Bonas, 2010). Whereas most *Xanthomonas* TALEs have repeats with 34 amino acids (Schornack et al., 2008; Lahaye & Bonas, 2011), some species/pathovars also contain TALEs with 35-aa repeats (e.g. Hax2 from *X. campestris*, which exclusively consists of 35-aa repeats, and AvrHah1 from *Xanthomonas gardneri*, which is a hybrid TALE consisting of both 34- and 35-aa repeats) (Schornack et al., 2008). Repeat sequences are nearly identical to each other, except for residues 12 and 13, which

have been dubbed repeat-variable diresidue (RVD) (Moscou & Bogdanove, 2009). The number of repeats and their distinct RVDs mediate the proteins' binding to double-stranded DNA in a sequence-specific manner, following a comma-free, non-overlapping and redundant recognition code where one RVD binds to one base pair (Boch et al., 2009; Moscou & Bogdanove, 2009). Whereas seven RVDs were sufficient for detectable reporter gene activation, full gene induction was only observed with eleven or more RVDs (Boch et al., 2009). Until now, a few dozen different RVDs have been found in nature. Yet, only five different RVDs account for more than 85% of all naturally occurring RVDs in *Xanthomonas* (Boch & Bonas, 2010; Wang et al., 2014): HD (His-Asp) corresponds to cytosine (C), NI (Asn-Ile) to adenine (A), NG (Asn-Gly) to thymine (T), NN (Asn-Asn) to purine bases (G and A), and NS (Asn-Ser) to all four bases. Using tailored reporter systems, recognition specificities of all theoretically possible RVDs were elucidated (Yang et al., 2014; Juillerat et al., 2015). These combinatorial approaches confirmed what was already suspected from the high-resolution three-dimensional structures. Analyses of protein-DNA co-crystals had suggested that the second RVD residue, which is in direct contact to the DNA base pair, is critical for determining the binding specificity, whereas the first RVD residue plays an important role in stabilizing the conformation of a loop in the protein structure and only indirectly influences the binding specificity (Deng et al., 2012; Mak et al., 2012).

In contrast to most xanthomonads, TALEs from *Ralstonia solanacearum*, called RipTALS (de Lange et al., 2013), contain 35-aa

repeats (Heuer et al., 2007). RipTALS were detected in more than 300 *R. solanacearum* strains (Heuer et al., 2007). Lang and colleagues (2013) have demonstrated that the Brg11 protein (also called RipTAL1) interacts with DNA via the same code as TALEs from *Xanthomonas* spp, even if RipTAL1 shares only 40% of sequence identity with AvrBs3 from *X. euvesicatoria* (28% identity between the terminal regions and 56% identity between the central domains) (Schornack et al., 2006). Related proteins, termed Bat proteins, were identified in the endosymbiotic bacterium *Burkholderia rhizoxinica*, but not in other *Burkholderia* species, and the Bat repeat domains were found to mediate sequence-specific DNA binding via the same code as TALEs, despite less than 40% sequence identity (de Lange et al., 2014).

All previous work on TALEs from *Xanthomonas* was performed with strains from clade-2, which includes the species *Xanthomonas citri*, *X. euvesicatoria*, *X. campestris*, and *X. gardneri* (Young et al., 2008). In contrast, virtually nothing is known about the structure and function of TALEs from clade-1 xanthomonads, which include the "early-branching species" *Xanthomonas albilineans*, *Xanthomonas sacchari*, *Xanthomonas hyacinthi*, *Xanthomonas theicola*, and *Xanthomonas translucens* (Parkinson et al., 2007). Whereas *X. albilineans* and *X. sacchari* were found to be devoid of Hrp T3SS and TALEs (Pieretti et al., 2009; Studholme et al., 2011), Southern blot hybridization had indicated the presence of up to eight TALEs per strain among different strains of *X. translucens* (Gonzalez et al., 2007). Recently, the presence of TALEs in *X. translucens* was confirmed by whole-genome sequencing of the pathotype strain

CFBP 2154 (Pesce et al., 2015). In this study, we isolated TALE genes from different pathovars of *X. translucens*, compared their structures and elucidated their contribution to pathogenicity of barley-pathogenic xanthomonads.

Materials and methods

Bacterial strains and growth conditions

Fifteen natural isolates of *X. translucens* were used in this study (Table VI.1). All of them come from the strain collection at the Université catholique de Louvain in Belgium (UPB strains) or from the Collection Française de Bactéries associées aux Plantes (CFBP strains) (http://www6.inra.fr/cirm/CFBP-Bacteries-associees-aux-Plantes). *Xanthomonas* strains were cultivated at 28 °C in PSA medium (10 g peptone, 10 g sucrose, 1 g glutamic acid, 16 g agar, l^{-1} H_2O). *Escherichia coli* DH10b (Durfee et al., 2008) bacteria, which were used for molecular cloning, were cultivated at 37°C in lysogenic broth (LB).

Rifampicin-resistant *Xanthomonas* mutants were selected upon plating on rifampicin-containing PSA medium at high cell density and one clone was chosen for further experiments. Plasmids were introduced into *E. coli* by thermo-transformation and into *X. translucens* by conjugation using pRK2013 as a helper plasmid in tri-parental mating (Figurski & Helinski, 1979). Antibiotics were added to the medium at the following final concentrations: rifampicin, 100 µg/ml; gentamicin, 20 µg/ml, kanamycin 50 µg/ml.

Table VI.1 Bacterial strains and plasmids used in this study.

Bacterial strains and plasmids	Description	Reference
pVO155	pUC119-derived suicid vector with *uida* cassette (Km^R)	Oke and Long, 1999
pVO155::Nter TALE	pVO155 containing a 375 bp *tale* gene fragment from CFBP 2054 (Km^R)	This study
pXTale	pBBR1MCS-5 containing a synthetic gene encoding the N- and C-terminal domain of the complete TALE from DMS 18974 (Gm^R)	This study
pXTale::RVDsTALE4	pXTale with repeat region of *tale*4 from UPB787 (Gm^R)	This study
CFBP 2053^R	*X. translucens* pv. *graminis* (Rf^R)	This study
CFBP 2054^R	*X. translucens* pv. *translucens* (Rf^R)	This study
UPB886^R	*X. translucens* pv. *hordei* (Rf^R)	This study
UPB787^R	*X. translucens* pv. *translucens* (Rf^R)	This study
UPB820^R	*X. translucens* pv. *hordei* (Rf^R)	This study
UPB787^RΔ*tale4*	UPB787^R containing pVO155::NterTALE	This study
UPB820^RΔ*tale6*	UPB820^R containing pVO155::NterTALE	This study
UPB820^RΔ*tale7*	UPB820^R containing pVO155::NterTALE	This study
CFBP 2054^RΔ*tale1*	CFBP 2054^R containing pVO155::NterTALE	This study
CFBP 2054^RΔ*tale6*	CFBP 2054^R containing pVO155::NterTALE	This study
UPB787^RΔ*hrcT*	UPB787^R containing pVO155::*hrcT*	This study
UPB820^RΔ*hrcT*	UPB820^R containing pVO155::*hrcT*	This study

Δ: mutated, bp: base pair. Resistances to the antibiotics kanamycin, gentamycin and rifampicin are annotated as Km^R, Gm^R and Rf^R, respectively.

Plant material and plant inoculations

Barley (*Hordeum vulgare* L.) plants of the cultivar Morex (six-rowed spring barley) were grown in growth chambers with cycles of 16 hours of light per day at 22 °C and 50% relative humidity. For virulence assays, leaves of three-week old plants were cut at about two centimeter below the leaf tip with sterile scissors that have been soaked in a bacterial suspension at optical density at 600 nm (OD_{600}) of 0.5, corresponding to 3×10^8 CFU/ml. Immediately after inoculation, plants were transferred for 24 hours into a home-built chamber that allowed to reach nearly saturated relative humidity. Symptoms were scored at 15 days post-inoculation, using five replications per condition. Statistical significance of the results was assessed using the Tukey honest significant difference test for post-ANOVA pairwise comparisons, set at 5% ($P < 0.05$). Water was used as negative control for all inoculation experiments.

Molecular biological methods

Genomic DNA from *X. translucens* strains as isolated using the Wizard® Genomic DNA Purification Kit and plasmid DNA was isolated with the Wizard® Plus SV Minipreps DNA Purification Systems (Promega, Madison, WI, U.S.A.), according to the instructions. Restriction enzymes were used according to the manufacturer's recommendations (New England Biolabs, Évry, France). Ligation reactions of DNA fragments into plasmid vectors were performed using T4 DNA ligase (Promega). Polymerase chain reactions (PCRs) were conducted in 20 µL volumes using GoTaq® G2

DNA Polymerase (Promega). Primer sequences are provided as supplemental material (Suppl. data VI.S1).

For Southern blots, 3 µg of genomic DNA were digested with high-fidelity restriction enzymes (*Bam*HI or *Sph*I; New England Biolabs, Évry, France) over night, and DNA fragments were separated electrophoretically at 50 mV for ~3 days. Upon capillary transfer on an Amersham™ Hybond™-N+ membrane for 24 hours, membranes were probed with a digoxigenin-labeled DNA probe according to the instructions in the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche Diagnostics GmbH, Mannheim, Germany). Two probes were prepared using digoxigenin-11-dUTP (Roche Diagnostics GmbH). The first probe was a 706-bp DNA fragment that corresponds to the end of the N-terminal TALE domain. As a second probe, a 375-bp DNA fragment corresponding to the beginning of the C-terminal domain was produced. Both probes were obtained by PCR, using genomic DNA of strain CFBP 2541 as template.

DNA sequencing of repeat arrays from *tale* genes

Central repeat arrays of *X. translucens tale* genes were amplified by PCR using oligonucleotide primers that were designed on the conserved flanking sequences of the two *tale* genes from *X. translucens* pv. *cerealis* CFBP 2541 (GenBank accession number CM003052). In order to improve yield and specificity, betaine was added to the PCR reaction at a final concentration of 10 mM (Rees et al., 1993). Upon electrophoretic separation, DNA fragments of the expected size were purified with the QIAquick Gel Extraction Kit (Qiagen SAS, Courtaboeuf, France) and cloned into pGEM®-T Easy

(Promega). DNA inserts of recombinant clones were sequenced after plasmid purification. Repeat arrays of *tale* genes from four strains were sequenced: UPB787, UPB820, CFBP 2054 and CFPB 2541.

Construction of *X. translucens* mutants and complementation assays

tale genes were knocked-out by insertion of a plasmid that contains a partial *tale* sequence by single homologous recombination. To facilitate recombination, 375-bp DNA fragments were PCR amplified from *X. translucens* strains CFPB 2541 and CFBP 2054 using *tale*-specific oligonucleotide primers containing restriction sites at their 5' end (Suppl. data VI.S1). PCR amplicons were digested with *Xba*I and *Bam*HI and cloned into the suicide vector pVO155 (Oke & Long, 1999). Recombinant *E. coli* clones were validated by PCR and DNA sequencing. pVO155 derivatives were then introduced into *X. translucens* strains (UPB787^R, UPB820^R and CFBP 2054^R) via tri-parental conjugation and single cross-over events were monitored by PCR on resulting *X. translucens* trans-conjugants (Table VI.1).

For complementation of *tale* mutants, a new "*tale* vector", pXtale, was developed for *X. translucens*. This vector allows to re-construct full-length *tale* genes which encode the authentic N- and C-terminal domains from *X. translucens*. First, DNA fragments corresponding to the N- and C-terminal domains of the TALE protein from *X. translucens* pv. *translucens* strain DSM 18974 (accession no. CCP38287) were synthesized (Biomatik, Wilmington, DE, U.S.A.) and cloned into the *Hin*dIII/*Xba*I-digested plasmid pBBR1MCS-5 (Kovach et al., 1995). In this construct, the N- and C-terminal

domains are separated by a 31-bp linker that contains two recognition sites of the type IIs endonuclease *Bsa*I, thus allowing scarless Golden Gate cloning of the repeat region in between (Engler et al., 2008). To facilitate Golden Gate cloning, the unique *Bsa*I site of the pBBR1MCS-5 vector backbone was removed by site-directed mutagenesis, using a mutagenic oligonucleotide and the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent technologies, USA). This vector has been called pXtale (Suppl. data VI.S2).

The repeat region of *tale4* from *X. translucens* pv. *translucens* strain UPB787 was amplified by PCR, cloned into pGEM®-T Easy, and then subcloned into pXtale via the two *BsaI* sites. After each step, constructs were confirmed by PCR and DNA sequencing.

Bioinformatic analyses

Database searches were performed using BLASTP or TBLASTN (http://blast.ncbi.nlm.nih.gov; Altschul et al. 1990). Global pairwise sequence alignments were performed using the Needlemann-Wunsch algorithm (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) (Needleman & Wunsch, 1970). Multiple sequence alignments were generated using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) (Edgar, 2004). The Artemis genome browser was used to analyze the genomic context of *tale* genes (http://www.sanger.ac.uk/resources/software/artemis/) (Rutherford et al., 2000). Phylogenetic trees were calculated using a web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences (http://phylogeny.lirmm.fr/) (Dereeper

et al., 2010). Sequence logos were generated at the website http://weblogo.berkeley.edu/logo.cgi (Crooks et al., 2004).

Results

Two distinct TALE families co-exist in strains of *Xanthomonas translucens*

As a first step to understand the structure and function of TALEs from *Xanthomonas*, we compared the publicly available sequences. Five genome sequences are deposited at GenBank and were searched for *tale* genes using TBLASTN and AvrBs3 as a query. No *tale* gene was found in the *X. translucens* pv. *graminis* strain ART-Xtg29, as reported previously (Wichmann et al., 2013). The draft genome sequences of strains DAR61454 and DSM 18974 were found to possess 11 and 7 fragments of *tale* genes, respectively, corresponding to peptide fragments of 11 to 270 amino acids for strain DAR61454 and fragments of 116 to 647 amino acids for strain DSM 18974. In addition, one complete *tale* gene was assembled for strain DSM 18974, corresponding to an 1111-aa TALE protein with 15.5 repeats. The high level of fragmentation is probably due to the repetitive character of the *tale* genes and the chosen sequencing strategy (Gardiner et al., 2014). Interestingly, two complete genes, *tale1* and *tale2*, but no extra *tale* gene fragments were found in *X. translucens* pv. *cerealis* strain CFBP 2541, which was attributed to the very high coverage and the significant number of single-nucleotide polymorphisms that distinguish individual repeats from each other (Pesce et al., 2015). Compared to TALEs from clade-2 xanthomonads, *X. translucens* TALEs form a separate branch when their phylogenetic

relationships are analyzed (Fig. VI.1). Yet, *X. translucens* TALEs are more closely related to TALEs from other xanthomonads than to the prototype RipTAL Brg11 from *R. solanacearum* (Cunnac et al., 2004).

Sequence comparisons of the two complete TALE proteins from strain CFBP 2541 revealed a surprisingly high divergence between them. The two N-terminal domains share only 77.6 % sequence identity (84.8% at the nucleotide level) and the two C-terminal domains share 80.4 % sequence identity (88.7% at the nucleotide level). This finding suggests that two distinct *tale* families may co-exist in one and the same strain of *X. translucens*. Fortunately, the first complete genome sequence for a strain of *X. translucens* was deposited at GenBank recently (acc. no. CP008714). Strain Xtu4699, which belongs to the pathovar undulosa, contains eight *tale* genes. Comparison of the N- and C-terminal domains of these eight TALE proteins revealed that they fall into two groups (Fig. VI.2).

The first one contains only one member, TAL-D for *X. translucens* pv. *undulosa* strain Xtu4699, which is very similar to the only completely assembled TALE protein from *X. translucens* strain DSM 18974. This group also contains TAL1 from CFBP 2541 (red square in Fig. VI.2). The three corresponding *tale* genes have a diagnostic feature at the 5' end of the coding sequence. These three genes lack the highly conserved *Bam*HI restriction site, which overlaps with the ATG start codon (ATGGATCC, *Bam*HI site underlined) in almost all *tale* genes. The second group comprises the remaining seven TALEs from strain DSM 18974 and TAL2 from CFBP 2541. The high diversity of sequence between the two groups

supports the idea of the co-existence of two distinct TALE families in different strains of *X. translucens*.

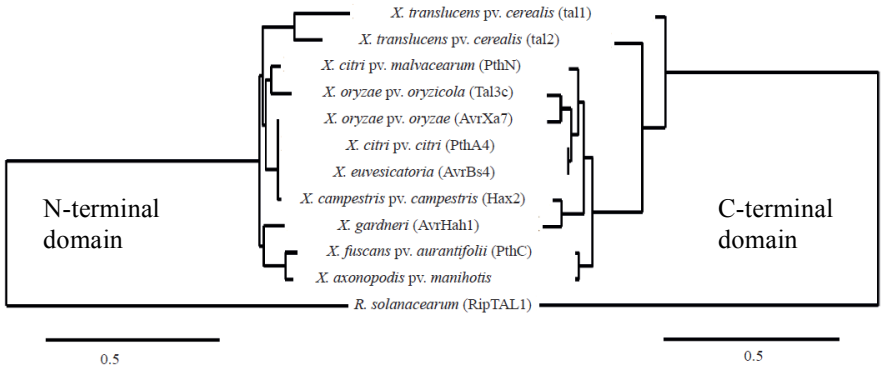


Figure VI.1 Phylogenetic trees derived from the N- and C-terminal domains of TALEs from *Xanthomonas* and *R. solanacearum*.
The phylogenetic tree on left side was constructed from the N-terminal domains of TALEs and the tree on the right side was constructed from the C-terminal domains of TALEs. Twelves TALEs from eleven species were used. Amino acids sequences were obtained from the NCBI database. *X. translucens* pv. *cerealis* (accession number CM003052), *X. citri* pv *malvacearum* (AF016221), *X. oryzae* pv. *oryzicola* (AF016221), *X. oryzae* pv. *oryzae* (AAG02079), *X. citri* pv. *citri* (AAM39311), *X. euvesicatoria* (CAA48680), *X. campestris* pv. *campestris* (AY993937), *X. gardneri* (ABP97430), *X. fuscans* pv. *aurantifolii* (AAY43358), X. *axonopodis* pv. *manihotis* (ADM80412) and *R. solanacearum* (AGG20296).

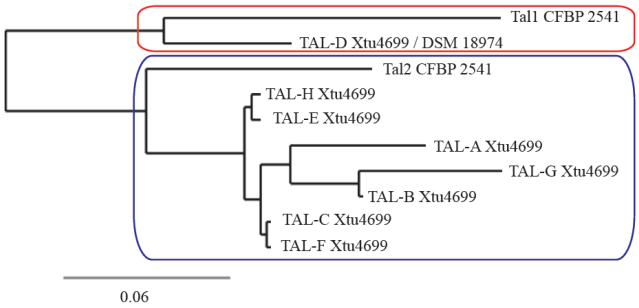


Figure VI.2 Phylogenetic tree derived from the concatenated N- and C-terminal domains of TALEs from *X. translucens*.
TAL-A to TAL-H correspond to the eight TALEs from *X. translucens* pv. *undulosa* strain Xtu4699 (accession number CP008714), Tal1 and Tal2 are from *X. translucens* pv. *cerealis* strain CFBP 2541 (acc. no. CM003052). The red box indicates group-1 members, which lack the *Bam*HI site at the ATG start codon. The blue box indicates group-2 members that contain two conserved *Bam*HI sites in the 5'- and 3'-regions of the coding sequence. TAL-D from strain Xtu4699 is identical to a TALE sequence from strain DSM 18974.

Most pathovars of *X. translucens* encode TALE proteins some of which are apparently conserved at the pathovar level

In order to elucidate the presence and repertoires of *tale* genes in the different pathovars of *X. translucens*, 14 representative strains from nine pathovars, including seven type/pathotype strains, were analyzed by Southern blot hybridization. Initial attempts to use a 725-bp C-terminal fragment of AvrXa7 from the *X. oryzae* pv. *oryzae* as a hybridization probe were not successful (Wonni et al., 2013). Therefore, two new "mixed probes" were developed from the two *tale* genes from strain CFBP 2541, which contains exactly one member of each of the two *X. translucens tale* gene families. PCR primers were designed that could anneal to both *tale* genes from CFBP 2541, thus ensuring a near-equimolar abundance of both gene fragments in the PCR-amplified hybridization probes. The first probe of 706_{tal1} / 676_{tal2} bp corresponds to the end of the N-terminal TALE domain ("N probe") and the second probe of 375 bp corresponds to the beginning of the C-terminal domain ("C probe") (Suppl.data VI.S3).

Southern blots were generated from genomic DNA that was digested with *Bam*HI (Fig. VI.3) or with *Sph*I (Suppl. data VI.S4). *Bam*HI is the restriction enzyme that is commonly used to analyze *tale* gene repertoires from *Xanthomonas*, since almost all genes contain a conserved *Bam*HI site at the ATG start codon and a second conserved site ~150 bp upstream of the stop codon (Boch & Bonas, 2010). Unfortunately, the first site is only conserved in one of the two *tale* families, and also the second site is not strictly conserved (Suppl. data. VI.S3). We therefore used *Sph*I as an alternative since all

X. translucens tale genes were found to possess at least one site in front of the repeat array (~40 bp before the first repeat), and another site downstream of the repeat array, although at variable positions (Suppl. data VI.S3). In contrast to conventional *tale* genes, Southern blot analyses do therefore not allow to determine the number of *X. translucens tale* repeats from the size of the hybridizing DNA bands (Fig. VI.3 and Suppl. data 3)

When *Bam*HI-digested genomic DNA was probed, between 0 and 8 *tale*-specific bands were detected using the N probe or the C probe (Fig. VI.3). Interestingly, pathogens of small grain cereals, such as pathovars *hordei*, *secalis*, *translucens* and *undulosa*, contained more *tale* genes (4 to 8 genes per strain) than strains of pathovars that infect grasses, i.e. *arrhenatheri*, *cerealis*, *graminis*, *phleipratensis*, which only contained 0 to 3 *tale* genes. Absence of tale genes appears to be typical for strains of the pathovar *graminis*, since also strains UPB437 and ART-Xtg29 were found to be devoid of *tale* genes (Gonzalez et al., 2007; Wichmann et al., 2013). Seven haplotypes were identified for the nine barley-pathogenic strains belonging to *X. translucens* pv. *hordei*, *X. translucens* pv. *translucens* and *X. translucens* pv. *undulosa.* While band sizes varied to a large extent it is noteworthy that all eight strains of the pathovars *hordei* and *translucens* shared two conserved bands of ~3.0 and ~3.3 kb in size (Fig. VI.3 blue scares), and another band of ~2.0 kb was found to be specific to strains of *X. translucens* pv. *hordei* (Fig. VI.3 red square).

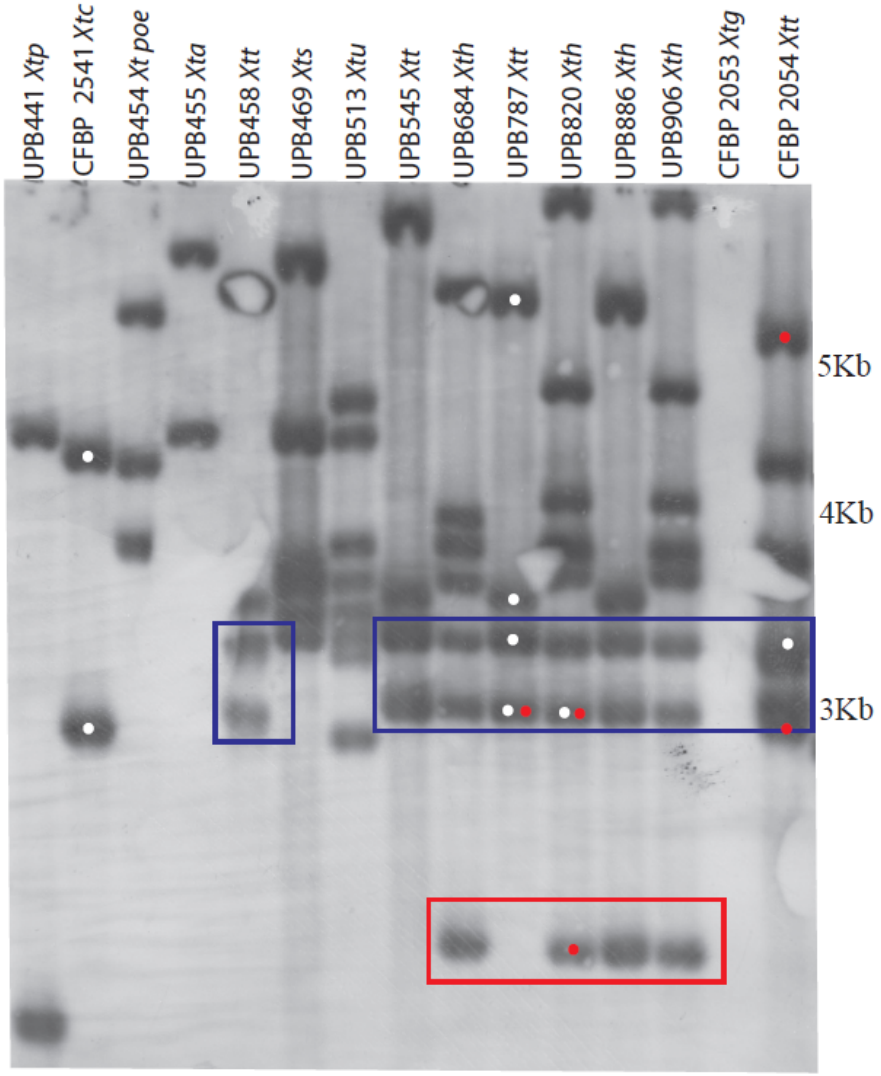


Figure VI.3 Repertoire of *tale* genes in several strains of *X. translucens*. *Bam*HI-digested genomic DNA was hybridized with a mixed probe corresponding to the 3'-terminal regions of the two *tale* genes from *X. translucens* pv. *cerealis* strain CFBP 2541. The following pathovars were used: *Xta*, pv. *arrhenatheri*; *Xtc*, pv. *cerealis*; *Xtg*, pv. *graminis*; *Xth*, pv. *hordei*; *Xtp*, pv. *phleipratensis*, *Xts*, pv. *secalis*; *Xtt*, pv. *translucens*; *Xtu*, pv. *undulosa*. Blue squares indicate conserved bands of the barley pathovars *translucens* and *hordei*. The red square indicates a conserved band in the four *X. translucens* pv. *hordei* strains. White dots and red dots correspond to sequenced and mutated *tale* genes, respectively.

Similar results were obtained when *Sph*I-digested DNA was probed with the C probe, which is the only of the two probes that would anneal to the repeat array (Suppl. data VI.S4). In conclusion, most *X. translucens* strains contain *tale* genes and barley-pathogenic strains appear to share some *tale* genes, as indicated by tale-specific DNA fragments of an identical size.

X. translucens TALEs are built from two distinct repeat motifs and contain new RVDs

Since the function of TALEs is linked to their repeat domain and since conserved band sizes are no proof for functional equivalence, we aimed at sequencing the repertoires of *tale* repeats arrays from different barley-pathogenic xanthomonads. Four strains from three pathovars were chosen for further analyses, UPB787, UPB820, CFBP 2054 and CFBP 2541. Repeat arrays were amplified by PCR using primers that would anneal downstream and upstream of the repeat array of both *X. translucens tale* families (Suppl. data VI.S3). PCR amplification of repetitive regions, as was previously shown for *tale* genes (Hommelsheim et al., 2014), is challenging and may result in artifacts due to undesired annealing of incompletely amplification products in subsequent PCR cycles. Indeed, we observed increasing difficulties to amplify *tale* repeat arrays with increasing number of *tale* genes per strain, as evidenced by an increase of background smear at the cost decreasing band intensities upon gel electrophoresis. Since no easy solution is available to circumvent these limitations and risks of artifacts, all PCR reactions were done at least two times and only sequenced that were obtained at least three times were

considered as being correct. Comparison of PCR and Southern blot experiments indicated that amplicon sizes did not correlate with the restriction fragment lengths, likely due to a limited conservation of the restriction sites (Suppl. data VI.S3). In any case, the complete repertoire of *tale* repeat arrays was obtained for two strains (CFBP 2541 and UPB787).

Upon cloning and sequencing, eight complete *tale* repeat arrays were obtained, all two from strain CFBP 2541, all four from strain UPB787, and one for each of the strains UPB820 and CFBP 2054 (Table VI.2 and Fig. VI.3, yellow stars). Both the CFBP 2054 TALE and TALE3 from UPB787 have 12.5 repeats and their RVD patterns are identical. The first six RVDs are also present in a TALE fragment from strain DSM 18974 (encoded by contig CAPJ01000425), which is a sibling of CFBP 2054. The *tale*7 from UPB820 and *tale4* from UPB787 constitute another couple with 11.5 repeats each and identical RVD patterns. For none of the new eight repeat arrays could a corresponding *tale* gene be found in the genome of Xtu4699. However, the *X. translucens* pv. *undulosa* strain appears to share at least two TALEs with strain DSM 18974. The RVD pattern of TalD (acc. no. AKK68069) is identical to the one from the DSM 18974 TALE (acc. no. CCP38287). And another TALE fragment from strain DSM 18974 that is encoded on contig CAPJ01000489 appears to correspond to *taleA* from Xtu4699 (acc. no. AKK66437) since they share a continuous stretch of ten RVDs. These comparisons indicate that different pathovars of barley pathogens have distinct TALE repertoires, yet they appear to share a few conserved TALEs.

X. translucens TALEs repeat arrays differ from TALEs of clade-2 xanthomonads. *X. translucens* TALEs are composed of two different repeat units consisting of either 34 or 35 amino acids.

Table VI.2 RVDs patterns of TALEs from four *X. translucens* strains.

Strain	RVD num.	RVDs
UPB787	16.5	NN-QD-NG-NN-HN-KG-NI-HD-NI-NH-NG- HN-HD-HD-NI-NN-HD
	14.5	NN-HD-HD-HD-NI-NI-NI-NN-HD-HD-NN-NN-NI-NN-HD
	12.5	NG-NN-HD-HD-NN-NI-HG-HD-ND-HG-NI-NN-HD
	11.5	NI-NG-HN-NK-HD-NH-HN-HD-HD-HD-HD-QD
UPB820	11.5	NI-NG-HN-NK-HD-NH-HN-HD-HD-HD-HD-QD
CFBP 2541	15.5	NN-NN-KI-NN-HD-NG-HD-NG-NG-NK-HD-HD-NN-QD-NG-QD
	11.5	NS-KI-NI-HD-NK-GI-HD-NK-HD-NN-HD-NK
CFBP 2054	12.5	NG-NN-HD-HD-NN-NI-HG-HD-ND-HG-NI-NN-HD

Repeat arrays of eight TALEs have been sequenced, corresponding to the complete TALomes of *X. translucens* pv. *cerealis* strain CFBP 2541 (2 TALEs) and *X. translucens* pv. *translucens* strain UPB787 (4 TALEs). Identical TALE RVD patterns were found in the 11.5-repeats TALEs from strains UPB787 and *X. translucens* pv. *hordei* UPB820 (red fonts) and in the 12.5-repeats TALEs from strains UPB787 and *X. translucens* pv. *translucens* CFBP 2054 (green color). In Southern blots, strains UPB820 and CFBP 2054 contain eight and seven TALE-specific DNA bands, respectively.

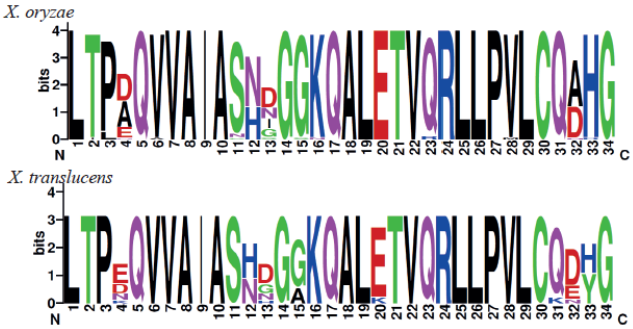
When taking into account one member of all the distinct TALEs from *X. translucens*, i.e. eight from Xtu4699, four from strain UPB787 and two from strain CFBP 2541, 17% of the repeats have 34 amino acids while the majority has 35 amino acids. The only known example of another hybrid TALE with both 34- and 35-aa repeats in the same protein is AvrHah1 from *Xanthomonas gardneri*, which consists of four 34-aa repeats and nine 35-aa repeats, plus one half-repeat at the C-terminal end of the repeat array (Schornack et al., 2008).

A comparison of repeat consensus sequences from clade-2 xanthomonads, as exemplified by *X. oryzae*, from *X. translucens*, from *R. solanacearum* and from the exceptional 35-aa repeats of Hax2 highlights the strong conservation of seven amino acids immediately before (Q_5VVAIAS$_{10}$) and after (G_{14}GKQALE$_{20}$) the RVD motif (Fig. VI.4). Not surprisingly, more variation is found at more distant positions, because these residues are not under strict packing constraints as the residues around the DNA-contacting RVD are (Suppl. Fig VI.S5). These distant residues have evolved lineage-specific patterns, such as the Pro-Tyr motif at positions 33 and 34 in the 35-aa repeats from *X. translucens* and *R. solanacearum*. Yet, both types of *X. translucens* repeats resemble more the other *Xanthomonas* repeats, as evidenced by conservation of the Leu-Thr-Pro motif at positions 1 to 3, and the Thr-Val-Gln-Arg-Leu-Leu-Pro-Val-Leu-Cys motif at positions 21 to 30.

Inspection of the RVDs from *X. translucens* TALE repeats and comparison with *Xanthomonas* clade-2 members TALEs (Boch & Bonas 2010) revealed an overrepresentation of HN, NH and NK motifs (8.6%, 5.3% and 2.4% for *X. translucens* versus 0.1%, 0.05% and 0.3% for *Xanthomonas*) and un underrepresentation of NS, N* and HG motifs (0.5%, 0% and 1.4% for *X. translucens* versus 8.1%, 6.8% and 5.2% for *Xanthomonas*). Moreover, eight new RVDs were observed in 14 TALEs from *X. translucens* that were not found before in *Xanthomonas*: QD (3.8%), KG, Y* (amino acid thirteen is missing in this repeat type), KI, YK, YD and GI (0.5%), in decreasing order of their frequency. Based on previous work using tailored reporter

systems (Yang et al., 2014; Juillerat et al., 2015; Miller et al., 2015) we predict the base preference for all of them (Table VI.3).

A



B

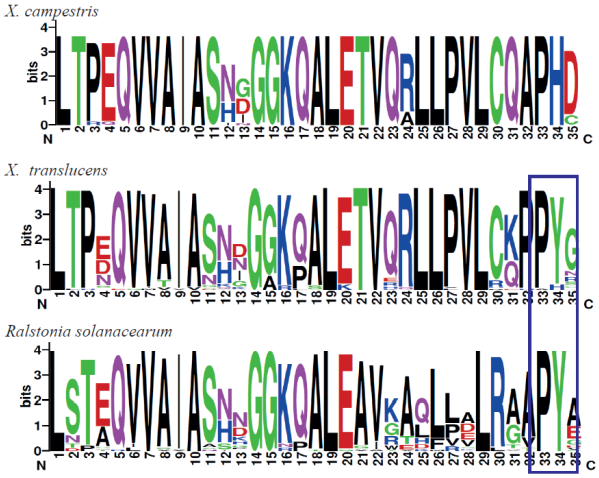


Figure VI.4 Consensus sequences of TALEs repeats.
A. Consensus sequences of 34-amino acids repeats. The *X. oryzae* sequence logo was generated from three TALEs (TalC from *X. oryzae* pv. *oryzae* strain BAI3, Tal9a from *X. oryzae* pv. *oryzae* strain PXO99 and Tal3c from *X. oryzae* pv. *oryzicola* strain BLS256), corresponding to 58 repeats. The *X. translucens* sequence logo was generated from the 12 repeats in the six sequenced TALEs of *X. translucens* pv. *cerealis* strain CFBP 2541 and *X. translucens* pv. *hordei* strain UPB787. B. Consensus sequences of 35-amino acids repeats. The *X. campestris* sequence logo was generated from the 21 repeats of Hax2. The *X. translucens* sequence logo was generated from 59 repeats in the strains CFBP 2541 and UPB787. The *Ralstonia solanacearum* sequence logo was generated from three TALEs of the *R. solanacearum* strains GMI1000 and PSI07, corresponding to 44 repeats. The blue square indicates a conserved sequence motif in the repeats of *X. translucens* and *R. solanacearum*.

Table VI.3 Predicted target nucleotides for new RVDs of *X. translucens*.

RVD	Predicted target	Juillerat *et al.*, 2015	Miller *et al.*, 2015
QD	C	C-a-t	C-
KG	T	A-C-G-T	T
KI	A	A-G-c	A
YK	G	G-a-t	G
YD	C	C-a-t	C
NF	A	A-c	A
GI	A	/	A

Predictions are based on previous work using reporter systems (Juillerat et al., 2015; Miller et al., 2015). Uppercase letters correspond to high-affinity binding and lowercase letters correspond to low-affinity binding in the reporter assay, respectively.

TALEs contribute to pathogenicity of *X. translucens*

It has been shown previously that the Hrp T3SS contributes more to the virulence of *X. translucens* pathovars *hordei* and *translucens*, which infect barley, than to the pathovar graminis, which infects grasses (Wichmann et al., 2013; Pesce et al., 2015). Interestingly, the barley pathogens possess TAL effectors whereas the grass pathogen does not. We therefore proposed that at least certain TAL effectors could contribute to the pathogenicity of *X. translucens*, as has been shown previously for other xanthomonads (Castiblanco et al., 2013; Streubel et al., 2013).

To test this hypothesis, knockout plasmids were constructed that contain a 375-bp internal DNA fragment originating from a *X. translucens tale* gene. Because of the observed sequence diversity among *X. translucens tale* genes, seven different knockout suicide plasmids were generated containing PCR amplicons from strains (four for CFBP 2054 and three for CFBP 2541). Four representative

knockout plasmids were conjugated into three barley-pathogenic *X. translucens* strains (CFBP 2054^R, UPB787^R and UPB820^R). Single recombination insertions were expected to knockout one *tale* gene per trans-conjugant. In total, 20 trans-conjugants were obtained. Southern blot analyses were used to identify the inactivated *tale* genes (data not shown). Altogether, five *tale* genes were mutated in the three strains (Fig. VI.3 red circles). Surprisingly, some trans-conjugants seemed to have incorporated the suicide plasmid at another locus. And three of the five mutated *tale* genes were hit more than once.

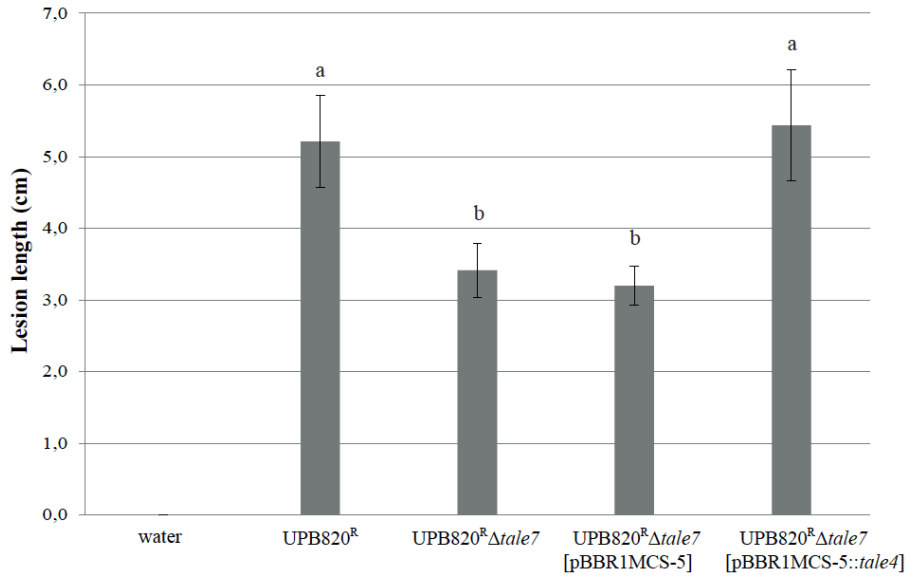
To evaluate the contribution of the mutated *tale* genes to the virulence of *X. translucens*, leaf clip assays were performed on barley leaves of the cultivar Morex. Mutations in three genes, *tale1* and *tale7* from strain CFBP 2054^R and *tale-8* from strain UPB820^R, had no obvious effect on virulence under our experimental conditions. In contrast, knockout mutations in *tale4* from *X. translucens* pv. *translucens* strain UPB787^R and in *tale7* from *X. translucens* pv. *hordei* strain UPB820^R resulted in significantly shorter lesions in comparison to the corresponding wild-type strains (Fig. VI.5A and B). Both genes belonged to a DNA fragment of identical size on a Southern blot, and sequencing of the corresponding *tale* repeat arrays had shown that both genes contain 11.5 repeats with the same RVDs pattern (Fig. VI.3 and suppl data VI.S4). For complementation assays, a *X. translucens* pv. *translucens*-derived *tale* gene was reconstructed by Golden Gate cloning, consisting of DSM 18974 (a.k.a. CFBP 2054)-derived sequences upstream and downstream of the repeat array, and the 11.5 repeats of *tale4* from UPB787^R. The repeat domain of *tale4* from UPB787 is strictly identical to *tale7* from

UPB820^{R}. This construction was able to complement the *tale7* mutant of strain UPB820^{R} with respect of symptom formation.

The virulence TALE4 from strain UPB787 does not turn the grass pathogen CFBP 2541^{R} into a barley pathogen

Since grass-pathogenic xanthomonads of the pathovar graminis were devoid of *tale* genes, we wondered if the virulence-mediating *tale* gene *tale4* from strain UPB787 could enable a graminis strain to cause symptoms on barley. To this end, *tale4* was introduced on a plasmid into the *X. translucens* pv. *graminis* strain CFBP 2541^{R}. Barley leaves of the cultivar Morex were infiltrated with the parental strain and the *tale*-containing trans-conjugant. Leaves remained symptomless until eight days after infiltration and no differences were observed between the two strains. Hence, the conserved 11.5-repeat *tale* gene from barley pathogens does not determine host specificity in itself.

A



B

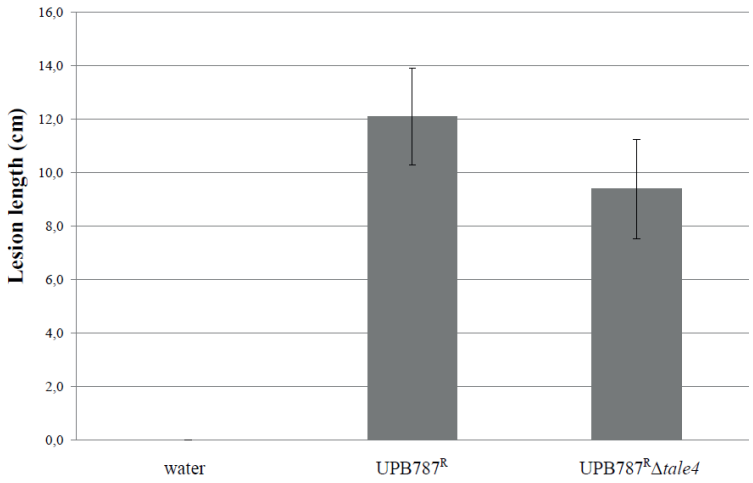


Figure VI.5 TALE7 from *X. translucens* pv. *hordei* UBP820 and TALE4 from *X. translucens* pv. *translucens* UBP787 contribute to symptom development on barley.

A. Barley leaves were inoculated with the wild-type strain UPB820^{R}, the *tale6* mutant (UPB820^{R}Δ*tale7*), the *tale6* mutant carrying the empty vector (UPB820^{R}Δ*tale7* [pBBR1MCS-5]) and the *tale6* mutant complemented with *tale4* from UPB787 (UPB820^{R}Δ*tale7* [pBBR1MCS-5::*tale4*]). B. Barley leaves were inoculated with the wild-type strain UPB787^{R} and the *tale4* mutant (UPB787^{R}Δ*tale4*). Lesion length was measured at 15 dpi. Data are the mean of five assays. Error bars represent +/- SD. Bars with same letters are not statistically different based on a Tukey's honest significant difference test (a = 0.05).

Discussion

In this study, we catalog the repertoires of TALEs from nine pathovars of *X. translucens*. Southern blot experiments permitted to detect between zero to eight *tale* gene fragments per strain. These numbers are similar to the number of *tale* genes in Citrus- and cotton-pathogenic xanthomonads and less than in rice-pathogenic xanthomonads from Asia (Gonzalez et al., 2007; Scholze & Boch, 2011). Interestingly, strains from small-grain cereals appear to contain more *tale* genes (four to eight) than strains that infect grasses (zero to four *tale* genes). It is tempting to speculate that these differences are linked to domestication and levels of agricultural intensification. Production of small-grain cereals is certainly more intensified than grass production. Similarly, rice production in Asia is more intensified than rice production in Africa, and Asian rice of the species *Oryza sativa* was domesticated before the African rice species *Oryza glaberrima* (Wang et al., 2014). And reminiscent of the situation with *X. translucens* from more or less intensified crops, rice-pathogenic xanthomonads from Asia that evolved on *O. sativa* and its predecessors have more *tale* genes that African rice pathogens that evolved on *O. glaberrima* and its predecessors (Gonzalez et al., 2007).

Comparison of *tale* RFLP profiles from *X. translucens* pathogens on barley and/or wheat, i.e. *translucens*, *hordei* and *undulosa*, indicated that some TALEs appear to be conserved at the pathovar level, as far as one can tell from Southern blot experiments. For instance, all *hordei* and *translucens* strains shared two *Bam*HI fragments of ~3.0 and ~3.3 kb. Other bands were common to the *hordei* strains but absent in the pathovars *translucens* and *undulosa*,

such as a *Bam*HI fragment of ~2.3 kb. Since the analyzed strains originate from very distant sites, they are unlikely to be clonal. Broad conservation of *tale* genes, as evidenced by identical restriction fragments, may indicate that they play an important role in the pathogenicity of the strains, such as contributing to colonization of certain host plants.

Eight repeat arrays from *X. translucens tale* genes were sequenced in this study, which brings the total number of publicly available repeat arrays to 19, including eight *tale* genes that were very recently deposited at GenBank. Among the eight new sequences, we present the first repeat arrays from the pathovars *hordei*, defined as barley-restricted pathogen. The sizes of the repeat arrays vary between 11.5 and 16.5 repeats. Using a reporter system it has been shown that 10.5 repeats are sufficient for full gene induction (Boch et al., 2009). Hence, all cloned repeat arrays are likely to belong to functional *tale* genes. Due to polymorphisms in the restriction sites for *Bam*HI and *Sph*I, which were considered to be widely conserved, repeat array sizes cannot be predicted from Southern blot analyses using these two restriction enzymes. The available sequences will help to develop new Southern schemes for *X. translucens tale* genes. For instance, genomic DNA could be digested with *Sph*I, which has a conserved site 37 bp upstream of the repeat array, and *Aat*II, which has a conserved site 88 bp downstream of the repeat array, followed by hybridization with a repeat-derived probe. Alternatively, repeat array sizes can be deciphered by PCR analyses, but this technique has strong limitations as well due to the repetitive character of the DNA sequence (Hommelsheim et al., 2014). The limitation become more evident the

more *tale* genes are present in a given strain, which may make analyses of cereal pathogens cumbersome. Finally, genomes may be sequenced completely using long-read next generation sequencing technologies, such as single-molecule nanopore sequencing (Wang et al., 2015).

Sequence comparison of complete *tale* genes and repeats arrays from three pathovars revealed that TALEs from *X. translucens* have unique structural features. First, most repeats consist of 35 amino acids which is rare for clade-2 xanthomonds and typical for TALEs from *R. solanacearum*. Second, most *X. translucens* TALEs contain both 34-aa and 35-aa repeats, which has only been observed once for AvrHah1 from *X. gardneri*. Third, the predominant 35-aa repeats from *X. translucens* share a short sequence motif, Pro_{33}-Gly_{34}, with the repeats from *R. solanacearum*. Yet, sequence similarities of *X. translucens* TALEs are higher with TALEs from clade-2 xanthomonads than with TALEs from *R. solanacearum*. Assuming that clade-1 and clade-2 strains acquired Hrp T3SS independently of each other (Jacobs et al., 2015), we also suspect an independent acquisition of TALEs in clade-1 xanthomonads. Indeed, the two species *X. hyacinthi* and *X. theicola*, which are the closest neighbors of *X. translucens* within clade-1, have a similar Hrp T3SS and possess *tale* genes (Fischer Le-Saux, personal communication). Interestingly, two distinct families of *tale* genes appear to co-exist in strains of *X. translucens*. For the eigth *tale* genes from strain Xtu4699, within-family sequence identity of the N- and C-terminal domains is at ~95% whereas the inter-family sequence identity is at only ~85%. Nevertheless, both families contain similar repeat modules of 34 or 35

amino acids. Complete sequences of more *tale* genes are required to analyze how stable the co-existence of both families is and how frequent hybrid genes may evolve due to homologous recombination.

Sequencing of repeat arrays also revealed the existence of hitherto undescribed RVDs. For some of them, recognition specificity can be predicted from reporter assays (Yang et al., 2014; Juillerat et al., 2015; Miller et al., 2015). Unfortunately, these systematic studies only involved canonical RVDs with two amino acid residues. But single-residue "RVDs" exist as well. For instance, a single asparagine (N) was predicted to mediate binding to cytosine and thymine, and a single histidine (H) was predicted to bind at least to thymine (Moscou & Bogdanove, 2009). The single tyrosine (Y) found in TALEs from *X. translucens* pv. *undulosa* may bind to cytosine and thymine as well, but this needs to be confirmed experimentally.

A mutational approach allowed us to test the contribution of individual *tale* genes to the pathogenicity of *X. translucens*. Four different *tale* genes, based on their DNA fragment sizes in Southern blots, were hit. For three of them, no phenotype was observed under our experimental conditions, which may indicate that these genes are not functional, function redundantly, have only a minor contribution to pathogenicity, or play a role in the interaction with other plants. Interestingly, one gene was knocked out in two strains from the pathovars *hordei* and *translucens* by insertion mutation, and both mutants behaved similarly in our pathogenicity assays. Similar observations were done in other pathosystems where typically only one out of several *tale* genes plays a major role in pathogenicity (Hutin et al., 2015). For complementation of the *tale* mutant, we relied

on authentic sequences from *X. translucens*. First, it was desirable to make sure that all sequence motifs, such as the type III secretion/translocation signal and the acidic activation domain, are functional in our pathosystem. Second, due to the distinct structural features of the repeats it had to be ensured that the repeat array is compatible with upstream and downstream sequences, which contain extra "pseudo repeats" that contribute to DNA binding (Gao et al., 2012). Third, due to the presence of new RVDs with uncertain specificity, available TALE designer kits were not an option and the repeat array had to be isolated from the genomic sequence via PCR, because gene synthesis was also not possible due to the repetitive character of this region. To facilitate reconstruction of *X. translucens tale* genes, we therefore constructed a new vector, pXtale, which allows Golden Gate cloning of *tale* repeat arrays. Using this approach, we were able to complement the *tale* mutant(s).

In conclusion, we provide an overview about *tale* genes in *X. translucens*, give new insight into the structure of *tale* genes and we isolated the first *tale* gene from a clade-1 *Xanthomonas* that contributes to pathogenicity. With only 11.5 repeats, this is the smallest *tale* gene for which a function, i.e. a virulence effect could be demonstrated. Other major virulence TALEs contain at least 17.5 repeats, e.g. Tal5 from *X. oryzae* pv. *oryzae* and PthA4 from *X. citri* pv. *citri* (Streubel et al., 2013; Hu et al., 2014). Even if uncertainties with RVD specificity are clarified, it will still be difficult to reliably predict target sites in the barley genome the relatively small number on RVDs results in many candidate target genes with similar low scores (data not shown). Therefore, transcriptomic analyses will be

required to identify the corresponding susceptibility gene (Wilkins et al., 2015).

References

Boch J, Bonas U. 2010. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-436.

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

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512.

Castiblanco LF, Gil J, Rojas A, Osorio D, Gutiérrez S, Muñoz-Bodnar A, Perez-Quintero AL, Koebnik R, Szurek B, López C, Restrepo S, Verdier V, Bernal AJ. 2013. TALE1 from *Xanthomonas axonopodis* pv. *manihotis* acts as a transcriptional activator in plant cells and is important for pathogenicity in cassava plants. *Mol Plant Pathol* **14**, 84-95.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A Sequence Logo Generator. *Genome Research* **14**, 1188–1190

Cunnac S, Occhialini A, Barberis P, Boucher C, Genin S. 2004. Inventory and functional analysis of the large Hrp regulon in *Ralstonia solanacearum*: identification of novel effector proteins translocated to plant host cells through the type III secretion system. *Mol Microbiol* **53**, 115-128.

de Lange O, Schreiber T, Schandry N, Radeck J, Braun KH, Koszinowski J, Heuer H, Strauß A, 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 Phytol* **199**, 773-786.

de Lange O, Wolf C, Dietze J, Elsaesser J, Morbitzer R, Lahaye T. 2014. Programmable DNA-binding proteins from *Burkholderia* provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res* **42**, 7436-7449.

Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N. 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720-723.

Dereeper A, Audic S, Claverie JM, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* **10**, 8.

Durfee T, Nelson R, Baldwin S, Plunkett G 3rd, Burland V, Mau B, Petrosino JF, Qin X, Muzny DM, Ayele M, Gibbs RA, Csörgo B, Pósfai G, Weinstock GM, Blattner FR. 2008. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* **190**, 2597-2606.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Engler C, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, e3647.

Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc Natl Acad Sci U S A* **76**, 1648-1652.

Gao H, Wu X, Chai J, Han Z. 2012. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* **22**, 1716-1720.

Gardiner DM, Upadhyaya NM, Stiller J, Ellis JG, Dodds PN, Kazan K, Manners JM. 2014. Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS One* **9**, e84995.

Gonzalez C, Szurek B, Manceau C, Mathieu T, Séré Y, Verdier V. 2007. Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. *Mol Plant Microbe Interact* **20**, 534-546.

Guo Y, Figueiredo F, Jones J, Wang N. 2011. HrpG and HrpX play global roles in coordinating different virulence traits of *Xanthomonas axonopodis* pv. *citri*. *Mol Plant Microbe Interact* **24**, 649-661.

Gürlebeck D, Thieme F, Bonas U. 2006. Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. *J Plant Physiol* **163**, 233-255.

He SY, Nomura K, Whittam TS. 2004. Type III protein secretion mechanism in mammalian and plant pathogens. *Biochim Biophys Acta* **1694**, 181-206.

Heuer H, Yin YN, Xue QY, Smalla K, Guo JH. 2007. Repeat domain diversity of *avrBs3*-like genes in *Ralstonia solanacearum* strains and association with host preferences in the field. *Appl Environ Microbiol* **73**, 4379-4384.

Hommelsheim CM, Frantzeskakis L, Huang M, Ülker B. 2014. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci Rep* **4**, 5052.

Hu Y, Zhang J, Jia H, Sosso D, Li T, Frommer WB, Yang B, White FF, Wang N, Jones JB. 2014. *Lateral organ boundaries 1* is a disease susceptibility gene for citrus bacterial canker disease. *Proc Natl Acad Sci U S A* **111**, E521-E529.

Hutin M, Pérez-Quintero AL, Lopez C, Szurek B. 2015. MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Front Plant Sci* **6**, 535.

Jacobs JM, Pesce C, Lefeuvre P, Koebnik R. 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front Plant Sci* **6**, 431.

Juillerat A, Pessereau C, Dubois G, Guyot V, Maréchal A, Valton J, Daboussi F, Poirot L, Duclert A, Duchateau P. 2015. Optimized tuning of TALEN specificity using non-conventional RVDs. *Sci Rep* **5**, 8150.

Kay S, Boch J, Bonas U. 2005. Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol Plant Microbe Interact* **18**, 838-848.

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM 2nd, Peterson KM. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**, 175-176.

Lahaye T, Bonas U. 2001. Molecular secrets of bacterial type III effector proteins. *Trends Plant Sci* **6**, 479-485.

Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716-719.

Miller JC., Zhang L., Xia DF., Campo JJ., Ankoudinova IV., Guschin DY., Rebar EJ. 2015. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat Meth* **15**, 465-471.

Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA pecognition by TAL effectors. *Science* **326**, 1501.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443-453.

Oke V, Long SR. 1999. Bacterial genes induced within the nodule during the *Rhizobium*–legume symbiosis. *Mol Microbiol* **32**, 837-849.

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D. 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int J Syst Evol Microbiol* **57**, 2881-2887.

Pesce C, Bolot S, Cunnac S, Portier P, Fischer-Le Saux M, Jacques MA, Gagnevin L, Arlat M, Noël LD, Carrère S, Bragard C, Koebnik R. 2015. High-quality draft genome sequence of the *Xanthomonas translucens* pv. *cerealis* pathotype strain CFBP 2541. *Genome Announc* **3**, e01574-14.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques MA, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M, Rott P. 2009. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* **10**, 616.

Rees WA, Yager TD, Korte J, von Hippel PH. 1993. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**, 137-144.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Scholze H, Boch J. 2010. TAL effector-DNA specificity. *Virulence* **1**, 428-432.

Scholze H, Boch J. 2011. TAL effectors are remote controls for gene activation. *Curr Opin Microbiol* **14**, 47-53.

Schornack S, Meyer A, Römer P, Jordan T, Lahaye T. 2006. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J Plant Physiol* **163**, 256-272.

Schornack S, Minsavage GV, Stall RE, Jones JB, Lahaye T. 2008. Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol* **179**, 546-556.

Streubel J, Pesce C, Hutin M, Koebnik R, Boch J, Szurek B. 2013. Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol* **200**, 808-819.

Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, Grant M. 2011. Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel)* **2**, 1050-1065.

Tampakaki AP, Skandalis N, Gazi AD, Bastaki MN, Sarris PF, Charova SN, Kokkinidis M, Panopoulos NJ. 2010. Playing the "Harp": evolution of our understanding of *hrp*/*hrc* genes. *Annu Rev Phytopathol* **48**, 347-370.

Wang CL, Qin TF, Yu HM, Zhang XP, Che JY, Gao Y, Zheng CK, Yang B, Zhao KJ. 2014. The broad bacterial blight resistance of rice line CBB23 is triggered by a novel transcription activator-like (TAL) effector of *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Pathol* **15**, 333-341.

Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, Zuccolo A, Song X, Kudrna D, Ammiraju JS, Cossu RM, Maldonado C, Chen J, Lee S, Sisneros N, de Baynast K, Golser W, Wissotski M, Kim W, Sanchez P, Ndjiondjop MN, Sanni K, Long M, Carney J, Panaud O, Wicker T, Machado CA, Chen M, Mayer KF, Rounsley S, Wing RA. 2014. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet* **46**, 982-988.

Wang Y, Yang Q, Wang Z. 2015. The evolution of nanopore sequencing. *Front Genet* **5**, 449.

Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R. 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol Plant Pathol* **14**, 576-588.

Wilkins KE, Booher NJ, Wang L, Bogdanove AJ. 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.

Wonni I, Cottyn B, Detemmerman L, Dao S, Ouedraogo L, Sarra S, Tekete C, Poussier S, Corral R, Triplett L, Koita O, Koebnik R, Leach J, Szurek B, Maes M, Verdier V. 2013. Analysis of *Xanthomonas oryzae* pv. *oryzicola* population in Mali and Burkina Faso reveals a high level of genetic and pathogenic diversity. *Phytopathology* **104**, 520-531.

Yang B, Sugio A, White FF. 2006. *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proc Natl Acad Sci U S A* **103**, 10503-10508.

Yang J, Zhang Y, Yuan P, Zhou Y, Cai C, Ren Q, Wen D, Chu C, Qi H, Wei W. 2014. Complete decoding of TAL effectors for DNA recognition. *Cell Res* **24**, 628-631.

Young JM, Park DC, Shearman HM, Fargier E. 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **31**, 366-377.

Supplemental data VI.S1 Oligonucleotides used in this study.

Name	Purpose	Sequence
pVO155-Fw	Verification of constructs in pVO155	5′-GAGATCCCCAGCCCGCCTAATG
pVO155-Rev2		5′-GCGCGTTACAAGAAAGCCGG
pBBR-Fw	Verification of constructs in pBBR1MC-5	5'-AGAAGCACACGGTCACACTG
pBBR-Rev3		5'-CCGGCTCGTATGTTGTGTGG
pBBR-BsaI-Fw	Elimination of *Bsa*I restriction site from pBB1MCS-5	5'-GAGACGGCCCGAGA**T**CTTCGGGAGCGCCTG
pBBR-BsaI-Rev		5'-CAGGCGCTCCCGAAG**A**TCTCGGGCCGTCTC
Xt-Tal-fw	Sequencing of *tale* repeats arrays in pGEM-T easy	5'-GTCCCGACCTATTGACCCGT
Xt-Tal-Rev		5'-CGCAATGCGCTGACTGGCGC
Xt-RVDseq-Fw1	Sequencing of *tale* repeats arrays in pXtale	5'-ACGTGGCGGCGTGAACGCGG
Xt-RVDseq-Rev1		5'-TCCGGCCTTGGCCGCGTTGA
Xt-Nt-Tal-FW	Generation of Southern probe corresponding to the N terminus of TALEs from *X. translucens*	5'-GATGGCTTGCCCGCTCG
Xt-Nt-Tal-Rev2		5'-CCTCCACCGCGGTCACG
Xt-Ct-Tal-Fw	Generation of Southern probe corresponding to the C terminus of TALEs from *X. translucens*	5'-GCCCAGTTATCCAGCCCTG
Xt-Ct-Tal -Rev2		5'-GATGCGGTGCCAACGCTG
Xt-tal-KO-Fw	Cloning of an internal tale fragment for mutation	5'-GCAGGATCCGCCCAGTTATCCAGCCCTG
Xt-tal-KO-Rev		5'-CGCTCTAGATGCGGTGCCAACGCTG
Xt-tal-Rep-Fwd3	Golden Gate cloning (*Bsa*I) of repeat arrays in pXtale	5'-CTAGGGGTCTCGGAGGCCGTGCATGCATCGCGCAAT GCGCTGACTGGCGC
Xt-tal-Rep-Rev1		5'-CTAGGGTCTCTCAGGGCTGGATAACTGGGC

Restriction sites of the enzymes used for cloning are underlined.

<u>Supplemental data VI.S2</u> Nucleotide sequence of the pXtale vector.

Vector based on pBBR1MCS-5 (accession number U25061) containing synthetic gene fragments for the N- and C-terminal domains of a *tale* gene from *X. translucens* pv. *translucens* DSM 18974 (accession number CCP38287). The coding region for the N-terminal domain is shown in blue color and the coding region for the C-terminal domain is shown in green color. The *tale* start codon is highlighted in green and the stop codon is highlighted in yellow. The three *Bsa*I sites are highlighted in grey and the mutated nucleotide of the vector's *Bsa*I site (encompassing nucleotides 4766 to 3 in pBBR1MCS-5) is in red color.

CCCAGGGCGTCCAGAACGGGCTTCAGGCGCTCCCGAAGATCTCGGGCCGTCTCTTGGGCTTG
ATCGGCCTTCTTGCGCATCTCACGCGCTCCTGCGGCGGCCTGTAGGGCAGGCTCATACCCCT
GCCGAACCGCTTTTGTCAGCCGGTCGGCCACGGCTTCCGGCGTCTCAACGCGCTTTGAGATT
CCCAGCTTTTCGGCCAATCCCTGCGGTGCATAGGCGCGTGGCTCGACCGCTTGCGGGCTGAT
GGTGACGTGGCCCACTGGTGGCCGCTCCAGGGCCTCGTAGAACGCCTGAATGCGCGTGTGAC
GTGCCTTGCTGCCCTCGATGCCCCGTTGCAGCCCTAGATCGGCCACAGCGGCCGCAAACGTG
GTCTGGTCGCGGGTCATCTGCGCTTTGTTGCCGATGAACTCCTTGGCCGACAGCCTGCCGTC
CTGCGTCAGCGGCACCACGAACGCGGTCATGTGCGGGCTGGTTTCGTCACGGTGGATGCTGG
CCGTCACGATGCGATCCGCCCCGTACTTGTCCGCCAGCCACTTGTGCGCCTTCTCGAAGAAC
GCCGCCTGCTGTTCTTGGCTGGCCGACTTCCACCATTCCGGGCTGGCCGTCATGACGTACTC
GACCGCCAACACAGCGTCCTTGCGCCGCTTCTCTGGCAGCAACTCGCGCAGTCGGCCCATCG
CTTCATCGGTGCTGCTGGCCGCCCAGTGCTCGTTCTCTGGCGTCCTGCTGGCGTCAGCGTTG
GGCGTCTCGCGCTCGCGGTAGGCGTGCTTGAGACTGGCCGCCACGTTGCCCATTTTCGCCAG
CTTCTTGCATCGCATGATCGCGTATGCCGCCATGCCTGCCCCTCCCTTTTGGTGTCCAACCG
GCTCGACGGGGGCAGCGCAAGGCGGTGCCTCCGGCGGGCCACTCAATGCTTGAGTATACTCA
CTAGACTTTGCTTCGCAAAGTCGTGACCGCCTACGGCGGCTGCGGCGCCCTACGGGCTTGCT
CTCCGGGCTTCGCCCTGCGCGGTCGCTGCGCTCCCTTGCCAGCCCGTGGATATGTGGACGAT
GGCCGCGAGCGGCCACCGGCTGGCTCGCTTCGCTCGGCCCGTGGACAACCCTGCTGGACAAG
CTGATGGACAGGCTGCGCCTGCCCACGAGCTTGACCACAGGGATTGCCCACCGGCTACCCAG
CCTTCGACCACATACCCACCGGCTCCAACTGCGCGGCCTGCGGCCTTGCCCCATCAATTTTT
TTAATTTTCTCTGGGGAAAAGCCTCCGGCCTGCGGCCTGCGCGCTTCGCTTGCCGGTTGGAC
ACCAAGTGGAAGGCGGGTCAAGGCTCGCGCAGCGACCGCGCAGCGGCTTGGCCTTGACGCGC
CTGGAACGACCCAAGCCTATGCGAGTGGGGGCAGTCGAAGGCGAAGCCCGCCCGCCTGCCCC
CCGAGCCTCACGGCGGCGAGTGCGGGGGTTCCAAGGGGGCAGCGCCACCTTGGGCAAGGCCG
AAGGCCGCGCAGTCGATCAACAAGCCCCGGAGGGGCCACTTTTTGCCGGAGGGGGAGCCGCG
CCGAAGGCGTGGGGGAACCCCGCAGGGGTGCCCTTCTTTGGGCACCAAAGAACTAGATATAG
GGCGAAATGCGAAAGACTTAAAAATCAACAACTTAAAAAAGGGGGGTACGCAACAGCTCATT
GCGGCACCCCCCGCAATAGCTCATTGCGTAGGTTAAAGAAAATCTGTAATTGACTGCCACTT
TTACGCAACGCATAATTGTTGTCGCGCTGCCGAAAAGTTGCAGCTGATTGCGCATGGTGCCG
CAACCGTGCGGCACCCTACCGCATGGAGATAAGCATGGCCACGCAGTCCAGAGAAATCGGCA
TTCAAGCCAAGAACAAGCCCGGTCACTGGGTGCAAACGGAACGCAAAGCGCATGAGGCGTGG
GCCGGGCTTATTGCGAGGAAACCCACGGCGGCAATGCTGCTGCATCACCTCGTGGCGCAGAT
GGGCCACCAGAACGCCGTGGTGGTCAGCCAGAAGACACTTTCCAAGCTCATCGGACGTTCTT
TGCGGACGGTCCAATACGCAGTCAAGGACTTGGTGGCCGAGCGCTGGATCTCCGTCGTGAAG
CTCAACGGCCCCGGCACCGTGTCGGCCTACGTGGTCAATGACCGCGTGGCGTGGGGCCAGCC
CCGCGACCAGTTGCGCCTGTCGGTGTTCAGTGCCGCCGTGGTGGTTGATCACGACGACCAGG
ACGAATCGCTGTTGGGGCATGGCGACCTGCGCCGCATCCCGACCCTGTATCCGGGCGAGCAG
CAACTACCGACCGGCCCCGGCGAGGAGCCGCCCAGCCAGCCCGGCATTCCGGGCATGGAACC
AGACCTGCCAGCCTTGACCGAAACGGAGGAATGGGAACGGCGCGGGCAGCAGCGCCTGCCGA

TGCCCGATGAGCCGTGTTTTCTGGACGATGGCGAGCCGTTGGAGCCGCCGACACGGGTCACG
CTGCCGCGCCGGTAGCACTTGGGTTGCGCAGCAACCCGTAAGTGCGCTGTTCCAGACTATCG
GCTGTAGCCGCCTCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGG
AGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCGTTTTT
ATCAGGCTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGC
CTGAGCAAACTGGCCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATA
AACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACC
GGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAG
CACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCA
TCGCAGTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAA
CAAAATATTAACGCTTACAATTTCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCG
ATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGAT
TAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGC
GCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGATC
AATGAGGCAATAGCGTCATCAACCATGCGATCTCCTCTTGGTCGAATACCGGGAAATCTTCC
GCTGCGCCTGCGAAGGGATCCGCATCTTGCTCCAAGAACGCAGCGCTGGACGCTGCCAGTTC
GCCGTCCGTCGGCATGCCAGGATCCGGGAGGCCGCCGATCCTGGTACGCTGGCGTTTTACGC
CCCAGCTACTGAGGGGGAGCAAAGGCGGTGCATCGCGCTGTTCGGGAACGAATACCTCGGCA
GCCTGCTGTGCAGAGGAGCGGTTGACAGAACTCTCTGATCGGGACCGTTTGCGGTTGCTTGC
CCGAGCCCGGTCTGCCTCGTGAATTGGGCTGGGCGCGTCCAGTTCACGCTCCAGGGAATCGG
CGAATGCATCCACGGACTCCTGGCTCGGAGTTTGAGCCGAAGCAGAGGACGGCTTCGCCCCT
TTCACCCCTAATGCTTGAAGGATGCGGTGCCAACGCTGGGAGGCTGGGGGAAGCGTTCCACT
GATGGCTTCGAGTTCTGTGACGCCCACACGGCGAAATAGCTGCAACAATCCTTCCCTGCTCA
TCTCAAACCGCGTCATGGCTTCATGAAATGCTTGCGCTGGGTGGGAGTGGCACTGGAAAAAC
CCCAGCACGCGAACCACTTGCGCGTGGTCGGCAACCAGATGGGCCGTGCCTTCGGGAACACG
GTTATGGGTTCTTGTGACCAATGCCGGCGCGTGCGGCAATCCCTTTTTCACCGCATCCAGCG
CAGGACGTCCGCCGAGGCAGGCCAAGGCGACGAGGCGGTCGTTGGTCAACGCGGCCAAGGCC
GGATCAGGTGAGACCCTCGTTGATGAGCTCGCAGGTCTCACCTCCACCGCGTTCACGCCGCC
ACGTTTTGCAATCTTGAGGAGTTGGCCTGTGACCAACTGTAACGGTGGAGCTCTCAACTCTT
CCGCCACCGTGAGCAAGGCCTCCAGGGCCCGTGCGCCCGACTTCTGTTTGCCGACGCCGACG
ATGTCTTCGTGTGTCGCCTCTGGCAACGCCGCAATCATCTCCGAATACCTGGCCGCTATGGT
CCCTAGCGCTGGCGGGTGTTTGCTGAGCTCAACGATGTGCGCGTGTGTAAACCCATGGCCGA
CCAGTGCCTCGTGGTGCTGCGCCACTGTCGAACGCACCTTCGGTTTGATCTTCTCCTGTTGC
TGCTGGCTGTAGCCGAACGTGCCTAGATCCACATGCGCGGCAGGCAAAGCGTCGGAGGTCTG
GGCAGAACGCCGCCGCGCCGCCGCTTTGGTGCGCGGCGGCGCGGCAGTGATCGCCGCACTTG
CTGAGGGCTGCGGGTCATCTACTGCACGCAGAGCCGATTGCACCTCATCCCCCTCTCCTGGA
GCAGACTCTGCACGAGCAGCGCCGAAGGAAGGCAATGAATCAAAAGGCGATCCAGCAAAAAG
CGACGGATCAATCTGCTGACGTAGCAGGCCGCTCAAGCTGCCCGCTGAAAACGCCGGCGAGC
GTGCAGGGGAGGGCGGCGGACTGGTCCGGGATACCATCCGCCGAGCGGGCAAGCCATCCAGG
GGACTGCTGGCAGGCGTAGACACCAACCGATCCGCGATGGGCTGGACCGCATCCGGCTGAGA
TCCGGCCTGAAGCTCGCGAGCAGTACTTGGCGTGCGGGAGCGAATAGGCTCCATCAGTCATT
CCTCTTTCAAGCTTATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTC
CCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA
ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTC
GGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC
GTATTGGGCGCATGCATAAAAACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCA
TCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATA
ATATTTGCCCATGGACGCACACCGTGGAAACGGATGAAGGCACGAACCCAGTTGACATAAGC

CTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAACCT
TGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGTTATGACTGTTTT
TTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTT
GATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAACGATGTTACGCAGCAGGGCAGTCGC
CCTAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATGTAGGCTCGGCCCTGA
CCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCA
CCTACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACTTGCTCCGTAGTAAGACATTC
ATCGCGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCC
CAGGTTTGAGCAGCCGCGTAGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACC
GGAGGCAGGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGT
GCTTATGTGATCTACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAA
GTTGGGCATACGGGAAGAAGTGATGCACTTTGATATCGACCCAAGTACCGCCACCTAACAAT
TCGTTCAAGCCGAGATCGGCTTCCCGGCCGCGGAGTTGTTCGGTAAATTGTCACAACGCCGC
CAGGTGGCACTTTTCGGGGAAATGTGCGCGCCCGCGTTCCTGCTGGCGCTGGGCCTGTTTCT
GGCGCTGGACTTCCCGCTGTTCCGTCAGCAGCTTTTCGCCCACGGCCTTGATGATCGCGGCG
GCCTTGGCCTGCATATCCCGATTCAACGGC

<u>Supplemental data VI.S3</u> Multiple sequence alignments of the coding region for the N- and the C-terminal domains of eleven *X. translucens* TALEs.

A. Alignment of coding regions for the N-terminal domains. B. Alignment of coding regions for the C-terminal domains. Eight *tale* genes are from *X. translucens* pv. *undulosa* strain Xtu4699 (tal-A to H) (acc. no. AKK66437), tal-DSM is from *X. translucens* pv. *translucens* strain DSM 18974 (acc. no. CCP38287) and tal1 and tal2 are from *X. translucens* pv. *cerealis* strain CFBP 2541 (acc. no. CM003052). Sequences similar to *Bam*HI sites (GGATCC) and *Sph*I sites (GCATGC) are shown in red and blue clore, respectively. Oligonucleotides used to generate Southern probes are highlighted in grey. Primer sequences used for amplification of repeats arrays are underlined.

A

```
tal2      ATGGATCCCATTCGTTCTCGCACGCCAATTCCTGCCCGCGAGCTTCAGGCCGGAGCCCAA
tal-B     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCCGGATCCCAA
tal-E     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCCGGATCTCAG
tal-C     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCCGGATCTCAG
tal-H     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCCGGATCTCAG
tal-F     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCCGGATCTCAG
tal-A     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCCGGATCCCAA
tal-G     ATGGATCCCATTCGTTCTCGCACGCCAAGTCCTGCCCGCGAGCTTCTGGCCGGATCCCAA
tal1      ATGGAGCCTATTCGCTCCCGCACGCCAAGTCCTGCTCGCGAGCTTCAGGCAGGATTTCAG
tal-D     ATGGAGCCTATTCGCTCCCGCACGCCAAGTACTGCTCGCGAGCTTCAGGCCGGATCTCAG
tal-DSM   ATGGAGCCTATTCGCTCCCGCACGCCAAGTACTGCTCGCGAGCTTCAGGCCGGATCTCAG
          ***** ** ***** ** ********** * **** ********** *** ***   **

tal2      CCGGATGGGGTCCAGCCGACTGCAGATCGGGGCGGGTCTCCGCCTGCTGGCAGCCCCCTG
tal-B     CCGGATGGGGTCCAGCCGACTGCAGATCCGAGGGTGTCTCCGCCTGCTGGCAGCCCCCTG
tal-E     CCGGATGCGGTCCAGCCCATTGCGGATCGGTTGGTGTCTCCGCCTGCTGGCAGCCCGCTG
tal-C     CCGGATGCGGTCCAGCCCATTGCGGATCGGTTGGTGTCTCCGCCTGCTGGCAGCCCGCTG
tal-H     CCGGATGCGGTCCAGCCCATTGCGGATCGGTTGGTGTCTCCGCCTGCTGGCAGCCCGCTG
tal-F     CCGGATGCGGTCCAGCCCATTGCGGATCGGTTGGTGTCTCCGCCTGCTGGCAGCCCGCTG
tal-A     CCGGATGGGGTCCAGCCGACTGCAGATCCGAGGGTGTCTCCGCCTGCTGGCAGCCCCCTG
tal-G     CCGGATGGGGTCCAGCCGACTGCAGATCCGAGGGTGTCTCCGCCTGCTGGCAGCCCGCTG
tal1      CTGGAGGCGGTCCAGCCCATTGCGGATCGGTTGGTGTCTATGCCTGCCAACAGTCCCCTG
tal-D     CCGGATGCGGTCCAGCCCATCGCGGATCGGTTGGTGTCTACGCCTGCAAGCAGTCCCCTG
tal-DSM   CCGGATGCGGTCCAGCCCATCGCGGATCGGTTGGTGTCTACGCCTGCCAGCAGTCCCCTG
          * *** * ********* *  ** **** *   * ****  ******   *** ** ***

tal2      GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGGTGCCGCCGCCCCCTGCAAGC
tal-B     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGCCGCCCCCTGCAAGC
tal-E     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGTCGCCCCCTGCAAGC
tal-C     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGTCGCCCCCTGCAAGC
tal-H     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGTCGCCCCCTGCAAGC
tal-F     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGTCGCCCCCTGCAAGC
tal-A     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGCCGCCCCCTGCAAGC
tal-G     GATGGCTTGCCCGCTCGGCGGACGATGTCCCGGACCCAGCTGCCGCCGCCCTCTGCAAGC
tal1      GATGGCTTGCCCGCTCGGCGGATGGAGTCCCGGGCCAGTCCGCCGCCCTCCCCTGCACGC
tal-D     GATGGCTTGCCCGCTCGGCGGATGGTATCCCGGACCAGTCCGCCGCCCTCCCCTGCACGC
tal-DSM   GATGGCTTGCCCGCTCGGCGGATGGTATCCCGGACCAGTCCGCCGCCCTCCCCTGCACGC
          ********************* *   ****** **     **** *  ** ***** **
```

```
tal2       GTGCCTGCGTTCTCAGCGGGCAGCTTCGGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-B      GTGCCTGCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-E      GTGCCTGCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-C      GTGCCTGCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-H      GTGCCTGCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-F      GTGCCTGCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-A      GGGCCTTCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal-G      GGGCCTGCGTTCTCAGCGGGCAGCTTCAGCGACCTGTTACGT---CAGGTTGATTCGTCG
tal1       TCGCCTGCGTTTTCAGCGGGCAGCTTGAGCGACATGCTGCGTCAGCAGATTGATCCGTCG
tal-D      TCGCCGGCGTTTTCAGCGGGCAGCTTGAGCGGCCTGCTACGTCAGCAGATTGATCCGTCG
tal-DSM    TCGCCGGCGTTTTCAGCGGGCAGCTTGAGCGGCCTGCTACGTCAGCAGATTGATCCGTCG
             ***  **** **************  *** * ** * ***   *** ***** *****

tal2       CTTTTTGATGCATCGCTTTTTGATTCTATGCCTGCCTTCGGCGCTCATCATGCAGCGGCT
tal-B      CTTTTTGATGCATCGTTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCG
tal-E      CTTTTTGATGCATCGTTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCT
tal-C      CTTTTTGATGCATCGTTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCT
tal-H      CTTTTTGATGCATCGTTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCG
tal-F      CTTTTTGATGCATCGTTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCT
tal-A      CTTTTTGATGCATCGTTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCT
tal-G      CTTTTTAATGCATCGCTTTTTGATTCGATGCCTGCCTTCGGCGCTCATCATGCACAGGCT
tal1       CTTTTTGCTGGATCGCCTTTTGATTTATTGCCTCCCTTCGCCGCTGCTCGTGCAGAGGCT
tal-D      CTTTTTGCTGGATCGCCTTTTGATTCATTGCCTTCCTTCGGCGCTGCTCGTGCAGAGTCT
tal-DSM    CTTTTTGCTGGATCGCCTTTTGATTCATTGCCTTCCTTCGGCGCTGCTCGTGCAGAGTCT
           ******  ** ****  ********   ***** ****** ****  ** ****  * *

tal2       GCCACAGGCGAGTTGGATGAGGTGCAATCGGCTCTGCGTGCAGCCGATGACCCGCAGCCC
tal-B      GCCACAGGCGAGTTGGATGAGGTGCAATCGGCTCTGCGTGCAGCCGATGACCCGCAGCCC
tal-E      GCCACAGGCGAGTTGGATGAGGTGCAATCGGCTCTGCGTGCAGCCGATGACCCGCAGCCC
tal-C      GCCACAGGCGAGTTGGATGAGGTGCAATCGGCTCTGCGTGCAGCCGATGACCCGCAGCCC
tal-H      GCCACAGGCGAGTTGGATGAGGTGCAATCGGCTCTGCGTGCAGCCGATGACCCGCAGCCC
tal-F      GCCACAGGCGAGTTGGATGAGGTGCAATCGGCTCTGCGTGCAGCCGATGACCCGCAGCCC
tal-A      GCCACAGGCGAGTTGGATGAGGCGCAATCGGCTCTGCGTGCAGTAGATGACCCGCAGCCC
tal-G      GCCACAGGCGAGTTGGATGAGGCGCAATCGGCTCTGCGTGCAGTAGATGACCCGCAGTCC
tal1       GCTCCAGGAGAGGGGGATGAGGTGCAATCGGGTCTGCGTACAGCAGATGACCCGCAGCCA
tal-D      GCTCCAGGAGAGGGGGATGAGGTGCAATCGGGTCTGCGTGCAGTAGATGACCCGCAGCCC
tal-DSM    GCTCCAGGAGAGGGGGATGAGGTGCAATCGGCTCTGCGTGCAGTAGATGACCCGCAGCCC
           **  **** ***  ******** ******** ******* ***  ************ *

tal2       CCCTCGCGCGTAGCTGTCACTGCCGC----------------------------ACGGCCA
tal-B      CCCGTGCGCGTAGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAACCGGCCCAGCGGCCA
tal-E      CCCGTGCGCGTAGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAGCCGGCCCAGCGGCCA
tal-C      CCCGTGCGCGTAGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAGCCGGCCCAGCGGCCA
tal-H      CCCGTGCGCGTAGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAGCCGGCCCAGCGGCCA
tal-F      CCCGTGCGCGTAGCTGTCACTGCCGCGCGGCCGCCGCGCGCCAAGCCGGCCCAGCGGCCA
tal-A      TCAGCAAGTGCGGCGATCACTGCCGC---GCCGCCCCCGCACCAAAGCGGCGGCGCGG---
tal-G      TCAGCAAGTGCGGCGATCACTGCCGC---GCCGCCGCGCACCAAAGCGGCGGCGCGG---
tal1       CCCGTGCGCGTAGCTGTCACTGCCGCGCGGCCGCAGCGCGCCAAGCCGGCCCGGCGGCCA
tal-D      TCAGCAAGTGCGGCGATCACTGCCGC---GCCGCCGCGCACCAAAGCGGCGGCGCGG---
tal-DSM    TCAGCAAGTGCGGCGATCACTGCCGC---GCCGCCGCGCACCAAAGCGGCGGCGCGG---
            *     * *  **  **********                              ***

tal2       CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCACATGTGGATCTAAGCACGTTC
tal-B      CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal-E      CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal-C      CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal-H      CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal-F      CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal-A      CGGTGTTCTGCCCAGACCTTGGACGCTTTGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal-G      CGGCGTTCTGCCCAGACCTTGGACGCTTCGCCTGCCGCGGATGTGGATCTAAGCACGTTC
tal1       CGGCGTGCTGCGCAAACCTCCGATGCTTCGCCTGCCGCACATGTGGATCTAAGCACGTTC
tal-D      CGGCGTTCTGCCCAGACCTCCGACGCTTTGCCTGCCGCGCATGTGGATCTAGGCACGTTC
tal-DSM    CGGCGTTCTGCCCAGACCTCCGACGCTTTGCCTGCCGCGCATGTGGATCTAGGCACGTTC
           *** ** **** ** ****  ** **** *********  *********** ********
```

```
tal2       GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGGAAGTGCGTTCGACAGTGGCGCAG
tal-B      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGACGGTGCGTTCGTCAGTGGCGCAG
tal-E      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGACGGTGCGTTCGACAGTGGCGCAG
tal-C      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGACGGTGCGTTCGACAGTGTCGCAG
tal-H      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGACGGTGCGTTCGACAGTGGCGCAG
tal-F      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGACGGTGCGTTCGACAGTGGCGCAG
tal-A      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGTGCGTTCGACAGTGGCGCAG
tal-G      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGACGGTGCGTTCGTCAGTGGCGCAG
tal1       GGCTACAGCCAGCAGCAGCAGGAGAAGATCAAACCGAAGGTGCGTTCGACAGTGGCGCAG
tal-D      GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGTGCGTTCGACATTGGCGCAG
tal-DSM    GGCTACAGCCAGCAGCAACAGGAGAAGATCAAACCGAAGGTGCGTTCGACAGTGGCGCAG
           ***************** ******************   ********* ** ** *****

tal2       TACCACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-B      CACCACGCCGCATTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-E      CACCACGCCGCATTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-C      CACCACGCCGCATTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-H      CACCACGCCGCATTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-F      CACCACGCCGCATTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-A      CACCACGCCGCATTGGTCGGCCATGGATTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-G      CACCACGCCGCATTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal1       CACCACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAGA
tal-D      CACCACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
tal-DSM    CACCACGAGGCACTGGTCGGCCATGGGTTTACACACGCGCACATCGTTGAGCTCAGCAAA
            ******  *** ************* ******************************* *

tal2       CACCAGGCAGCGCTAGGGACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCGTTGCCA
tal-B      CACCCGGCAGCGTTAGGAACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCATTGCCA
tal-E      CACCCGCCAGCGCTAGGGACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCATTGCCA
tal-C      CACCCGGCAGCGTTAGGAACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCATTGCCA
tal-H      CACCCGGCAGCGTTAGGAACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCATTGCCA
tal-F      CACCCGGCAGCGTTAGGAACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCATTGCCA
tal-A      CACCCGCCAGCGCTAGGGACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCGTTGCCA
tal-G      CACCCGGCAGCGTTAGGAACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCATTGCCA
tal1       CACGTGGCAGCGTTAGGGACCATCGCTGGCAGGTATCAAGACATAATCGCGGTGTTGCCA
tal-D      CACCCGCCAGCGCTAGGGACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCGTTGCCA
tal-DSM    CACCCGCCAGCGCTAGGGACCATAGCGGCCAGGTATTCGGAGATGATTGCGGCGTTGCCA
           ***  * ***** **** ***** ** * *******   ** ** ** ****  ******

tal2       GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal-B      GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal-E      GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal-C      GAGGCCGCACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal-H      GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal-F      GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal-A      GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACATTGTGCGGGCGCACGGACCCTG
tal-G      GAGGCCACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCGGGCGCACGGGCCCTG
tal1       GAGGCGAAACACGAAGACATCGTCGGCGTTGGCAAACAGAAGTCGGGCGCACGGGCCCTG
tal-D      GAGGCGACACACGAAGACATCGTCGGCGTCGGCAAACAGAAGTCGGGCGCACGGGCCCTG
tal-DSM    GAGGCGACACACGAAGACATCGTCGGCGTCGGCAAACAGAAGTCGGGCGCACGGGCCCTG
           *****   *************** ***** ********     ************ *****

tal2       GAGGCGTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-B      GAGGCCTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-E      GATGCCTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-C      GAGGCCTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-H      GAGGCCTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-F      GAGACCTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-A      GAGGTCTTGCTCATGGTGGTGCAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAAGC
tal-G      GAGGCCTTGCTCATGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal1       GAGGCCTTGCTAACGGTGGCGGAAGATTTGAGAGGTCCACCGTTACAGTTGGACACAGGC
tal-D      GAGGCCTTGCTCACGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
tal-DSM    GAGGCCTTGCTCACGGTGGCGGAAGAGTTGAGAGCTCCACCGTTACAGTTGGTCACAGGC
           **    ***** * ***** * **** ******* ***************** **** **
```

```
tal2      CAACTCCTCAGGATTGCAAAACGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-B     CAACTCCTCAAGATTGCAAAACGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-E     CAACTCCTCAAGATTGCAAAACGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-C     CAACTCCTCAAGATTGCAAAAGGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-H     CAACTCCTCAAGATTGCAAAACGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-F     CAACTCCTCAAGATTGCAAAAGGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-A     CAACTCCTCAAGATTGCAAAACGTGGTGGCGTGACTGCGGTGGAGGCCGTGCATGCATCG
tal-G     CAACTCCTCAAGATTGCAAAACGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal1      CAACTCCTCAAGATTGCAAAACGTGGTGGCGTGACCGCGGTGGAGGCCGTGCATGCATCG
tal-D     CAACTCCTCAAGATTGCAAAACGTGGCGGCGTGAACGCGGTGGAGGCCGTGCATGCATCG
tal-DSM   CAACTCCTCAAGATTGCAAAACGTGGCGGCGTGAACGCGGTGGAGGCCGTGCATGCATCG
          ********** ********** **** *******  ************************

tal2      CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-B     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-E     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-C     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-H     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-F     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-A     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-G     CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal1      CGCAATGCGCTGACTGGCGCCCCCCTGCAC
tal-D     CGCAATGCGCTGACTGGCGCCCCACTGCAC
tal-DSM   CGCAATGCGCTGACTGGCGCCCCCCTGCAC
          ******************** ******
```

B

```
tal1      AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCGTTTTGACCAACGACCGC
tal2      AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCGTTGGCCGCGTTGAACAACGACCGC
tal-D     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCGCGTTGACCAACGACCGC
tal-DSM   AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCGCGTTGACCAACGACCGC
tal-C     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCGCGTTGACCAACGACCGC
tal-A     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCGC
tal-F     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCGC
tal-E     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCGCGTTGACCAACGACCGC
tal-H     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGGCCGC
tal-B     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCGCGTTGACCAACGACCGC
tal-G     AGTATTTTTGCCCAGTTATCCAGCCCTGATCCGGCCTTGGCCACGTTGACCAACGACCGC
          ********************************** ******   **** ***** ****

tal1      CTCGTCGCCTTGGCCTGCCTCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal2      CTTGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCAGTGAAAAAGGGATTG
tal-D     CTCGTCGCCTTGGCCTGCCTCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-DSM   CTCGTCGCCTTGGCCTGCCTCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-C     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-A     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-F     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-E     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAAGGATTG
tal-H     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-B     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
tal-G     CTCGTCGCCTTGGCTTGCATCGGCGGACGTCCTGCGCTGGATGCGGTGAAAAAGGGATTG
          ** *********** *** ************************ ******** ******
```

```
tal1      CCGCGCGCGCTGGCATTGATCACAAGAACCCGTAACCTTGTTCCCGAAGGCACGGCCCAT
tal2      CCGCACGCGCCGGCATTGATCGCAAGAACCCATAACCGTGTTCCCGACGGCAAGGCCCAT
tal-D     CCGCACGCGCCGGCATTGGTCACAAGAACCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-DSM   CCGCACGCGCCGGCATTGGTCACAAGAACCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-C     CCGCACGCGCCGGCATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-A     CCGCACGCGCCGGCATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-F     CCGCACGCGCCGGCATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-E     CCGCACGCGCCGGCATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-H     CCGCACGCGCCGGCATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-B     CCGCACGCGCCGGCATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
tal-G     CCGCACGCGCCGGAATTGATCACAAGAGTCCATAACCGTGTTCCCGAAGGCACGGCCCAT
          **** ***** ** **** ** *****  ** ***** ********* **** *******

tal1      CTGGTTGCCGATGCCGCGCAAGTGGTTCGCGTGCTGGGGTTTTTCCAGTGCCACTCCCAC
tal2      CTGGTTGCCGATGCCGCGCAAGTGGTTCGCGTGCTGAGTTTTTTTCAGTGCCACTCCCAC
tal-D     CTGGTTGCCGACCACGCGCAAGTGGTTCGCGTGCTGGGGTTTTTCCAGTGCCACTCCCAC
tal-DSM   CTGGTTGCCGACCACGCGCAAGTGGTTCGCGTGCTGGGGTTTTTCCAGTGCCACTCCCAC
tal-C     CTGGTTGCCGACCTCGCGCAAGTGGTTCGCGTGCTGAGTTTTTTCCAGTGCCACTCTCAC
tal-A     CTGGTTGCCGACCTCGCGCAAGTGGTTCGCGTGCTGAGTTTTTTCCAGTGCCACTCTCAC
tal-F     CTGGTTGCCGACCTCGCGCAAGTGGTTCGCGTGCTGAGTTTTTTCCAGTGCCACTCTCAC
tal-E     CTGGTTGCCGACCTCGCGCAAGTGGTTCGCGTGCTGAGTTTTTTCCAGTGCCACTCTCAC
tal-H     CTGGTTGCCGACCTCGCGCAAGTGGTTCGCGTGCTGAGTTTTTTCCAGTGCCACTCTCAC
tal-B     CTGGTTGCCGACCACGCGCAAGTGGTTCGCGTGCTGAGTTTTTTCCAGTGCCACTCCCAG
tal-G     CTGGTTGCCGACCACGCGCAAGTGGTTCGCGTGCTGGGGTTTTTCCAGTGCCACTCCCAG
          ***********   ********************** * ***** *********** **

tal1      CCAGCGCAAGCATTTCATGAAGCCATGACGCGGTTTGAGATGAGCAGTGAAGGATTGTTG
tal2      CCAGCGCAAGCATTTGATGAAGCGATGATGCAATTCGGGATGAGTGGGCGCCGGTTGTTA
tal-D     CCAGCGCAAGCATTTCATGAAGCCATGACGCGGTTTGAGATGAGCAGGGAAGGATTGTTG
tal-DSM   CCAGCGCAAGCATTTCATGAAGCCATGACGCGGTTTGAGATGAGCAGGGAAGGATTGTTG
tal-C     CCAGCGCAAGCATTTGATGAAGCGATGAGGCAATTCGGGATGAGCAGGCACGGGTTGTTA
tal-A     CCAGCGCAAGCATTTGATGAAGCGATGAGGCAATTCGGGATGAGCAGGCACGGGTTGTTA
tal-F     CCAGCGCAAGCATTTGATGAAGCGATGAGGCAATTCGGGATGAGCAGGCACGGGTTGTTA
tal-E     CCAGCGCAAGCATTTGATGAAGCGATGAGGCAATTCGGGATGAGCAGGCACGGGTTGTTA
tal-H     CCAGCGCAAGCATTTGATGAAGCGATGAGGCAATTCGGGATGAGCAGGCACGGGTTGTTA
tal-B     CGAGGGCAAGTATTTCATGAAGCGATGAAGCGGTTTGAGATGAGCAGGGAAGGATTGTTG
tal-G     CGAGGGCAAGTATTTCATGAAGCGATGAAGCGGTTTGAGATGAGCAGGGAAGGATTGTTG
          * ** ***** **** ******* **** **  ** * ******  *     * *****

tal1      CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCCGCAGTGGAACGCTCCCCCCA
tal2      CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCACACAGTGGAACGCTTCCCCCA
tal-D     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTTCCCCCA
tal-DSM   CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTTCCCCCA
tal-C     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
tal-A     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
tal-F     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
tal-E     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
tal-H     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
tal-B     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
tal-G     CAGCTATTTCGCCGTGTGGGCGTCACAGAACTCGAAGCCATCAGTGGAACGCTCCCCCCA
          ************************************   ************ ******

tal1      GCCTCCCAGCGTTGGCACCGCATCCTTCAAGCATTAGGGATGAAGGGGGCGAAACCATCC
tal2      GCCTCCCAGCGTTGGCACCGCATCCTGCAAGCATCAGGGATGAAAAGGGGGAAACCGTCC
tal-D     GCCTCCCAGCGTTGGCACCGCATCCTTCAAGCATTAGGGGTGAAAGGGGCGAAGCCGTCC
tal-DSM   GCCTCCCAGCGTTGGCACCGCATCCTTCAAGCATTAGGGGTGAAAGGGGCGAAGCCGTCC
tal-C     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGGGGCGAAACCGCCC
tal-A     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGGGGCGAAACCGCCC
tal-F     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGCGGCGAAACCGCCC
tal-E     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGGGGCGAAACCGTCC
tal-H     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGGGGCGAAACCGTCC
tal-B     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGGGGCGAAACCGCCC
tal-G     GCCTCCCAGCGTTGGGACCGCATGCTTCAAGCATCAGGGAGGAAAGGGGCGAAACCGCCC
          *************** ******* ** ******* ****  ***   ** *** **  **
```

```
tal1       TCTGCTTCAGCTCAAACTCCTAGCCAGGAGTCCTTGCATGCATTCGCCGATTCCCTGAAG
tal2       TCTGCTTCGGCTCAAACTCCGGGCCAGGAGTCCTTGCATGCATTCGCCGATTCCCTTGAA
tal-D      TCTGCTTCGGCTCAAACTCCGAGCCAGGAGTCCGTGGATGCATTCGCCGATTCCCTGGAG
tal-DSM    TCTGCTTCGGCTCAAACTCCGAGCCAGGAGTCCGTGGATGCATTCGCCGATTCCCTGGAG
tal-C      TCTGCTTCGGCTCAAACTCAGGGCCAGGAGTCCTTGGATGCATTCGCCGATTCCCTGGAA
tal-A      TCTGCTTCGGCTCAAACTCAGGGCCAGGAGTCCTTGGATGCATTCGCCGATTCCCTGGAA
tal-F      TCTGCTTCGGCTCAAACTCAGGGCCAGGAGTCCTTGGATGCATTCGCCGATTCCCTGGAA
tal-E      TCTGCTTCGGCTCAAACTCCAAGCCAGGAGTCCGTGGATGCATTCGCCGATTCCCTGGAA
tal-H      TCTGCTTCGGCTCAAACTCCAAGCCAGGAGTCCGTGGATGCATTCGCCGATTCCCTGGAA
tal-B      TCTGCTTCGGCTCAAACTCAGGGCCAGGAGTCCTTGGATGCATTCGCCGATTCCCTGGAA
tal-G      TCTGCTTCGGCTCAAACTCAGGGCCAGGAGTCCTTGGATGCATTCGCCGATTCCCTGGAA
           ******** **********   *********** ** ******************  *

tal1       CGTGAACTGGACGCGCCCAACCCAATTCACGAGGCAGACCGGGCTCGGGCAAGCAA---C
tal2       CGTGAACTGGAGGCGTCCAGCCCAATGCGCCAGACAGGCCAGACTCTGGCAAGCAGCTCC
tal-D      CGTGAACTGGACGCGCCCAGCCCAATTCACGAGGCAGACCGGGCTCGGGCAAGCAA---C
tal-DSM    CGTGAACTGGACGCGCCCAGCCCAATTCACGAGGCAGACCGGGCTCGGGCAAGCAA---C
tal-C      CGTGAACTGGACGCGCCCAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAG---C
tal-A      CGTGAACTGGACGCGCCCAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAG---C
tal-F      CGTGAACTGGACGCGCCCAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAG---C
tal-E      CGTGAACTGGACGCGCCCAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAG---C
tal-H      CGTGAACTGGACGCGCACAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAG---C
tal-B      CGTGAACTGGACGCGCCCAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAA---C
tal-G      CGTGAACTGGACGCGCCCAGCCCAATGCACCAGGCAGGCCAGACTCTGGCAAGCAG---C
           *********** ***  ** ****** * * ** *** ** * *** ********    *

tal1       CGCAAACGGTCCCGATCGGATAGTGTTGTCAACCACTCCTCTGCACAGCAGGCTGCCGAG
tal2       CGCAAACGGTCCCGATCGGATAGTGTTGTCAACCGCTCCTCTGCACAGCAGGCTGCCGAG
tal-D      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCTGCACAGCAGGCTGCCGAG
tal-DSM    CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCTGCACAGCAGGCTGCCGAG
tal-C      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCCGCACAGCAGGCTGCCGAG
tal-A      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCCGCACAGCAGGCTGCCGAG
tal-F      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCCGCACAGCAGGCTGCCGAG
tal-E      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCCGCACAGCAGGCTGCCGAG
tal-H      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCCGCACAGCAGGCTGCCGAG
tal-B      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCTGCACAGCAGGCTGCCGAG
tal-G      CGCAAACGGTCCCGATCAGAGAGTTCTGTCAACCGCTCCTCCGCACAGCAGGCTGCCGAG
           ***************** ** ***  ******** ****** ******************

tal1       GTGTTCGTTCCCGAACAGCGCGATGCAGCGCGTCTGCTCCCCCTCAGTAGCTGGGGTGTA
tal2       GTGTTCGTTCCCGAACAGCGCGATGCACCGCGTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-D      GTATTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGCGTA
tal-DSM    GTATTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGCGTA
tal-C      GTATTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-A      GTGTTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-F      GTGTTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-E      GTGTTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-H      GTGTTCGTTCCCGAACAGCGCGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-B      GTGTTCGTTCCCGAACAGCGGGATGCACCGCCTTTGCTCCCCCTCAGTAGCTGGGGTGTA
tal-G      GTGTTCGTTCCCGAACAGCGCGATGCACCGCTTTTGCTCCCCCTCAGTAGCTGGGGCGTA
           ** ***************** ****** *** * ********************** ***

tal1       AAACGCCAGCGTACCAGGATCGGGGGCGGCCTCCCGGATCCTGGTACGCCCACGAACGCC
tal2       AAACGCCAGCGTACGAGGATCGGGGGCGGCCTCCCGGATCCTGGTACGCCGACGGACGGC
tal-D      AAACGCCAGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCATGCCGACGGACGGC
tal-DSM    AAACGCCAGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCATGCCGACGGACGGC
tal-C      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCACGCCGACGCACGGC
tal-A      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCACGCCGACGCACGGC
tal-F      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCACGCCGACGCACGGC
tal-E      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCACGCCGACGCACGGC
tal-H      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCACGCCGACGCACGGC
tal-B      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATCCTGGCACGCCGACGCACGGC
tal-G      AAACGCCGGCGTACCAGGATC---GGCGGCCTCCCGGATACTGGCATGCCGACGGACGGC
           ******* ****** ******   *************** **** * *** *** *** *
```

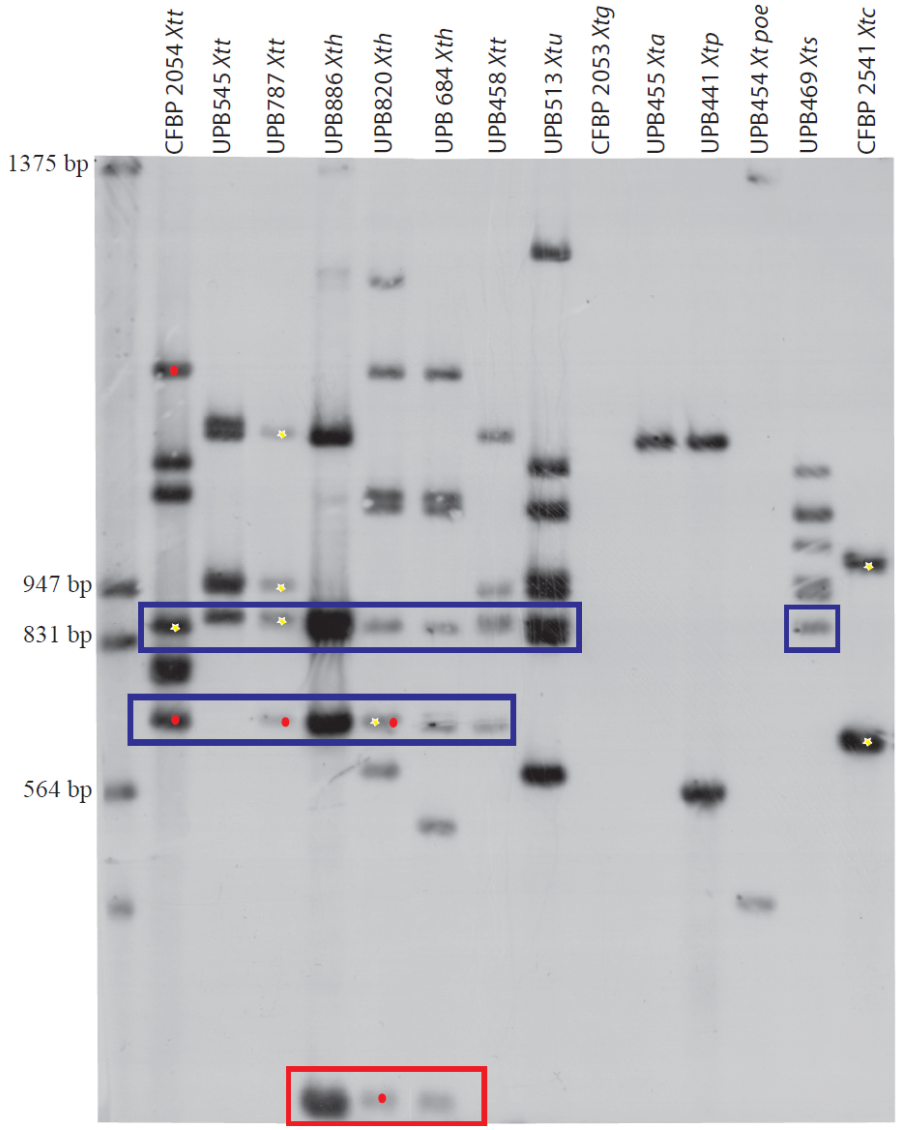
```
tal1      GACCTGGCAGCTTCCAGCGCTGCGTTCTTGGAACAAGCTGCAGCCTTCTTCGGAGGCGCA
tal2      GAACTGGCAGCGTCCAGCGCGGCGCTCTGGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-D     GAACTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGATCCCTTCGCAGGCGCA
tal-DSM   GAACTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGATCCCTTCGCAGGCGCA
tal-C     GACCTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-A     GACCTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-F     GACCTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-E     GACCTGGCAGCGTCCAGCGGTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-H     GACCTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-B     GACCTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
tal-G     GAACTGGCAGCGTCCAGCGCTGCGTTCTTGGAGCAAGATGCGGACCCCTTCGCAGGCGCA
          ** ******** *******  *** *** *** **** *** *    ***** *******

tal1      GGGGAAGATCCCCTGAGATTCGAAGAAGAGGAAATCGCATGGTTGATGACGCTATTGCCT
tal2      GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGATGGAGCTATTGCCT
tal-D     GCGGAAGATTTCCCGGTATTCGACCAAGAGGAGATCGCATGGTTGATGACGCTATTGCCT
tal-DSM   GCGGAAGATTTCCCGGTATTCGACCAAGAGGAGATCGCATGGTTGATGACGCTATTGCCT
tal-C     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAAGGAGCTATTGGCT
tal-A     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAAGGAGCTATTGGCT
tal-F     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAAGGAGCTATTGGCT
tal-E     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAAGGAGCTATTGGCT
tal-H     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAAGGAGCTATTGGCT
tal-B     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAAGGAGCTATTGGCT
tal-G     GCGGAAGATTTCCCGGCATTCGACCAAGAGGAGATCGCATGGTTGAGGGAGCTATTGGCT
          * *******  ** *  ******  ******* ************* *  ******* **

tal1      CATTGA
tal2      CATTGA
tal-D     CATTGA
tal-DSM   CATTGA
tal-C     CATTGA
tal-A     CATTGA
tal-F     CATTGA
tal-E     CATTGA
tal-H     CATTGA
tal-B     CATTGA
tal-G     CATTGA
          ******
```

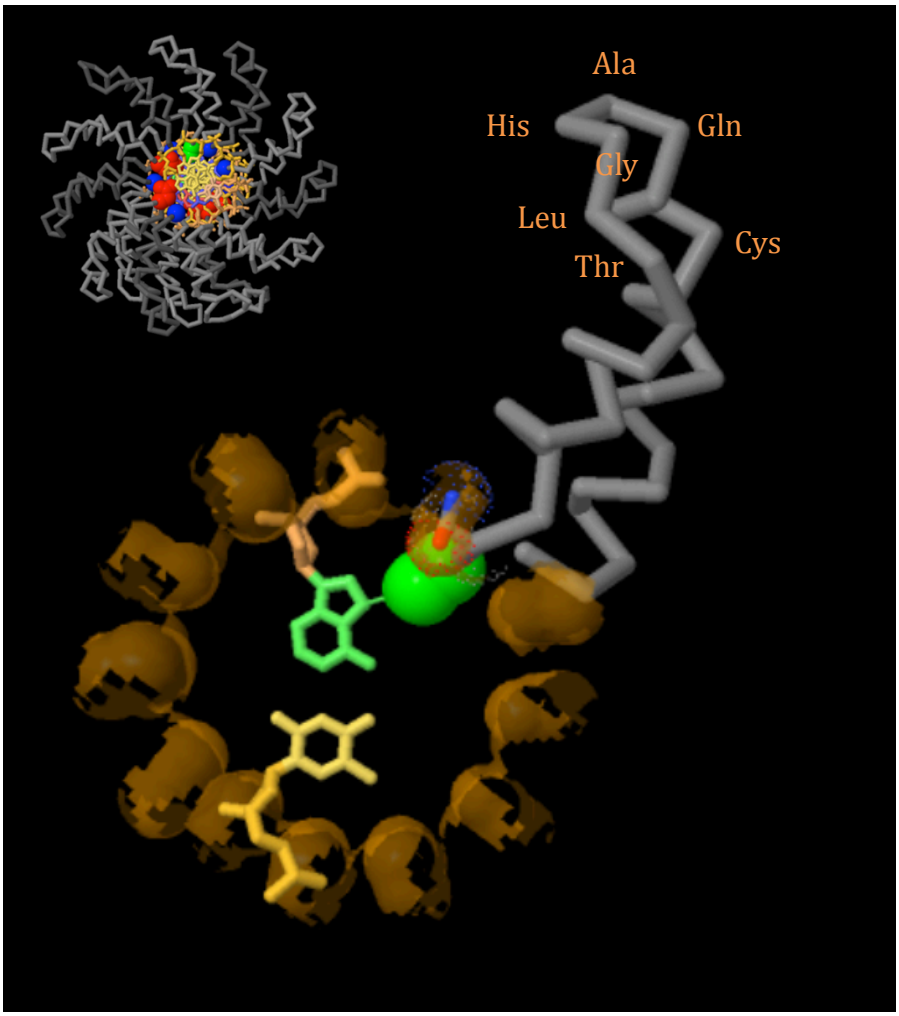
<u>Supplemental data VI.S4</u> Repertoire of *tale* genes in several strains of *X. translucens*.

*Sph*I-digested genomic DNA was hybridized with a mixed probe corresponding to the 3'-terminal regions of the two *tale* genes from *X. translucens* pv. *cerealis* strain CFBP 2541. The following pathovars were used: *Xta*, pv. *arrhenatheri*; *Xtc*, pv. *cerealis*; *Xtg*, pv. *graminis*; *Xth*, pv. *hordei*, *Xts*, pv. *secalis*; *Xtt*, pv. *translucens*; *Xtu*, pv. *undulosa, Xtp,* pv. *arrhenatheri*. Blue squares indicate conserved bands of the barley pathovars *translucens* and *hordei*. The red square indicates a conserved band in the four *X. translucens* pv. *hordei* strains. White stars and red dots correspond to sequenced and mutated *tale* genes, respectively.



Supplemental data VI.S5 Three-dimensional representation of a TALE repeat.

Van der Waals surfaces of the DNA backbone are in beige color. The contacted adenine base and the corresponding serine of the second RVD residue (repeat position 13) are in green color. The inset in the upper left corner shows the repeat array wrapped around the DNA molecule, seen along the axis of the DNA molecule. The same perspective is used for the larger view. The visualization is based on PDB entry 3V6T (crystal structure of the DNA-bound dHax3 TAL effector at 1.85 Å resolution) and was generated with the JSmol javascript applet (http://sourceforge.net/projects/jsmol/). Amino acids abbreviations correspond to threonine (Thr), leucine (Leu), glycine (Gly), histidine (His), alanine (Ala), glutamine (Gln) and cysteine (Cys).



Chapter VII

Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*

Previous work had demonstrated that TAL effector-mediated activation of certain *SWEET* genes promotes bacterial leaf blight, caused by *Xanthomonas oryzae* pv. *oryzae* (Antony et al., 2010; Chen et al., 2012; Chu et al., 2006, Yang et al., 2006, Yu et al., 2013). But it was unclear to which extent other members of the SWEET protein family could also support bacterial virulence. In order to test the different *SWEET* paralogs from rice for their potential to act as susceptibility genes for *X. oryzae* pv. *oryzae*, we established an experimental system that allows to elucidate whether activation of individual plant genes can contribute to virulence (or mediate resistance). This system relies on the construction of artificial *TAL* genes, which are then introduced into *Xanthomonas* strains. This system, which we developed in the well-characterized *X. oryzae* / rice pathosystem, is expected to find wide application in the field of bacterial phytopathology. Once candidate target genes for TAL effectors are identified by transcriptomic (microarrays, RNAseq) and/or bioinformatics analyses, they can be easily tested for their contribution to certain host responses, i.e. susceptibility or resistance. This system will therefore be useful to identify major host plant genes controlling the proliferation *X. translucens* in barley plants.

This work was published.

Streubel J*, Pesce C*, Hutin M, Koebnik R, Boch J, Szurek B. 2013. Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. *New Phytol.* 200 (3): 808–819. doi: 10.1111/nph.12411.

*These authors contributed equally to this work.

Five phylogenetically close rice *SWEET* genes confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*

Streubel, J.[1]*, Pesce, C. [2,3]*, Hutin, M.[2], Koebnik, R.[2], Boch, J.[1] and Szurek, B[2].

*These authors contributes equally to this work.

[1]Department of Genetics, Martin Luther University Halle-Wittenberg, Weinbergweg 10, D-06120, Halle (Saale), Germany

[2]UMR 186 IRD-UM2-Cirad "Résistance des Plantes aux Bioagresseurs", BP 64501, 34394 Montpellier cedex 5, France

[3]Applied Microbiology Phytopathology, Earth and Life Institute, Université Catholique de Louvain, Louvain-la Neuve, Belgium

My contribution to this work consisted of the construction of plasmids encoding artificial TALEs, construction of *Xanthomonas* strains expressing artificial TALEs, GUS assays, plants inoculations, *in planta* growth assays and RNA isolation for qRT-PCR.

Abstract

- Bacterial plant-pathogenic *Xanthomonas* strains translocate transcription activator-like (TAL) effectors into plant cells to function as specific transcription factors. Only a few plant target genes of TAL effectors have been identified, so far. Three plant *SWEET* genes encoding putative sugar transporters are known to be induced by TAL effectors from rice-pathogenic *Xanthomonas oryzae* pv. *oryzae* (*Xoo*).
- We predict and validate that expression of *OsSWEET14* is induced by a novel TAL effector, Tal5, from an African *Xoo* strain. Artificial TAL effectors (ArtTALs) were constructed to individually target 20 *SWEET* orthologs in rice. They were used as designer virulence factors to study which rice *SWEET* genes can support *Xoo* virulence.
- The Tal5 target box differs from those of the already known TAL effectors TalC, AvrXa7 and PthXo3, which also induce expression of *OsSWEET14* suggesting evolutionary convergence on key targets. ArtTALs efficiently complemented an *Xoo talC* mutant demonstrating that specific induction of *OsSWEET14* is the key target of TalC. ArtTALs that specifically target individual members of the rice *SWEET* family revealed three known and two novel *SWEET* genes to support bacterial virulence.
- Our results demonstrate that five phylogenetically close SWEET proteins, which presumably act as sucrose transporters, can support *Xoo* virulence.

Introduction

Plant pathogens threaten the production of food crops worldwide. The bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) causes Bacterial leaf blight (BLB), one of the most devastating diseases of rice (Nino-Liu et al., 2006), which is a primary food plant for half of the world's population. Virulence of *Xoo* relies on injection of a cocktail of effector proteins into plant cells via a specialized type III secretion system (White & Yang, 2009). TAL (transcription activator-like) effectors constitute a major *Xoo* effector family as many as 19 genes per strain (Salzberg et al., 2008). TAL effectors employ tandem, nearly identical 34-amino-acid repeats to directly bind specific DNA sequences in promoter regions. Guided by a repeat variable diresidue (RVD) specific for one of the four DNA bases, or combinations thereof, each repeat binds one base pair in a contiguous sequence (Boch et al., 2009; Moscou & Bogdanove, 2009; Boch & Bonas, 2010). The DNA-binding domain, together with a C-terminal activation domain, enables TAL effectors to function as specific transcription factors that induce expression of target plant genes to the benefit of the pathogen (Boch & Bonas, 2010). The modular TAL effector architecture also allows one to artificially rearrange repeats to generate artificial TAL effectors with novel, predictable DNA specificities (Boch et al., 2009). Artificial TAL effectors (also called designer TAL effectors) have been used to analyze the specificities and efficiencies of RVDs (Boch *et al*., 2009; Christian *et al*., 2012; Cong *et al*., 2012; Streubel *et al.*, 2012) and induce target gene expression in different organisms (Morbitzer et al., 2010; Geißler et al., 2011; Miller et al., 2011; Zhang et al., 2011; Bultmann et al.,

2012; Li et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013). Consequently, TAL effectors have become widely used in biotechnology as specific activators, repressors, and nucleases in many different organisms (Bogdanove & Voytas, 2011; Mussolino & Cathomen, 2012).

Several TAL effectors contribute significantly to bacterial virulence (Boch & Bonas, 2010), but in general little is known about their biological virulence targets. Because virulence of most *Xoo* strains mainly relied on the presence of individual TAL effectors and thus the ability to induce specific host genes, these targets have been termed susceptibility genes (White & Yang, 2009). The rice *SWEET*/*nodulin-3* gene family holds the best-studied examples of TAL effector virulence targets so far. Two *SWEET* genes, *Os8N3* and *Os11N3* (also called *OsSWEET11* and *OsSWEET14*, respectively), have been identified as targets of four different TAL effectors from various *Xoo* strains (Chu et al., 2006a; Yang et al., 2006; Antony et al., 2010; Chen et al., 2012). A third *SWEET* gene, *Xa25*, was also shown to be induced after infection with *Xoo* strain PXO339 (Liu et al., 2011), and it is possible that *Xa25* is induced by an hitherto not identified TAL effector. The SWEET proteins form a heterogeneous family, which is divided into at least four clades in plants (Chen et al., 2010). SWEETs are membrane proteins that are involved in diverse functions, such as pollen development (Guan et al., 2008), senescence (Quirino et al., 1999), and microbe-plant interactions (Gamas et al., 1996; Chu et al., 2006a,b; Yang et al., 2006; Antony et al., 2010; Liu et al., 2011). *Arabidopsis* and rice genomes include 17 and 22 paralogs, respectively. By contrast, chordates and arthropods contain

only one and *Caenorhabditis elegans* seven *SWEET* genes, respectively. This suggests that *SWEETs* may have undergone higher diversification in plants to potentially accompany the evolution of the vascular system (Baker et al., 2012). Recent data demonstrated that some members of the SWEET family facilitate sucrose and glucose efflux from the phloem parenchyma cells to the apoplast (Chen et al., 2010, 2012). This gave rise to the hypothesis that the pathogen-mediated induction of certain *SWEET* genes might increase sugar availability in the apoplast, and therefore plays a role in *Xoo* nutrition. Members of the *SWEET* gene family in *Arabidopsis* are also induced after infection with *Pseudomonas syringae* strain DC3000 (Chen et al., 2010), which indicates that this plant gene family has a general function for pathogen proliferation in the host.

Here we describe that a novel TAL effector, Tal5 from *Xoo*, also targets the rice *OsSWEET14* gene. This observation provides compelling evidence that evolution of TAL effectors can converge on the same or highly related virulence targets. The importance of SWEETs for susceptibility towards *Xoo* prompted us to systematically study the role of this gene family for *Xoo* virulence. We constructed a series of artificial TAL effectors, which were designed to induce expression of individual *SWEET* genes in rice, and analyzed the ability of these TAL effectors to function as artificial virulence factors for *Xoo*. Our findings demonstrated that only the particular subgroup of rice SWEETs that are predicted to be sucrose transporters can be exploited as *Xoo* virulence targets. They also indicated that diverse SWEET family members probably fulfill different physiological functions. Finally, our approach demonstrates that artificial TAL

effectors are suitable tools to study if candidate host genes are bona fide virulence targets for *Xanthomonas*.

Materials and Methods

Bacterial strains and growth conditions

The bacterial strains used in this study were *Escherichia coli* DH5α (Stratagene, La Jolla, CA), *Agrobacterium tumefaciens* GV3101 (Van Larebeke et al., 1974) and *Xoo* strains BAI3 (Gonzalez et al., 2007), BAI3Δ*talC* (Yu et al., 2011) and PXO99^{A} (Hopkins et al., 1992). *E. coli* cells were cultivated at 37°C in lysogenic broth (LB) medium, *Xoo* strains at 28°C in PSA medium (10 g peptone, 10 g sucrose, 1 g glutamic acid, 16 g agar, l^{-1} H_2O) medium and *A. tumefaciens* GV3101 at 28°C in YEB medium. Plasmids were introduced into *E. coli* and *A. tumefaciens* strains by electroporation and into *Xoo* by conjugation using pRK2013 as a helper plasmid in triparental mating (Figurski & Helinski, 1979). Rifampicin- and gentamicin- resistant clones were selected upon plating on PSA medium and one isolate was chosen for further experiments. Antibiotics were added to the medium at the following final concentrations: rifampicin, 100 μg ml^{-1}; gentamycin, 20 μg ml^{-1}.

Plant material and plant inoculations

Experiments were performed under glasshouse conditions under cycles of 12h of light at 28°C and 80% relative humidity (HR) and 12h of dark at 25°C and 70% RH. *Oryza sativa* ssp. *japonica* cv. Nipponbare and the near-isogenic *O. sativa* ssp. *indica* lines IR24 and

IRBB13 were used for virulence assays, growth curves and quantitative reverse polymerase chain reaction (qRT-PCR). Leaves of 3-wk-old plants were infiltrated with a bacterial suspension with an optical density at 600 nm (OD_{600}) of 0.5 using a needleless syringe, as previously described (Reimers & Leach, 1991), and symptoms (water-soaked lesions) were scored 6 d postinoculation (dpi). Leaf-clip inoculation was performed on 4- to 5-wk-old rice plants using a bacterial suspension with an OD_{600} of 0.2 (Kauffman et al., 1973), and sizes of lesions were measured 15 dpi. Statistical significance of the results was assessed using the Tukey honest significant difference test for post–ANOVA pairwise comparisons, set at 5% ($P < 0.05$).

Nicotiana benthamiana plants were grown under 16 h of light, 40%-60% RH, at 23 : 19°C, day : night in the growth chamber. Leaves of 4- to 6-wk-old plants were inoculated with *A. tumefaciens* strains using a needleless syringe.

β-Glucuronidase (GUS) reporter constructs and GUS assays

β-Glucuronidase assays from plant samples were essentially performed as described by Boch *et al.* (2009). Briefly, coding regions of *talC* and *tal5* were inserted into pGWB2 under control of a *35S* promoter (Nakagawa et al., 2007). A PCR-amplified 341 bp fragment containing the *OsSWEET14* promoter region was cloned into pENTR/D-TOPO (Life Technologies GmbH, Darmstadt, Germany) and then fused to the *uidA* reporter gene by LR recombination into pGWB3 (Nakagawa et al., 2007; Yu et al., 2011).

To assay reporter activity, *A. tumefaciens* strains delivering TAL effector constructs and GUS reporter constructs were resuspended in

infiltration medium (0.1 M 2-[N-morpholino] ethanesulfonic acid [MES], 1 M $MgCl_2$, 0.1 M acetosyringone), mixed in equal amounts and inoculated into *N. benthamiana* leaves at a total OD_{600} of 0.8. Two days postinfiltration, leaf discs were sampled and GUS activities were quantified using 4-methyl-umbelliferyl-β-D-glucuronide (MUG). Protein concentrations were determined using Bradford assays. Data were compiled from triplicate samples originating from different plants.

RNA isolation and qRT- PCR

Leaves of 3-wk-old rice plants of the varieties Nipponbare, IR24 or IRBB13 were infiltrated with water or with the different *Xoo* strains using an OD_{600} of 0.5. At 1 dpi, leaf segments from three plants were ground into a fine powder using the Qiagen TissueLyser system (30 rps for 30s). Rice total RNA was isolated using the Qiagen RNeasy kit following the manufacturer's recommendations (QIAGEN). cDNA was generated from 2µg RNA using the Fermentas first-strand cDNA synthesis kit following the manufacturer's recommendations (Thermo Fisher Scientific Inc., Waltham, MA, USA). Real-time PCR was performed using the iCycler (Bio-Rad, München, Germany) and 10 µl (50%) ABsolute qPCR SYBR Green Fluorescein Mix (ABgene Limited, Hamburg, Germany), 1 pmol forward, 1 pmol reverse oligonucleotide primer and 12.5 ng template cDNA. For each gene, a minimum of two independent biological replicates were analyzed, each with two technical qPCR replicates. The specificity of the primer pairs was checked by a melting curve and the PCR product was documented by agarose gel electrophoresis. The amplification

efficiency for each primer pair was analyzed using a standard curve plot of a dilution series. cDNA amounts were normalized using actin as a reference gene. All RT primer sequences and corresponding amplification efficiencies are provided in Supporting data VII.S3.

In planta growth assays

Leaves of 3-wk-old IR24 or IRBB13 plants were inoculated with a bacterial suspension of *Xoo* strains using an OD_{600} of 0.2. A 1 cm^2 infected leaf segment was collected 6 d after infiltration, and ground into a fine powder using the Qiagen TissueLyser system (30 rps for 30 s). Ground material was resuspended in 1 ml sterile water, and 5 µl drops of a dilution series were spotted as triplicates onto selective PSA plates containing rifampicin and gentamycin. This experiment was performed three times.

Expression analysis by Western blotting

The *X. oryzae* pv. *oryzae* BAI3D*talC* strains carrying a plasmid encoding an artificial or native TAL effector, or empty vector, were grown in liquid PSA medium supplemented with 20 µg ml^{-1} gentamicin at 30°C. Cells of 1 ml of a bacterial suspension at an OD600 of 0.2 were harvested and the TAL effector expression was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting using an anti-Flag antibody (Sigma-Aldrich).

Results

Tal5 from *Xoo* strain MAI1 is a novel TAL effector that targets *OsSWEET14*

We analyzed the TAL effector repertoire of an African *Xoo* isolate from Mali, MAI1, which harbors a relatively small set of only eight *TAL* genes (Gonzalez et al., 2007). We screened a MAI1 genomic DNA cosmid library for clones containing *TALs* and subcloned eight *Bam*HI fragments of different sizes containing putative *TAL* genes. Sequencing revealed that one of them, named *tal5*, encodes a new TAL effector with a unique order of 17.5 DNA-binding repeats (Fig.VII.1a).

Using a program to scan rice promoters for potential TAL effector binding sites (Pérez-Quintero *et al*., 2013), we predicted which candidate rice genes targeted by Tal5 (Suppl. data VII.S4). Interestingly, the top score was obtained for a binding site located within the promoter region of the gene *Os11g31190* (hereafter *OsSWEET14*), which is a member of the *SWEET*/*nodulin-3* gene family (Baker et al., 2012). Remarkably, *OsSWEET14* is a host susceptibility gene that was previously reported as a target of the major virulence TAL effectors AvrXa7, PthXo3 and TalC from *Xoo* strains PXO86, JXO1^{A} and BAI3, respectively (Chu et al., 2006a; Antony et al., 2010; Yu et al., 2011). The target boxes for TalC and AvrXa7/PthXo3 are upstream of and overlapping the predicted *OsSWEET14* TATA box, respectively. By contrast, the predicted Tal5 box is located downstream of the TATA box (Fig. VII.1b).

(a)

NN HD NN HD NG HD HD NG HD NI NI NN HD HD N* NG HD NI

(b)

>pOsSWEET14

TGTGCAGCTATATTGCCTATTGGTGTCCAGGGTCACACACCATAAGGGCATGCATGTCAGCAGCTGGT

CATCATGTGTGCCTTTTCATTCCCTTCTTCCTTCCTAGCACTATATAAACCCCCTCCAACCAGGTGCT

AAGCTCATCAAGCCTTCAAGCAAAGCAAACTCAAGTAGTAGCTGATTACCAGCTCTTCTCTCTTCTCA

TTGAGAAGAGGGAATTAAGTTTTGATCTCTGCTTTATTGCCTGATCATCCTCTTGTTACTTGCAAGCA

AGAACAGTAGTGTACTGTGCCTCATTGATCTCCTCCCACCAAACTCTCTCTCTCTCTCTCATATTCCG

AGCTAGCTAGTTAATCAAGATCTTGCTGCA***ATG***

Figure VII.1. Tal5 is a novel transcription activator-like (TAL) effector predicted to target *OsSWEET14*.

(a) Hypervariable amino acids at positions 12 and 13 (repeat variable diresidue) of the 17.5 repeats of Tal5. A single amino acid code was used for each amino residue (* represents a missing 13th residue). (b) The 370 bp sequence of the *OsSWEET14* promoter region is depicted by the TalC box in bold (Yu et al*.,* 2011), the AvrXa7 box underlined (Römer et al*.,* 2010) and the candidate Tal5 box highlighted in gray. The artificial TAL effector (ArtTAL) boxes for ArtTAL14-1 and ArtTAL14-2 are indicated by dotted and dashed frames, respectively. The 5′ terminal nucleotide of each target box is indicated in white font. The predicted start codon of *OsSWEET14* is indicated in bold italic font.

To test if *OsSWEET14* is a direct virulence target of Tal5, we performed two complementary assays. First, we used a *talC* deletion mutant of *Xoo* BAI3, BAI3Δ*talC*, which is impaired in *OsSWEET14* induction, resulting in fewer symptoms and reduced proliferation of the bacteria *in planta* compared with the parental strain (Yu et al., 2011). Leaves of the susceptible rice variety IR24 were infiltrated with derivatives of BAI3Δ*talC* containing the empty cloning vector or plasmids that encode *talC* or *tal5*. Water and wildtype bacteria of strain BAI3 served as negative and positive controls, respectively. qRT-PCR analyses showed that *OsSWEET14* was induced by both TalC and Tal5, but not upon inoculation of rice leaves with BAI3Δ*talC* carrying the empty vector (Fig. VII.2b). Second, we used *Xoo* PXO99^{A}, which is not able to cause disease on *xa13* rice lines

(e.g. IRBB13) as a result of a deletion in the PthXo1 target box which prevents PthXo1-mediated induction of the susceptibility gene *OsSWEET11* (Yang et al., 2006). Similar to the previous experiment, only *Xoo* PXO99^{A} containing *tal5* or *talC* induced *OsSWEET14* expression in *xa13* rice lines (Fig. VII.2b).

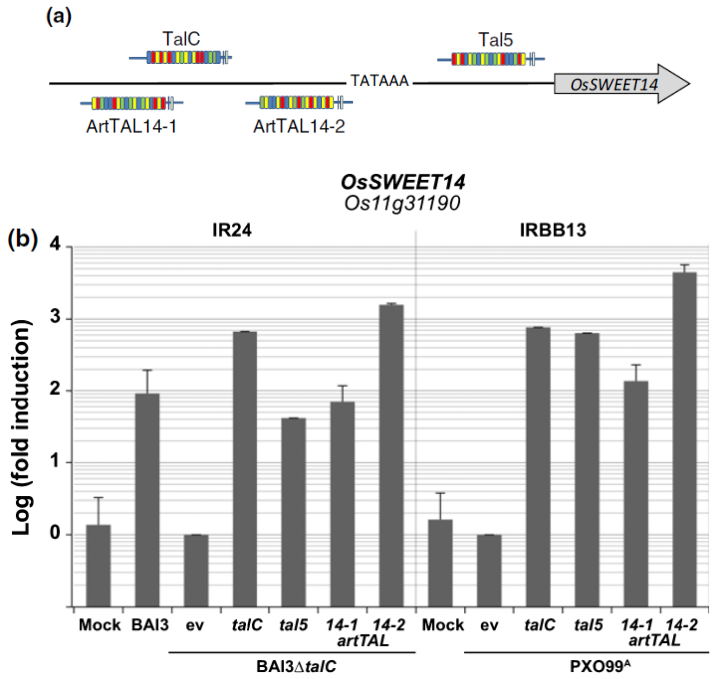


Figure VII.2 (a,b). *OsSWEET14* is induced by Tal5 and artificial transcription activator-like (ArtTAL) effectors.
(a) Schematic representation of the *OsSWEET14* promoter region and associated natural TAL and ArtTAL effectors next to their target boxes. The coding sequence is represented by an open arrow and the TATA box is indicated. (b) *OsSWEET14* expression levels. Leaves of *Oryza sativa* IR24 variety were infiltrated with water (Mock), *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strain BAI3, and *Xoo* BAI3Δ*talC* derivatives carrying an empty vector (ev) or plasmids containing the TAL effector genes *tal5*, *talC*, *artTAL14-1* or *artTAL14-2*. Rice IRBB13 leaves were infiltrated with water or with *Xoo* strain PXO99^{A} derivatives carrying the same constructs as used for BAI3Δ*talC*. *OsSWEET14* transcript abundances were determined 1 d postinoculation (dpi) by quantitative reverse transcription polymerase chain reaction (qRT-PCR). The bars represent a minimum of two biological samples taken from a collection of three leaves. The fold change was calculated in comparison to leaves treated with BAI3Δ*talC* (ev) or PXO99^{A} (ev). Error bars represent ± SD. Actin was used as a reference gene to normalize cDNA amounts.

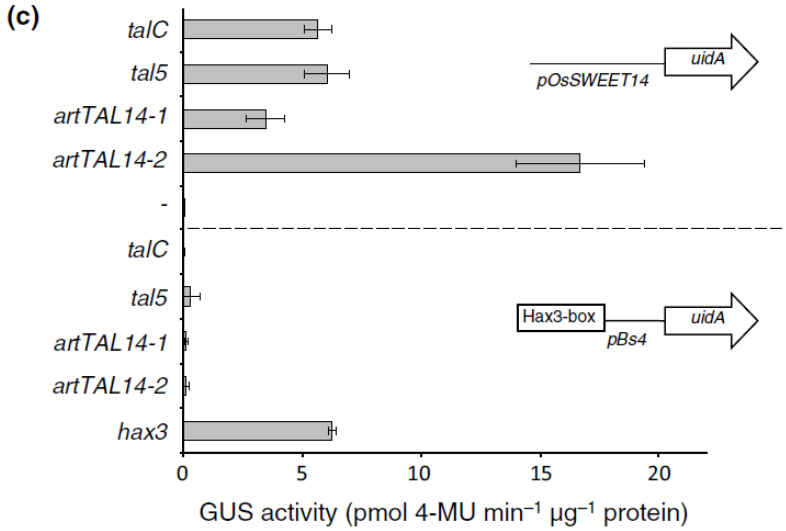


Figure VII.2 (c). *OsSWEET14* is induced by Tal5 and artificial transcription activator-like (ArtTAL) effectors.
(c) The *OsSWEET14* promoter is recognized by Tal5 and ArtTALs. A 341 bp fragment of the *OsSWEET14* promoter region was cloned into a GUS reporter vector and codelivered via *Agrobacterium tumefaciens* into *Nicotiana benthamiana* leaves, along with *35S*-driven *talC*, *tal5*, *artTAL14-1*, *artTAL14-2* or *hax3* (error bars indicate ± SD; $n = 3$). MU, 4-methyl-umbelliferone. The Hax3-box cloned in front of the minimal *pBs4* promoter served as the specificity control.

To analyze the functionality of the putative target box for Tal5, a 341 bp fragment containing the *OsSWEET14* promoter region was cloned in front of a promoterless *uidA* reporter gene (Yu et al., 2011). The reporter construct and *tal5* under control of a constitutive promoter were cotransformed into *N. benthamiana* using *Agrobacterium*. GUS assays showed that the *OsSWEET14* promoter fragment was recognized by Tal5, thus leading to reporter gene activity (Fig. VII.2c).

Finally, to analyze whether or not Tal5-mediated *OsSWEET14* induction affects infection, IR24 and IRBB13 rice lines were challenged with BAI3Δ*talC* and PXO99^{A} derivatives, respectively, and disease symptoms were scored upon leaf infiltration and leaf

clipping. As shown in Figure VII.3a, *tal5* partially complemented the BAI3Δ*talC* mutant strain. Similarly, introduction of *tal5* into PXO99^{A} resulted in increased aggressiveness in IRBB13 lines (Fig. VII.3b). Altogether, our data indicate that Tal5 can function as a virulence factor in *Xoo* by up-regulating *OsSWEET14*, and potentially facilitates propagation of *Xoo* strain MAI1 in rice leaves.

Artificial TAL effectors targeting *OsSWEET14* effectively substitute for the virulence contributions of TalC and PthXo1

Although TAL effectors have a high target site specificity, they nevertheless induce the expression of a significant number of genes in host cells (Marois et al., 2002; Kay et al., 2009). While some of them are directly induced, others are up-regulated indirectly. Collectively, most of the (directly and indirectly) induced genes probably represent collateral targets that are not related to the infection process. Therefore, it is difficult to know which of these genes is responsible for the virulence effect of a given TAL effector and thus represents a bona fide host susceptibility gene.

We aimed to develop a functional test to determine if *OsSWEET14* is indeed a valid virulence target for *Xoo*. Artificial TAL effectors can be constructed with a designed repeat array and a tailored DNA-binding specificity to induce expression of specified target genes (Morbitzer et al., 2010; Bogdanove & Voytas, 2011). We constructed two artificial TAL effectors (ArtTAL14-1 and ArtTAL14-2) that target sequences in the *OsSWEET14* promoter and that are different from those that are recognized by the four natural TAL effectors. (Fig. VII.1b and Fig. VII.2a), for more information on the

design of the artificial TAL effectors, see suppl. methods VII.S1 and S2. The rationale was that both sets, ArtTALs and natural TAL effectors, will induce expression of *OsSWEET14*, but it is unlikely that they have common collateral targets. Indeed, the prediction of target sites in rice for the ArtTALs, Tal5, and TalC supported that they likely share no target genes besides *OsSWEET14* (suppl. data VII.S5). We numbered the artificial TAL effectors according to their target *OsSWEET* paralog (i.e. ArtTAL14 for *OsSWEET14*).

As envisaged, both ArtTALs induced expression of *OsSWEET14* when delivered by *Xoo* BAI3Δ*talC* and PXO99^{A}, respectively (Fig. VII.2b). In addition, both ArtTALs also recognized the *OsSWEET14* promoter in transient GUS reporter studies in *N. benthamiana* (Fig. VII.2c). We then tested whether the ArtTALs contribute to the virulence of *Xoo*, similar to the natural TAL effectors targeting *OsSWEET14*. To this end, *Xoo* BAI3Δ*talC* derivatives, complemented with plasmid-borne copies of *artTAL14-1* or *artTAL14-2,* were infiltrated into leaves of the rice line IR24. Strikingly, both strains caused typical disease symptoms (water-soaked lesions), basically identical to *Xoo* strains BAI3 and BAI3Δ*talC* containing *talC* or *tal5* (Fig. VII.3a). Quantitative leaf-clipping assays using these strains showed that both ArtTALs partially restored virulence of the *Xoo* BAI3Δ*talC* strain, reaching levels similar to those observed upon complementation with *talC*. Similarly, both *artTAL* genes conferred virulence to *Xoo* strain PXO99^{A} when assayed by leaf infiltration or leaf clipping of otherwise resistant *xa13* rice plants (Fig. VII.3b).

Next, we quantified bacterial populations upon leaf infection with *Xoo* strains BAI3 and BAI3Δ*talC* derivatives (Fig. VII.3c). At 6 dpi,

the sizes of bacterial populations of all BAI3Δ*talC* strains harboring natural or artificial TALs were significantly larger than those of BAI3Δ*talC* carrying the empty vector.

Together, these data show that the artificial TAL effectors were functionally equivalent to natural ones and further demonstrate that induction of *OsSWEET14* is sufficient to confer virulence to the *Xoo* strains tested. These data also emphasize that induction of *OsSWEET14* can fully compensate the lack of *OsSWEET11* expression in the PXO99^{A}-IRBB13 combination. Together with previous studies (Yang et al., 2006; Antony et al., 2010), these findings strongly suggest that these two SWEETs have a similar function in respect of *Xoo* virulence.

A series of artificial TAL effectors induce expression of individual *SWEET* genes and unmask new potential virulence targets of *Xoo*

To investigate whether additional *SWEET* genes may act as virulence targets of *Xoo*, we adopted the same strategy as used for *OsSWEET14* (Figs. VII.2 and VII.3) for the remaining 21 *OsSWEET* rice paralogs (Table VII.1). Generally, two ArtTALs (named according to the *OsSWEET* paralog) were designed for each rice *SWEET* gene, targeting specific sequences in the corresponding promoter regions, with the exception of *OsSWEET7d* and *OsSWEET7e*, for which no specific ArtTALs could be generated (Table VII.1). We predicted the top 20 possible target sequences for each ArtTAL and found that they target almost exclusively nonoverlapping sets of genes. Next, we evaluated the capacity of each of these ArtTALs to confer virulence to the *Xoo* strain BAI3Δ*talC*. Only ArtTALs targeting *OsSWEET11* to

OsSWEET15 led to the production of typical water-soaked lesions (Fig. VII.4a). Strikingly, no disease symptoms were observed upon infiltration of plants with *Xoo* strains producing ArtTALs targeting any of the other *SWEET* paralogs (Table VII.1), as exemplified for *OsSWEET1a*, *OsSWEET6b* and *OsSWEET7a* in Fig. VII.4a.

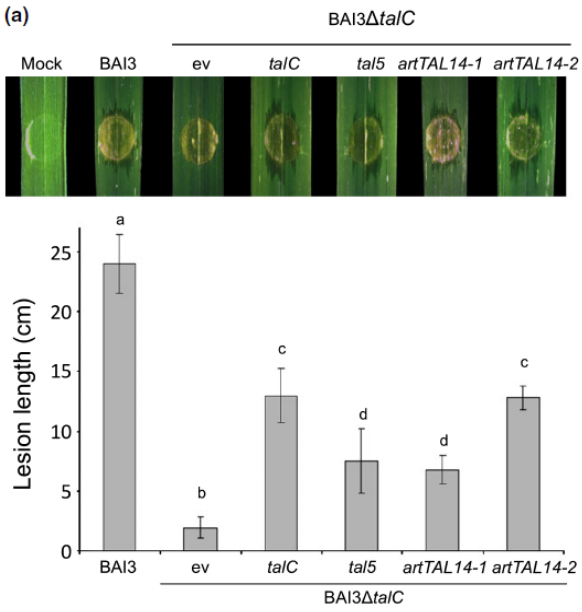
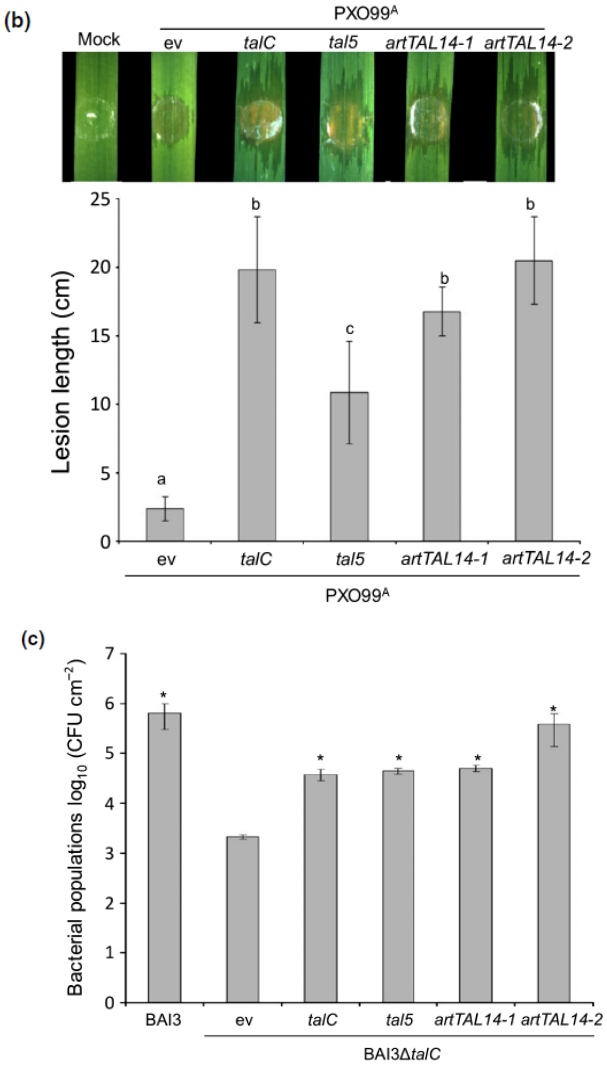


Figure VII.3 (a). Tal5 and artificial transcription activator-like (TAL) effectors act as virulence factors.
Phenotypes of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strains on *Oryza sativa* leaves. (a) Leaves of IR24 plants were inoculated with strains BAI3 or BAI3Δ*talC* carrying empty vector (ev) or TAL effectors *talC*, *tal5*, *artTAL14-1*, or *artTAL14-2*. Leaves were photographed 6 d postinoculation (dpi). Lesion length was measured at 15 d after leaf-clip inoculation. Data are the mean of eight measurements. Error bars represent ± SD. Bars with same letters are not statistically different based on a Tukey's honest significant difference test ($\alpha = 0.05$).



<u>Figure VII.3 (b,c).</u> Tal5 and artificial transcription activator-like (TAL) effectors act as virulence factors.

Phenotypes of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strains on *Oryza sativa* leaves. (b) Leaves of IRBB13 plants were inoculated with strain PXO99^{A} carrying empty vector (ev) or TAL effectors *talC*, *tal5*, *artTAL14-1*, or *artTAL14-2*. Leaves were photographed 6 d postinoculation (dpi). Lesion length was measured at 15 d after leaf-clip inoculation. Data are the mean of eight measurements. Error bars represent ± SD. Bars with same letters are not statistically different based on a Tukey's honest significant difference test ($\alpha = 0.05$). (c) Bacterial growth in IR24 rice leaves. Bacterial counts in leaf segments infiltrated with *Xoo* strains BAI3 or BAI3Δ*talC* carrying empty vector (ev) or TAL effectors *talC*, *tal5*, *artTAL14-1*, or *artTAL14-2* were analyzed 6 dpi. Error bars represent SD ($n = 3$). An asterisk indicates a significant difference in Student's *t*-test at $P < 0.01$ when comparing samples with BAI3Δ*talC* (ev).

To exclude the possibility that lack of symptom formation was the result of a failure of *SWEET* gene induction, we checked whether or not the ArtTAL proteins are stably produced in *Xoo* and the target genes in rice are induced. Indeed, Western blot analysis indicated that all ArtTALs accumulated at significant level and showed protein integrity in *Xoo* (suppl. data VII.S1). Quantitative RT-PCR experiments showed that most *SWEET* genes were induced in an ArtTAL-dependent manner, except for *OsSWEET2c*, *OsSWEET5* and *OsSWEET7c* (suppl. data VII.S2). *OsSWEET6a* and *OsSWEET7b* were not analyzed, because no specific amplification products could be detected although several pairs of primers have been tested.

Interestingly, all five rice *SWEET* paralogs whose activation was associated with the development of typical disease symptoms belong to clade III of the *SWEET* gene family, which comprises *OsSWEET11* (*Os8N3*/*Xa13*), *OsSWEET12*, *OsSWEET13* (*Xa25*), *OsSWEET14*, and *OsSWEET15*. *OsSWEET11* and *OsSWEET14* are known major *Xoo* susceptibility genes of rice (Yang et al., 2006; Antony et al., 2010), which confirms the suitability of our strategy. Furthermore, we evaluated the basal and induced *OsSWEET* gene expression levels (suppl. data VII.S6), which suggests that clade III *SWEETs* are generally not induced at higher levels than nonclade III ones. However, we cannot rule out the possibility that a few of them act as susceptibility genes when induced at higher levels and/or under different conditions.

Table VII.1 Artificial transcription activator-like (ArtTAL) effector repeat variable diresidue (RVD), target sequences and *in planta* activity

Gene ID[a]	Gene name[b]	ArtTAL[c]	RVD Sequence[d]	Target box[e]	Position from ATG[f]	Water soaking[g]	qRT-PCR[h]
Os01g65880	*OsSWEET1a*	ArtTAL1a-1	NG-NG-HD-NG-HD-NG-NI-HD-NI-HD-NI-HD-NI-NK-NG-NK-NG	TTTCTCTACACACCAGTGT	−221	–	+
		ArtTAL1a-2	HD-HD-HD-NG-NG-NI-NG-NI-NK-HD-NI-HD-NI-HD-NI	TCCCTTATATAAGCACACA	−184	–	+
Os05g35140	*OsSWEET1b*	ArtTAL1b-1	NI-NG-NI-NG-NG-NI-NI-HD-HD-NK-NK-HD-NI-NK-NG-NG	TATATATAACCGGGCAGTT	−260	–	+
		ArtTAL1b-2	HD-HD-NI-HD-NI-HD-HD-HD-NG-HD-NG-HD-HD-HD-NG-HD	TCCACACCCTCTCCCTC	−291	–	+
Os01g36070	*OsSWEET2a*	ArtTAL2a-1	HD-NI-HD-NK-NK-HD-NK-NI-HD-NG-NI-NG-HD-HD-HD-HD	TCACGGCGACATATCCCCC	−215	–	+
		ArtTAL2a-2	NI-NG-NI-NG-HD-NG-NG-NK-HD-NK-NK-NI-HD-NN-NK-NN-NI-HD	TATATCTTGCGGACGGGAC	−181	–	–
Os01g50460	*OsSWEET2b*	ArtTAL2b-1	HD-NG-HD-NG-HD-NI-HD-NG-NK-NI-HD-HD-NK-NI-HD-NG-HD-NK	TCTCTCACTGACCGACTCG	−110	–	+
		ArtTAL2b-2	NI-NG-NI-NI-NG-NI-NK-HD-HD-NI-HD-NK-NK-NI-NI-NG-NG-NG	TATAATAGCCCACGGAATTT	−81	–	+
Os01g40960	*OsSWEET2c*	ArtTAL2c	HD-NI-NG-NG-HD-NI-HD-NK-NG-NK-NK-HD-NK-NG-NG-NK-NI	TCATTCCACGTCGGCGTTGA	−91	–	–
Os05g12320	*OsSWEET3a*	ArtTAL3a-1	HD-HD-HD-NI-HD-NI-NG-NG-NI-NI-HD-NI-NK-HD-HD-NI-NG	TCCCACATTAAACAGCCAT	−117	–	+
		ArtTAL3a-2	NI-NI-NI-NG-NI-NK-HD-HD-NG-NK-HD-HD-NI-NG-NI-NG-NG	TAAATAGCCTGCCCATATT	−89	–	+
Os01g12130	*OsSWEET3b*	ArtTAL3b-1	NK-HD-NG-NG-HD-NI-NG-HD-HD-NI-HD-HD-NI-NG-NG-HD-NG	TGCTTCATCCAACCCATTCT	−143	–	+
		ArtTAL3b-2	NI-NG-NI-NI-NI-NG-NI-NK-NI-HD-NI-HD-HD-HD-NI-NI	TATAAATAGACAGACCCAC	−92	–	+
Os02g19820	*OsSWEET4*	ArtTAL4-1	NI-NG-NI-HD-NG-NI-HD-NG-NI-HD-NI-HD-NG-NG-NK-HD-NI-NK	TATACTACTACACTTGCAG	−155	–	+
		ArtTAL4-2	HD-NK-HD-NK-NG-HD-NG-HD-NK-NI-HD-NG-HD-HD-NI-HD-NG-NI	TCGCGTCTCGACTCCACTA	−185	–	+
Os05g51090	*OsSWEET5*	ArtTAL5	NN-NK-NI-NN-NN-NN-HD-NI-NN-NI-NN-NN-HD-HD-NK-HD-HD-NI	TGGAGGGCAGAGGCCGCCA	−283	–	–
Os01g42110	*OsSWEET6a*	ArtTAL6a-1	HD-NI-HD-NK-HD-HD-NG-NI-NG-NI-NI-HD-NI-NI-HD-HD-NI-HD	TCACGCCTATATAAACAAACCAC	−126	–	na
		ArtTAL6a-2	NI-HD-NG-HD-NK-NI-NG-HD-NK-NI-NG-HD-NK-HD-HD-NG-HD-HD	TACTCGATCGATCGCCTCC	−163	–	na
Os01g42090	*OsSWEET6b*	ArtTAL6b-1	NG-NG-HD-NI-HD-NG-NN-NG-NN-NG-NG-HD-NN-NG-NG-NN-NG-NG	TTTCACTGTGTTCCTTGTT	−123	–	+
		ArtTAL6b-2	NG-HD-NN-HD-NN-HD-NK-NG-HD-NG-NI-NG-NI-NI-NK-HD	TTCGCGCGTCTATATAAGC	−106	–	nd
Os09g08030	*OsSWEET7a*	ArtTAL7a-1	NI-HD-NN-HD-NN-NN-NG-NI-NG-NI-NG-HD-HD-NI-NN-NI	TACGCGGTATATATCCAGA	−175	–	+
		ArtTAL7a-2	NI-NG-NI-NG-NI-NG-HD-HD-NI-NK-NI-NK-HD-NG-HD-HD-NK-HD	TATATATCCAGAGCTCCGC	−168	–	–
Os09g08440	*OsSWEET7b*	ArtTAL7b	HD-NK-NK-NI-HD-NG-HD-NK-NK-NI-NI-NG-NK-NK-NG-NK-HD	TCGGACTCCGGAATGGTGGC	−239	–	na
Os12g07860	*OsSWEET7c*	ArtTAL7c	HD-NI-HD-HD-NK-NK-NK-NI-HD-NG-HD-HD-NK-NI-HD-NI-HD-HD	TCACCGGGACTCCGACACC	−222	–	–
Os08g42350	*OsSWEET11, Os8N3, Xa13*	ArtTAL11-1	HD-HD-NI-NI-HD-NG-NI-NG-NI-NI-HD-NI-HD-NG-NN	TCCAACTATATAAACACTG	−211	+/–	+
		ArtTAL11-2	NN-HD-NI-NN-NG-NG-NK-NG-NI-NK-HD-NI-HD-NG-NG	TGCAGTTGTAGTAGCACTT	−166	+	+
Os03g22590	*OsSWEET12*	ArtTAL12-1	NI-NI-NI-NG-NI-HD-HD-NI-NK-NK-NI-HD-NK-NI-NG-NK-HD	TAAATACCAGGGACGATGCC	−188	–	–
		ArtTAL12-2	NI-NG-NI-NI-NI-NG-NN-NG-NG-NN-NN-HD-NG-HD-NG-HD-HD	TATAAATGTTGGGCTCTCC	−220	+	+
Os12g29220	*OsSWEET13, Os12N3, Xa25*	ArtTAL13-1	NI-NK-NK-NI-NK-NK-HD-NI-NI-HD-HD-NI-NI-NK-NG-NK-NI-NG	TAGGAGCCAACCAAGTGAT	−269	–	–
		ArtTAL13-2	HD-HD-NG-NI-NG-NG-NI-NI-NK-HD-HD-NI-HD-HD-NI-NI-NI-HD-NG	TCCTATATAAAGCACCACAACT	−230	+	+
Os11g31190	*OsSWEET14, Os11N3*	ArtTAL14-1	HD-NI-HD-NI-HD-HD-HD-NI-NG-NI-NI-NN-NN-NN-HD-NI-NG	TCACACACCATAAGGGCAT	−335	+	+
		ArtTAL14-2	HD-NI-NG-NN-NG-NN-NG-NN-HD-HD-NG-NG-NG-HD-NI-NG-NG	TCATGTGTGCCTTTTCATT	−300	+	+
Os02g30910	*OsSWEET15*	ArtTAL15-1	HD-NI-HD-HD-NG-HD-NG-HD-NG-HD-HD-HD-NG-NI-NG-NI	TCACCTCTCTCCCTATA	−242	+	+
		ArtTAL15-2	NI-NG-NI-NG-NI-NG-NI-NN-NK-NI-NN-HD-HD-NG-NK-HD-HD-NG	TATATATAGGAGCCTGCCT	−227	+/–	+
Os03g22200	*OsSWEET16*	ArtTAL16-1	NI-NG-NG-NK-NG-NG-NI-NI-NG-NK-NG-HD-NI-HD-NG-NK-HD-NG	TATTGTTAATGTCACTGCT	−147	–	+
		ArtTAL16-2	NG-NN-NG-NN-HD-NG-NI-NG-NI-NG-NI-NG-NG-HD-NK-NG-HD-NK	TTGTGCTATATTCGTCG	−112	–	+

a MSU Version 7. b Based on the nomenclature and clade classification from Chen *et al.* (2010). c For each gene two ArtTAL effectors binding to distinct target sequences were designed. d Amino acids in one letter code. e The 5′ terminal nucleotide of each target box is indicated in bold. f Position of ArtTAL binding box including the initial T counted from ATG (not included) in bp. g Leaves of *Oryza sativa* cultivar IR24 were infiltrated using a blunt syringe and scored at 6 d postinoculation (dpi). No, partial, and water-soaked lesion are indicated as (–), (+/–) and (+), respectively. h *OsSWEET* expression analysis using quantitative reverse transcription polymerase chain reaction (qRT-PCR). Induction, +; no induction, –; no amplification product, na; not determined, nd.

To demonstrate that the observed gain of virulence phenotypes can be solely attributed to the induction of the targeted *SWEET* genes, we evaluated the induction profile of each of the five clade-III paralogs by qRT-PCR during infection of plants with *Xoo* strains delivering different ArtTALs. Among the five candidate susceptibility genes we identified, only *OsSWEET14* was activated upon infiltration of rice leaves with *Xoo* wildtype strain BAI3. None of them was induced upon infiltration with water or *Xoo* strain BAI3Δ*talC* (Fig. VII.4b). In addition, only the expected target *SWEET* gene was induced upon infection of *Xoo* strains delivering the cognate ArtTAL, indicating that the observed phenotypes are not the result of induction of collateral susceptibility genes.

Finally, we quantified bacterial populations *in planta* in order to further demonstrate the function of the clade-III *SWEET* genes as virulence targets of *Xoo*. As shown in Figure VII.5, BAI3Δ*talC* derivatives carrying ArtTALs inducing any of the five clade-III SWEET genes grew to significantly larger population sizes than BAI3Δ*talC* or BAI3Δ*talC* carrying ArtTALs that target the other SWEET clade members.

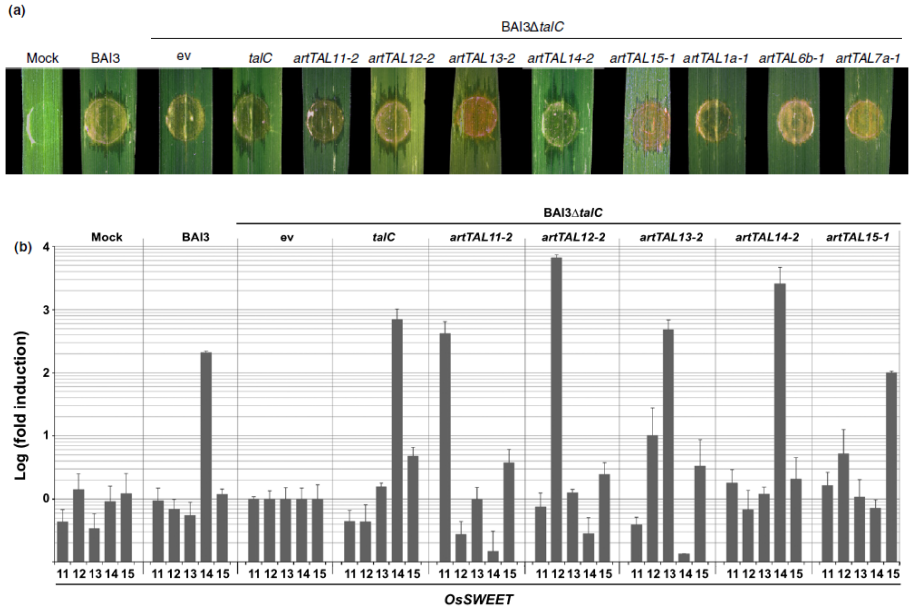


Figure VII.4. Artificial transcription activator-like (TAL) effectors that specifically induce clade-III *OsSWEET* genes confer virulence to *Xanthomonas oryzae* pv. *oryzae* (*Xoo*).

(a) Phenotypes of *Xoo* strains on rice leaves. *Oryza sativa* cultivars IR24 were inoculated with water (Mock), *Xoo* BAI3, or *Xoo* BAI3Δ*talC*carrying empty vector (ev), *talc* or artificial transcription activator-like (*artTAL*) effectors targeting the five clade-III *OsSWEET11* to *OsSWEET15* genes, as well as *OsSWEET1a, OsSWEET6b* and *OsSWEET7a*. (b) Transcript abundances of clade-III *OsSWEET* genes *SWEET11* to *SWEET15* were determined 1dpi by quantitative reverse transcription polymerase chain reaction (qRT-PCR). The bars represent a minimum of two biological samples taken from three leaves. The fold change was calculated in comparison to BAI3Δ*talC* (ev). Error bars represent ± SD. Actin was used as a reference gene to normalize cDNA amounts.

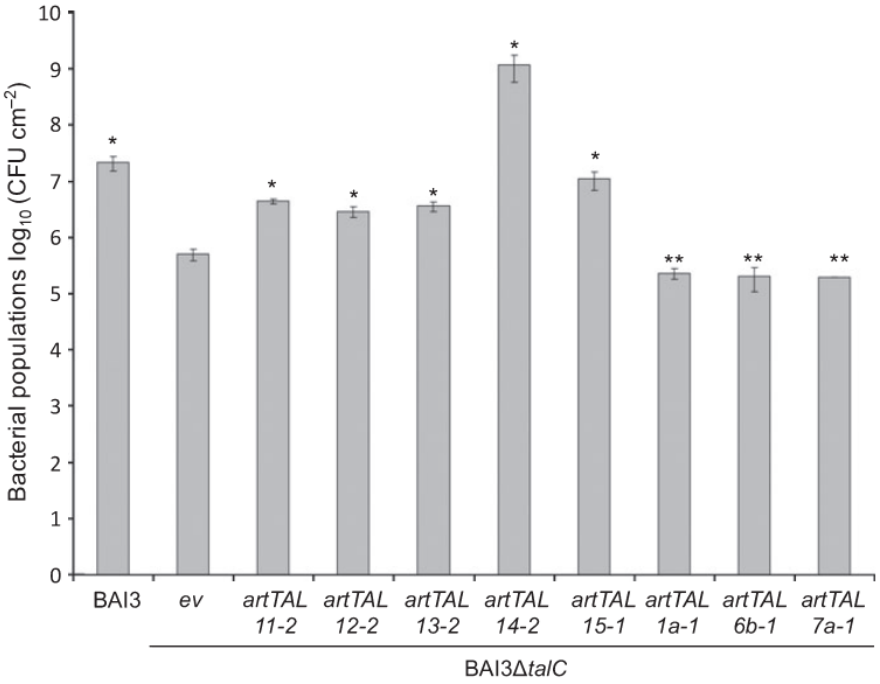


Figure VII.5. Artificial transcription activator-like (TAL) effectors that induce clade-III *OsSWEET* genes restore bacterial growth in rice leaves. *Oryza sativa* IR24 leaves were infiltrated with *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) BAI3, *Xoo*BAI3Δ*talC* carrying empty vector (ev), or *artTAL11-2*, *artTAL12-2*, *artTAL13-2*, *artTAL14-2*, *artTAL15-1*, *artTAL1a-1*, *artTAL6b-1*, and *artTAL7a-1*, which target *OsSWEET11*, *OsSWEET12*, *OsSWEET13*, *OsSWEET14*, *OsSWEET15*, *OsSWEET1a*, *OsSWEET6b* and *OsSWEET7a*, respectively. Bacterial populations were scored 6 d postinfiltration. Data are the arithmetic means from three plants and three independent experiments. Error bars represent ± SD. *Significant difference in Student's *t*-test at $P < 0.01$ when comparing samples with BAI3Δ*talC* (ev); **, no differences were detected at $P < 0.01$.

Discussion

Modern agriculture relies on effective control of pathogens. Better knowledge about host molecular determinants underlying plant diseases is thus essential to develop new strategies leading to plant resistance. However, only a few host genes have been characterized until now that are critical to promote bacterial diseases. These include members of the *SWEET/nodulin-3* family which were recently

identified as essential alternative *Xoo* rice virulence targets, also called susceptibility genes (White & Yang, 2009). Current models suggest that bacteria hijack individual SWEETs that function as sugar transporters, leading to the accumulation of nutrients in the apoplast where *Xoo* multiplies. Previous studies identified *OsSWEET11* and *OsSWEET14* as susceptibility genes, as their induction by cognate native TAL effectors provides virulence to strains carrying them (Talbot, 2010).

Here, we further corroborate the key role of *OsSWEET14* in bacterial virulence as a direct target of natural *Xanthomonas* TAL effectors. Tal5, a novel major virulence TAL effector from a Malian strain of *Xoo*, recognizes a distinct sequence within the *OsSWEET14* promoter sequence, which differs from those that are recognized by strains from the Philippines and from Burkina Faso. This functional convergence of virulence factors illustrates the evolutionary pressure that is imposed on *Xoo* strains of different geographic origins and genetic lineages to evolve *TAL* genes that induce this particular *SWEET* gene.

In a recent report, Li *et al.* (2013) identified *OsSWEET12* as a third potential susceptibility gene. Upon infection of rice with *Xoo* mutant strains defective for *OsSWEET11* activation, the authors showed that complementation with two independent ArtTALs activating *OsSWEET12* leads to a gain of virulence. We significantly extend this study by providing additional important experiments. We analyzed bioinformatically that the ArtTALs target no other common genes besides *OsSWEET14* (suppl. data VII. S5). We verified ArtTALs stability (suppl. data. VII.S1) and assessed their binding

specificity to their target promoter using GUS reporter assays (Fig. VII.2c). Importantly, we demonstrate that ArtTAL-mediated *OsSWEET* activation is not the result of collateral expression of other clade III paralogs (Fig. VII.4). Finally, we extended this strategy to the full *SWEET* gene family, systematically addressing the function of 20 of the 22 *OsSWEET* paralogs. These studies highlight the functionality of ArtTALs as useful tools to investigate the function of host genes during the plant-pathogen interaction. In total, we generated 36 ArtTALs, 26 of which (72%) were functional, and induced their target plant genes upon delivery by *Xanthomonas*. Few ArtTALs were not functional in our study, which might be because of nonoptimal choice of the target region or a noninducible status of the target gene. Overall, our findings emphasize that ArtTALs are suitable tools to induce host genes of choice and that bacterial delivery is a highly efficient means to do so. In summary, we confirm that *OsSWEET11*, *OsSWEET12*, and *OsSWEET14* are essential bona fide host virulence targets of *Xoo*; we formally validate the function of *OsSWEET13*/*Xa25* as a fourth susceptibility gene, as up to this only indirect evidence has been available (Liu et al., 2011); we identify *OsSWEET15* as novel potential susceptibility gene which can support the development of disease symptoms triggered by *Xoo*; and, in total, we analyzed 20 *OsSWEET* paralogs only five of which could be demonstrated to function as susceptibility genes.

Our data suggest that, in addition to the three already identified *SWEET* genes (*OsSWEET11*, *OsSWEET12* and *OsSWEET14*), two more (*OsSWEET13* and *OsSWEET15*) may be exploited by *Xoo,* presumably to satisfy the pathogen's nutritional needs. *OsSWEET13* is

activated by *Xoo* strain PXO339, but the corresponding TAL effector(s) remain(s) to be identified (Liu *et al*., 2011). By contrast, no native TAL effectors have been reported so far for *OsSWEET12* and *OsSWEET15*. Interestingly, screening GenBank for TAL effectors predicted to recognize DNA boxes within their promoter indicated that the TAL protein AAW76267 of *Xoo* strain KACC10331 (Lee et al., 2005) and Tal7b/Tal8b of *Xoo* strain PXO99^{A} (Salzberg et al., 2008) may target *OsSWEET12* and *OsSWEET15*, respectively (Grau et al., 2013). Although transcriptomic data confirmed an increase in *OsSWEET15* expression in plants challenged with PXO99^{A} as compared to water (Grau et al., 2013), *OsSWEET15* induction might not be sufficient for growth on IRBB13 (Fig. VII.3b), indicating that a threshold expression level might be required to support infection. Indeed, the target box for Tal7b/Tal8b appears to be suboptimal, as a G is found at the 5' end of the predicted DNA box, which is known to be less efficient for gene activation (Boch et al., 2009).

Our study highlighted the fact that five *OsSWEET* genes may be manipulated by *Xoo* to favor disease. Interestingly, these five *SWEET* genes, *OsSWEET11* to *OsSWEET15*, all belong to phylogenetic clade III (Fig. VII.6). Several *A. thaliana* SWEETs, as well as rice OsSWEET11 and OsSWEET14, are both low-affinity glucose and sucrose transporters (Chen et al., 2010, 2012). OsSWEET11 and OsSWEET14 are bidirectional sugar transporters and differ in their transport direction from the SUT/SUC high-affinity H^{+} sugar symporters of plants (Kühn & Grof, 2010; Chen et al., 2012). Therefore, increased expression of *SWEETs* alone might lead to a net efflux of sugars from the plant cells. Under physiological conditions,

SWEET sugar transporters facilitate flux of sucrose from the phloem parenchyma cells into the apoplast, whereas SUT/SUC transporters load sucrose from the apoplast into the phloem companion cells and

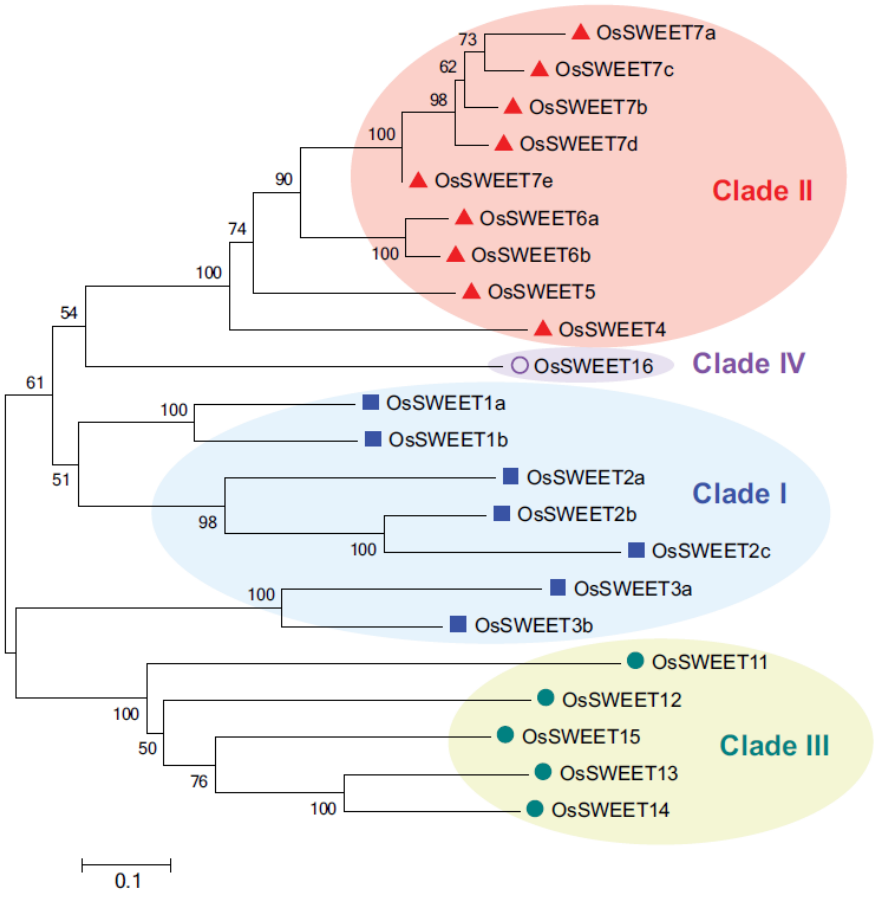


Figure VII.6. Phylogenetic tree of OsSWEET proteins.
Amino acid sequences were obtained from UniProt (http://www.uniprot.org/; Bairoch *et al.*, 2005) and aligned using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/; Edgar, 2004) with default parameters. The sequence of OsSWEET7a was corrected, as UniProt reports an erroneous tandem duplication of 54 amino acids at the N terminus. Alignments were analyzed by the neighbor-joining method using MEGA5.10 (http://www.megasoftware.net; Tamura *et al.*, 2011), assuming a Poisson substitution model, uniform rates among sites, and pairwise deletions for missing data and gaps. The numbers at the nodes represent bootstrap percentage values based on 1000 replications. Proteins of clade I, blue squares; clade II, red triangles; clade III, green dots; clade IV, purple circle.

sieve elements (Kühn & Grof, 2010; Braun, 2012). Our data imply that all clade-III OsSWEETs share a function that can support multiplication of *Xoo* and that this function is absent from the other SWEET paralogs in rice. It is tempting to speculate that this common function is the export of a particular sugar, possibly sucrose.

At present, it is unclear whether *Xoo* relies on SWEETs only at certain infection stages. In a study involving *Xoo* strain BAI3, only vascular spreading, but not local growth of *Xoo* is prevented in the absence of *SWEET* induction (Yu et al., 2011). Curiously, the locally restricted rice pathogen *X. oryzae* pv. *oryzicola* is also able to benefit from TAL effectors that mediate *SWEET* gene induction, although no native *SWEET*-targeting TAL effectors have been reported in *Xoc* (Verdier et al., 2012). Other pathogens, such as *P. syringae* pv. *tomato* DC3000 and the fungal pathogens *Golovinomyces cichoracearum* and *Botrytis cinerea*, also induce the expression of *SWEET* genes in *Arabidopsis*, including those that facilitate sugar transport (Chen et al., 2010, 2012). This suggests that different pathogens exploit plant sugar transporters by different means to gain access to carbohydrate nutrients.

The hallmark of our approach relies on its great potential to discover as yet unknown plant susceptibility genes. ArtTALs perfectly mimic the action of natural TAL effectors and will therefore enable to the major plant genes controlling bacterial proliferation in any TAL-mediated plant disease to be pinpointed, such as Citrus canker and cassava bacterial blight (Duan et al., 1999; Rao et al., 2002). In addition, searches for unresponsive alleles will offer new resources to achieve sustainable plant resistance by marker-assisted breeding, thus

circumventing transgenic approaches. Strategies aimed at producing rice lines with improved and broad resistance against BLB already incorporate the concept of gain of "resistance by defective susceptibility", as successfully done by pyramiding the unresponsive allele of *OsSWEET11*, *xa13* (Rao et al., 2002).

In addition to *OsSWEET11*, hijacking of *OsSWEET14* by *Xoo* appears to be a widespread strategy to provoke disease. As reported by Li *et al.* (2012), 32 out of 40 *Xoo* strains collected worldwide were impaired for virulence when inoculated on TALEN-edited *OsSWEET14* rice lines mutated in the AvrXa7/PthXo3 box. Our results on Tal5 and ArtTAL14s, increase the number of experimentally proven native and artificial TAL effectors activating *OsSWEET14* to six, which recognize five different target boxes (Fig. VII.1). This finding illustrates the impressive potential of TAL effectors to induce genes by binding to differently located target sequences. Therefore, it will be essential to take the natural diversity of TAL effector repertoires in *Xoo* populations into account before exploiting naturally or generating unresponsive susceptibility target gene variants.

References

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

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Natale DA, O'Donovan C, Redaschi N, Yeh LS. 2005. The Universal Protein Resource (UniProt). *Nucleic Acids Research* **33**: D154-159.

Baker RF, Leach KA, Braun DM. 2012. SWEET as sugar: new sucrose effluxers in plants. *Molecular Plant* **5**: 766-768.

Boch J, Bonas U. 2010. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annual Review of Phytopathology* **48**: 419-436.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**: 1509-1512.

Bogdanove AJ, Voytas DF. 2011. TAL effectors: customizable proteins for DNA targeting. *Science* **333**: 1843-1846.

Braun DM. 2012. SWEET! The pathway is complete. *Science* **335**: 173-174.

Bultmann S, Morbitzer R, Schmidt CS, Thanisch K, Spada F, Elsaesser J, Lahaye T, Leonhardt H. 2012**.** Targeted transcriptional activation of silent *oct4* pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Research* **40**: 5368-77.

Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, Guo WJ, Kim JG, Underwood W, Chaudhuri B*,* Chermak D, Antony G, White FF, Somerville SC, Mudgett MB, Frommer WB. 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**: 527-532.

Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB. 2012. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**: 207-211.

Christian ML, Demorest ZL, Starker CG, Osborn MJ, Nyquist MD, Zhang Y, Carlson DF, Bradley P, Bogdanove AJ, Voytas DF. 2012. Targeting G with TAL effectors: A comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS ONE* **7**: e45383.

Chu Z, Fu B, Yang H, Xu C, Li Z, Sanchez A, Park YJ, Bennetzen JL, Zhang Q, Wang S. 2006a. Targeting *xa13*, a recessive gene for bacterial blight resistance in rice. *Theoretical and Applied Genetics* **112**: 455-461.

Chu Z, Yuan M, Yao J, Ge X, Yuan B, Xu C, Li X, Fu B, Li Z, Bennetzen JL*,* Zhang Q, Wang S. 2006b. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes and Development* **20**: 1250-1255.

Cong L, Zhou R, Kuo YC, Cunniff M, Zhang F. 2012 Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature Communications* **3**: 968.

Duan YP, Castañeda A, Zhao G, Erdos G, Gabriel DW. 1999. Expression of a single, host-specific, bacterial pathogenicity gene in plant cells elicits division, enlargement, and cell death. *Molecular Plant–Microbe Interactions* **12**: 556-560.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Figurski D, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2 is dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Science, USA* **76**: 1648-1652.

Gamas P, Niebel Fde C, Lescure N, Cullimore J. 1996. Use of a subtractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. *Molecular Plant–Microbe Interactions* **9**: 233-242.

Geißler R, Scholze H, Hahn S, Streubel J, Bonas U, Behrens SE, Boch J. 2011**.** Transcriptional activators of human genes with programmable DNA-specificity. *PLoS ONE* **6**: e19509.

Gonzalez C, Szurek B, Manceau C, Mathieu T, Sere Y, Verdier V. 2007. Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. *Molecular Plant–Microbe Interactions* **20**: 534-546.

Grau J, Wolf A, Reschke M, Bonas U, Posch S, Boch J. 2013. Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Computational Biology* **9**: e1002962.

Guan YF, Huang XY, Zhu J, Gao JF, Zhang HX, Yang ZN. 2008. *RUPTURED POLLEN GRAIN1*, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in Arabidopsis. *Plant Physiology* **147**: 852-863.

Hopkins CM, White FF, Choi SH, Guo A, Leach JE. 1992. Identification of a family of avirulence genes from *Xanthomonas oryzae* pv. *oryzae*. *Molecular Plant–Microbe Interactions* **5**: 451-459.

Kauffman HE, Reddy AS, Hsieh SP, Merca SD. 1973. An improved technique for evaluating resistance of rice varieties to *Xanthomonas oryzae*. *Plant Disease Management Reports* **57** : 537-541.

Kay S, Hahn S, Marois E, Wieduwild R, Bonas U. 2009. Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3Δrep16. *The Plant Journal* **59**: 859-871.

Kühn C, Grof CP. 2010. Sucrose transporters of higher plants. *Current Opinion in Plant Biology* **13**: 288-298.

Lee BM, Park YJ, Park DS, Kang HW, Kim JG, Song ES, Park IC, Yoon UH, Hahn JH, Koo BS, Lee GB, Kim H, Park HS, Yoon KO, Kim JH, Jung CH, Koh NH, Seo JS, Go SJ. 2005. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Research* **33**: 577-586.

Li T, Liu B, Spalding MH, Weeks DP, Yang B. 2012**.** High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature Biotechnology* **30**: 390-392.

Li T, Huang S, Zhou J, Yang B. 2013**.** Designer TAL effectors induce disease susceptibility and resistance to *Xanthomonas oryzae* pv. *oryzae* in rice. *Molecular Plant* **6**: 781-789.

Liu Q, Yuan M, Zhou Y, Li X, Xiao J, Wang S. 2011. A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. *Plant, Cell and Environment* **34**: 1958-1969.

Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, Joung JK. 2013. Robust, synergistic regulation of human gene expression using TALE activators. *Nature Methods* **10**: 243-245.

Marois E, Van den Ackerveken G, Bonas U. 2002. The *Xanthomonas* type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. *Molecular Plant–Microbe Interactions* **15**: 637-646.

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Ahang HS, Holmes MC, Zhang L, Gregory PD, Redar EJ. 2011. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology* **29**, 143-148.

Morbitzer R, Römer P, Boch J, Lahaye T. 2010. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proceedings of the National Academy of Science, USA* **107**: 21617-21622.

Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**: 1501.

Mussolino C, Cathomen T. 2012. TALE nucleases: tailored genome engineering made easy. *Current Opinion in Biotechnology* **23**: 644-650.

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T. 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering* **104**: 34-41.

Nino-Liu DO, Ronald PC, Bogdanove AJ. 2006. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Molecular Plant Pathology* **7**: 303-324.

Pérez-Quintero A, Rodriguez-R L, Dereeper A, Lopez C, Koebnik R, Szurek B, Cunnac S. 2013. An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS ONE* **8**: e68464.

Perez-Pinera P, Ousterout DG, Brunger JM, Farin AM, Glass KA, Guilak F, Crawford GE, Hartemink AJ, Gersbach CA. 2013. Synergistic and tunable gene activation in human cells by combinations of synthetic transcription factors. *Nature Methods* **10:** 239-242.

Quirino BF, Normanly J, Amasino RM. 1999. Diverse range of gene activity during *Arabidopsis thaliana* leaf senescence includes pathogen-independent induction of defense-related genes. *Plant Molecular Biology* **40**: 267-278.

Rao KK, Lakshminarasu M, Jena KK. 2002. DNA markers and marker-assisted breeding for durable resistance to bacterial blight disease in rice. *Biotechnology Advances* **20**: 33-47.

Reimers PJ, Leach JE. 1991. Race-specific resistance to *Xanthomonas oryzae* pv. *oryzae* conferred by bacterial blight resistance gene *Xa-10* in rice (*Oryza sativa*) involves accumulation of a lignin-like substance in host tissues. *Physiological and Molecular Plant Pathology* **38**: 39-55.

Römer P, Recht S, Strauß T, Elsaesser J, Schornack S, Boch J, Wang S, Lahaye T. 2010. Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytologist* **187**: 1048-1057.

Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, Furutani A, Ochiai H, Delcher AL, Kelley D, Madupu R, Puiu D, Radune D, Shumway M, Trapnell C, Aparna G, Jha G, Pandey A, Patil PB, Ishihara H, Meyer DF, Szurek B, Verdier V, Koebnik R, Dow JM, Ryan RP, Hirata H, Tsuyumu S, Won Lee S, Seo YS, Sriariyanum M, Ronald PC, Sonti RV, Van Sluys MA, Leach JE, White FF, Bogdanove AJ. 2008. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99^{A}. *BMC Genomics* **9**: 204.

Streubel J, Blücher C, Landgraf A, Boch J. 2012. TAL effector RVD specificities and efficiencies. *Nature Biotechnology* **7**: 593-595.

Talbot NJ 2010. Cell biology: Raiding the sweet shop. *Nature* **468**: 510-511.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Van Larebeke N, Engler G, Holsters M, Van den Elsacker S, Zaenen I, Schilperoort RA, Schell J. 1974. Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability. *Nature* **252**: 169-170.

Verdier V, Triplett LR, Hummel AW, Corral R, Cernadas RA, Schmidt CL, Bogdanove AJ, Leach JE. 2012. Transcription activator-like (TAL) effectors targeting *OsSWEET* genes enhance virulence on diverse rice (*Oryza sativa*) varieties when expressed individually in a TAL effector-deficient strain of *Xanthomonas oryzae*. *New Phytologist* **196**: 1197-1207.

White FF, Yang B. 2009. Host and pathogen factors controlling the rice-*Xanthomonas oryzae* interaction. *Plant Physiology* **150**: 1677-1686.

Yang B, Sugio A, White FF. 2006. *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proceedings of the National Acadamy of Science, USA* **103**: 10503-10508.

Yu Y, Streubel J, Balzergue S, Champion A, Boch J, Koebnik R, Feng J, Verdier V, Szurek B. 2011. Colonization of rice leaf blades by an african strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice Nodulin-3 *Os11N3* gene. *Molecular Plant–Microbe Interactions* **24**: 1102-1113.

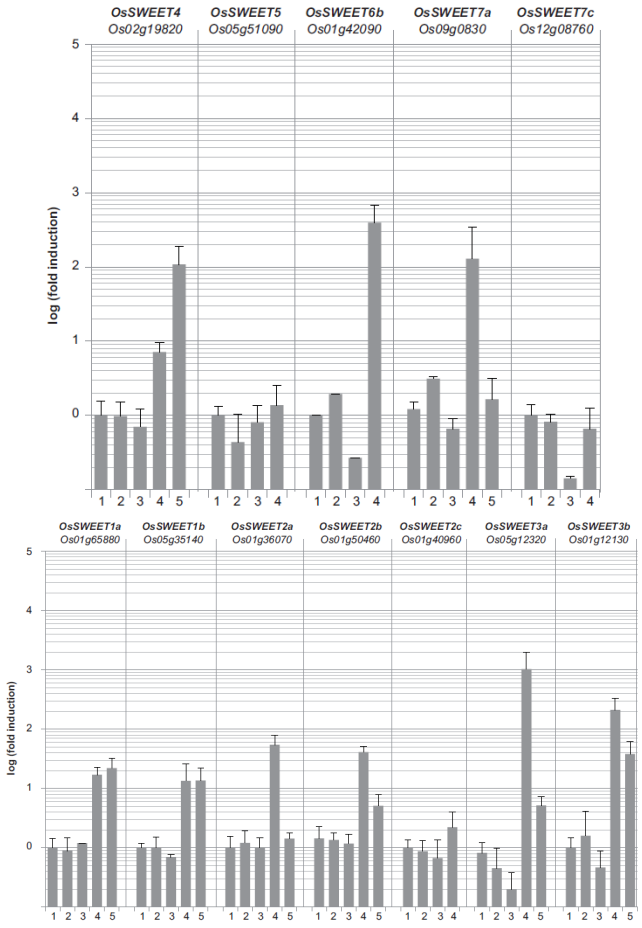
Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnology* **29**: 149-153.

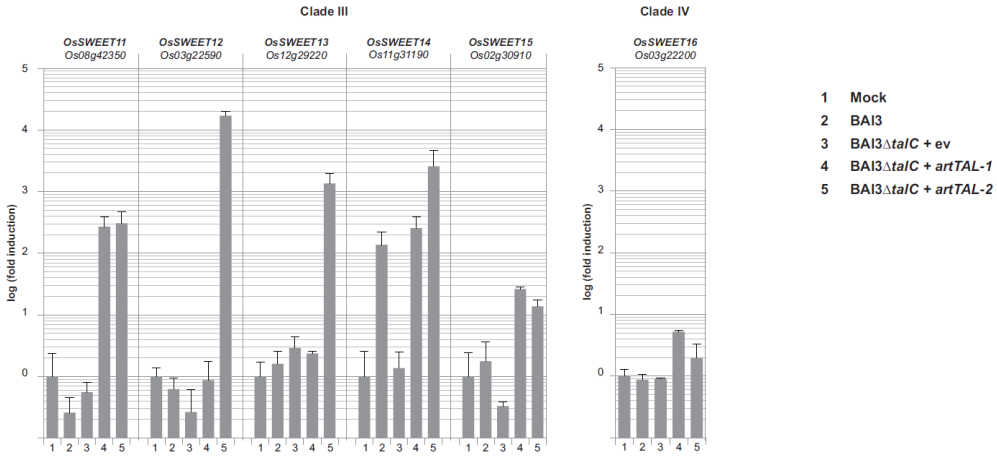
<u>Supplemental data VII.S1</u>. ArtTAL protein expression analysis in *Xanthomonas oryzae* pv. *oryzae*.

For each *OsSWEET* gene (except for *OsSWEET7d* and *OsSWEET7e*, see text), one or two artificial TAL effectors were generated. Artificial TAL effector constructs as well as *tal5* and *talC* were inserted into the broad host vector pSKX1 allowing a *lac* promoter-driven constitutive expression and the fusion to a C-terminal FLAG epitope. *Xanthomonas oryzae* pv. *oryzae* strains BAI3Δ*talC* carrying the corresponding TAL effector construct were grown in PSA medium. Cells of 1 ml of a bacterial suspension at an OD_{600} of 0.2 were harvested and analysed by SDS-PAGE followed by immunoblotting. The TAL effectors are ordered according to the different *OsSWEET* gene clades. The expected size for TAL effectors with 17.5 repeats is 125 kDa and for TAL effectors with 21.5 or 22.5 repeats is 135 or 140 kDa, respectively.

<u>Supplemental data VII.S2</u>. Quantification of *OsSWEET* transcript levels

Three-weeks old *Oryza sativa* plants from the cultivar Nipponbare or IR24 (*OsSWEET14*) were inoculated with water (Mock, 1) or with *Xanthomonas oryzae* pv. *oryzae* strains BAI3 (2), BAI3Δ*talC* carrying an empty vector (3), or BAI3Δ*talC* carrying the corresponding artificial TAL effector (4: *artTAL-1*; 5: *artTAL-2*). Please notice that for genes *Os01g40960, Os05g50190*, *Os01g42090*, and *Os12g07860*, only one artificial TAL effector was analysed. Genes *Os01g42110* (*OsSWEET6a*) and *Os09g08440* (*OsSWEET7b*) were not analyzed, because no specific amplification product was detected. The expression level of the gene *Os01g42090* in the sample treated with ArtTAL6b-2 was not determined. Transcript levels of the respective *OsSWEET* genes were determined one day post inoculation by quantitative RT-PCR. The bars represent the mean of two to three biological samples. The fold change was calculated in comparison to water and plotted on a logarithmic scale. Error bars represent the standard deviation. Actin was used as a reference gene to normalize cDNA amounts. The *OsSWEET* genes are ordered in the different clades according to the phylogenetic tree shown in fig. VII.6.




Clade III
Clade IV
OsSWEET11
Os08g42350
OsSWEET12
Os03g22590
OsSWEET13
Os12g29220
OsSWEET14
Os11g31190
OsSWEET15
Os02g30910
OsSWEET16
Os03g22200
log (fold induction)
0
1
2
3
4
5
1 2 3 4 5
1 Mock
2 BAI3
3 BAI3ΔtalC + ev
4 BAI3ΔtalC + artTAL-1
5 BAI3ΔtalC + artTAL-2

Supplemental data VII.S3 Sequences of oligonucleotides used for qRT-PCR.

Gene ID	Gene name	Forward primer 5´-> 3	Reverse primer 5´-> 3	E[1]	E in %
Os01g65880	*OsSWEET1a*	TCTCCATCTGCATGTACGCC	TGCATTCTTCCGTGCGTGTC	1.817	90.83%
Os05g35140	*OsSWEET1b*	ACTGCCTCCTCTCTGCATG	GTAGATGAACCACGACGTGC	1.590	79.48%
Os01g36070	*OsSWEET2a*	TGGCGCTGGGATTGCAG	GACGCAACGCTGAGGTAAC	1.931	96.57%
Os01g50460	*OsSWEET2b*	ATGATATCTCCTGCTTTGCTG	CCGAATCACAACTCCCATGAC	1.894	94.69%
Os01g40960	*OsSWEET2c*	GGCAACAAGAGCGAGCAGC	TAGAAGATGAAGAGACAAATGTAGG	2.026	101.28%
Os05g12320	*OsSWEET3a*	CCTTCCTGATATACGTTTTATTGTG	TTACGTATATGATGGTTGTGTATTG	1.907	95.37%
Os01g12130	*OsSWEET3b*	CGTGTAGCAGTAGGAATTCTAG	TGCGTAAACCGTGCGTATG	1.895	94.77%
Os02g19820	*OsSWEET4*	GGGGTCGGTGGAGCAGTAC	CCGTTGGGGATCGTGATG	1.840	92.12%
Os05g50190	*OsSWEET5*	CGTCCGCAATGTCGTTGG	GCTCTTTGTCGCGATCAC	1.756	87.82%
Os01g42090	*OsSWEET6b*	CCAACAAGAAGCGCCTGAG	ACGGAGACGACACTGCC	1.767	88.33%
Os09g08030	*OsSWEET7a*	TGGCGTACTCTTAGGCGTG	CTGGATAGTGATATCGATGGTG	1.952	97.61%
Os12g07860	*OsSWEET7c*	TAGCGGTCGGTGTGCTTTTG	CTGACAGGGGCGACGATG	1.893	94.67%
Os08g42350	*OsSWEET11*	AGTCGACGGGAGGGTACAG	TTCGGGTACATGACGTAGGG	1.897	94.85%
Os03g22590	*OsSWEET12*	CATTGGTCTTTGCTGTGGGG	CGGTCGGTGGCCTTGAC	1.957	97.86%
Os12g29220	*OsSWEET13*	GGCCTGTCCCTGCAGCATC	CCCGAACACCCCCACGTTC	1.909	95.47%
Os11g31190	*OsSWEET14*	CTACCTGGCCCCACTGC	GTGCGCACCACCAGCC	1.963	98.15%
Os02g30910	*OsSWEET15*	TGTCCATGGAACGCAGCAC	CAGCCGAGGACGTGGAC	1.878	93.91%
Os03g22200	*OsSWEET16*	GATGGGGCATCAGGTGGAG	GGTCCATCAGCTTCATCAAAATTC	1.781	89.04%
	Actin	TGTGTGTGACAATGGAACTGGC	GAGTCCAACACGATACCAGTTG	1.884	94.22%

[1]Efficiency was calculated with $E=10^{-1}/m$; m was calculated by using a standard curve plot of a dilution series.

Supplemental data VII.S4 Talvez ranking of Tal5 target candidates upon whole promoterome search.

Prediction rank[1]	Gene ID[2]	Talvez prediction score	Position from ATG[3]	Predicted target box[4]	Annotation
1	*Os11g31190*	15.043	-235	**T**AAGCTCATCAAGCCTTCA	nodulin MtN3 family protein, putative, expressed
2	*Os02g29150*	13.856	-69	**T**CCACTCCCCAAACCCTCG	OsFBO11 - F-box and other domain containing protein, expressed
3	*Os06g48680*	13.634	-97	**T**CCACTCCTCGCGCCTTCA	expressed protein
4	*Os01g11720*	13.551	-470	**T**GCACTCATAAAACCCTAA	putative protein containing Pfam profile: PF03004, Transposase_24
5	*Os07g46560*	13.517	-227	**T**GAGCTCATAAAGCCTTCA	seven in absentia protein family domain containing protein, expressed
6	*Os01g21542*	13.467	-81	**C**GCGCTCCTTAAGCCCTTA	retrotransposon protein, putative, unclassified, expressed
7	*Os03g05250*	13.373	-459	**T**ACCATCCTCAAATCCTCA	expressed protein
8	*Os08g11450*	13.263	-219	**T**CCACTCCTCCACCACTCA	expressed protein
9	*Os03g54750*	13.244	-215	**C**ACACTCCTCACACCACCA	COBRA-like 3 protein precursor, putative, expressed
10	*Os01g04340*	13.223	-219	**C**CCACTCCTCGACCCTTCA	hsp20/alpha crystallin family protein, putative, expressed

[1]Talvez prediction for the transcription activator-like (TAL)-DNA box pair in a whole proterome screening

[2]MSU version 7

[3]Position of TAL binding box including the 5´terminal nucleotide counted from ATG (not included)

[4]The 5' terminal nucleotide of each target box is indicated in bold

<u>Supplemental data VII.S5</u> Target candidates of artifical TAL effectors ArtTAL14-1 and ArtTAL 14-2 upon whole promoterome search.

ArtTAL	Prediction rank[1]	Gene ID[2]	Talvez prediction score	Position from ATG[3]	Predicted target box[4]	Annotation
ArtTAL14-1	1	*Os11g31190*	19.309	-335	**T**CACACACCATAAGGGCA	nodulin MtN3 family protein, putative, expressed
ArtTAL14-1	2	*Os04g17890*	15.363	-459	**T**CAAAAACCATAAAAACAT	retrotransposon protein, putative, unclassified, expressed
ArtTAL14-1	3	*Os04g17850*	15.363	-459	**T**CAAAAACCATAAAAACAT	retrotransposon protein, putative, unclassified, expressed
ArtTAL14-1	4	*Os08g07940*	15.210	-387	**T**CATACACCATACGAGCAT	disease resistance protein RPM1, putative, expressed
ArtTAL14-1	5	*Os11g09130*	14.944	-240	**T**CACACGCCATAAAAAAAT	receptor-like protein kinase 5 precursor, putative, expressed
ArtTAL14-1	6	*Os04g03210*	14.097	-444	**C**TACACACCATAAGAAAAT	receptor kinase, putative, expressed
ArtTAL14-1	7	*Os07g07860*	14.028	-106	**T**CACACACCATCAAGCTAT	LTPL76 - Protease inhibitor/seed storage/LTP family protein precursor, expressed
ArtTAL14-1	8	*Os05g03934*	14.008	-127	**C**CACACACCACACACACAT	expressed protein
ArtTAL14-1	9	*Os04g17810*	13.858	-212	**T**CAAAAACCACAAAAACAT	retrotransposon protein, putative, unclassified
ArtTAL14-1	10	*Os03g02514*	13.412	-346	**T**CAAACACCACAAGGCCAC	hydrolase, alpha/beta fold family protein, putative, expressed
ArtTAL14-1	11	*Os05g35274*	13.306	-406	**C**CAAACCCCACAAAAACAT	RNA recognition motif containing protein, putative, expressed

ArtTAL14-1	12	*Os02g17000*	13.285	-36	**C**ACCAAACCATAAGAACAT	OsSub14 - Putative Subtilisin homologue, expressed
ArtTAL14-1	13	*Os04g38290*	13.225	-234	**T**CACACACAACACGCACAT	expressed protein
ArtTAL14-1	14	*Os03g10310*	13.216	-357	**T**CACAAACCCTACGCACAT	expressed protein
ArtTAL14-1	15	*Os12g30000*	13.135	-476	**T**CATACAACATAAGAAAAT	expressed protein
ArtTAL14-1	16	*Os06g12760*	13.107	-124	**C**CACACACCATACACATAT	hypothetical protein
ArtTAL14-1	17	*Os04g47960*	13.038	-315	**T**CCCCAACCATAAAACCAT	expressed protein
ArtTAL14-1	18	*Os02g03940*	12.887	-160	**C**CGCACACCATAAAAACCC	expressed protein
ArtTAL14-1	19	*Os06g08740*	12.711	-443	**T**CACACACCATCAACGCAG	expressed protein
ArtTAL14-1	20	*Os03g61590*	12.659	-456	**T**CACACACCAAAAGAAAAA	expressed protein
ArtTAL14-2	1	*Os11g31190*	18.951	-300	**T**CATGTGTGCCTTTTCATT	nodulin MtN3 family protein, putative, expressed
ArtTAL14-2	2	*Os05g36220*	14.795	-154	**T**CACATGAACCTTTTCATT	transposon protein, putative, CACTA, En/Spm sub-class, expressed
ArtTAL14-2	3	*Os11g41780*	14.795	-176	**T**CACATGAACCTTTTCATT	transposon protein, putative, CACTA, En/Spm sub-class, expressed
ArtTAL14-2	4	*Os07g05850*	14.795	-154	**T**CACATGAACCTTTTCATT	transposon protein, putative, CACTA, En/Spm sub-class, expressed
ArtTAL14-2	5	*Os10g38314*	14.704	-156	**T**CATATCTCACTTTTCATT	glutathione S-transferase, N-terminal domain containing protein, expressed
ArtTAL14-2	6	*Os12g02830*	14.331	-189	**C**CATCTCTGCCTTTTCCTT	SNF7 domain containing protein, putative, expressed
ArtTAL14-2	7	*Os05g44530*	13.739	-217	**C**CATATATGATTTTTCATT	BTB3 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain, expressed
ArtTAL14-2	8	*Os02g09820*	13.305	-396	**C**CCCATGTGCCTTTTAATT	zinc finger, C3HC4 type domain

						containing protein, expressed
ArtTAL14-2	9	*Os03g22050*	13.268	-383	**T**AATGTGTACCTTACCATT	CAMK includes calcium/calmodulin depedent protein kinases, expressed
ArtTAL14-2	10	*Os05g12140*	13.081	-211	**T**CCTAAATGCATTTTCATT	Leucine Rich Repeat family protein, expressed
ArtTAL14-2	11	*Os01g29850*	13.075	-433	**T**CATATGCCCCCTTTCACT	expressed protein
ArtTAL14-2	12	*Os05g25160*	13.075	-194	**T**CATATGCCCCCTTTCACT	transposon protein, putative, CACTA, En/Spm sub-class, expressed
ArtTAL14-2	13	*Os11g10880*	13.067	-86	**T**CATACCTGCCTTTTCCCT	expressed protein
ArtTAL14-2	14	*Os09g32190*	12.956	-482	**C**CATACACACCTTTTAATT	tRNA-splicing endonuclease positive effector-related, putative, expressed
ArtTAL14-2	15	*Os04g03890*	12.955	-367	**T**TATGTATGACTTTTAATT	cytochrome P450, putative, expressed
ArtTAL14-2	16	*Os12g30580*	12.780	-127	**T**CATCTGTCCCATTTCCTT	expressed protein
ArtTAL14-2	17	*Os08g43670*	12.743	-263	**T**CATGCGTACCCTTTCCCT	RING-H2 finger protein ATL2B, putative, expressed
ArtTAL14-2	18	*Os03g26260*	12.704	-451	**C**AATGTGCGCCATTTCATT	NB-ARC domain containing protein, expressed
ArtTAL14-2	19	*Os09g19330*	12.697	-343	**T**AACATCTACCTTTTCACT	stripe rust resistance protein Yr10, putative, expressed
ArtTAL14-2	20	*Os07g06830*	12.632	-374	**T**CACGTGTGTTTTTTCATT	gibberellin receptor GID1L2, putative, expressed

[1]Talvez prediction for the transcription activator-like (TAL)-DNA box pair in a whole proterome screening
[2]MSU version 7
[3]Position of TAL binding box including the 5´terminal nucleotide counted from ATG (not included)
[4]The 5' terminal nucleotide of each target box is indicated in bold

<u>Supplemental data VII.S6</u>. Basal and induced *OsSWEET* gene expression levels.

Clade	Gene ID[1]	Gene name	Sample[2]	2ΔCt[3]
I	*Os01g65880*	*OsSWEET1a*	Water	0.043941
			ArtTAL1a-1	1.285167
			ArtTAL1a-2	1.520956
I	*Os05g35140*	*OsSWEET1b*	Water	0.084838
			ArtTAL1b-1	1.562180
			ArtTAL1b-2	5.531725
I	*Os01g36070*	*OsSWEET2a*	Water	0.046305
			ArtTAL2a-1	3.675345
			ArtTAL2a-2	0.075831
I	*Os01g50460*	*OsSWEET2b*	Water	0.489184
			ArtTAL2b-1	17.895277
			ArtTAL2b-2	1.849142
I	*Os01g40960*	*OsSWEET2c*	Water	0.000115
			ArtTAL2c-1	0.000307
I	*Os05g12320*	*OsSWEET3a*	Water	0.000143
			ArtTAL3a-1	0.619059
			ArtTAL3a-2	0.001398
I	*Os01g12130*	*OsSWEET3b*	Water	0.232770
			ArtTAL3b-1	1.360545
II			ArtTAL3b-2	8.662907
	Os02g19820	*OsSWEET4*	Water	0.078664
			ArtTAL4-1	1.456052
			ArtTAL4-2	37.445649
II	*Os05g50190*	*OsSWEET5*	Water	0.000033
			ArtTAL5-1	0.000049
II	*Os01g42090*	*OsSWEET6b*	Water	0.003852
			ArtTAL6b-1	4.999260
II	*Os09g08030*	*OsSWEET7a*	Water	0.017625
			ArtTAL7a-1	0.059922
			ArtTAL7a-2	0.000550
II	*Os12g07860*	*OsSWEET7c*	Water	0.000804
			ArtTAL7c-1	0.000544
III	*Os08g42350*	*OsSWEET11*	Water	0.001342
			ArtTAL11-1	0.744088
			ArtTAL11-2	0.364775

III	*Os03g22590*	*OsSWEET12*	Water	0.000795
			ArtTAL12-1	0.000703
			ArtTAL12-2	12.130544
III	*Os12g29220*	*OsSWEET13*	Water	0.011003685
			ArtTAL13-1	0.029532199
			ArtTAL13-2	14.0792389
III	*Os11g31190*	*OsSWEET14*	Water	0.000946
			ArtTAL14-1	0.159974
			ArttAL14-2	2.450459
III	*Os02g30910*	*OsSWEET15*	Water	0.049225
			ArtTAL15-1	1.519168
			ArtTAL15-2	1.190815
IV	*Os03g22200*	*OsSWEET16*	Water	0.000183
			ArtTAL16-1	0.003405
			ArtTAL16-2	0.000287

[1]MSU version 7

[2]The expression levels of the respective genes were analyzed in RNA samples from *Oryza sativa* cultivar Nipponbare or IR24 plants that were inoculated with water or with *Xanthomonas oryzae* pv. *oryzae* BAI3Δ*talC* carrying the respective ArtTAL

[3]2ΔCt = Ct (gene of interest) – Ct (Actin) with the amplification efficiencies set to 2; values represent the mean of at least two biological replicates.

Supporting Information Methods VII. S1

Generation of *talC* and *tal5* expression constructs.

For transient expression in *Nicotiana benthamiana*, the coding region of *talC* was PCR-amplified, inserted into pENTR/D-TOPO (Life Technologies GmbH), and transferred by GATEWAY LR-recombination to the binary vector pGWB2 which facilitates a constitutive *35S*-driven expression *in planta* and adds an N-terminal GFP tag to TalC. For expression in *Xanthomonas*, the coding region of the N- and C-terminal parts of *talC* were amplified as Golden Gate-compatible modules. N- and C-terminal modules were inserted together with a repeat region dummy module (*hax3* repeat module) into a Golden Gate-compatible pBBR1MCS-5 derivative, pSKX1, which facilitates a *lac* promoter-driven expression and addition of a C-terminal FLAG epitope. The dummy repeat region was then exchanged with a fragment encoding the *talC* repeat region by classic cloning using *Stu*I and *Aat*II. DNA sequencing confirmed the integrity of the final construct. The coding region of *tal5* was extracted from cosmid clone 6B7 (obtained upon screening of a MAI1 genomic DNA library for clones carrying *TAL* genes by colony hybridization, Y. Yu, V. Verdier and B. Szurek, *unpublished data*) and inserted into pSKX1 and pGWB2 for expression in *Xanthomonas* and delivery by *Agrobacterium*, respectively, via classic cloning using *Bam*HI. One *Bam*HI restriction site is overlapping with the ATG start codon and the second one is located 150 bp upstream of the stop codon of the *tal5* gene. Most of the *talC* coding region including the repeat region was exchanged with the corresponding part of *tal5* in the pSKX1-derivative using *Bam*HI. Sequencing confirmed the right insertion into the destination vector.

Supporting Information Methods VII.S2

Choice of target sites and generation of artificial TAL effectors.

Based on the sequence information given by the Rice Genome Annotation Project (version 7; http://rice.plantbiology.msu.edu/) the promoter regions (500 to 40 bp upstream of the ATG) of the *OsSWEET* genes were analyzed to find two appropriate artificial TAL effector (ArtTAL) binding sites per *OsSWEET* gene. Potential binding sites were selected according to several criteria:

a. ArtTAL binding sites should comprise at least 19 bp (including the initial T), thus corresponding to at least 17.5 repeats to ensure a high specificity.
b. ArtTAL binding site should explicitly be highly specific within the *OsSWEET* family.
c. Similar to many natural TAL effector binding sites, ArtTAL binding sites should include or be near to the TATA element.
d. ArtTAL binding sites should be at least 40 to 60 bp upstream of the potential translational start site.
e. There should be no additional ATG between the 3´ end of the ArtTAL binding sites and the potential translational start site.
f. The corresponding artificial TAL effector should contain at least 3 to 4 properly spaced strong RVDs (HD, NN).

The specificity of the chosen binding sites was checked using the Talvez software (http://bioinfo.mpl.ird.fr/cgi-bin/talvez/talvez.cgi; Pérez-Quintero *et al*., 2013). The absence of common collateral target sites (besides of the *OsSWEET* gene) was evaluated among the top-20 predicted targets of each ArtTAL. The ArtTALs were generated as previously described by using the *hax3* backbone and between 17.5 and 21.5 repeats to specify the DNA target sequence. The broad host range vector pSKX1 was used to express the ArtTALs under control of a *lac* promoter. The fusion to a C-terminal FLAG epitope allows detection of the ArtTALs by Western blotting.

Chapter VIII

Discussion and Perspectives

The objective of the thesis was to better understand the interaction between cereals and *X. translucens*, the causal agent of bacterial leaf streak. At the beginning of this project, pathogenicity mechanisms of this clade-1 *Xanthomonas* species were poorly studied at the molecular level. Consequently, this project begins with several questions.

As a first step, it was necessary to establish reproducible and preferably quantitative pathogenicity assays for *X. translucens* on barley. The objective was to propose new standard approaches of inoculation under laboratory conditions. For that, fives methods have been tested. Leaf clipping approach which allows to follow the spread of the disease have been used in all the experiment of this thesis due to it quantitative and reproducible characteristics. Symptoms are scored over a period of 15 days after inoculation. The second method is the stem inoculation, an alternative method that is used to score pathogens which are not spreading along the leaf blade. With this technique, symptoms such as appearance of translucent streaks are scored seven days after inoculation.

In parallel, a new approach to observe the bacterial colonization *in planta* has been created. For this purpose, GFP-labeled fluorescent *X. translucens* strains have been constructed and inoculated into barley leaves. Confocal imaging of clipped leaves demonstrated that *X. translucens* pv. *hordei* colonizes the xylem tissues.

These optimized and new laboratory tools form the basis for follow-up studies on the *X. translucens* pathogenicity. The host range of cereals-pathogenic pathovars could be redefined based on the two proposed inoculation techniques. Preliminary data obtained by confocal imaging should be confirmed and extended by using an additional set of bacterial strains and barley varieties. Better identification of the host range and accurate *in planta* localization during the infection process are helpful to clarify the pathovar classification of *X. translucens* strains.

In order to obtain new insights into the molecular determinants provoking disease, representative (pathotype) strains of *X. translucens* pv. *cerealis*, *X. translucens* pv. *graminis* and *X. translucens* pv. *translucens* have been sequenced using Illumina technology. However, while sequencing of the *X. translucens* pv. *translucens* strain CFBP 2054 was in progress, a draft genome sequence of the same strain (DSM 18974) was released from GenBank. When this project was initiated, not genome sequence was publicly available. Since then, several new *X. translucens* genome sequences were deposited in a public database. Currently, nine sequences are available at GenBank, including two from this work, corresponding to seven pathovars. These resources are helpful to identify new candidate virulence factors of *X. translucens*. Moreover, they form an important base for taxonomic classification, for the development of diagnostic detection tool and for the implementation of epidemiological surveillance tool.

Indeed, no rigorous molecular approach exists for the classification of *X. translucens* at the subspecies level. Genome-wide

comparison of a set of housekeeping genes or average nucleotide identity (ANI) analyses can determine if two strains belong to the same species or subspecies (Konstantinidis & Tiedje, 2005). More efficient methods for the classification of pathovars are required. Possibly, pathovars (as defined by standard inoculation tests) and subspecies (as defined by ANI or multi-locus sequence analyses [MLSA]) are correlated. If so, simple DNA-based techniques could be used to infer pathovars, data that will be important for quarantine issues. More analyses and works on representative strains are required before recommendations can be given. The increasing availability of genome sequences in public databases will certainly help to clarify the differences between pathovars.

Genome sequences of *X. translucens* are also useful to establish new and precise detection tools. An easy, fast and cheap method to identify pathovars will be important for epidemiological surveillance and to rapidly trigger quarantine measurements, if appropriate.

Finally, genome sequences of a representative set of strains, such as pathotype strains, allow to select genetic markers for high-resolution molecular typing. One class of important and well-known genetic markers are variable numbers of tandem repeats (VNTR). VNTRs are short nucleotide sequences organized in repeats, the number of which is used to study genetic diversity. Over the last years, several VNTR schemes have been published for plant-pathogenic bacteria, including species and pathovars of *Xanthomonas* (*X. citri, X. oryzae*, *X. arboricola*, and *X. axonopodis* pv. *manihotis*) (Ngoc et al., 2009; Trujillo et al., 2014; Poulin et al., 2015; Pruvost at al., 2014; Cesbron et al., 2014). As a diagnostic and epidemiological

tool, a VNTR scheme targeting both cereals- and grasses-pathogenic xanthomonads will be of high interest, and our genome sequences are expected to contribute to such a scheme.

Based on genomic data, comparative genomics is an important approach to identify candidate virulence factors. Our analysis has focused on the *hrp* gene cluster, which is known to be involved in pathogenicity in several *Xanthomonas* species. Comparison of five *X. translucens hrp* gene clusters revealed that all of them have the same genetic organization. The key regulatory genes *hrpG* and *hrpX* flank the *hrp* core cluster while these two regulatory genes are encoded next to each other at a distant genetic locus in clade-2 xanthomonads This organization is similar to the cluster from *R. solanacearum* and also from the two other clade-1 species *X. theicola* and *X. hyacinthi*. Another distinct characteristic of *X. translucens hrp* clusters is the presence of a conserved gene at the border of the cluster. This gene is strongly induced by HrpG and was therefore called *hgiA* for HrpG-induced gene. Homologs of *hgiA* have been found in all the clade-1 *hrp* clusters (*X. theicola* and *X. hyacinthi*) and also in *Burkholderia andropogonis* and in *Collimonas fungivorans*. Whenever *hgiA* is present no homolog of the T3SS translocon genes *hrpF* (from *X. euvesicatoria*) (Rossier et al., 2000, Büttner & Bonas, 2002) or *popF* (from *R. solanacearum*) (Meyer et al., 2006) was found. This mutual exclusivity led us speculate that *hgiA* encodes a functional analog of HrpF.

Since nothing was known about the contribution of the T3SS to pathogenicity of barley-pathogenic xanthomonads, mutants in a well-conserved *hrp* gene, *hrcT*, and in *hgiA* were constructed. Leaf-clip

inoculations demonstrated that *hrp* locus is required for disease development and bacterial multiplication of *X. translucens* in barley. Interestingly, *hgiA* mutants showed the same phenotype as *hrp* mutants, i.e. loss of virulence on the host plant barley and loss of HR on a non-host plant (pepper). These results support the hypothesis that *hgiA* plays a role in the delivery of the type III effector proteins into plant cells. Since *hgiA* is the first gene of a two-gene operon, the insertion mutation in *hgiA* likely has a polar effect on the downstream *hpaH* gene. Surprisingly, the loss of virulence observed in barley was partially complemented by *hgiA* but also by *hpaH*. These results show that both of these genes seem to be required for efficient effector translocation into plant cell. The translocon complex in the plant plasma membrane is maybe not composed of a single protein but could be formed by a protein complex. More experiments are required to support the role of HgiA as a translocon component. For instance, it could be tested if HgiA itself is secreted by *X. translucens* or when expressed in a clade-2 *Xanthomonas*, such as the model strain *X. euvesicatoria* 85-10. It could also be tested if the *hgiA* mutant can still secrete type III effectors into the medium, which would be evidence that it is a more downstream step of translocation, i.e. the delivery into the plant cell, which is affected by the mutation, but not the secretion out of the bacterium itself.

After became evident that type III secretion critically contributes to pathogenicity, we wondered whether or not type III effectors from the TALE family are important for virulence. When 14 *X. translucens* strains corresponding to nine pathovars were analyzed, the diversity with respect to number of TALEs was not very pronounced. Between

zero to eight TALEs were found per strain. Strains with the largest number of TALEs belong to the small grain cereals pathogens (*X. translucens* pvs. *hordei*, *translucens, undulosa,* and *secalis*). By contrast, grasses pathogens possess only between zero to three TALEs. Several hypotheses could explain this observation. First, it could be that this difference is linked to domestication of the host plants. Indeed, it has been published that African isolates of the rice pathogen *X. oryzae* pv. *oryzae* contain fewer TALEs than Asian strain of *X. oryzae* pv. *oryzae*. This finding coincides with the fact the African rice species *Oryzae glaberrima* was much later domesticated than the Asian rice species *Oryzae sativa* (~3500 years *versus* ~8000 years) (Wang et al., 2014). It is conceivable to assume that cereals were earlier and stronger domesticated than grasses, which would coincide with the number of larger number of TALEs in the cereals pathogens versus the grass pathogens. The second hypothesis is related to the way of infection. In the case of grasses pathogens, *X. translucens* penetrates directly to the vascular system of the host plant due to lesions caused by scything, mowing or grazing ruminants. Since bacteria have direct access to the xylem they may need fewer TALEs. On the other hand, the infection way of cereals pathogens is different and bacteria have to colonize the host plant after penetration into the plant tissue probably via stomata or hydathodes, thus requiring more virulence determinants such as T3Es. To back these results and to better understand the link between the number of TALEs (which can be obtained per southern blot, for example), pathovars and mode of infection, more strains of *X. translucens* need to be studied.

The structure of TALEs is known to be conserved among *Xanthomonas* species. They are composed of three domains with well-conserved N- and C-terminal domains and the central domain which is composed of nearly identical repeats (Guo et al., 2011). Interestingly, the N- and the C-terminal domains of *X. translucens* TALEs are distinct to those of the other TALES from clade-2 xanthomonads. And TALEs from *X. translucens* reveal unique features, such as a mix of 34- and 35-amino acids repeats and the presence of new hypervariable dipeptides. Until now, only one chimeric TALE with mixed repeats has been on *X. gardneri*. Based on these two structural characteristics, we speculate that TALEs in clade-1 xanthomonads were acquired independently of TALEs in clade-2 xanthomonads. This observation is linked to the presence of a distinct *hrp* cluster in *X. translucens* compared to clade-2 *Xanthomonas* species. These findings lead to an intriguing question about evolution and acquisition of pathogenicity factors in *Xanthomonas*.

Based on these data on *X. translucens* and also on our work on *X. cannabis* pv. *cannabis* published this years (Jacob et al., 2015; appendix A), we developed a model of a stepwise evolution of pathogenicity in *Xanthomonas*, consisting of the successive acquisition of genes for:

1) a type II secretion system and cell-wall degrading enzymes
2) *hrpG* and *hrpX* regulatory genes in the clade-2 species
3) *hrp* T3SS genes and core effector genes
4) accessory type III effector genes

This model fits to most clade-2 xanthomonads except for *X. cannabis* pv. *cannabis* strains which possess *hrpX* and *hrpG* but which do not have *hrp* T3SS genes and type III effectors. At least two explanations are possible. For instance, it is possible that the T3SS genes and core effector genes got lost in the *X. cannabis* pv. *cannabis* lineage after their acquisition. Or the second hypothesis is the independently and later acquisition of the *hrpG* and *hrpX* genes into the *X. cannabis* strains (appendix A).

For the case of clade-1 species, the evolution of pathogenicity determinants differs slightly from the model explained before. Clade-1 xanthomonads can be sub-divided into two groups: one group without a *hrp* cluster (*X. albilineans* and *X. sacchari*) and a second group with *hrp* genes that includes *X. translucens, X. theicola* and *X. hyacinthi.* Based on the genomic organization of the *hrp* cluster, group 2 seems to have acquired *hrpX* and *hrpG* together with the *hrp* cluster, followed by the acquistion of type III effectors. Sequencing of more *Xanthomonas* strains, including underrepresented species and pathovars and non-pathogenic environmental isolates, will help to better understand evolution of pathogenic bacteria and may point to new ways to control diseases caused by *Xanthomonas* on several host plants.

RFLP profiles based on *tale* genes from a few *X. translucens* pathogens on barley (pvs. *translucens* and *hordei*) indicated that some TALEs appear to be conserved. Other bands were common to the *hordei* strains but absent in the pathovar *translucens*. Since the analyzed strains originate from distant geographic sites, they are unlikely to be clonal. Hence, broad conservation of these *tale* genes

indicates that they play a major role in the pathogenicity during the interaction with certain host plants. In order to better understand the evolution of *tale* genes and their collective contribution to pathogenicity, it is necessary to obtain the complete sequences of *tale* genes from a larger set of representative *X. translucens* strains. Even if we succeeded to obtain the full sequence of the two *tale* genes from *X. translucens* pv. *cerealis* strain CFBP 2541 thanks to extremely high coverage of Illumina sequence reads, it is usually impossible to assemble these genes from high-throughput short-read sequencing technologies. For instance, none of the other Illumina-generated *X. translucens* genome sequences at GenBank contains the full set of *tale* genes. Quite the contrary, these genome sequences contain only *tale* gene fragments for the N- and C-terminal domains and a few short contigs with very few repeats of the central protein domain, which are on top of that at risk of being misassembled. Recently, a new sequencing technology generating long reads was introduced by Pacific Biosciences (PacBio). Indeed, this technology has been proven to lead to a complete *X. translucens* genome sequence, including the assembly of all eight *tale* genes from *X. translucens* pv. *undulosa* strain Xtu4699. Doubtlessly, PacBio sequencing will contribute to clarify whether or not the TALE repertoire of *X. translucens* strains influences the host range of a given pathovar.

A mutagenesis approach allowed us to identify a *tale* gene that contributes to pathogenicity in barley-pathogenic *X. translucens* strains. Knockout of this gene in one *X. translucens* pv. *translucens* strain and in one *X. translucens* pv. *hordei* strain caused reduced symptoms after leaf clipping of barley plants. With only 11.5 repeats,

this gene is the smallest natural *tale* gene found so far with a virulence function. In order to find the target gene(s) induced by this TALE, RNAseq transcriptomic experiments using two strains of *X. translucens* belonging to different pathovars are in progress and are expected to uncover strong candidates for the corresponding susceptibility gene. Candidate susceptibility genes can be easily tested using the strategy that we developed for the test of clade-III *SWEET* genes in rice (Streubel, Pesce et al., 2013).

After identification of susceptibility genes in barley plants, the model crop of our study, homologs of these genes could be searched in the genomes of other cereals, such as wheat, oat, … If these genes are present in another crop it would be interesting to test if they play a role as a susceptibility gene as well. First, quantitative RT-PCR could be used to see whether or not such gene is induced upon infection by *X. translucens*. Then, mutants in *tale* genes and/or artificial TALEs could be used to test whether induction of susceptibility gene homologs in other crops is required for disease development.

Identification of such a susceptibility gene in barley and other cereals would be an important step for plant breeding because it would allow screening different varieties of barley and other crops for polymorphisms within the EBE box in the promoter region of the target gene. Recently, a similar knowledge-based molecular screen uncovered a broad-spectrum resistance allele to bacterial blight from wild rice (Hutin et al., 2015). Identification of resistant varieties that cannot be infected by strains containing the major virulence TALE would help to breed new barley lines with a natural resistance against *X. translucens*.

As an alternative, a transgenic approach could be considered that would lead to modifications in the targeted EBE in the promoter region of the susceptibility gene. Genome editing based on TALENs (TALE-nucleases) or the CRISPR/Cas9 system has been established for cereals, including wheat and barley plants (Gurushidze et al., 2014). Consequently, small mutations within the EBE box of the susceptibility gene could be generated and modified plants could be evaluated for resistance against strains containing corresponding TALE(s). Such strategy has been successfully applied in rice for the well-known susceptibility gene *OsSWEET14* (Li et al., 2012).

Chapter IX

Conclusion

With this study, new insight was obtained into the pathogenicity of *X. translucens* on barley plants. We show the important role of the type III secretion system of *X. translucens* pv. *translucens* which is encoded by a noncanonical *hrp* gene cluster and which plays a key role in disease development and bacterial colonization of barley leaves. Then we identified one major virulence TALE, which is conserved among two pathovars of barley-pathogenic xanthomonads. And we propose a model for the evolution of pathogenicity in the genus *Xanthomonas* which points to a slightly distinct evolution of clade-2 xanthomonads and clade-1 xanthomonads such as *X. translucens*. The main perspectives of this study is the identification of the target susceptibility gene in barley for the major virulence TALE, which will help to breed barley varieties that are resistant against bacterial leaf streak caused by *X. translucens*.

References

Adhikari TB, Gurung S, Hansen JM, Bonman JM. 2012. Pathogenic and genetic diversity of *Xanthomonas translucens* pv. *undulosa* in North Dakota. *Phytopathology* **102**, 390-402.

Alfano JR, Collmer A. 1997. The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *J Bacteriol* **179**, 5655-5662.

Alizadeh A, Rahimian H. 1989. Bacterial leaf streak of *Gramineae* in Iran. *Bulletin OEPP/EPPO Bulletin* **19**: 113-117.

Ames NP, Rhymer CR. 2008. Issues surrounding health claims for barley. *J Nutr* **138**, 1237S-1243S.

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

Arlat M, Gough CL, Barber CE, Boucher C, Daniels MJ. 1991. *Xanthomonas campestris* contains a cluster of *hrp* genes related to the larger *hrp* cluster of *Pseudomonas solanacearum*. *Mol Plant Microbe Interact* **4**, 593-601.

Beattie GA. 2007. Plant-associated bacteria: survey, molecular phylogeny, genomics and recent advances. In: Plant-Associated Bacteria (ed. Samuel S. Gnanamanickam), pp. 1-55; Springer, Dordrecht, The Netherlands (ISBN 978-1-4020-4536-3).

Bingle LE, Bailey CM, Pallen MJ. 2008. Type VI secretion: a beginner's guide. *Curr Opin Microbiol* **11**, 3-8.

Blocker A, Komoriya K, Aizawa SI. 2003. Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc Natl Acad Sci U S A* **100**, 3027-3030.

Boch J. 2011. TALEs of genome targeting. *Nat Biotech* **29**, 135-136.

Boch J, Bonas U. 2010. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-436.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512.

Bogdanove AJ, Beer SV, Bonas U, Boucher CA, Collmer A, Coplin DL, Cornelis GR, Huang HC, Hutcheson SW, Panopoulos NJ, Van Gijsegem

F. 1996. Unified nomenclature for broadly conserved *hrp* genes of phytopathogenic bacteria. *Mol Microbiol* **20**, 681-683.

Bogdanove AJ, Schornack S, Lahaye T. 2010. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol* **13**, 394-401.

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

Bragard C, Singer E, Alizadeh A, Vauterin L, Maraite H, Swings J. 1997. *Xanthomonas translucens* from small grains: diversity and phytopathological relevance. *Phytopathology* **87**, 1111-1117.

Bull CT, De Boer SH,.Denny TP, Firrao G, Fischer-Le Saux M, Saddler GS,. Scortichini M, Stead DE, Takikawa Y. 2012. List of New Names of Plant Pathogenic Bacteria (2008-2010). *J. Plant Pathology* **94**, 21-27.

Buonaurio R. 2008. Infection and plant defense response during plant bacterial interaction. In: *Plant Microbe Interactions* (Editors: E. Ait Barka and C. Clément), pp 169-197, Research Signpost, Kerala, India (ISBN: 978-81-308-0212-1).

Burkinshaw BJ, Strynadka NC. 2014. Assembly and structure of the T3SS. *Biochim Biophys Acta* **1843**, 1649-1663.

Büttner D, He SY. 2009. Type III protein secretion in plant pathogenic bacteria. *Plant Physiol* **150**, 1656-1664.

Büttner D, Nennstiel D, Klüsener B, Bonas U. 2002. Functional analysis of HrpF, a putative type III translocon protein from *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* **184**, 2389-2398.

Cascales E. 2008. The type VI secretion toolkit. *EMBO Rep* **9**, 735-741.

Cesbron S, Pothier J, Gironde S, Jacques MA, Manceau C. 2014. Development of multilocus variable-number tandem repeat analysis (MLVA) for *Xanthomonas arboricola* pathovars. *J Microbiol Methods* **100,** 84-90.

Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu XQ, Guo WJ, Kim JG, Underwood W, Chaudhuri B, Chermak D, Antony G, White FF, Somerville SC, Mudgett MB, Frommer WB. 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527-532.

Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB. 2012. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**, 207-211.

Chu Z, Fu B, Yang H, Xu C, Li Z, Sanchez A, Park YJ, Bennetzen JL, Zhang Q, Wang S. 2006. Targeting *xa13*, a recessive gene for bacterial blight resistance in rice. *Theor Appl Genet* **112**, 455-461.

Cohn M, Bart RS, Shybut M, Dahlbeck D, Gomez M, Morbitzer R, Hou BH, Frommer WB, Lahaye T, Staskawicz BJ. 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.

Cornelis GR, Van Gijsegem F. 2000. Assembly and function of type III secretory systems. *Annu Rev Microbiol* **54**, 735-774.

Cui H, Wang Y, Xue L, Chu J, Yan C, Fu J, Chen M, Innes RW, Zhou JM. 2010. *Pseudomonas syringae* effector protein AvrB perturbs *Arabidopsis* hormone signaling by activating MAP kinase 4. *Cell Host Microbe* **7**, 164-175.

De Lange O, Schreiber T, Schandry N, Radeck J, Braun KH, Koszinowski J, Heuer H, Strauß A, 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 Phytol **199**, 773-786.

De Lange O, Wolf C, Dietze J, Elsaesser J, Morbitzer R, Lahaye T. 2014. Programmable DNA-binding proteins from *Burkholderia* provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res* **42**, 7436-7449.

De Lange O, Wolf C, Thiel P, Krüger J, Kleusch C, Kohlbacher O, Lahaye T. 2015. DNA-binding proteins from marine bacteria expand the known sequence diversity of TALE-like repeats. *Nucleic Acids Res*, *in press*.

Delepelaire P. 2004. Type I secretion in gram-negative bacteria. *Biochim Biophys Acta* **1694**, 149-161.

Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N. 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720-723.

Dowson WJ. 1939. On the systematic position and generic names of the Gram negative bacterial plant pathogens. *Zentralbl Bakteriol Parasitenk Infektionskr Hyg Abt II* **100**, 177-193.

Doyle EL, Booher NJ, Standage DS, Voytas DF, Brendel VP, Vandyk JK, Bogdanove AJ. 2012. TAL Effector-Nucleotide Targeter (TALE-NT)

2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40**, W117-W122.

Doyle EL, Hummel AW, Demorest ZL, Starker CG, Voytas DF, Bradley P, Bogdanove AJ. 2013. TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat –1. *PLoS One* **8**, e82120.

Duveiller E. 1994. Bacterial leaf streak or black chaff of cereals. *EPPO Bulletin* **24**, 135-158.

Duveiller E, Bragard C. 1992. Comparison of immunofluorescence and two assays for detection of *Xanthomonas campestris* pv. *undulosa* in seeds of small grains. *Plant Disease* **76**, 999-1003.

Duveiller E, Bragard C, Maraite H. 1991. Bacterial diseases of wheat in the warmer areas - reality or myth? In: *Proceedings of the Wheat for Nontraditional Warm Areas International Conference* (ed. D. Saunder), pp. 189-202; UNDP/CIMMYT, El Batan, Mexico.

Duveiller E, van Ginkel M, Thijssen M. 1992. Genetic analysis of resistance to bacterial leaf streak caused by *Xanthomonas campestri*s pv. *undulosa* in bread wheat. *Euphytica* **66**, 35-43.

Duveiller E, Fucikovsky L, Rudolph K, eds. 1997. The Bacterial Diseases of Wheat. Concepts and Methods of Disease Management. Mexico, D.F.: CIMMYT (ISBN 968-6923-73-X).

Duveiller E. 1989. Research on '*Xanthomonas translucens*' of wheat and triticale at CIMMYT. *EPPO Bulletin* **19**, 97-103.

Flor HH. 1942. Inheritance of pathogenicity in *Melamspora lini*. *Phytopathology* **32**, 653-669.

Forster RL, Mihuta-Grimm L, Schaad NW. 1986. Black chaff of wheat and barley. *Current Information Series* No. 784, Idaho Agricultural Extension Service, Moscow, Idaho, U.S.A.

Giblot-Ducray D, Marefat A, Gillings MR, Parkinson NM, Bowman JP, Ophel-Keller K, Taylor C, Facelli E, Scott ES. 2009. Proposal of *Xanthomonas translucens* pv. *pistaciae* pv. nov., pathogenic to pistachio (*Pistacia vera*). *Syst Appl Microbiol* **32**, 549-557.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh

A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92-100.

Gonzalez C, Szurek B, Manceau C, Mathieu T, Séré Y, Verdier V. 2007. Molecular and pathotypic characterization of new *Xanthomonas oryzae* strains from West Africa. *Mol Plant Microbe Interact* **20**, 534-546.

Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL. 2006. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu Rev Microbiol* **60**, 425-449.

Grau J, Boch J, Posch S, 2013. TALENoffer: genome-wide TALEN off-target prediction. *Bioinformatics* **29**, 2931-2932.

Groisman EA, Ochman H. 1993. Cognate gene clusters govern invasion of host epithelial cells by S*almonella typhimurium* and *Shigella flexneri*. *EMBO J* **12**, 3779-3787.

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

Gürlebeck D, Thieme F, Bonas U. 2006. Type III effector proteins from the plant pathogen *Xanthomonas* and their role in the interaction with the host plant. *J Plant Physiol* **163**, 233-255.

Gurushidze M, Hensel G, Hiekel S, Schedel S, Valkov V, Kumlehn J. 2014. True-breeding targeted gene knock-out in barley using designer TALE-nuclease in haploid cells. *PLoS One* **9**, e92046.

Hahn J, Luttinger A, Dubnau D. 1996. Regulatory inputs for the synthesis of ComK, the competence transcription factor of *Bacillus subtilis*. *Mol Microbiol* **21**, 763-775.

Hancock, JF. 2004. Plant Evolution and the Origin of Crop Species. CABI Publishing. ISBN 0- 85199-685-X.

Hauben L, Vauterin L, Swings J, Moore ER. 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol* **47**, 328-335.

Hauck P, Thilmony R, He SY. 2003. A *Pseudomonas syringae* type III effector suppresses cell wall-based extracellular defense in susceptible *Arabidopsis* plants. *Proc Natl Acad Sci U S A* **100**, 8577-8582.

Hayward AC. 1993. The host of *Xanthomonas*. In: *Xanthomonas* (eds. J.G. Swings and E.L. Civerolo), pp. 51-54; Chapman & Hall, London, United Kingdom.

Hockett. EA, Nilan RA. 1985. Genetics. In: Barley (ed. D.C. Rasmusson), pp. 187-230; American Society of Agronomy-Crop Science Society of America-Soil Science Society of America, Madison, Wisconsin, U.S.A.

Hu Y, Zhang J, Hongge J, sosso D, Li T, Frommer WB, Yang B, White FF, Wang N, Jones JB. 2014. *Lateral organ boundaries 1* is a disease susceptibility gene for citrus bacterial canker disease. *Proc Natl Acad Sci U S A* **111**, E521-E529.

Hutin M, Sabot F, Ghesquière A, Koebnik R, Szurek B. 2015. A knowledge-based molecular screen uncovers a broad spectrum *OsSWEET14* resistance allele to bacterial blight from wild rice. *Plant J*, *in press*.

International Barley Genome Sequencing Consortium (IBGSC). 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**: 711-716.

Iyer AS, Mccouch SR. 2004. The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Mol Plant Microbe Interact* **17**, 1348-1354.

Jacobs JM, Pesce C, Lefeuvre P, Koebnik R. 2015. Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*. *Front Plant Sci* **6**, 431.

Jankele R, Svoboda P. 2014. TAL effectors: tools for DNA Targeting. *Brief Funct Genomics* **13**, 409-419.

Jenkins CL, Starr MP. 1985. Formation of halogenated aryl-polyene (xanthomonadin) pigments by the type and other yellow-pigmented strains of *Xanthomonas maltophilia*. *Ann Inst Pasteur Microbiol* **136B**, 257-264.

Jones LR, Johnson AG, Reddy CS. 1917. Bacterial blight of barley. *J Agric Res* **11**, 625-643.

Jones JD, Dangl JL. 2006. The plant immune system. *Nature* **444**, 323-329.

Juillerat A, Bertonati C, Dubois G, Guyot V, Thomas S, Valton J, Beurdeley M, Silva GH, Daboussi F, Duchateau P. 2014. *BurrH*: a new modular DNA binding protein for genome engineering. *Sci Rep* **4**, 3831.

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

Koebnik R, Krüger A, Thieme F, Urban A, Bonas U. 2006. Specific binding of the *Xanthomonas campestris* pv. *vesicatoria* AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J Bacteriol* **188**, 7652-7660.

Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**, 2567-2572.

Korobko AP, Wondimagegne E, Anisimoff BV. 1985. Bacterial stripe and black chaff of wheat in Ethiopia. Scientific Phytopathological Laboratory, Ambo, Ethiopia.

Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, Sukhan A, Galán JE, Aizawa SI. 1998. Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602-605.

Lahaye T, Bonas U. 2001. Molecular secrets of bacterial type III effector proteins. *Trends Plant Sci* **6**, 479-485.

Leo JC, Grin I, Linke D. 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos Trans R Soc Lond B Biol Sci* **367**, 1088-1101.

Leyns F, De Cleene M, Swings J, De Ley J. 1984. The host range of the genus *Xanthomonas*. *Bot Rev* **50**, 308-356.

Li L, Atef A, Piatek A, Ali Z, Piatek M, Aouida M, Sharakuu A, Mahjoub A, Wang G, Khan S, Fedoroff NV, Zhu JK, Mahfouz MM. 2013. Characterization and DNA-binding specificities of *Ralstonia* TAL-like effectors. *Mol Plant* **6**, 1318-1330.

Li RF, Lu GT, Li L, Su HZ, Feng GF, Chen Y, He YQ, Jiang BL, Tang DJ, Tang JL. 2014. Identification of a putative cognate sensor kinase for the two-component response regulator HrpG, a key regulator controlling the expression of the *hrp* genes in *Xanthomonas campestris* pv. *campestris*. *Environ Microbiol* **16**, 2053-2071.

Li T, Liu B, Spalding MH, Weeks DP, Yang B. 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotech* **30**, 390-392.

Li T, Liu B, Chen CY, Yang B. 2014. TALEN utilization in rice genome modifications. *Methods* **69**, 9-16.

Li YR, Che YZ, Zou HS, Cui YP, Guo W, Zou LF, Biddle EM, Yang CH, Chen GY. 2011. Hpa2 required by HrpF to translocate *Xanthomonas oryzae* transcriptional activator-like effectors into rice for pathogenicity. *Appl Environ Microbiol* **77**, 3809-3818.

Lindgren PB, Peet RC, Panopoulos NJ. 1986. Gene cluster of *Pseudomonas syringae* pv. "*phaseolicola*" controls pathogenicity of bean plants and hypersensitivity of nonhost plants. *J Bacteriol* **168**, 512-522.

Lory S. 1992. Determinants of extracellular protein secretion in gram-negative bacteria. *J Bacteriol* **174**, 3423-3428.

Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716-719.

Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA, Toth I, Salmond G, Foster GD. 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* **13**, 614-629.

Marefat A, Ophel-Keller K, Scott ES, Sedgley M. 2006a. The use of ARMS PCR in detection and identification of xanthomonads associated with pistachio dieback in Australia. *Eur J Plant Pathol* **116**, 57-68.

Marefat A, Scott ES, Ophel-Keller K, Sedgley M. 2006b. Genetic, phenotypic and pathogenic diversity among xanthomonads isolated from pistachio (*Pistacia vera*) in Australia. *Plant Pathol* **55**, 639-649.

Masuch G, Schoene K, Paul VH. 1989. Histological investigations on the pathogenesis of *Xanthomonas campestris* pv. *graminis* to *Lolium multiflorum*. *EPPO Bulletin* **19**, 73-80.

Meckler JF, Bhakta MS, Kim MS, Ovadia R, Habrian CH, Zykovich A, Yu A, Lockwood SH, Morbitzer R, Elsäesser J, Lahaye T, Segal DJ, Baldwin EP. 2013. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* **41**, 4118-4128.

Mehta YR. 1990. Management of *Xanthomonas translucens* pv. *undulosa* and *hordei* through cereal seed testing. *Seed Sci Technol* **18**, 467.

Meyer D, Cunnac S, Guéneron M, Declercq C, Van Gijsegem F, Lauber E, Boucher C, Arlat M. 2006. PopF1 and PopF2, two proteins secreted by the type III protein secretion system of *Ralstonia solanacearum*, are translocators belonging to the HrpF/NopX family. *J Bacteriol* **188**, 4903-4917.

Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501.

Mulatu B, Grando S. 2006. Barley Research and Development in Ethiopia. *Proceedings of the 2nd National Barley Research and Development Review Workshop*. *Holetta Agricultural Research Center; Holetta, Ethiopia*. pp. 187-204.

Muñoz Bodnar A, Bernal A, Szurek B, López C. 2013. Tell me a tale of TALEs. *Mol Biotechnol* **53**, 228-235.

Naushad S, Adeolu M, Wong S, Sohail M, Schellhorn H, Gupta R. 2015. A phylogenomic and molecular marker based taxonomic framework for the order *Xanthomonadales*: proposal to transfer the families *Algiphilaceae* and *Solimonadaceae* to the order *Nevskiales* ord. nov. and to create a new family within the order *Xanthomonadales*, the family *Rhodanobacteraceae* fam. nov., containing the genus *Rhodanobacter* and its closest relatives. *Antonie van Leeuwenhoek* **107**, 467-485.

Ngoc LB, Verniere C, Vital K, Guerin F, Gagnevin L, Brisse S, Ah-You N, Pruvost O. 2009. Development of 14 minisatellite markers for the citrus canker bacterium, *Xanthomonas citri* pv. *citri*. *Mol Ecol Resour* **9,** 125-127.

Niepold F, Anderson D, Mills D. 1985. Cloning determinants of pathogenesis from *Pseudomonas syringae* pathovar *syringae*. *Proc Natl Acad Sci U S A* **82**, 406-410.

Noël L, Thieme F, Gäbler J, Büttner D, Bonas U. 2003. XopC and XopJ, two novel type III effector proteins from *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* **185**, 7092-7102.

Oerke EC. 2006. Crop losses to pests. *J Agric Sci* **144**, 31-43

Parkinson N, Aritua V, Heeney J, Cowie C, Bew J, Stead D. 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int J Syst Evol Microbiol* **57**, 2881-2887.

Parkinson N, Cowie C, Heeney J, Stead D. 2009. Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int J Syst Evol Microbiol* **59**, 264-274.

Pereira ALG, Campacci CA, Zagatto AG, Watanabe K. 1983. Occurrence of *Xanthomonas campestris* pv. *cerealis* on Wheat (*Triticum aestrivum*) in Parana State. *Biologico Sao Paulo* **49,** 9-13.

Pérez-Quintero AL, Rodriguez-R LM, Dereeper A, López C, Koebnik R, Szurek B, Cunnac S. 2013. An improved method for TAL effectors

DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS One* **8**, e68464.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques MA, Lauber E, Manceau C, Mangenot S, Poussier S, Segurens B, Szurek B, Verdier V, Arlat M, Rott P. 2009. The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* **10**, 616.

Pieretti I, Cociancich S, Bolot S, Carrère S, Morisset A, Rott P, Royer M. 2015. Full genome sequence analysis of two isolates reveals a novel *Xanthomonas* species close to the sugarcane pathogen *Xanthomonas albilineans*. Genes (Basel) **6**, 714-733.

Popp J, Pető K, Nagy J. 2013. Pesticide productivity and food security. A review. *Agron Sustain Dev* **33**, 243-255.

Poulin L, Grygiel P, Magne M, Gagnevin L, Rodriguez-R LM, Forero Serna N, Zhao S, El Rafii M, Dao S, Tekete C, Wonni I, Koita O, Pruvost O, Verdier V, Vernière C, Koebnik R. 2015. New multilocus variable-number tandem-repeat analysis tool for surveillance and local epidemiology of bacterial leaf blight and bacterial leaf streak of rice caused by *Xanthomonas oryzae*. *Appl Environ Microbiol* **81**, 688-698.

Pritchard L, Birch PR. 2014. The zigzag model of plant-microbe interactions: is it time to move on? *Mol Plant Pathol* **15**, 865-870.

Pruvost O, Magne M, Boyer K, Leduc A, Tourterel C, Drevet C, Ravigné V, Gagnevin L, Guérin F, Chiroleu F, Koebnik R, Verdier V, Vernière C. 2014. A MLVA genotyping scheme for global surveillance of the citrus pathogen *Xanthomonas citri* pv. *citri* suggests a worldwide geographical expansion of a single genetic lineage. *PLoS One* **9,** e98129.

Rademaker JLW, Norman DJ, Forster RL, Louws FJ, Schultz MH, De Bruijn FJ. 2006. Classification and identification of *Xanthomonas translucens* isolates, including those pathogenic to ornamental asparagus. *Phytopathology* **96**, 876-884.

Rafiqi M, Bernoux M, Ellis JG, Dodds PN. 2009. In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Semin Cell Dev Biol* **20**, 1017-1024.

Rairdan G, Moffett P. 2007. Brothers in arms? Common and contrasting themes in pathogen perception by plant NB-LRR and animal NACHT-LRR proteins. *Microbes Infect* **9**, 677-686.

Reddy, C.S., Godkin, J. & Johnson, A.G. 1924. Bacterial blight of rye. *J Agric Res* **28**, 1039-1040.

Römer P, Hahn S, Jordan T, Strauß T, Bonas U, Lahaye T. 2007. Plant pathogen recognition mediated by promoter activation of the pepper *Bs3* resistance gene. *Science* **318**, 645-648.

Römer P, Recht S, Strauß T, Elsaesser J, Schornack S, Boch J, Wang S, Lahaye T. 2010. Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, *Xanthomonas oryzae* pv. *oryzae*. *New Phytol* **187**, 1048-1057.

Rosqvist R, Forsberg A, Wolf-Watz H. 1991. Intracellular targeting of the *Yersinia* YopE cytotoxin in mammalian cells induces actin microfilament disruption. *Infect Immun* **59**, 4562-4569.

Rossier O, Van Den Ackerveken G, Bonas U. 2000. HrpB2 and HrpF from *Xanthomonas* are type III-secreted proteins and essential for pathogenicity and recognition by the host plant. *Mol Microbiol* **38**, 828-838.

Ryan RP, Vorhölter FJ, Potnis N, Jones JB, Van Sluys MA, Bogdanove AJ, Dow JM. 2011. Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat Rev Microbiol* **9**, 344-355.

Schaad NW, Vidaver AK, Lacy GH, Rudolph K, Jones JB. 2000. Evaluation of proposed amended names of several pseudomonads and xanthomonads and recommendations. *Phytopathology* **90**, 208-213.

Scholze H, Boch J. 2010. TAL effector-DNA specificity. *Virulence* **1**, 428-432.

Shan Q, Zhang Y, Chen K, Zhang K, Gao C. 2015. Creation of fragrant rice by targeted knockout of the *OsBADH2* gene using TALEN technology. *Plant Biotechnol J* **13**, 791-800.

Shekhawat GS, Patel PN. 1978. Histology of barley plant and rice leaf infected with *Xanthomonas translucens* f. sp. *hordei*. *J Phytopathol* **93**, 105-112.

Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LM, Araya JE, Baia GS, Baptista CS, Barros MH, Bonaccorsi ED, Bordin S, Bové JM, Briones MR, Bueno MR, Camargo AA, Camargo LE, Carraro DM, Carrer H, Colauto NB, Colombo C, Costa FF, Costa MC, Costa-Neto CM, Coutinho LL, Cristofani M, Dias-Neto E, Docena C, El-Dorry H, Facincani AP, Ferreira AJ, Ferreira VC, Ferro JA, Fraga JS, França SC, Franco MC, Frohme M, Furlan LR, Garnier M, Goldman

GH, Goldman MH, Gomes SL, Gruber A, Ho PL, Hoheisel JD, Junqueira ML, Kemper EL, Kitajima JP, Krieger JE, Kuramae EE, Laigret F, Lambais MR, Leite LC, Lemos EG, Lemos MV, Lopes SA, Lopes CR, Machado JA, Machado MA, Madeira AM, Madeira HM, Marino CL, Marques MV, Martins EA, Martins EM, Matsukuma AY, Menck CF, Miracca EC, Miyaki CY, Monteriro-Vitorello CB, Moon DH, Nagai MA, Nascimento AL, Netto LE, Nhani A Jr, Nobrega FG, Nunes LR, Oliveira MA, de Oliveira MC, de Oliveira RC, Palmieri DA, Paris A, Peixoto BR, Pereira GA, Pereira HA Jr, Pesquero JB, Quaggio RB, Roberto PG, Rodrigues V, de M Rosa AJ, de Rosa VE Jr, de Sá RG, Santelli RV, Sawasaki HE, da Silva AC, da Silva AM, da Silva FR, da Silva WA Jr, da Silveira JF, Silvestri ML, Siqueira WJ, de Souza AA, de Souza AP, Terenzi MF, Truffi D, Tsai SM, Tsuhako MH, Vallada H, Van Sluys MA, Verjovski-Almeida S, Vettore AL, Zago MA, Zatz M, Meidanis J, Setubal JC. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151-157.

Smith EF, Jones LR, Reddy CS. 1919. The black chaff of wheat. *Science* **50**, 48.

Stella S, Molina R, Bertonatti C, Juillerrat A, Montoya G. 2014. Expression, purification, crystallization and preliminary X-ray diffraction analysis of the novel modular DNA-binding protein BurrH in its apo form and in complex with its target DNA. *Acta Crystallogr F Struct Biol Commun* **70**, 87-91.

Studholme DJ, Wasukira A, Paszkiewicz K, Aritua V, Thwaites R, Smith J, Grant M. 2011. Draft genome sequences of *Xanthomonas sacchari* and two banana-associated *Xanthomonads* reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel)* **2**, 1050-1065.

Sugio A, Yang B, Zhu T, White FF. 2007. Two type III effector genes of *Xanthomonas oryzae* pv. *oryzae* control the induction of the host genes *OsTFIIAγ1* and *OsTFX1* during bacterial blight of rice. *Proc Natl Acad Sci U S A* **104**, 10720-10725.

Szurek B, Marois E, Bonas U, Van Den Ackerveken G. 2001. Eukaryotic features of the *Xanthomonas* type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper. *Plant J* **26**, 523-534.

Talbot NJ. 2010. Cell biology: Raiding the sweet shop. *Nature* **468**, 510-511.

Tang X, Xiao Y, Zhou JM. 2006. Regulation of the type III secretion system in phytopathogenic bacteria. *Mol Plant Microbe Interact* **19**, 1159-1166.

Thieme F, Koebnik R, Bekel T*, et al.*, 2005. Insights into Genome Plasticity and Pathogenicity of the Plant Pathogenic Bacterium Xanthomonas campestris pv. vesicatoria Revealed by the Complete Genome Sequence. *J Bacteriol* **187**, 7254-66.

Tian D, Wang J, Zeng X, Gu K, Qiu C, Yang X, Zhou Z, Goh M, Luo Y, Murata-Hori M, White FF, Yin Z. 2014. The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *Plant Cell* **26**, 497-515.

Triplett LR, Verdier V, Campillo T, Van Malderghem C, Cleenwerck I, Maes M, Deblais L, Corral R, Koita O, Cottyn B, Leach JE. 2015. Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie van Leeuwenhoek* **107**, 869-881.

Trujillo CA, Arias-Rojas N, Poulin L, Medina CA, Tapiero A, Restrepo S, Koebnik R, Bernal AJ. 2014. Population typing of the causal agent of cassava bacterial blight in the Eastern Plains of Colombia using two types of molecular markers. *BMC Microbio*l **14**, 161.

Tseng TT, Tyler BM, Setubal JC. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* **9**, S2.

Upadhyay SK, Kumar J, Alok A, Tuli R. 2013. RNA-guided genome editing for target gene mutations in wheat. *G3 (Bethesda)* **3,** 2233-2238.

Vauterin L, Hoste B, Kersters K, Swings J. 1995. Reclassification of *Xanthomonas*. *Int J Syst Bacteriol* **45**, 472-489.

Vauterin L, Swings J, Kersters K. 1991. Grouping of *Xanthomonas campestris* pathovars by SDS-PAGE of proteins. *J Gen Microbiol* **137**, 1677-1687.

Verdier V, Restrepo S, Mosquera G, Jorge V, Lopez C. 2004. Recent progress in the characterization of molecular determinants in the *Xanthomonas axonopodis* pv. *manihotis*–cassava interaction. *Plant Mol Biol* **56**, 573-584.

Vivian A, Arnold DL. 2000. Review bacterial effector genes and their role in host-pathogen interactions. *J Plant Pathol* **82**, 163-178.

Wallin JR. 1946. Parasitism of *Xanthomonas translucens* (J.J. and R.) Dowson on grasses and cereals. *Iowa St Coll J Sci* **20**, 171-193.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* **32**, 947-951.

Weber E, Ojanen-Reuhs T, Huguet E, Hause G, Romantschuk M, Korhonen TK, Bonas U, Koebnik R. 2005. The type III-dependent Hrp pilus is required for productive interaction of *Xanthomonas campestris* pv. *vesicatoria* with pepper host plants. *J Bacteriol* **187**, 2458-2468.

Wengelnik K, Rossier O, Bonas U. 1999. Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. *vesicatoria* result in constitutive expression of all *hrp* genes. *J Bacteriol* **181**, 6828-6831.

Wengelnik K, Van den Ackerveken G, Bonas U. 1996. HrpG, a key *hrp* regulatory protein of *Xanthomonas campestris* pv. *vesicatoria* is homologous to two-component response regulators. *Mol Plant Microbe Interact* **9**, 704-712

White FF, Potnis N, Jones JB, Koebnik R. 2009. The type III effectors of *Xanthomonas*. *Mol Plant Pathol* **10**, 749-766.

Wichmann F, Vorhölter FJ, Hersemann L, Widmer F, Blom J, Niehaus K, Reinhard S, Conradin C, Kölliker R. 2013. The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol Plant Pathol* **14**, 576-588.

Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF Jr, Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D Sr, Chapman P, Clendenning J, Deatherage G, Gillet W, Grant C, Kutyavin T, Levy R, Li MJ, McClelland E, Palmieri A, Raymond C, Rouse G, Saenphimmachak C, Wu Z, Romero P, Gordon D, Zhang S, Yoo H, Tao Y, Biddle P, Jung M, Krespan W, Perry M, Gordon-Kamm B, Liao L, Kim S, Hendrick C, Zhao ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**, 2317-2323.

Wright DA, Li T, Yang B, Spalding MH. 2014. TALEN-mediated genome editing: prospects and perspectives. *Biochem J* **462**, 15-24.

Yang B, Sugio A, White FF. 2005. Avoidance of host recognition by alterations in the repetitive and C-terminal regions of AvrXa7, a type III effector of *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Microbe Interact* **18**, 142-149.

Yang B, Sugio A, White FF. 2006. *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proc Natl Acad Sci U S A* **103**, 10503-10508.

Yang Y, Gabriel DW. 1995. Intragenic recombination of a single plant pathogen gene provides a mechanism for the evolution of new host specificities. *J Bacteriol* **177**, 4963-4968.

Young JM, Park DC, Shearman HM, Fargier E. 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst Appl Microbiol* **31**, 366-377.

Young JM, Saddler GS, Takikawa Y, De Boer SH, Vauterin L, Gardan L, Gvozdyak RI, Stead, DE. 1996. Names of plant pathogenic bacteria 1864-1995. *Rev Plant Pathol* **75**, 721-763.

Yu Y, Streubel J, Balzergue S, Champion A, Boch J, Koebnik R, Feng J, Verdier V, Szurek B. 2011. Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice *Nodulin-3 Os11N3* gene. *Mol Plant Microbe Interact* **24**,1102-1113.

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

Zillinsky FJ, Borlaug NE. 1971. Progress in developing triticale as an economic crop. *Int Maize Wheat Improv Cen Res Bull* **17**, 18-21.

APPENDIX A

Comparative genomics of a cannabis pathogen reveals insight into the evolution of pathogenicity in *Xanthomonas*

Jonathan Jacobs[1], Céline Pesce[1,2], Pierre Lefeuvre[3], and Ralf Koebnik[1*]

[1] UMR Interactions – Plantes – Microorganismes – Environnement (IPME), IRD-Cirad-UM, Montpellier, France [2] Earth and Life Institute, Applied Microbiology Phytopathology, Université Catholique de Louvain, Louvain-la-Neuve, Belgium [3] Cirad, UMR Peuplements Végétaux et Bioagresseurs en Milieu Tropical (PVBMT), Pôle de Protection des Plantes, Saint-Pierre, Ile de la Réunion, France

Running title: Evolution of pathogenicity in *Xanthomonas*

Keywords: comparative genomics, *Xanthomonas*, hemp, cell-wall degrading enzymes, type II secretion system, type III secretion system, *hrp genes*, PIP box

I contributed to this work by constructing bacterial strains and performing inoculation and motility assays.

Abstract

Pathogenic bacteria in the genus *Xanthomonas* cause diseases on over 350 plant species, including cannabis (*Cannabis sativa* L.). Because of regulatory limitations, the biology of the *Xanthomonas*-cannabis pathosystem remains largely unexplored. To gain insight into the evolution of *Xanthomonas* strains pathogenic to cannabis, we sequenced the genomes of two geographically distinct *Xanthomonas* strains, NCPPB 3753 and NCPPB 2877, which were previously isolated from symptomatic plant tissue in Japan and Romania. Comparative multilocus sequence analysis of housekeeping genes revealed that they belong to Group 2, which comprises most of the described species of *Xanthomonas*. Interestingly, both strains lack the Hrp Type III secretion system and do not contain any of the known Type III effectors. Yet their genomes notably encode two key Hrp pathogenicity regulators HrpG and HrpX, and *hrpG* and *hrpX* are in the same genetic organization as in the other Group 2 xanthomonads. Promoter prediction of HrpX-regulated genes suggests the induction of an aminopeptidase, a lipase and two polygalacturonases upon plant colonization, similar to other plant-pathogenic xanthomonads. Genome analysis of the distantly related *Xanthomonas maliensis* strain 97M, which was isolated from a rice leaf in Mali, similarly demonstrated the presence of HrpG, HrpX and a HrpX-regulated polygalacturonase, and the absence of the Hrp Type III secretion system and known Type III effectors. Given the observation that some *Xanthomonas* strains across distinct taxa do not contain *hrpG* and *hrpX*, we speculate a stepwise evolution of pathogenicity, which involves (i) acquisition of key regulatory genes and cell wall-

degrading enzymes, followed by (ii) acquisition of the Hrp type III secretion system, which is ultimately accompanied by (iii) successive acquisition of type III effectors.

Introduction

Plant pathogenic bacteria in the genus *Xanthomonas* collectively cause major losses worldwide on over 350 plant species, including crops such as banana, tomato, pepper, sugar cane and many cereals. Over 20 *Xanthomonas* species are divided into two main phylogenetic groups based on 16S rDNA and *gyrB* sequence analysis (Hauben et al., 1997; Parkinson et al., 2007) and subdivided into pathovars loosely corresponding to host specificity. Group 1, also known as the early branching group, comprises highly diverse *Xanthomonas* species including important sugarcane and cereal pathogens (e.g. *Xanthomonas sacchari*, *Xanthomonas albilineans* and *Xanthomonas translucens*). Group 2, the largest and best-described group, includes species such as *Xanthomonas oryzae*, *Xanthomonas citri*, *Xanthomonas vasicola*, *Xanthomonas euvesicatoria*, *Xanthomonas axonopodis* and *Xanthomonas campestris* (Hauben et al., 1997; Parkinson et al., 2007). This diverse genus of bacteria infects and associates with many plant hosts, but individual strains typically posses very restricted host ranges limited to a single genus.

Xanthomonas spp. employ a suite of virulence factors to colonize plant tissue, including adhesins, cell wall-degrading enzymes, extracellular polysaccharide and protein secretion systems (Büttner and Bonas, 2010). The Hrp (hypersensitive response and pathogenicity) Type III secretion system (T3SS) is a major virulence trait found in most pathogenic *Xanthomonas* spp. and serves as a molecular syringe to deliver effector proteins into host cells to suppress defenses and modulate plant physiology to promote pathogen growth (White et al., 2009). Plants also evolved resistance proteins

that recognize pathogen avirulence effectors and inhibit infection often via a hypersensitivity response (HR), a form of programmed cell death (Bent and Mackey, 2007). A majority of sequenced pathogenic *Xanthomonas* strains have limited host ranges likely due to the plant recognition of Type III (T3)-secreted avirulence effectors (White et al., 2009). In *Xanthomonas* spp., HrpX, an AraC-type regulator, is the transcriptional activator of the genes encoding the T3SS and many of its associated effectors (Koebnik et al., 2006; Tang et al., 2006). HrpG, an OmpR-family and major pathogenicity regulator, positively regulates expression of *hrpX* (Tang et al., 2006). Mutant strains lacking either *hrpX* and *hrpG* are unable to activate expression of the T3SS and thus are non pathogenic (Wengelnik et al., 1996; Tang et al., 2006; Mole et al., 2007). The importance of the T3SS and many T3-secreted effectors during infection is heavily studied, but the evolutionary history of the acquisition of genes encoding the T3SS, associated T3-secreted effectors and regulators, HrpX and HrpG, remains unclear.

Hemp or cannabis (*Cannabis sativa* L.) is a major, global cash crop with many applications such as seed for human consumption, oil, fiber for clothing or ropes, pulp for paper, plastic and composite material (www.hemp.com). Since 2010 worldwide hemp production has dramatically increased, and recent surges of hemp production in the United States, China, Australia, Canada and many other countries have made hemp a multi-million dollar industry (www.hemp.com, FAO 2013). A draft genome is now available for *C. sativa* cv. Purple Kush (van Bakel et al., 2011), potentially providing a base for molecular and evolutionary understanding of this plant species. Hemp

plant production is limited by bacteria, fungi, nematodes and viruses (McPartland et al., 2000), but because of regulatory constraints, little is known about hemp diseases such as bacterial leaf spot of cannabis caused by *Xanthomonas* species.

Symptoms associated with *Xanthomonas* bacterial leaf spot include water-soaking lesions followed by necrosis accompanied by a yellow halo (Severin, 1978; Netsu et al., 2014). The host range of these *Xanthomonas* strains appears to be quite large unlike most xanthomonads (Severin, 1978; Netsu et al., 2014). Under laboratory conditions, these bacteria caused symptoms on a wide range of plants including cannabis, tomato, mulberry, geranium and *Ficus erecta* (Severin, 1978; Netsu et al., 2014). These strains further trigger an HR on tobacco, but do not elicit any response after inoculation on common bean (Severin, 1978; Netsu et al., 2014). The factors that contribute to pathogenicity and host range of cannabis-infecting *Xanthomonas* are unknown.

To gain insight into the evolution and pathogenicity of bacterial pathogens of cannabis, we sequenced two geographically distinct *Xanthomonas* strains, NCPPB 3753 and 2877, which were previously isolated from symptomatic hemp leaf tissue from Japan and Romania, respectively (Severin, 1978; Netsu et al., 2014). We tested their ability to infect barley, a previously unreported, compatible monocot host. We determined with comparative whole genome analysis based on average nucleotide identity (ANI) and multilocus sequence analysis (MLSA) the relationship of these cannabis strains to each other and other xanthomonads. We provide evidence that NCPPB 3753 and NCPPB 2877 form a unique species in the genus *Xanthomonas* herein

called *Xanthomonas cannabis*. We further describe likely virulence traits encoded by their genomes. Most notably these strains lack a Hrp T3SS but posses the major *hrp* virulence regulators HrpX and HrpG. Based on our comparative genomic analysis in *X. cannabis*, we provide a putative model for acquisition of the T3SS, T3-secreted effectors and the *hrp* regulators in *Xanthomonas* spp.

Results / Methods / Discussion

Phenotypic Evaluation

Two representative strains of *X. cannabis* (also known as *Xanthomonas campestris* pv. *cannabis*), isolated from symptomatic hemp leaves (*C. sativa* L.), were chosen for genome sequencing. Type strain NCPPB 2877 was isolated by I. Sandru at the Lovrin station in the Timiș județ (Romania) in 1974 (Severin, 1978), and strain NCPPB 3753 was isolated by Y. Takiwawa in the Kanuma region of Tochigi Prefecture (Japan) in 1982 (isolate SUPP546; Netsu et al., 2014). *X. cannabis* strains were previously reported to cause disease on many dicot host plants. To determine if *X. cannabis* could infect a monocot host, we inoculated barley (*Hordeum vulgare* L. cv. Morex) leaves with the cannabis strains by infiltration. Overnight cultures grown using PSA (Tsuchiya et al., 1982) or NB medium (Sigma-Aldrich, USA) were pelleted and resuspended in water. Plant leaves were infiltrated by a needleless syringe with a water-bacterial suspension or water as a control. Leaves developed necrosis around the zone of infiltration followed by leaf yellowing (Figure 1A, data only shown for strain NCPPB 3753). These symptoms closely resembled the leaf spot symptoms on *C. sativa* observed by previous characterizations

(Netsu et al., 2014; Severin, 1978). Similar symptoms were observed with lower inoculum (OD600 = 0.05 and 0.1). Tomato is a compatible host for *X. cannabis* (Severin, 1978), and therefore we decided to test *X. cannabis* virulence of pepper, another solanaceous plant. Pepper leaves were infiltrated with strain NCPPB 3753 as with barley. Pepper plants displayed water-soaking lesions 48 hours post inoculation (Figure 1B). Both strains elicited an HR when inoculated on tobacco (Figure 1C), but the nature of this HR remains to be determined.

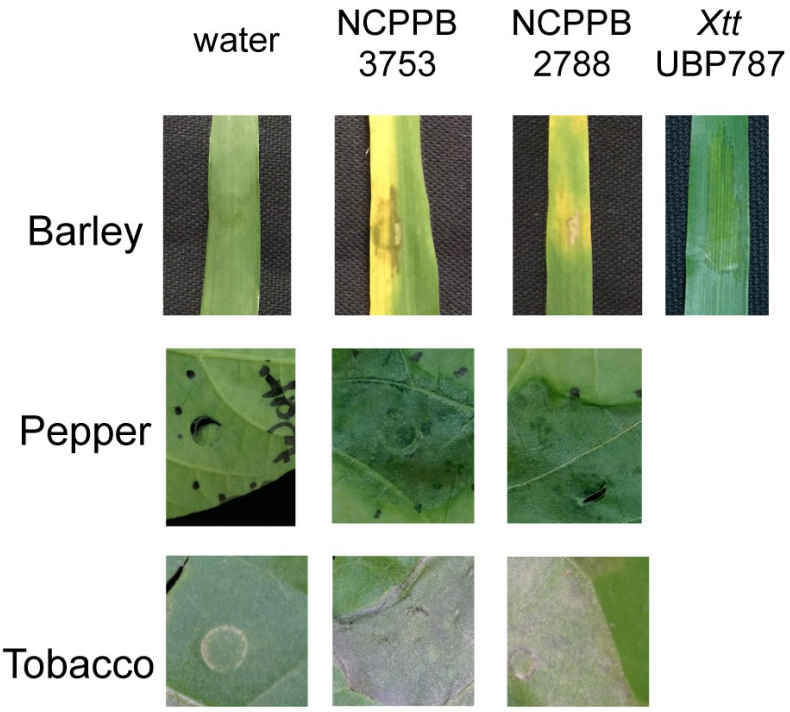


Figure 1. Phenotypic analysis of *X. cannabis* on different plants.
Symptom development and HR were evaluated on three plants: A) barley, B) pepper and C) tobacco. Leaves were inoculated by needleless infiltration with water-bacterial suspensions of *X. cannabis* NCPPB 3753 (OD_{600} = 0.5) or water as a control. Images were taken A&C) 48 hours post inoculation or B) four days post inoculation. The barley pathogen *X. translucens* pv. *translucens* strain UPB787 served as a positive control for symptoms on barley. Plants were grown in a growth chamber at 22°C, 50% humidity and 16 hours of light.

Genome Sequencing and Annotation

The genomes of strains NCPPB 2877 and NCPPB 3753 were sequenced using the Illumina Hi-Seq2500 platform (Fasteris SA, Switzerland). The shotgun sequencing yielded 2,921,175 100-bp paired-end reads (730 Mb) for strain NCPPB 2877 and 2,464,521 paired-end reads (616 Mb) for strain NCPPB 3753, with insert sizes ranging from of 250 bp to 1.5 kb. Draft genome sequences were assembled using the Edena algorithm v3.131028 (Hernandez et al., 2014), yielding 257 contigs ≥200 bp (N_{50} = 38,306 bp) with 69 x coverage for strain NCPPB 2877 and 260 contigs (N_{50} = 35,229 bp) with 73 x coverage for strain NCPPB 3753. For comparison, draft genome sequences were also assembled using the Velvet algorithm v1.1.04 (Zerbino and Birney, 2008), yielding 564 contigs ≥200 bp (N_{50} = 15,608 bp) for strain NCPPB 2877 and 469 contigs (N_{50} = 20,963 bp) for strain NCPPB 3753. Because of their better quality, Edena-derived contigs were annotated with GeneMarkS+ release 2.9 (revision 452131) (Borodovsky and Lomsadze, 2014), as implemented in the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/ annotation_prok/), which predicted a total of 4,095 genes within 4,756,730 bp for strain NCPPB 2877 and 4,160 genes within 4,837,471 bp for strain NCPPB 3753. These whole genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession no. JSZE00000000 (NCPPB 2877) and JSZF00000000 (NCPPB 3753). The versions described in this paper are the first versions, JSZE01000000 and JSZF01000000.

Comparison of the Two Genome Sequences

ANI provides a robust method to determine bacterial species definition based on whole genome sequence comparison and is considered the new standard for species definition (Konstantinidis and Tiedje, 2005; Figueras et al., 2015). To determine if NCPPB 2877 and NCPPB 3753 are the same species, the ANI was calculated for both genome sequences using JSpecies (Richter and Rosselló-Móra, 2009). BLAST-based comparison revealed 99.2% ANI for the 92.5% sequences that could be aligned, and MUMmer-based comparison revealed 99.1% ANI for the 96.9% sequences that could be aligned, thus confirming that both strains belong to the same species.

Using our web-based pipeline for prediction of satellites (http://www.biopred.net/VNTR/), we then evaluated whether or not both strains belong to a clonal complex. For satellite prediction, the following parameters were chosen (Zhao et al., 2012): algorithm, TRF; region length, 30 to 1,000 bp; unit length, 5 to 12 bp; and at least 6 tandem repeats with a similarity of at least 80% among the repeats. In total, 45 microsatellites were predicted, 35 of which were found to be present in both genome sequences. For 34 of them, repeat numbers could be derived; while one locus was not informative because it was located at the end of two contigs and thus not completely assembled in NCPPB 3753. To provide further evidence that the calculated repeat numbers were meaningful, the corresponding loci were also analyzed in the Velvet-based genome assemblies. Strikingly, there was not a single discrepancy between the Edena- and Velvet-based data, except for the fact that some satellite loci were not completely assembled by

Velvet while they were complete in the Edena assembly. For the complete loci, 28 loci (82%) were different between the two strains with respect to repeat numbers. For the six loci with identical repeat numbers DNA sequence analysis revealed that five of them were identical due to homoplasia. Thus, both strains differ by almost all (97%) of their completely assembled microsatellite loci, a finding that indicates that both strains do not belong to a clonal complex.

Taxonomic Position of the Two Cannabis Pathogens

Comparison of 16S rDNA sequences is a method of choice to elucidate the taxonomic positions of bacterial strains, and was previously used to analyze and delineate 20 species of *Xanthomonas* (Hauben et al., 1997). It was found that the genus *Xanthomonas* exhibited a relatively high level of 16S rDNA sequence identity, with on average 14 single-nucleotide polymorphisms (SNPs) between two different *Xanthomonas* species (Hauben et al., 1997). The 16S rDNA sequences of both cannabis pathogens were found to be identical. When we compared the 16S rDNA sequence of the cannabis pathogens with those of the 20 *Xanthomonas* type strains, the cannabis pathogen grouped with Group 2 strains, which contains the majority of characterized *Xanthomonas* species. Interestingly, GenBank comparison revealed that another recently sequenced strain that was isolated from symptomatic bean plants in Rwanda, Nyagatare, contains the same 16S rDNA sequence (Aritua et al., 2015).

Previously *X. cannabis* strains were also called *X. campestris* pv. *cannabis* based on the similarity of the 16S rDNA sequence to *X. campestris*, but it has been suggested that the name should be

changed to *X. cannabis* (Netsu et al., 2014). Since the resolution of the 16S rDNA sequence is very low within Group 2 strains (Hauben et al., 1997) and often only distinguishes a species by one or two SNPs, we performed whole-genome comparisons including one representative strain per species for which genome sequences were available (Figure 2). The pairwise ANI of the two cannabis strains against any of the representative strains was below 90%, regardless of which algorithm (BLAST or MUMmer) was used, indicating that these two strains belong to an unique and distinct *Xanthomonas* species (Figure 2). We suggest that *X. cannabis* is the appropriate name for this bacterial species based on our ANI analysis and as previously suggested by Netsu and colleagues (2014).

Guided by the observation that their 16S rDNA sequences were identical to that of the Nyagatare strain, we compared the genomes of *X. cannabis* NCPPB 3753 and NCPPB 2877 and *X.* sp. Nyagatare. JSpecies calculations revealed that the two cannabis strains were 96.3% to 96.4% identical to the Nyagatare strain, when calculated over the 88.6% to 91.0% of the genome sequence that could be aligned by the more robust MUMmer algorithm (Richter and Rosselló-Móra, 2009). These values are slightly above the ≈95–96% transition zone, above which strains can be considered to belong to the same taxonomically circumscribed prokaryotic species (Konstantinidis and Tiedje, 2005). Therefore the Nyagatare strain most likely belongs to the *X. cannabis* species. It would be interesting to perform functional studies to determine the similarities and differences between these closely related *X. cannabis* strains because

the Nyagatare strain is a reported bean pathogen and *X. cannabis* strains elicit no response when inoculated on bean (Netsu et al., 2014).

Partial sequencing of the *gyrB* and other housekeeping genes for MLSA grouped all *Xanthomonas* species into four major MLSA subgroups (Parkinson et al., 2007; Young et al., 2008). We used MLSA of seven housekeeping genes (*atpD*, *dnaK*, *efp*, *glnA*, *gyrB*, *lepA*, and *rpoD*) and ANI calculations to understand the relationship of the cannabis and Nyagatare strains compared to other xanthomonads. The *X. cannabis* NCPPB 2877 and NCPPB 3753 and Nyagatare strains appear to form a distinct grouping according to MLSA (Figure 2). Based on the comparisons of ANI and of seven concatenated internal portions of the genes, we overall conclude that the cannabis and Nyagatare strains belong to a single species, *X. cannabis* because: 1) they are above the 95-96% ANI threshold for species definition and 2) they appear to belong to a new MLSA clade (Figure 2). This new clade corresponds to the novel species-level clade (slc) 1, as suggested by Parkinson and co-workers, which also contains the pathovars *esculenti* and *zinniae* (Parkinson et al., 2009). Following the pathovar designation, we suggest to name the two cannabis strains *X. cannabis* pv. *cannabis*, and the bean-pathogenic Nyagatare strain *X. cannabis* pv. *phaseoli*.

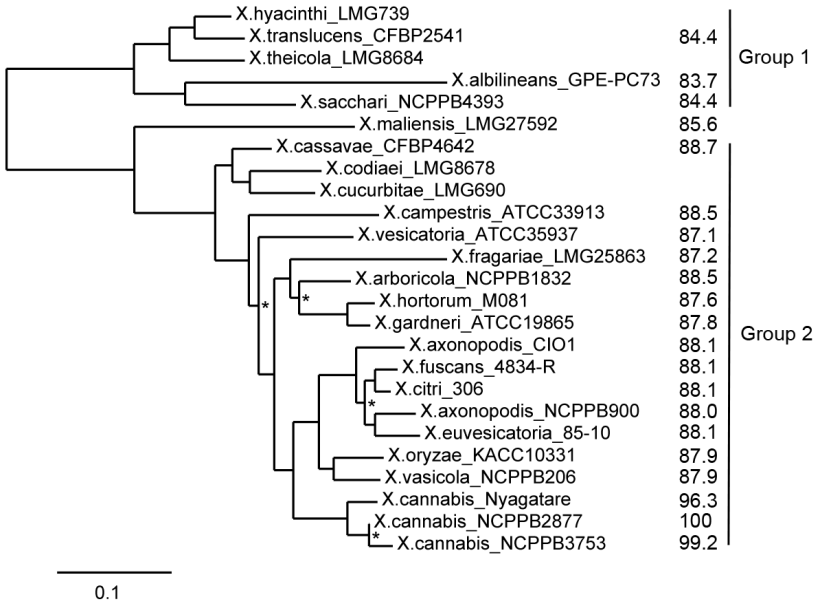


Figure 2. ANI and MLSA-based phylogenetic tree of xanthomonads.
Phylogenetic analysis was performed on the Phylogeny.fr platform (Dereeper et al., 2008). Sequences were aligned with MUSCLE (v3.7) configured for highest accuracy (MUSCLE with default settings). After alignment, ambiguous regions (i.e. containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) using default parameters. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0). The HKY85 substitution model was selected assuming an estimated proportion of invariant sites and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data. Reliability for internal branch was assessed using the aLRT test (SH-Like). Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3). All nodes were supported by bootstrap values above 0.95, except for those marked with a blue asterisk (0.923) or a red asterisk (<0.9). Calculated ANI values in comparison to NCPPB 2877 and the two major phylogenetic groups are indicated on the right side.

Comparison of Pathogenicity-related Gene Clusters

Several gene clusters are considered to be important for pathogenicity of xanthomonads and their possible contribution to host- and tissue-specificity has been studied previously (Lu et al., 2008). We therefore

analyzed whether and to which extent these gene clusters are conserved in the cannabis strains.

The *rpf* (regulation of pathogenicity factors) gene cluster plays a role in the intercellular signal-response system that links synthesis and perception of the diffusible signal factor (DSF) cis-11-methyl-2-dodecenoic acid to the synthesis of extracellular enzymes, extracellular polysaccharide, and biofilm dispersal (Dow, 2008). The genetic organization of the cannabis *rpf* gene cluster resembles that of *X. euvesicatoria* strain 85-10, yet it contains an additional gene, *rpfI*, downstream of gene XCV1913, which is not present in *X. euvesicatoria* strain 85-10. Moreover, the proteins RpfF, responsible for DSF synthesis, and the two-component regulatory system RpfC/RpfG, which is involved in DSF perception and signal transduction (Dow, 2008), are highly conserved in the cannabis pathogens.

The *gum* gene clusters encode proteins that are involved in the exopolysaccharide (EPS) biosynthesis (Becker et al., 1998). The core *gum* gene cluster, consisting of *gumB* to *gumM*, all transcribed in the same direction, is entirely conserved in the cannabis pathogen. As in other xanthomonads, the first gene, *gumB*, is located downstream of a proline-specific tRNA gene. The *gumA* gene which is located upstream of the tRNA gene and downstream *pheS* and *pheT* is most likely not a *bone fide gum* gene. The accessory *gumN* gene, which for instance is disrupted in *X. euvesicatoria* strain 85-10 and *X. oryzae* pv. *oryzicola* strain BLS256, appears to be intact in the cannabis pathogen. *gumO* and *gumP*, annotated as a 3-oxoacyl-(acyl carrier protein) synthase and a metal-dependent hydrolase, respectively, are

located directly downstream of *gumN*. The contribution of these two genes to EPS biosynthesis is questionable since it is not present in several xanthomonads infecting monocotyledons, such as *X. sacchari*, *X. translucens* and *X. oryzae*, and it is also not present in the recently sequenced *Xanthomonas maliensis* strain 97M, which was isolated from rice leaves in Mali. On the other side, this unique conservation in bacteria colonizing eudicots could suggest a role in host specificity at the level of a division.

Lipopolysaccharide (LPS) is another bacterial polysaccharide, which is firmly attached to the outer membrane, and an aberrant structure of the LPS O-chain has been linked to virulence defects (Mhedbi-Hajri et al., 2011). The gene cluster responsible for LPS biosynthesis is always present between the highly conserved *etfA* and *metB* genes of *Xanthomonas*. A remarkably high degree of variation both in number and in identity of LPS genes has been found in different xanthomonads, even within a single species or pathovar (Patil et al., 2007). In the two cannabis pathogens, 14 genes are predicted to participate in the LPS biosynthesis. The gene content and genetic organization is largely identical to a few other *Xanthomonas* strains belonging to different species: *Xanthomonas gardneri* strain ATCC 19865, *Xanthomonas vesicatoria* strain ATCC 35937, and strains of *X. campestris* (B100, CN14, CN15, CN16, JX, Xca5, and 756C); except that the *X. campestris* strains have a short, additionally predicted gene, *wxcH*, just downstream of the *etfA* gene, which is not present (or frame-shifted) in the other species. Interestingly, the relatively closely related Nyagatare strain shares only the four *etfA*-proximal genes, *wxcK* to *wxcN*, and the two *metB*-proximal genes,

wzm and *wzt*, with the cannabis strains, while between them three large genes without orthologs in other xanthomonads are predicted. As observed before on a smaller set of *Xanthomonas* genome sequences (Patil et al., 2007; Lu et al., 2008), the cannabis pathogens underscore the fact that the LPS gene cluster varies highly between and within species, suggesting multiple horizontal gene transfers and re-assortments. Yet, it is unclear what evolutionary pressure drives this variation – is it escape from recognition by plant immune receptors, and/or is it escape from recognition by bacteriophages?

Motility is an important feature of bacteria that is governed by flagella-based swimming and/or pilus-based twitching or gliding (Rossez et al., 2015). The flagellum is a complex machinery, the biosynthesis of which depends on dozens of genes, collectively called *fle*, *flg*, *flh* and *fli* genes, which are organized in large gene clusters. The cannabis strain NCPPB 2877 contains the full gene complement of flagellar biosynthesis. In contrast, strain NCPPB 3753 appears to have lost two large regions in its two genes clusters. One deletion encompasses the genes *flgG* to *flgJ* and the 5' portion of *flgK*. Downstream of *flgF*, remnants of an IS element were present in the genome assembly. The second, even larger deletion goes from the 3' portion of *fliK* until the end of *flhA*, totaling to more than 11 genes. Absence of flagella under standard culture conditions was confirmed for strain NCPPB 3753 by a swimming assay in semisolid medium (Figure 3). Briefly overnight cultures grown in liquid NB medium were fixed to an OD600 = 1.0 and stabbed into semi-solid NYGA medium (0.3% agar). Motility was defined as observed turbid growth outside the stab zone of inoculation in the semi-solid medium. Tubes

inoculated with NCPPB 3753 were turbid after four days of growth (Figure 3), which demonstrates that this strain is motile. Growth of NCPPB 2877 was confined to the stab zone of inoculation (Figure 3), which supports our hypothesis that NCPPB 2877 is non-motile because it lacks a flagellum.

Loss of flagella in xanthomonads is not without precedent (Darrasse et al., 2013). When Darrasse and co-workers tested 300 *Xanthomonas* strains representing different species and pathovars, five percent of the tested strains turned out to carry a deletion in the flagellar cluster and were non-motile. Co-isolation of flagellate and non-flagellate variants from an outbreak suggested that flagellar motility is not essential for fitness within the plant and that mixed populations could be a strategy to avoid detection by the plant defense system (Darrasse et al., 2013). Indeed, many plants have evolved receptor-like kinases (e.g. FLS2) that detect peptide epitopes of the FliC subunit of the flagellum (e.g. flg22 and flgII-28) (Sun et al., 2006), and it has been reported that otherwise isogenic flagellin-deficient *Xanthomonas* bacteria have a colonization advantage over flagellated bacteria in citrus leaves (Shi et al., 2015). Interestingly, when we analyzed the FliC flagellin sequence of the two cannabis strains, we detected a flg22 variant that deviates from the consensus sequence by as much as nine amino acid residues and is identical to the flg22 sequences from other xanthomonads, such as *X. oryzae* and *X. vasicola* (Supplemental Figure S1). In contrast, the related Nyagatare strain has a flg22 sequence that only deviates from the consensus by four residues and is similar to the elicitation-active flagellins that are recognized by FLS2 in Arabidopsis (Sun et al.,

2006). These finding suggest that the cannabis strains use at least two out of several strategies to escape from recognition by the plant immune surveillance system, either by loss of the flagellum or by allelic variation of flagellin epitope(s) (Rossez et al., 2015).

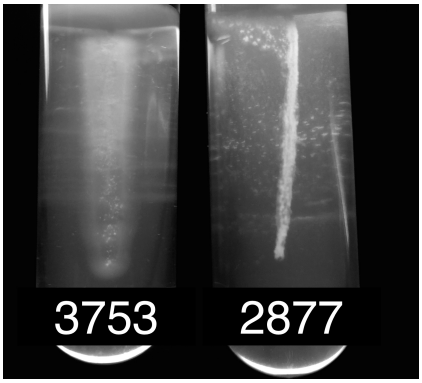


Figure 3. Motility of *X. cannabis* strains.
Bacterial motility was determined by stab inoculation in motility medium of either strain NCPPB 3753 or NCPPB 2877. Motility was evaluated qualitatively by turbid growth (motile) compared to localized, fixed growth (non-motile) around the zone of inoculation. Bacteria were grown in semi-solid NYGA agar medium (0.3%) as previously described (Sun et al., 2006).

The cannabis pathogen contains two different Type II protein secretion systems, the Xcs and the Xps system, as found in several other xanthomonads (Lu et al., 2008). The *xcs* gene cluster, consisting of 12 genes, *xcsC* to *xcsN*, is located downstream of a TonB-dependent receptor gene and upstream of a GntR-family regulator gene, which is also the case for *X. euvesicatoria* strain 85-10 and strains of *X. campestris*. The related Nyagatare strain contains another four genes between *xcsN* and the GntR-family regulator gene, a feature that is shared with the *X. citri* pv. *citri* strain 306 and the *Xanthomonas arboricola* strain 3004. Interestingly, the *xcs* gene

cluster is also present in the *X. maliensis* strain 97M, but absent in the Group 1 strains *X. albilineans* GPE PC-73, *X. sacchari* NCPPB 4393, and *X. translucens* pv. *cerealis* CFBP 2541. The second Type II secretion system, encoded by the *xps* gene cluster, consists of 11 genes, *xpsE* to *xpsN*, followed by *xpsD*, and is found between the adhesion gene *xadA* and the *pnuC* gene. This position in the bacterial chromosome is largely conserved among *Xanthomonas* strains belonging to different Groups, such as the *X. sacchari* strain NCPPB 4393, the *X. translucens* pv. *cerealis* strain CFBP 2541, the *X. maliensis* strain 97M, and strains of Group 2.

Many Gram-negative plant-pathogenic bacteria have evolved another protein secretion system, the T3SS, which plays a pivotal role in the pathogen-host interaction (Büttner, 2012). The Hrp system serves as a molecular syringe that injects T3 effector proteins into the cytosol of the host's cells at the benefit of the bacterium. Yet, many injected effectors can also betray the pathogen to the host plant, triggering a strong defense response, which typically is accompanied by an effector-triggered HR (White et al., 2009). Notably the early-branching Group 1 sugarcane pathogen *X. albilineans* does not have an Hrp T3SS, a finding that could be attributed to the reduced genome size of this pathogen (Pierretti et al., 2009). Since then, new genome sequences became available, including other Group 1 strains belonging to the species *X. sacchari* (Studholme et al., 2011). Some of these strains do not appear to have undergone extensive genome reduction, yet, they do not encode a T3SS. When we analysed the gene content of the cannabis strains we realized that they do not have any of the *hrc*, *hrp* or *hpa* genes encoding the T3SS. We were also not

able to detect any traces of a T3-effector gene when using the set of described genes as queries (http://www.xanthomonas.org/t3e.html). In contrast, the Nyagatare strain has a full gene complement for the Hrp T3SS and also possesses a couple of effector proteins (Aritua et al., 2015). This is an interesting finding for several reasons. First, how do the cannabis pathogens suppress plant defense, a function that became more and more linked to the activity of T3-effectors (e.g. Canonne et al., 2011; Schulze et al., 2012; Sinha et al., 2013; Strok et al., 2015; Li et al., 2015)? Second, which molecular entities are triggering the HR in *Nicotiana tabacum* if not type III effectors? Interestingly, a type II-secreted pectate lyase, XagP, from *X. axonopodis* pv. *glycines* was previously found to be associated with HR induction on tobacco and pepper, but not on cucumber, sesame and tomato, thus giving a prime example that cell wall-degrading enzymes could trigger an HR (Kaewnum et al., 2006). We therefore speculate that similar or other secreted enzymes that disturb the integrity of the plant cell wall could be responsible for HR triggered by the cannabis pathogen independent of a T3SS. Last but not least, these Group 2 genome sequences may serve as a valuable resource to predict new T3 effectors. As a relative of other xanthomonads it will provide a valuable negative training or filtering set for prediction algorithms.

Analysis of the HrpX Regulon

Previous work has shown that the Hrp T3SS and many of its secreted effectors are controlled by a regulatory cascade, consisting of HpaR2/HpaS, HrpG and HrpX (Büttner and Bonas, 2010; Li et al., 2014). However, HrpG and HrpX regulate pathogenesis beyond the

T3SS alone (Guo et al., 2011). We therefore wondered if these components are present in *Xanthomonas* strains that do not possess the Hrp T3SS, and if so, what genes would be controlled by these components. BLAST searches revealed that all four regulatory genes are conserved in the cannabis pathogens (Table 1). *hpaR2* and *hpaS* are also conserved in the *X. maliensis* strain and in strains from Group 1, such as *X. albilineans*, *X. sacchari* and *X. translucens*. Their location is always between the *glmS* and *glmU* gene, with some extra genes for an efflux pump between *glmS* and *hpaR2*. In contrast, *hrpG* and *hrpX* were not found in strains of *X. albilineans* or *X. sacchari*, but they are present in *X. translucens*. Interestingly, however, *hrpG* and *hrpX* from *X. translucens* cluster together with the rest of the T3SS *hrp* gene cluster while all other xanthomonads have *hrpG* and *hrpX* at a distinct genomic location between *radA* and *hsp70* (Wichmann et al., 2013). Only for *X. maliensis* we could not determine their genomic organization due to the highly fragmented genome assembly.

Within the Hpa-Hrp regulatory cascade, HrpX is the most downstream component that directly induces the synthesis of pathogenicity factors, such as the Hrp T3SS, T3 effectors and cell wall-degrading enzymes, by binding to a conserved *cis* element, called PIP (plant-induced promoter) box, within the promoter regions of the corresponding genes (Koebnik et al., 2006). For gene activation, both the PIP box as well as a properly spaced –10 promoter motif are required (Furutani et al., 2006). We therefore analyzed the genome sequences of the cannabis pathogens for the presence of the promoter motif TTCGB-N_{15}-TTCGB-$N_{30\text{-}32}$-TYNNNT (B represents C, G, or T;

Y represents C or T) (Koebnik et al., 2006). Using this conservative query, four genes were strongly predicted to belong to the HrpX regulon, two polygalacturonases (PehA, synonymous with PglA, and PehB, synonymous with PglB; Hsiao et al., 2008; Wang et al., 2008), one aminopeptidase and one lysophospholipase (LPL). Indeed, the two polygalacturonases have been previously shown to be regulated by HrpX in *X. campestris* (Wang et al., 2008). For *X. citri*, microarray transcriptome studies have shown previously that three genes with a canonical PIP box and –10 promoter motif (PehA, PehB, LPL, Table 1) are positively regulated by HrpX (Guo et al., 2011). Finally, our own RNAseq data show that the LPL is under control of HrpX in an African *X. oryzae* strain (unpublished data). We highly suspected that the four predicted genes with a canonical PIP box and properly spaced –10 promoter motif were indeed under control of HrpX in the cannabis strains.

Genome mining revealed that these four genes are present in most xanthomonads from Groups 1 and 2 (Table 1). Yet, in some lineages, one or the other gene apparently got lost. For instance, *X. sacchari*, *X. maliensis*, *X. arboricola*, *X. axonopodis* pv. *manihotis*, *X. oryzae* and *X. vasicola* lack one or the other polygalacturonase, and the aminopeptidase is absent from *X. alibilineans* and *X. oryzae*. In a few cases, these genes appear to have suffered from pseudogenization (Table 1) but this needs confirmation by targeted DNA sequencing due to the risk of sequence errors in some of the draft genome sequences.

Table 1. Presence of type III secretion systems, type III effectors, *hrp*/*hpa* regulatory genes, and homologs of predicted HrpX regulon members of *X. cannabis* pv. *cannabis* in representative strains of *Xanthomonas*.

Group	Species	Strain	GenBank accession no.	Hrp T3SS	T3Es	HrpG HrpX	HpaS HpaR2	PehA	PehD	Amino-peptidase	LPL
1	*X. albilineans*	GPE PC73	FP565176	no	no	no	YES	YES	YES	no	YES
1	*X. sacchari*	NCPPB 4393	AGDB01000001	no	no	no	YES	YES	no	YES	YES
1	*X. translucens* pv. *cerealis*	CFBP 2541	JWHD01000001	YES	YES	YES	no[a]	YES	YES	Ψ	YES
	X. maliensis	97M	AQPR01000001	no	no	YES	YES	no	PIP	PIP[5]	PIP[1]
2	*X. campestris* pv. *campestris*	ATCC 33913	AE008922	YES	YES	YES	YES	PIP	PIP[1]	PIP[1]	PIP
2	*X. vesicatoria*	ATCC 35937	AEQV01000004	YES	YES	YES	YES	PIP	PIP	PIP[1]	PIP
2	*X. vasicola* pv. *vasculorum*	NCPPB 206	AKBM01000001	YES	YES	YES	YES	PIP	no	PIP[1]	PIP
2	*X. oryzae* pv. *oryzae*	KACC110331	AE013598	YES	YES	YES	no[b]	Ψ	YES	no	PIP
2	*X. fuscans* subsp. *fuscans*	4834-R	FO681494	YES	YES	YES	YES	PIP	PIP	PIP[1]	PIP
2	*X. citri* pv. *citri*	306	AE008923	YES	YES	YES	YES	PIP	PIP	PIP[1]	PIP
2	*X. axonopodis* pv. *manihotis*	CIO1	AKCZ01000001	YES	YES	YES	YES	PIP	no	PIP[1]	PIP-Ψ
2	*X. axonopodis* pv. *vasculorum*	NCPPB 900	JPHD01000001	YES	YES	YES	no[c]	PIP	PIP	PIP[1]	PIP
2	*X. euvesicatoria*	85-10	AM039952	YES	YES	YES	YES	PIP	PIP	PIP[1]	PIP
2	*X. cassavae*	CFBP 4642	ATMC01000001	YES	YES	YES	YES	PIP	PIP	PIP[2]	PIP[3]
2	*X. arboricola* pv. *celebensis*	NCPPB 1832	JPHC01000001	YES	YES	YES	YES	PIP	no	PIP[1]	PIP[1]
2	*X. hortorum* pv. *carotae*	M081	AEEU01000001	YES	YES	YES	YES	PIP	PIP	PIP[1]	PIP
2	*X. gardneri*	ATCC 19865	AEQX01000001	YES	YES	YES	YES	PIP	PIP	PIP[1]	PIP
2	*X. fragariae*	LMG 25863	AJRZ01000001	YES	YES	YES	YES	PIP	PIP	Ψ	Ψ
2	*X. cannabis* pv. *phaseoli*	Nyagatare	JRQI01000001	YES	YES	YES	YES	PIP	PIP	PIP	PIP
2	*X. cannabis* pv. *cannabis*	NCPPB 2877	JSZE01000001	no	no	YES	YES	PIP	PIP	PIP	PIP
2	*X. cannabis* pv. *cannabis*	NCPPB 3753	JSZF01000001	no	no	YES	YES	PIP	PIP	PIP	PIP

"YES" indicates presence of a homolog, "no" indicates absence of the protein(s). "ψ" indicates pseudogenes, i.e. the gene is split into two or more fragments. "PIP" indices the presence of the gene along with a canonical PIP box and a properly spaced –10 promoter motif. Numbers in square brackets indicate the number of single-nucleotide variants with respect to the canonical PIP box and –10 promoter motif. [1] *hpaR2* is eroded while an N-terminally truncated form of *hpaS* is present in *X. translucens* pv. *cerealis* strain CFBP 2541. Remnants are found in two other *X. translucens* strains (ART-Xtg27 and DSM 18974). Interestingly, the *hpaR2*/*hpaS* locus is intact in *X. translucens* strain DAR61454. This is an example of ongoing gene erosion in one species of *Xanthomonas*. [2] The *hpaR2*/*hpaS* locus got destroyed by an IS element in *X. oryzae* pv. *oryzae* strain KACC 10331. Yet, *hpaR2*/*hpaS* is present in *X. oryzae* pv. *oryzicola* strain BLS256. Remnants are found in the African *X. oryzae* pv. *oryzae* strain NAI8 and in *X. oryzae* strains from the United States. This finding illustrates ongoing gene erosion in another species of *Xanthomonas*. [3] *hpaR2* is absent while an N-terminally truncated form of *hpaS* is present in *X. axonopodis* pv. *vasculorum* strain NCPPB 900.

Since these four genes appear to be under control of HrpX in the cannabis pathogens, we looked for evidence that the same regulation occurs in the other xanthomonads. We therefore compared the upstream regions of these four genes in a representative set of *Xanthomonas* strains. Strikingly, we found PIP boxes and properly spaced –10 promoter motifs for most of the genes in most of the *Xanthomonas* strains (Figure 4, Supplemental Figure S2), except for the Group 1 strains *X. albilineans*, *X. sacchari* and *X. translucens*. Multiple sequence alignments of the promoter regions of the three genes show that the PIP boxes are conserved in sequence, context and position (PIP boxes ~140- 150 bp before start codon of *pehA*, ~90 bp before start codon of the aminopeptidase gene, and ~210-250 bp before start codon of *pehB*) indicate that the PIP boxes of the polygalacturonase and aminopeptidase genes evolved early after separation of Group 2 from Group 1. In contrast, the PIP boxes of the lysophospholipase gene do not align with each other and reveal four subgroupings, which are compatible with a MLSA-based phylogeny of Group 2 strains (Supplemental Figure S2). This finding could suggest that the PIP box evolved several times independently at early times after separation of the four MLSA subgroups, or that the surrounding sequences evolved too extensive and thus do not allow to align the PIP boxes using standard parameters.

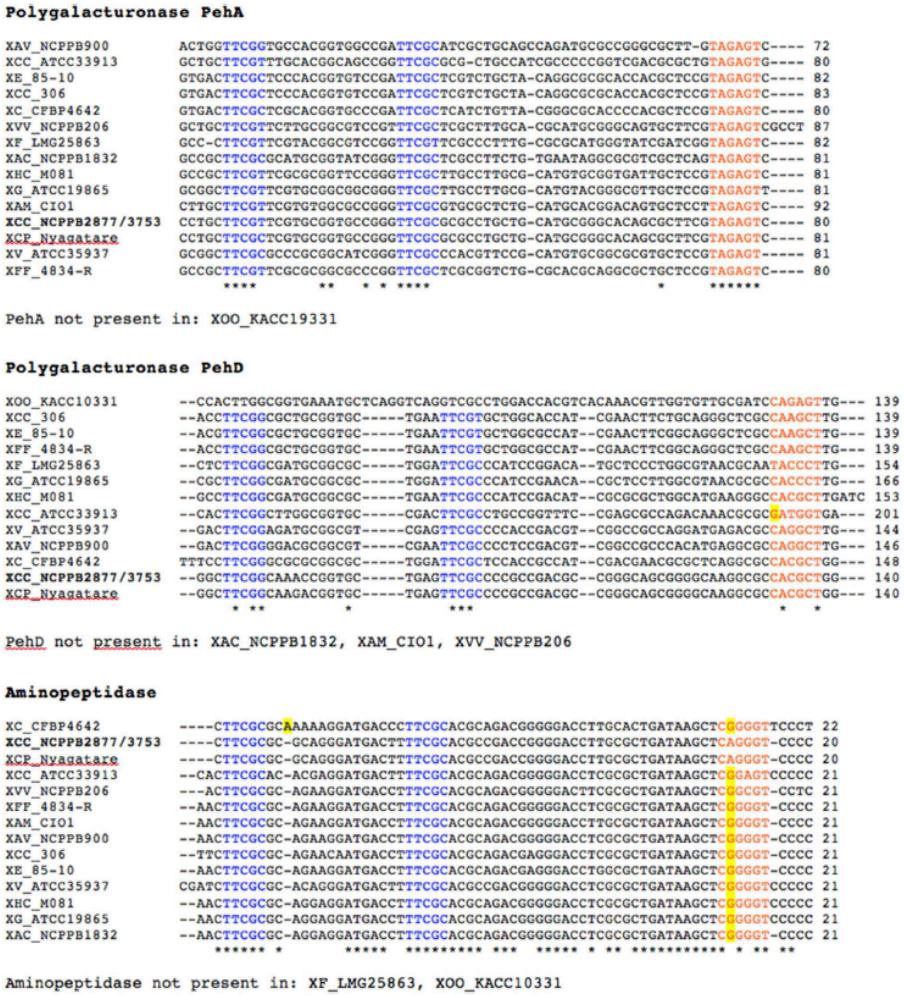


Figure 4. Promoter sequences of predicted HrpX regulon members and their homologs in strains of *Xanthomonas*.

Promoter regions encompassing 350 bp upstream of the translational start codon of a representative set of *Xanthomonas* strains were aligned by MUSCLE. PIP half boxes are shown in blue and the –10 promoter motif is shown in orange. Distance to the translational start codon is indicated on the right side of each sequence block. Deviations from the PIP consensus sequence are highlighted in yellow. The following *Xanthomonas* strains were analyzed: XAC (*X. arboricola* pv. *celebensis*) NCPPB 1832, XAM (*X. axonopodis* pv. *manihotis*) CIO1, XAV (*X. axonopodis* pv. *vasculorum*) NCPPB 900, XCC (*X. campestris* pv. *campestris*) ATCC 33913, XCC (*X. cannabis* pv. *cannabis*) NCPPB 2877, XCP (*X. cannabis* pv. *phaseoli*) Nyagatare, XC (*X. cassavae*) CFBP 4642, XCC (*X. citri* pv. *citri*) 306, XE (*X. euvesicatoria*) 85-10, XF (*X. fragariae*) LMG 25863, XFF (*X. fuscans* subsp. *fuscans*) 4834-R, XG (*X. gardneri*) ATCC 19865, XHC (*X. hortorum* pv. *carotae*) M081, XOO (*X. oryzae* pv. *oryzae*) KACC 19331, XVV (*X. vasicola* pv. *vasculorum*) NCPPB 206, and XV (*X. vesicatoria*) ATCC 35937.

To test if HrpX and HrpG could activate expression of the putative targets, we quantified gene expression of pehA, one of the four genes with a PIP box, in various mutant backgrounds of X. cannabis NCPPB 3753. HrpG* from X. euvesicatoria 85-10 is a HrpG variant that mimics the active form of HrpG (Wengelnik et al., 1999). Mutant strains with hrpG* induce expression of hrpX and further genes targeted by HrpX, which includes primarily promoters with PIP boxes (Wengelnik et al., 1999). We therefore transformed *X. cannabis* cells with either pBBR1MCS-5 (empty vector), pBBR1MCS-5::*hrpX* or pBBR1MCS-5::*hrpG**, as previously described (Koebnik et al, 2006). *hrpG** from X. euvesicatoria was cloned into pBBR1MCS-5, as described for *hrpX* (Wengelnik et al., 1999; Koebnik et al, 2006). Positive transformants were confirmed by PCR (data not shown).

Quantitative PCR (qPCR) with SYBR-green was used to determine the relative expression of putative HrpX-target *pehA*, *pehB*, and the gene encoding the putative LPL in *X. cannabis* NCPPB 3753 pBBR1MCS-5::*hrpG** and NCPPB 3653 pBBR1MCS-5::*hrpX* compared to NCPPB 3753 pBBR1MCS-5 (empty vector) as a control. *atpD* was used as a normalization control. *pehA*, *pehB* and the gene encoding the putative LPL were dramatically induced in NCPPB 3753 ectopically expressing *hrpG** or *hrpX*, compared to the empty vector control (Figure 5).

We did not detect differential expression of the putative aminopeptidase in our *hrp* regulatory variant backgrounds under the conditions mentioned above (data not shown). We only tested one set of primers for this gene under one condition, and further experiments need to be performed to determine if this negative result suggests that

this gene is not a target of HrpX. We overall conclude that the promoter of *pehA*, *pehB* and LPL are *bona fide* targets of HrpX, validating our bioinformatic analysis. This supports the hypothesis that HrpG and HrpX regulate pathogenesis beyond T3SS alone (Tang et al., 2006). Further investigation of the role of HrpX and HrpX targets remains to be determined, but *X. cannabis* is a potentially interesting model to understand the role of HrpX beyond the T3SS.

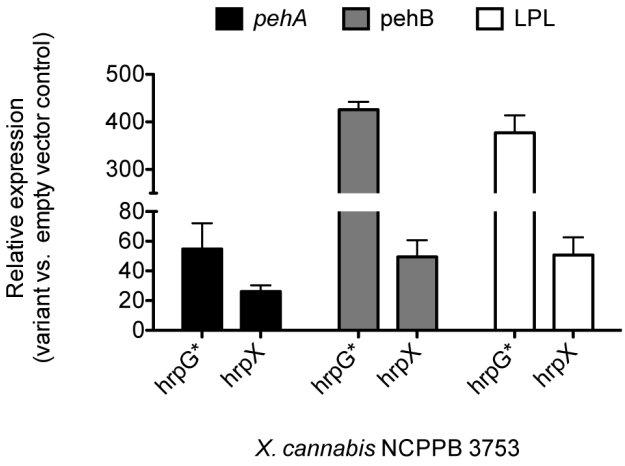


Figure 5. Gene expression analysis of genes with HrpX-inducible PIP boxes.

Bacteria were grown overnight in liquid NB (Sigma Aldrich, USA) supplemented with gentamycin (20 µg l^{-1}) and transferred to fresh 10 mL NB media with gentamycin for a final OD_{600} = 0.5. Bacteria were incubated for three hours, shaking at 28 °C. Transcriptional profiles and RNA was preserved with 5% phenol in ethanol as previously described (Jahn et al., 2008; Jacobs et al., 2012). Bacterial RNA was extracted with Trizol (Invitrogen, USA), cleaned up with Zymogen RNA concentrator (Zymo Research, USA) and treated with Turbo DNase (Invitrogen, USA) following manufacturer's protocols. RNA (1 µg per sample) was reverse transcribed with Superscript III (Invitrogen, USA) following the manufacturer's recommendation. qPCR with SYBR MESA BLUE Master Mix (Eurogentec, Belgium) was performed following the manufacturer's protocol on a Roche LightCycler 480 Real-Time PCR instrument with reaction parameters of 10-min polymerase activation at 95 °C, then 40 cycles, with an individual cycle consisting of 15 s at 95 °C and 1 min at 60 °C. Relative gene expression analysis of *pehA*, *pehB* and the gene encoding the putative lysophospholipase (LPL) were calculated by the delta-delta-Ct method for NCPPB 3753 variants ectopically expressing *hrpG** or *hrpX* compared to NCPPB 3753 pBBR1-MCS5. Expression of *atpD* was used as a normalization internal control gene. Primers are listed in Supplemental Table S1.

We did not detect differential expression of the putative aminopeptidase in our *hrp* regulatory variant backgrounds under the conditions mentioned above (data not shown). We only tested one set of primers for this gene under one condition, and further experiments need to be performed to determine if this negative result suggests that this gene is not a target of HrpX. We overall conclude that the promoter of *pehA*, *pehB* and LPL are *bona fide* targets of HrpX, validating our bioinformatic analysis. This supports the hypothesis that HrpG and HrpX regulate pathogenesis beyond T3SS alone (Tang et al., 2006). Further investigation of the role of HrpX and HrpX targets remains to be determined, but *X. cannabis* is a potentially interesting model to understand the role of HrpX beyond the T3SS.

One of the polygalacturonase genes, *pehA*, is not only under control of HrpX but was found to be regulated by Clp and RpfF in *X. campestris* (Hsiao et al., 2008). Since Clp and RpfF are conserved over all clades it is tempting to speculate that this layer of regulation has evolved before separation of the clades. However, comparison of the *pehA* promoter regions in a representative set of *Xanthomonas* strains revealed that the predicted Clp-binding site is not conserved (Supplemental Figure S3). Notably, multiple sequence alignment instead revealed a conserved sequence motif the position of which is shifted by 8 bp with respect to the predicted Clp-binding site (Supplemental Figure S3). In fact, this conserved box is as similar to the canonical Clp box as is the previously predicted Clp-binding site (Dong and Ebright, 1992), and its position is compatible with the mapping of the Clp-binding site by *lacZ* reporter fusions (Hsiao et al., 2008). Probably, the Clp-binding site was slightly mispredicted at that

time due to the absence of sufficient sequence data from diverse *Xanthomonas* strains, which help to uncover regulatory DNA elements.

Conclusions

Stimulated by two new genome sequences from cannabis-pathogenic xanthomonads, we explored the world of pathogenicity determinants in the genus *Xanthomonas*. A plethora of research data as well as our own analyses let us speculate about a stepwise evolution of pathogenicity in *Xanthomonas*. We developed a model (Figure 6) of evolution and acquisition of *Xanthomonas* pathogenicity factors based on a model proposed by Lu and colleagues (2008). Given the observation that some *Xanthomonas* strains across distinct taxa do not contain *hrpG* and *hrpX*, we speculate a stepwise evolution of pathogenicity (Figure 6), which involves (i) acquisition of key regulatory genes and cell wall-degrading enzymes, followed by (ii) acquisition of the Hrp type III secretion system, which is ultimately accompanied by (iii) successive acquisition of type III effectors. In parallel, as soon as *hrpG* and *hrpX* were acquired, a subset of genes, which can contribute to pathogenicity, evolved PIP boxes in their promoter regions thus ensuring their efficient expression during plant colonization.

Basic pathogenicity factors, such as the Xps type II secretion system and its secreted cell wall-degrading enzymes, as well as the *rpf* gene cluster, were probably already present in the ancestor of *Xanthomonas* and *Xylella*. Perhaps, these components were already expressed in response to environmental conditions, as we can still

observe for the genes that are controlled by RpfF and/or Clp. When the xanthomonads had separated into distinct genetic clades (Group 1 vs. Group 2), *hrpG* and *hrpX* were acquired, and perhaps evolved to be cross-regulated by the HpaR2/HpaS two-component regulatory system. It is conceivable that the *X. translucens* lineage acquired *hrpX* and *hrpG* together with the Hrp type III secretion system since the two regulatory genes are physically linked the to *hrp* gene cluster (Wichman et al., 2013), reminiscent of the situation in *Ralstonia solanacearum*. Notably, the other early-branching xanthomonads, such as *X. albilineans* and *X. sacchari*, did not acquire *hrpX* nor *hrpG*. In contrast, all other xanthomonads have the oppositely transcribed *hrpG* and *hrpX* genes in the same synteny group between *radA* and *hsp70*, suggesting a unique acquisition event.

Interestingly, there are xanthomonads that possess *hrpG* and *hrpX* but not the Hrp Type III secretion system nor any type III effectors, such as the cannabis pathogen, the new *X. maliensis* clade (Triplett et al., 2015), and a recently sequenced strain of *X. arboricola* (Ignatov et al., 2015). This could be taken as evidence that the Hrp system was either 1) acquired later or 2) got lost in some lineages. We favor the hypothesis that the T3SS and core effectors were gained in the case of the Nyagatare strain and not lost from *X. cannabis* NCPPB 2788 and NCPPB 3753. There seems to be no indication of any remnant features of the T3SS or effectors in the two *X. cannabis* genomes. Moreover the Nyagatare strain possesses the T3SS and only core effectors, which appear to be located on a pathogenicity island flanked by insertion/transposable elements. More systematic genome

sequencing of underexamined genetic lineages will shed light on this evolutionary puzzle.

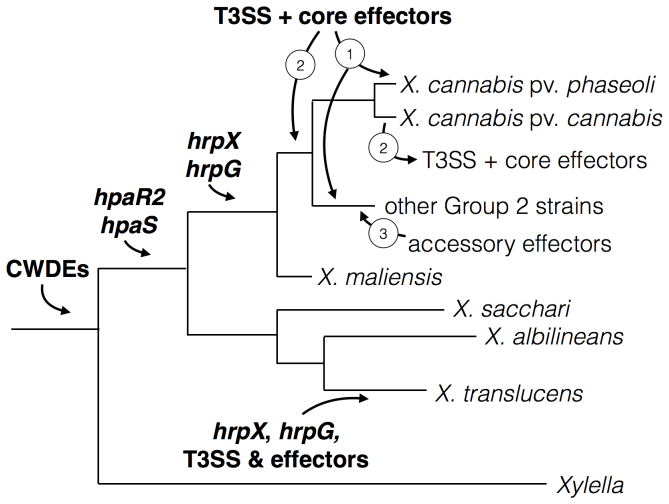


Figure 6. Model for virulence gene and regulator acquisition in *X. cannabis*.

The acquisition of various virulence traits was likely sequential in *Xanthomonas* spp., which represented by a schematic phylogenetic tree. Cell wall-degrading enzymes (CWDEs) is an ancient trait found in both *Xylella* and *Xanthomonas*, which suggests they predate the separation of these two plant-pathogenic genera. Genes encoding the regulators HpaR2 and HpaS were subsequently acquired before the separation of the *Xanthomonas* Group 2 and early branching Group 1. Then the regulatory genes *hrpX* and *hrpG* were gained by Group 2 strains and further promoters (e.g. PIP boxes) evolved to adapt to these virulence regulators. This is further supported by the absence of PIP boxes in front of non-*hrp* genes in Group 1 strains (*X. albilineans* and *X. sacchari*). Because the organization of *hrpX* and *hrpG* and the T3SS is more similar to *R. solanacearum*, different than Group 2 xanthomonads and lacking in most Group 1 species (*X. sacchari* and *X. albilineans*), we suspect an independent acquisition of this system in *X. translucens*. *hrpG* and *hrpX* are present in a similar location in all sequenced Group 2 species. We hypothesize that the T3SS and core effectors were either (1) acquired independently by individual pathovars or (2) acquired an earlier point and lost in some pathovars. After the acquisiton of the T3SS, we posit that accessory effectors were acquired (3) by horizontal gene transfer to alter host range and/or promote susceptibility by suppressing plant immunity.

Acknowledgments

This work benefited from a grant from the National Science Foundation to J.J., a grant from the Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture (093604) to C.P., and a grant from the Agence Nationale de la Recherche (ANR-2010-BLAN-1723) to R.K. We are grateful to Lionel Moulin, IRD Montpellier, for advice on ANI analyses.

References

Aritua, V., Musoni, A., Kato, F., Abang, M. M., Buruchara, R., Sapp, M., et al. (2015). The draft genome sequence of *Xanthomonas species* strain Nyagatare, isolated from diseased bean in Rwanda. *FEMS Microbiol. Lett.*, in press. doi: 10.1093/femsle/fnu055.

Aslam, S. N., Newman, M. A., Erbs, G., Morrissey, K. L., Chinchilla, D., Boller, T., et al. (2008). Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Curr. Biol.* 18, 1078–1083. doi: 10.1016/j.cub.2008.06.061.

Becker, A., Katzen, F., Pühler, A., and Ielpi, L. (1998). Xanthan gum biosynthesis and application: a biochemical/genetic perspective. *Appl. Microbiol. Biotechnol.* 50, 145–152.

Bent, A. F., and Mackey, D. (2007). Elicitors, effectors, and *R* genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45, 399–436. doi: 10.1146/annurev.phyto.45.062806.094427.

Bolot, S., Munoz Bodnar, A., Cunnac, S., Ortiz, E., Szurek, B., Noël, L. D., et al. (2013). Draft genome sequence of the *Xanthomonas cassavae* type strain CFBP 4642. *Genome Announc.* 1:e00679-13. doi: 10.1128/genomeA.00679-13.

Borodovsky, M., and Lomsadze, A. (2014). Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Curr. Protoc. Microbiol.* 32:7. doi: 10.1002/9780471729259.mc01e07s32.

Büttner, D. (2012). Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 76, 262–310. doi: 10.1128/MMBR.05017-11.

Büttner, D., and Bonas, U. (2010). Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiol. Rev.* 34, 107-133. doi: 10.1111/j.1574-6976.2009.00192.x.

Canonne, J., Marino, D., Jauneau, A., Pouzet, C., Brière, C., Roby, D., et al. (2011). The *Xanthomonas* type III effector XopD targets the *Arabidopsis* transcription

factor MYB30 to suppress plant defense. *Plant Cell* 23, 3498–3511. doi: 10.1105/tpc.111.088815.

Clarke, C. R., Chinchilla, D., Hind, S. R., Taguchi, F., Miki, R., Ichinose, Y., et al. (2013). Allelic variation in two distinct *Pseudomonas syringae* flagellin epitopes modulates the strength of plant immune responses but not bacterial motility. *New Phytol.* 200, 847–860. doi: 10.1111/nph.12408.

Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. *fuscans* strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469. doi: 10.1093/nar/gkn180.

Dong, Q., and Ebright, R. H. (1992). DNA binding specificity and sequence of *Xanthomonas campestris* catabolite gene activator protein-like protein. *J. Bacteriol.* 174, 5457–5461.

Dow, M. (2008). Diversification of the function of cell-to-cell signaling in regulation of virulence within plant pathogenic xanthomonads. *Sci. Signal.* 1:pe23. doi: 10.1126/stke.121pe23.

Felix, G., Duran, J. D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* 18, 265–276. doi: 10.1046/j.1365-313X.1999.00265.x.

Figueras, M. J., Beaz-Hidalgo, R., Hossain, M. J., and Liles, M. R. (2014). Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc*. 2:e00927-14. doi: 10.1128/genomeA.00927-14.

Furutani, A., Nakayama, T., Ochiai, H., Kaku, H., Kubo, Y., and Tsuge, S. (2006). Identification of novel HrpXo regulons preceded by two cis-acting elements, a plant-inducible promoter box and a –10 box-like sequence, from the genome database of *Xanthomonas oryzae* pv. *oryzae*. *FEMS Microbiol. Lett.* 259, 133–141. doi: 10.1111/j.1574-6968.2006.00265.x

Gardiner, D. M., Upadhyaya, N. M., Stiller, J., Ellis, J. G., Dodds, P. N., Kazan, K., *et al.* (2014). Genomic analysis of *Xanthomonas translucens* pathogenic on wheat and barley reveals cross-kingdom gene transfer events and diverse protein delivery systems. *PLoS One* 9:e84995 doi: 10.1371/journal.pone.0084995.

Guo, Y., Figueiredo, F., Jones, J., and Wang, N. (2011). HrpG and HrpX play global roles in coordinating different virulence traits of *Xanthomonas axonopodis* pv. *citri*. *Mol. Plant Microbe Interact.* 24, 649–661. doi: 10.1094/MPMI-09-10-0209.

Hauben, L., Vauterin, L., Swings, J., and Moore, E. R. (1997). Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int. J. Syst. Bacteriol.* 47, 328–335. doi: 10.1099/00207713-47-2-328.

Hernandez, D., Tewhey, R., Veyrieras, J. B., Farinelli, L., Østerås, M., François, P., et al. (2014). *De novo* finished 2.8 Mbp *Staphylococcus aureus* genome assembly from 100 bp short and long range paired-end reads. *BioInformatics* 30, 40–49. doi: 10.1093/bioinformatics/btt590.

Hsiao, Y. M., Zheng, M. H., Hu, R. M., Yang, T. C., and Tseng, Y. H. (2008). Regulation of the pehA gene encoding the major polygalacturonase of *Xanthomonas campestris* by Clp and RpfF. *Microbiology* 154, 705–713. doi: 10.1099/mic.0.2007/012930-0.

Ignatov, A. N., Kyrova, E. I., Vinogradova, S. V., Kamionskaya, A. M., Schaad, N. W., Luster, D. G. (2015). Draft genome sequence of *Xanthomonas arboricola* strain 3004, a causal agent of bacterial disease on barley. *Genome Announc.* 3:e01572-14. doi: 10.1128/genomeA.01572-14.

Jacobs, J. M., Babujee, L., Meng, F., Milling, A., and Allen, C. (2012). The *in planta* transcriptome of *Ralstonia solanacearum*: conserved physiological and virulence strategies during bacterial wilt of tomato. *MBio* 3:e00114-12. doi: 10.1128/mBio.00114-12.

Jahn, C., Charkowski, A., and Willis, D. K. (2008). Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J. Microbiol. Methods*. 75, 318–324. doi: 10.1016/j.mimet.2008.07.004.

Kaewnum, S., Prathuangwong, S., and Burr, T. J. (2006). A pectate lyase homolog, *xagP*, in *Xanthomonas axonopodis* pv. *glycines* is associated with hypersensitive response induction on tobacco. *Phytopathology* 96, 1230–1236. doi: 10.1094/PHYTO-96-1230.

Koebnik, R., Krüger, A., Thieme, F., Urban, A., and Bonas, U. (2006). Specific binding of the *Xanthomonas campestris* pv. vesicatoria AraC-type transcriptional activator HrpX to plant-inducible promoter boxes. *J. Bacteriol.* 188, 7652–7660. doi: 10.1128/JB.00795-06.

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A*. 102, 2567–2572. doi: 10.1073/pnas.0409727102.

Li, R. F., Lu, G. T., Li, L., Su, H. Z., Feng, G. F., Chen, Y., et al. (2014). Identification of a putative cognate sensor kinase for the two-component response regulator HrpG, a key regulator controlling the expression of the *hrp* genes in *Xanthomonas campestris* pv. *campestris*. *Environ. Microbiol.* 16, 2053–2071. doi: 10.1111/1462-2920.12207.

Li, S., Wang, Y., Wang, S., Fang, A., Wang, J., Liu, L., et al. (2015). The type III effector AvrBs2 in *Xanthomonas oryzae* pv. *oryzicola* suppresses rice immunity

and promotes disease development. *Mol. Plant Microbe Interact.*, in press. doi: 10.1094/MPMI-10-14-0314-R.

Lu, H., Patil, P., Van Sluys, M. A., White, F. F., Ryan, R. P., Dow, J. M., et al. (2008). Acquisition and evolution of plant pathogenesis-associated gene clusters and candidate determinants of tissue-specificity in *Xanthomonas*. *PLoS One* 3:e3828. doi: 10.1371/journal.pone.0003828.

McPartland, J. M., Clarke, R. C., and Watson, D. P. (2000). Hemp Diseases and Pests: Management and Biological Control: An Advanced Treatise. CABI Publishing, Oxfordshire, U.K. doi 10.1079/9780851994543.0000.

Mhedbi-Hajri, N., Jacques, M. A., and Koebnik, R. (2011). Adhesion mechanisms of plant-pathogenic *Xanthomonadaceae*. *Adv. Exp. Med. Biol.* 715, 71–89. doi: 10.1007/978-94-007-0940-9_5.

Mole, B. M., Baltrus, D. A., Dangl, J. L., and Grant, S. R. (2007). Global virulence regulation networks in phytopathogenic bacteria. *Trends Microbiol*. 15, 363–371. doi: 10.1016/j.tim.2007.06.005.

Netsu, O., Kijima, T., and Takikawa, Y. (2014). Bacterial leaf spot of hemp caused by *Xanthomonas campestris* pv. *cannabis* in Japan. *J. Gen. Plant Pathol.* 80, 164–168. doi: 10.1007/s10327-013-0497-8.

Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J., and Stead, D. (2007). Phylogenetic analysis of *Xanthomonas* species by comparison of partial gyrase B gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2881–2887. doi: 10.1099/ijs.0.65220-0.

Parkinson, N., Cowie, C., Heeney, J., and Stead, D. (2009). Phylogenetic structure of *Xanthomonas* determined by comparison of *gyrB* sequences. *Int. J. Syst. Evol. Microbiol.* 59, 264–274. doi: 10.1099/ijs.0.65825-0.

Patil, P. B., Bogdanove, A. J., and Sonti, R. V. (2007). The role of horizontal transfer in the evolution of a highly variable lipopolysaccharide biosynthesis locus in xanthomonads that infect rice, citrus and crucifers. *BMC Evol. Biol.* 7:243. doi: 10.1186/1471-2148-7-243.

Pieretti, I., Royer, M., Barbe, V., Carrere, S., Koebnik, R., Cociancich, S., et al. (2009). The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited *Xanthomonadaceae*. *BMC Genomics* 10:616. doi: 10.1186/1471-2164-10-616.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106.

Rossez, Y., Wolfson, E. B., Holmes, A., Gally, D. L., and Holden, N.J. (2015). Bacterial flagella: twist and stick, or dodge across the kingdoms. *PLoS Pathog.* 11:e1004483. doi: 10.1371/journal.ppat.1004483.

Schulze, S., Kay, S., Büttner, D., Egler, M., Eschen-Lippold, L., Hause, G., et al. (2012). Analysis of new type III effectors from *Xanthomonas* uncovers XopB and XopS as suppressors of plant immunity. *New Phytol.* 195, 894–911. doi: 10.1111/j.1469-8137.2012.04210.x.

Severin, V. (1978). Ein neues Bakterium an Hanf – *Xanthomonas campestris* pathovar *cannabis*. *Arch. Phytopathol. Pflanzenschutz* 14, 7–15.

Shi, Q., Febres, V. J., Jones, J. B., and Moore, G. A. (2015). Responsiveness of different citrus genotypes to the *Xanthomonas citri* ssp. *citri*-derived pathogen-associated molecular pattern (PAMP) flg22 correlates with resistance to citrus canker. *Mol. Plant Pathol.*, in press. doi: 10.1111/mpp.12206.

Sinha, D., Gupta, M. K., Patel, H. K., Ranjan, A., and Sonti, R. V. (2013). Cell wall degrading enzyme induced rice innate immune responses are suppressed by the type 3 secretion system effectors XopN, XopQ, XopX and XopZ of *Xanthomonas oryzae* pv. *oryzae*. *PLoS One* 8:e75867. doi: 10.1371/journal.pone.0075867.

Stork, W., Kim, J. G., and Mudgett, M. B. (2015). Functional analysis of plant defense suppression and activation by the *Xanthomonas* core type III effector XopX. *Mol. Plant Microbe Interact.* 28, 180–194. doi: 10.1094/MPMI-09-14-0263-R.

Studholme, D. J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., et al. (2011). Draft genome sequences of *Xanthomonas sacchari* and two banana-associated xanthomonads reveal insights into the *Xanthomonas* group 1 clade. *Genes (Basel)* 2, 1050–1065. doi: 10.3390/genes2041050.

Sun, W., Dunning, F. M., Pfund, C., Weingarten, R., and Bent, A. F. (2006). Within-species flagellin polymorphism in *Xanthomonas campestris* pv *campestris* and its impact on elicitation of *Arabidopsis FLAGELLIN SENSING2*-dependent defenses. *Plant Cell* 18, 764–779. doi: 10.1105/tpc.105.037648.

Tang, X., Xiao, Y., and Zhou, J. M. (2006). Regulation of the type III secretion system in phytopathogenic bacteria. *Mol. Plant Microbe Interact.* 19, 1159–1166. doi: 10.1094/MPMI-19-1159.

Triplett, L. R., Verdier, V., Campillo, T., Van Malderghem, C., Cleenwerck, I., Maes, M., et al. (2015). Characterization of a novel clade of *Xanthomonas* isolated from rice leaves in Mali and proposal of *Xanthomonas maliensis* sp. nov. *Antonie Van Leeuwenhoek* 107, 869–881. doi: 10.1007/s10482-015-0379-5.

Tsuchiya, K., Mew, T. W., and Wakimoto, S. (1982). Bacteriological and pathological characteristics of wild types and induced mutants of *Xanthomonas campestris* pv. *oryzae*. Phytopathology 72, 43–46.

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T.R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12:R102. doi: 10.1186/gb-2011-12-10-r102.

Wang, L., Rong, W., and He, C. (2008). Two *Xanthomonas* extracellular polygalacturonases, PghAxc and PghBxc, are regulated by type III secretion regulators HrpX and HrpG and are required for virulence. *Mol. Plant Microbe Interact.* 21, 555–563. doi: 10.1094/MPMI-21-5-0555.

Wengelnik, K., Van den Ackerveken, G., and Bonas, U. (1996). HrpG, a key hrp regulatory protein of *Xanthomonas campestris* pv. *vesicatoria* is homologous to two-component response regulators. *Mol. Plant Microbe Interact*. 9, 704–712. doi: 10.1094/MPMI-9-0704.

Wengelnik, K., Rossier, O., and Bonas, U. (1999). Mutations in the regulatory gene *hrpG* of *Xanthomonas campestris* pv. *vesicatoria* result in constitutive expression of all *hrp* genes. *J. Bacteriol*. 181, 6828–6831.

White, F. F., Potnis, N., Jones, J. B., and Koebnik, R. (2009). The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* 10, 749–766. doi: 10.1111/j.1364-3703.2009.00590.x.

Wichmann, F., Vorhölter, F., Hersemann, L., Widmer, F., Blom, J., Niehaus, K., et al. (2013). The noncanonical type III secretion system of *Xanthomonas translucens* pv. *graminis* is essential for forage grass infection. *Mol. Plant Pathol*. 14, 576–588. doi: 10.1111/mpp.12030.

Young, J. M., Park, D. C., Shearman, H. M., and Fargier, E. (2008). A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31, 366–377. doi: 10.1016/j.syapm.2008.06.004.

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107.

Zhao, S., Poulin, L., Rodriguez-R, L. M., Forero-Serna, N., Liu, S. H., Wonni, I., et al. (2012). Development of a variable number of tandem repeats typing scheme for the bacterial rice pathogen *Xanthomonas oryzae* pv. *oryzicola*. *Phytopathology* 102, 948–956. doi: 10.1094/PHYTO-04-12-0078-R.

Supplemental Figure 1. Flg22 epitope variants found in various strains of *Xanthomonas*.

On top, the prototype elicitor-active flg22 peptide from *Pseudomonas aeruginosa* is shown (Felix et al., 1999). Below, homologous sequences from different species and pathovars of *Xanthomonas* are aligned. Residues that deviate from the prototype sequence are in red. The peptide that corresponds to eliciting flagellin variants of *X. campestris* pv. *campestris* is highlighted in green, while the peptide that corresponds to non-eliciting flagellin variants highlighted in yellow (Sun et al., 2006).

QRLSTGSRINSAKDDAAGLQIA	Prototype flg22 peptide from *Pseudomonas aeruginosa* (Felix et al., 1999)
QQLSSGKRITSAAVDAAGLAIS	*X. translucens* pv. *cerealis* CFBP 2541
QQLSSGKRITSASADAAGLAIS	*X. campestris* pv. *campestris* B161
QQLSSGKRITSASVDAAGLAIS	*X. campestris* pv. *campestris* 8004, ATCC 33913, B100, B112, B186, etc.
QQLSSGKRITSASVDAAGLAIS	*X. gardneri* ATCC 19865
QQLSSGKRITSASVDAAGLAIS	*X. translucens* pv. *translucens* DSM 18974
QQLSSGKRITSASVDAASMAIS	*X. translucens* DAR61454
QQLSSGKRITSFAVDAAGGAIA	*X. arboricola* pv. *pruni* MAFF 301420, MAFF 301427, Xap33
QQLSSGKRITSFAVDAAGGAIA	*X. axonopodis* pv. *citrumelo* F1
QQLSSGKRITSFAVDAAGGAIA	*X. axonopodis* pv. *vasculorum* NCPPB 900
QQLSSGKRITSFAVDAAGGAIA	***X. cannabis* pv. *cannabis* NCPPB 2877, NCPPB 3753**
QQLSSGKRITSFAVDAAGGAIA	*X. euvesicatoria* 85-10, 66b, 83M
QQLSSGKRITSFAVDAAGGAIA	*X. hortorum* pv. *carotae* M081
QQLSSGKRITSFAVDAAGGAIA	*X. oryzae* pv. *oryzae* MAFF 311018, PXO86, PXO99^{A}, etc.
QQLSSGKRITSFAVDAAGGAIA	*X. oryzae* pv. *oryzicola* CFBP 7342, BLS256
QQLSSGKRITSFAVDAAGGAIA	*X. perforans* 91-118
QQLSSGKRITSFAVDAAGGAIA	*X. vasicola* pv. *holcicola* NCPPB 989, NCPPB 1060, NCPPB 1241, etc.
QQLSSGKRITSFAVDAAGGAIA	*X. vasicola* pv. *musacearum* NCPPB 2005, NCPPB 4379, NCPPB 4392, etc.
QQLSSGKRITSFAVDAAGGAIA	*X. vasicola* pv. *vasculorum* NCPPB 206, NCPPB 1326, NCPPB 1381, etc.
QQLSSGKRITSFSVDAAGGAIA	*X. fragariae* LMG 25863
QQLSSGKRITSFSVDAAGGAIA	*X. oryzae* pv. *oryzae* KACC 10331
QQLSSGKRITSFSVDAAGGAIA	*X. vesicatoria* 15b, 53M, ATCC 35937
QQLSSGKRIVSASVDAAGLAIS	*X. cassavae* CFBP 4642
QQLSSGKRIVSASVDAAGLAIS	*X. maliensis* M97
QQLSTGKRVNSAKDDAAAINIG	*X. albilineans* GPE PC73
QRLSSGLRINSAKDDAAGLAIS	*X. arboricola* pv. *celebensis* NCPPB 1630, NCPPB 1832
QRLSSGLRINSAKDDAAGLAIS	*X. axonopodis* pv. *phaseoli* NCPPB 556, NCPPB 670, NCPPB 2665, etc.
QRLSSGLRINSAKDDAAGLAIS	*X. axonopodis* pv. *punicae* LMG 859
QRLSSGLRINSAKDDAAGLAIS	*X. campestris* pv. *campestris* B94, B110, B111, B127, B305, etc.
QRLSSGLRINSAKDDAAGLAIS	*X. campestris* pv. *raphani* 756C, CFBP 5828
QRLSSGLRINSAKDDAAGLAIS	***X. cannabis* pv. *phaseoli* Nyagatare**
QRLSSGLRINSAKDDAAGLAIS	*X. citri* pv. *citri* 306, A^{w}12879, NCPPB 3610, etc.
QRLSSGLRINSAKDDAAGLAIS	*X. citri* pv. *glycines* 8ra
QRLSSGLRINSAKDDAAGLAIS	*X. citri* pv. *malvacearum* GSPB1386941, GSPB2388
QRLSSGLRINSAKDDAAGLAIS	*X. citri* pv. *mangiferaeindicae* LMG 941
QRLSSGLRINSAKDDAAGLAIS	*X. fuscans* subsp. *aurantifolii* ICPB 10535, ICPB 11122
QRLSSGLRINSAKDDAAGLAIS	*X. sacchari* R1
*.**:* *: * ***. *.	

Supplemental Figure 2. Comparison of promoter sequences of the *Xanthomonas* lysophospholipase gene.

Promoter regions encompassing 350 bp upstream of the translational start codon of a representative set of *Xanthomonas* strains were aligned by MUSCLE. PIP half boxes are shown in blue and the –10 promoter motif is shown in orange. Distance to the translational start codon is indicated on the right side of the lower sequence block. Deviations from the PIP consensus sequence are highlighted in yellow. For the set of analyzed strains, compare with Figure 4.

Supplemental Figure 3. Comparison of promoter sequences of the *Xanthomonas pehA* gene.

Promoter regions encompassing 350 bp upstream of the translational start codon of a representative set of *Xanthomonas* strains were aligned by MUSCLE. Consensus CAP/CLP binding boxes, according to Dong and Ebright, 1992, are indicated above the multiple sequence alignment, first aligned with the CLB binding box as determined by Hsiao et al., 2005, for X. campestris pv. campestris (highlighted in yellow, six conserved residues are shown in green), and then aligned with the CLB binding box as proposed by us (conserved residues are shown in red). The PIP half boxes are shown in blue and the -10 promoter motif is shown in orange. Distance to the translational start codon is indicated on the right side of the lower sequence block. For better comparison with Figure 4, PIP half boxes are shown in blue and the –10 promoter motif is shown in orange. Distance to the translational start codon is indicated on the right side of the lower sequence block. Deviations from the PIP consensus sequence are highlighted in yellow. For the set of analyzed strains, compare with Figure 4.

```
CAP/CLP box             aanTGT---GAnnnnnnTCACAntt
CAP/CLP box                       aanTGTGAnnnnnnTCACAntt
XCC_ATCC33913       TGACCACGT---GGATCTGCGCACTTGGTAGCAACAAGGAAGC-GCGGTGGCATGCTGCTTCGT
XE_85-10            CACCCGTGTCGT-GTTTTGCGCGCTGCGTTGCAAGGAGCGCGCAGCGCGCGGGCGTGACTTCGC
XCC_306             CACCCGTGTCGT-GTTTTGCGCGCTGCGTTGCAAGGAGCGCGCAGCGCGCGGGCGTGACTTCGC
XC_CFBP4642         CACCTGTGGCGT-GTTTTGCGCGCTGCGTTGCAAGGAGCGCGCAGCGCGCGGGCGTGACTTCGC
XVV_NCPPB206        CAATCACAC-ACAGCTCTGCGAACTGCATTGCAACGACGACATAGTGGCGCAACGCTGCTTCGT
XF_LMG25863         TGAGTACGC-ACCGTTCTGCGTGCTGCATTGCAATGCCGACATACCAGCGCAGTGCC-CTTCGT
XAC_NCPPB1832       AAGGCGCGC-GCGGTTCTGCGCGCTGCGTTGCAATGACGACATCCCGGCGCAGGGCCGCTTCGC
XHC_M081            AAATCGCGC-ACCGTTCTGCGCGCTGCGTTGCAATGCCGAAACACCGTCGCGGTGCCGCTTCGT
XG_ATCC19865        AAATTGCGC-AAAGTTCTGCGCGCTACGTTGCAATGACGACACACCGCTGCAGTGCGGCTTCGT
XAM_CIO1            CAATCGCGTCGCGTTTTTGCGCAGTGCATTGCAACGCCGATACACCGGCGTAGGCTTGCTTCGT
XCC_NCPPB2877/3753  TGGGCGCATCGCGTTTTTGCGCAGTGCATAGCAACGACGCGACACCGGCGTAGGCCTGCTTCGT
XCP_Nyagatare       CGGGCGCATCGCGTTTTTGCGCAGTGCATTGCAACGACGCGACACCGGGGTAAGCCTGCTTCGC
XV_ATCC35937        CGATCACGCAGC-GTTTTGCGCGTTGCATTGCAATAGCGACACAGCGGCGCAGGGCGGCTTCGC
XFF_4834-R          AGATTGCGCGTTAGTTGTGCGCACTCCATTGCAACGACGACACAGCGGCGCAGTGCCGCTTCGT
                                  * ****   *   * ****                          *****

XCC_ATCC33913       TTGCACGGCAGCCGGTTCGCGCGCTGCCATCGCCCCCGGTCGACGCGCTGTAGAGTG---- 80
XE_85-10            TCCCACGGTGTCCGATTCGCTCGTCTGCTACAGGCGCGCACCACGCTCCGTAGAGTC---- 82
XCC_306             TCCCACGGTGTCCGATTCGCTCGTCTGCTACAGGCGCGCACCACGCTCCGTAGAGTC---- 83
XC_CFBP4642         TCGCACGGTGCCCGATTCGCTCATCTGTTACGGGCGCACCCCACGCTCCGTAGAGTC---- 80
XVV_NCPPB206        TCTTGCGGCGTCCGTTTCGCTCGCTTTGCACGCATGCGGGCAGTGCTTCGTAGAGTCGCCT 87
XF_LMG25863         TCGTACGGCGTCCGGTTCGTTCGCCCTTTGCGCGCATGGGTATCGATCGGTAGAGTC---- 82
XAC_NCPPB1832       GCATGCGGTATCGGGTTCGCTCGCCTTCTGTGAATAGGCGCGTCGCTCAGTAGAGTC---- 81
XHC_M081            TCGCGCGGTTCCGGGTTCGCTTGCCTTGCGCATGTGCGGTGATTGCTCCGTAGAGTC---- 81
XG_ATCC19865        TCGTGCGGCGGCGGGTTCGCTTGCCTTGCGCATGTACGGGCGTTGCTCCGTAGAGTT---- 81
XAM_CIO1            TCGTGTGGCGCCGGGTTCGCGTGCGCTCTGCATGCACGGACAGTGCTCCTTAGAGT----- 92
XCC_NCPPB2877/3753  TCGTGCGGTGCCGGGTTCGCGCGCCTGCTGCATGCGGGCACAGCGCTTCGTAGAGTC---- 80
XCP_Nyagatare       TCGTGCGGTGCCGGGTTCGCGCGCCTGCTGCATGCGGGCACAGCGCTTCGTAGAGTC---- 81
XV_ATCC35937        GCCCGCGGCATCGGGTTCGCCCACGTTCCGCATGTGCGGCGCGTGCTCCGTAGAGT----- 81
XFF_4834-R          TCGCGCGGCGCCCGGTTCGCTCGCGGTCTGCGCACGCAGGCGCTGCTCCGTAGAGTC---- 80
                          **   * * ****                          *     ******
```

Supplemental Table S1. Primers used for qPCR.

Primer	Sequence	Gene target
2JMJ_0053	CAACCATGTGTTCGTCGCTG	*pehA* (Forward)
2JMJ_0054	CTGAGTGCCGCCTGGATG	*pehA* (Reverse)
2JMJ_0055	AATCGCTCACCACCGAAGTG	Lysophospholipase (Forward)
2JMJ_0056	CGGTCCACACCTTGTCGAAG	Lysophospholipase (Reverse)
2JMJ_0059	CGACCTGTTTTACCCCGGAA	*pehD* (Forward)
2JMJ_0060	GCCGATGTACGAATAGGCCA	*pehD* (Reverse)
2JMJ_0065	GAACGCGCCATTTCGGTG	*atpD* (Forward)
2JMJ_0066	GTCACCGGCTTCGTCGAT	*atpD* (Reverse)

Appendix B

Scientific communications

Oral presentations

- Pesce C., Streubel J., Szurek B., Koebnik R., and Boch J. (January 2012) Systematic functional analysis of *SWEET / nodulin-3* family members as susceptibility genes of rice to *Xanthomonas oryzae* pv. *oryzae*. **10èmes Rencontres Plantes-Bactéries** - Aussois (France)
- Pesce C., Streubel J., Koebnik R., Boch J., and Szurek B. (June 2012) *SWEET / nodulin-3* family members as susceptibility genes of rice to *Xanthomonas oryzae* pv. *oryzae*. **EMBO-Workshop "Plants-Microbes Interactions"** - Norwich (UK)
- Streubel J., Pesce C., Hutin M., Ortiz E., Pérez-Quintero A., Rodríguez-R L.M., Cunnac S., Koebnik R., Boch J., and Szurek B. (July 2012). Several rice *SWEET / nodulin-3* family members can function as virulence targets for *Xanthomonas* TAL effectors. **4th Xanthomonas Genomics Conference** - Angers (France)
- Boch J., Boureau T., Brin C., Cunnac S., David P., Feng J.X., Hajri A., Hutin M., Koebnik R., Mo W.L., Pesce C., Poussier S., Streubel J., Szurek B., Tang J.L., Tang W., Tran T.T., Verdier V., Wu F., and Zhao S. (August 2013). Contribution of type III/TAL effectors to pathogenicity. **APS-MSA Joint Meeting** - Austin, TX (U.S.A.)
- Pesce C., Streubel J., Koebnik R., Boch J., and Szurek B. (October 2012) Systematic functional analysis of *SWEET / nodulin-3* family members as susceptibility genes of rice to *Xanthomonas oryzae* pv. *oryzae*. **Effectome Meeting** - Lauret (France)
- Streubel J., Pesce C., Hutin M., Koebnik R., Boch J., and Szurek B. (October 2013) Rice *SWEETs* as virulence targets for *Xanthomonas* TALEs. **Annual Meeting of the COST Action SUSTAIN** - Birnam (Scotland)
- Pesce C., Bragard C., and Koebnik R. (October 2013) Virulence determinants involved in the *Xanthomonas translucens* / cereals interaction (TAL effectors roles). **Annual Meeting of the French Network on Xanthomonads** - Moulis (France)
- Pesce C., Streubel J., Hutin M., Szurek B., Koebnik R., and Boch J. (November 2013) Five phylogenetically close rice *SWEET* genes

confer TAL effector-mediated susceptibility to *Xanthomonas oryzae* pv. *oryzae*. **LAMFU-Manihot Biotec Meeting** - Sasaima (Colombia)

- Pesce C., Bragard C., and Koebnik R. (November 2013) Virulence determinants involved in the *Xanthomonas translucens* / cereals interaction (TAL effectors roles). **LAMFU-Manihot Biotec Meeting** - Sasaima (Colombia)
- Pesce C., Bragard C., and Koebnik R. (April 2014) Virulence determinants involved in the *Xanthomonas translucens* / cereals interaction (TAL effectors roles). **Workshop on "Cellular dynamics of effector action and recognition" of the COST Action SUSTAIN** - Toulouse (France)
- Pesce C., Berthelot E., Bragard C., and Koebnik R. (October 2014) Virulence determinants involved in the *Xanthomonas translucens* / cereals interaction - **Annual Meeting of the French Network on Xanthomonads** - Briandes (France)
- Pesce C., Bragard C., and Koebnik R. (October 2014) Virulence determinants involved in the *Xanthomonas translucens* / cereals interaction (TAL effector). **Effectome Meeting** - Lauret (France)
- Pesce C., Streubel J., Berthelot E., Hutin M., Boch J., Szurek B., Bragard C., and Koebnik R. (July 2015) Transcriptional Activator-Like (TAL) effectors: genetic tools and their roles in pathogenicity during *Xanthomonas translucens* / cereals interaction. **Seminar at CIAT** - Cali (Colombia)
- Pesce C., Berthelot E., Bragard C., and Koebnik R. (July 2015) *Xanthomonas translucens* – a role of type III effectors in pathogenicity. **5th *Xanthomonas* Genomics Conference** - Bogota (Colombia)
- Jacobs J.M., Pesce C., Bragard C., and Koebnik R. (July 2015) Yes we Can – Cannabis as a promising host to study pathogenicity of xanthomonads. **5th *Xanthomonas* Genomics Conference** - Bogota (Colombia)
- Jacobs J.M., Pesce C., Bragard C., and Koebnik R. (August 2015) Pots on pot: Comparative genomics of two cannabis pathogens reveals insights into the evolution of *Xanthomonas* pathogenicity. **APS-MSA Joint Meeting** - Pasadena, California (USA)
- Jacobs J.M., Pesce C., Vancheva T., Lefeuvre P., Barak-Cunningham J., and Koebnik R. (August 2015) Evolution of pathogenicity beyond type III effectors in *Xanthomonas*. **Workshop on "Evolutionary genomics of plant pathogens" of the COST Action SUSTAIN -** Kiel (Germany)

Poster presentation

- Pesce C., Streubel J., Ortiz E., Szurek B., and Boch J. (June 2012) Systematic functional analysis of *SWEET / nodulin-3* family members as susceptibility genes of rice to *Xanthomonas oryzae* pv. *oryzae* **EMBO-Workshop "Plants-Microbes Interactions"** - Norwich (UK)
- Pesce C., Bragard C., and Koebnik R. (January 2014) *Xanthomonas translucens* – Do type III effectors and particularly TAL effectors contribute to pathogenicity? **11èmes Rencontres Plantes-Bactéries** - Aussois (France)
  * Jury price of the French Society of Phytopathology
- Pesce C., Bragard C., and Koebnik R. (April 2014). *Xanthomonas translucens* – Do type III effectors and particularly TAL effectors contribute to pathogenicity? **Workshop on "Cellular dynamics of effector action and recognition" of the COST Action SUSTAIN** - Toulouse (France)
- Pesce C., Bragard C., and Koebnik R. (June 2014) *Xanthomonas translucens* – Do type III effectors and particularly TAL effectors contribute to pathogenicity? **EUCARPIA Cereals Section – I·T·M·I Joint Conference -** Wernigerode (Germany)
- Pesce C., Berthelot E., Snelling J., Tisserat N., Bragard C., and Koebnik R. (October 2014) Genomics-based effector mining in *Xanthomonas translucens* - Do type III effectors and particularly TAL effectors contribute to pathogenicity? **Effectome meeting** - Lauret (France)
- Pesce C., Berthelot E., Snelling J., Tisserat N., Bragard C., and Koebnik R. (October 2014) Genomics based effector mining in *Xanthomonas translucens* - Do type III effectors and particularly TAL effectors contribute to pathogenicity? **Annual Meeting of the COST Action SUSTAIN** - Zakopane (Poland)

Mentoring of students

Edwige Berthelot – Master's student. Thesis title: Identification and conservation of new type III effectors in *Xanthomonas translucens*. March 3 – June 27, 2014.