New Phytologist

Research

Two ancestral genes shaped the *Xanthomonas campestris* TAL effector gene repertoire

Nicolas Denancé[1]*, Boris Szurek[2]*, Erin L. Doyle[3,4], Emmanuelle Lauber[1], Lisa Fontaine-Bodin[2], Sébastien Carrère[1], Endrick Guy[1], Ahmed Hajri[5], Aude Cerutti[1], Tristan Boureau[5], Stéphane Poussier[5], Matthieu Arlat[1], Adam J. Bogdanove[3,6] and Laurent D. Noël[1]

[1]LIPM, Université de Toulouse, INRA, CNRS, UPS, F-31326 Castanet-Tolosan Cedex, France; [2]IRD, Cirad, Univ. Montpellier, IPME, Montpellier, France; [3]Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA 50011, USA; [4]Department of Biology, Doane University, Crete, NE 68333, USA; [5]IRHS, INRA, AGROCAMPUS-Ouest, Université d'Angers, SFR 4207 QUASAV, 49071 Beaucouzé Cedex, France; [6]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA

Authors for correspondence:
Laurent D. Noël
Tel: +33 5 6128 5047
Email: laurent.noel@inra.fr

Boris Szurek
Tel: +33 4 6741 6212
Email: boris.szurek@ird.fr

Received: *1 September 2017*
Accepted: *6 March 2018*

New Phytologist (2018) **219:** 391–407
doi: 10.1111/nph.15148

Key words: black rot, *Brassica rapa*, *Brassicaceae*, Hax, TALE, *Xanthomonas campestris*.

Summary

- *Xanthomonas* transcription activator-like effectors (TALEs) are injected inside plant cells to promote host susceptibility by enhancing transcription of host susceptibility genes. TALE-encoding (*tal*) genes were thought to be absent from *Brassicaceae*-infecting *Xanthomonas campestris* (*Xc*) genomes based on four reference genomic sequences.
- We discovered *tal* genes in 26 of 49 *Xc* strains isolated worldwide and used a combination of single molecule real time (SMRT) and *tal* amplicon sequencing to yield a near-complete description of the TALEs found in *Xc* (*Xc* TALome).
- The 53 sequenced *tal* genes encode 21 distinct DNA binding domains that sort into seven major DNA binding specificities. *In silico* analysis of the *Brassica rapa* promoterome identified a repertoire of predicted TALE targets, five of which were experimentally validated using quantitative reverse transcription polymerase chain reaction. The *Xc* TALome shows multiple signs of DNA rearrangements that probably drove its evolution from two ancestral *tal* genes. We discovered that Tal12a and Tal15a of *Xcc* strain Xca5 contribute together in the development of disease symptoms on susceptible *B. oleracea* var. botrytis cv Clovis.
- This large and polymorphic repertoire of TALEs opens novel perspectives for elucidating TALE-mediated susceptibility of *Brassicaceae* to black rot disease and for understanding the molecular processes underlying TALE evolution.

Introduction

Phytopathogenic bacteria of the *Xanthomonas* genus include 20 species that cause various diseases on > 400 different host plants. *Xanthomonas* species provide excellent models for genomic studies (Ryan & Dow, 2011; Mansfield *et al.*, 2012; Jacques *et al.*, 2016), and numerous *Xanthomonas* complete genome sequences have been obtained (da Silva *et al.*, 2002; Lee *et al.*, 2005; Thieme *et al.*, 2005; Salzberg *et al.*, 2008; Vorhölter *et al.*, 2008; Pieretti *et al.*, 2009; Bogdanove *et al.*, 2011; Tao *et al.*, 2012). Each *Xanthomonas* species is further divided into pathovars based on host specificity and mode of infection. For instance, *X. campestris* (*Xc*) includes three pathovars isolated on *Brassicaceae* worldwide – *campestris* (*Xcc*), *incanae* (*Xci*) and *raphani* (*Xcr*) – as well as a nonpathogenic group (*Xc*NP) (Fargier & Manceau, 2007).

Like many other pathogenic Gram negative bacteria, *Xc* possesses a type III secretion system (T3SS) that is required for injection of various type III effectors inside plant cells, thus contributing to virulence (Büttner & Bonas, 2010). Transcription activator-like effectors (TALEs), encoded by *tal* genes, are transcription factors that are injected through the T3SS by many *Xanthomonas* to promote virulence. Some *Ralstonia* and *Burkholderia* express TALE-like proteins the function and mode of action of which is starting to be deciphered (de Lange *et al.*, 2013, 2014). *Xanthomonas* TALEs are addressed to the plant nucleus where they activate gene expression upon direct binding to target sequences in the corresponding promoters (Schornack *et al.*, 2013). TALEs share a highly conserved modular structure: an N-terminal T3SS translocation domain, a central repeat domain, two C-terminal nuclear localization signals and an activation domain (Boch & Bonas, 2010). The repeat region mediates DNA-binding specificity, with each repeat specifying one DNA base, with some degeneracy. The base-specificity of each repeat is conferred by a single residue, the second in a pair of hypervariable residues at positions 12 and 13, known as the repeat variable di-residue (RVD) (Boch *et al.*, 2009; Moscou & Bogdanove, 2009;

*These authors contributed equally to this work.

Deng *et al.*, 2012; Mak *et al.*, 2013). Therefore, the RVD sequence of a TALE can be used to computationally predict the DNA sequences (effector-binding elements, EBEs) which the protein might bind, and allow the identification of candidate target genes. Recently it was shown that TALEs are able to drive transcription from EBEs located on either strands of target promoters (Streubel *et al.*, 2017; Wang *et al.*, 2017). Analysis of gene ontologies of candidate targets may shed light on the host physiological responses mediated by a given TALE (Noël *et al.*, 2013).

Xanthomonas genomes contain numbers of *tal* genes ranging from zero (e.g. in *X. vasicola* pv *musacearum*) to more than two dozen (e.g. in some strains of *X. oryzae* pv *oryzicola*, *Xoc*). A few TALE proteins have been reported to significantly contribute to pathogen virulence, albeit to varying extents. Such TALEs include: the PthA series from *X. citri* pv *citri* (*Xcc*) which cause canker on citrus (Swarup *et al.*, 1991); AvrXa7, PthXo1, PthXo2, PthXo3, Tal5 and TalC from *X. oryzae* pv *oryzae* (*Xoo*) strains, which cause bacterial leaf blight on rice (Hutin *et al.*, 2015a); Tal2g from *Xoc*, which promotes bacterial leaf streak on rice (Cernadas *et al.*, 2014); and $TALE1_{Xam}$, which is required for *X. axonopodis* pv *manihotis* (*Xam*) virulence on cassava (Castiblanco *et al.*, 2013). Other TALEs have been shown to contribute in more subtle ways to *Xanthomonas* virulence, such as Avrb6 from the cotton pathogen *X. citri* ssp *malvacearum* (*Xcm*) and AvrHah1 from the tomato pathogen *X. gardneri* (Yang *et al.*, 1994, 1996; Schwartz *et al.*, 2017), PthXo6 and PthXo7 from *Xoo* (Sugio *et al.*, 2007) and $TAL20_{Xam668}$ from *Xam* (Cohn *et al.*, 2014). Hax2, Hax3 and Hax4 from *Xcc* strain Xca5 collectively promote necrosis on radish (Kay *et al.*, 2005). TALE-induced plant targets are defined as susceptibility (*S*) genes if they contribute to disease development (White & Yang, 2009). Examples include sugar transporter genes in bacterial blights of rice, cassava and cotton (Yang *et al.*, 2006; Antony *et al.*, 2010; Yu *et al.*, 2011; Streubel *et al.*, 2013; Cohn *et al.*, 2014; Cox *et al.*, 2017), a sulfate transporter gene in bacterial leaf streak of rice (Cernadas *et al.*, 2014), three members of the *Lateral Organ Boundaries* (*LOB*) gene family of transcription factors in citrus canker (Hu *et al.*, 2014; Zhang *et al.*, 2017), and a bHLH transcription factor that upregulates a pectate lyase gene during bacterial spot of tomato (Schwartz *et al.*, 2017). Besides activating *S* genes, TALEs may act as avirulence (Avr) proteins by activating an executor *R* (Resistance) gene (Boch *et al.*, 2014; Zhang *et al.*, 2015). Examples include AvrBs3 from *X. euvesicatoria* (*Xe*), which activates the pepper *Bs3* gene (Bonas *et al.*, 1989; van den Ackerveken *et al.*, 1996; Römer *et al.*, 2007). Other TALE-activated executor genes include *Xa27*, *Xa10* and *Xa23* in rice (Gu *et al.*, 2005; Tian *et al.*, 2014; Wang *et al.*, 2015), and *Bs4C* from pepper (Strauss *et al.*, 2012). Another type of resistance to *Xanthomonas* involves impairment of TALE activation of *S* genes. Such resistance by loss-of-susceptibility is found in rice, where polymorphisms in *SWEET* gene promoters have been found that prevent their induction by corresponding TALEs (Chu *et al.*, 2006; Yang *et al.*, 2006; Hutin *et al.*, 2015b). Thus, identifying TALE targets can be key to informed breeding or engineering of crops for resistance or decreased susceptibility to *Xanthomonas* (Hutin *et al.*, 2015a; Zhang *et al.*, 2015).

Depending on the species or pathovar, *tal* genes can be either chromosomal or plasmid-borne. They are often localized on mobile insertion cassettes (MICs) (Ferreira *et al.*, 2015) that may promote their transposition and lateral transfer. Little is known about the mechanism underlying the evolution and diversification of *tal* genes, largely because of the high conservation of these DNA sequences. However, it is assumed that gene duplication and accumulation of point mutations play important roles in diversification (Bogdanove *et al.*, 2011). Another mechanism might be through recombination facilitated by the repetitive structure of the central region, leading to variants differing by deletion, insertion or rearrangement of repeat units, as suggested from artificial manipulation experiments (Pérez-Quintero *et al.*, 2015). For example, the generation of repeat region variants with internal deletions has been witnessed *in vitro* as a result of inter- or intragenic recombination (Yang & Gabriel, 1995a; Yang *et al.*, 2005; Lau *et al.*, 2014; Booher *et al.*, 2015), illustrating the high potential for *tal* genes to diversify *in planta* (Vera Cruz *et al.*, 2000).

Here, we report on the identification of a *Xanthomonas campestris* repertoire of 53 *tal* genes encoding TALEs with 21 RVD combinations. The relationships among *tal* genes provide evidence of recombination events having driven the evolution of the TAL repertoire (TALome) from two ancestral *tal* genes. We present evidence for a host species-dependent, cooperative contribution to disease development by two TALEs of *Xcc* strain Xca5. Our results reveal a previously unrecognized prevalence of TALEs in *Xc* strains pathogenic to Brassicaceae, and open the door to the elucidation of Brassicaceae susceptibility to black rot disease.

Materials and Methods

Bacterial strains, plasmids and growth conditions

The *Xanthomonas campestris* (*Xc*) strains used in this study are listed in Supporting Information Table S1 and can be obtained from the CIRM-CFBP collection. All Xca5 strain derivatives were described previously (Kay *et al.*, 2005). *Xc* cells were grown at 28°C in MOKA medium (Blanvillain *et al.*, 2007). *Escherichia coli* cells were grown on Luria–Bertani medium at 37°C. For solid media, agar was added at a final concentration of 1.5% (w/v). Antibiotics were used at the following concentrations: 50 $\mu g\ ml^{-1}$ rifampicin, 50 $\mu g\ ml^{-1}$ kanamycin, 5 $\mu g\ ml^{-1}$ tetracycline, 50 $\mu g\ ml^{-1}$ ampicillin and 40 $\mu g\ ml^{-1}$ spectinomycin.

Xc genotyping methods

Amplified fragment length polymorphism (AFLP) analysis of *Xc* gDNA was performed with the *Sac*I and *Msp*I restriction enzymes as described previously (Ah-You *et al.*, 2007). The presence/absence of fragments was determined using GeneMapper (Applied Biosystems) with the following criteria: size between 60 and 500 bp; peak area > 1000; peak high > 800 relative fluorescence units, no signal in negative controls. The Sac+T/Msp+C primer combination was not analyzed.

Phylogenetic analyses

For the AFLP analysis, phylogenetic distances were calculated from Dice similarity coefficients with 5000 bootstraps and used to construct a weighted neighbor-joining tree using the DARwin software package (http://darwin.cirad.fr/). The robustness of the tree was assessed by bootstrap analysis (5000 resamplings). Average nucleotide identities based on Blast (ANIb) across *Xc* genome sequences were calculated using Jspecies (Richter & Rossello-Mora, 2009).

Dot blot, Southern blot and PCR-based detection of *tal* gene content in *Xc*

gDNA was prepared from liquid cultures as described (Wizard Genomic DNA Purification kit, Promega) and used for dot blot, Southern and PCR analyses. Primers used to detect *tal* sequences were 5′-GGACTAGTCCAGAGCATTGTTGCCCAGTTATCTC-3′ and 5′-CCGCTCGAGCGGGTTCGGTGACGCCCACTCT-3′. Dot blot hybridizations were performed as described (Hajri *et al.*, 2009). Southern hybridization was performed using the 865-bp 5′ region of *tal22a* (also named *hax2*) as probe, amplified from Xca5 genomic DNA using the primers 5′-GGTCTCGATGGATCCCATTCGTTCGCGC-3′ and 5′-CCAGAGTCAGCGTTCAGGGGGGCACCCGT-3′.

Cloning of native *tal* gene repeat domains and sequence analysis

Amplification of *tal* repeat regions was performed using the forward primers 5′-tttggtctcAAGGTTGCCAGAGGCGACACACGAAGACG-3′ or 5′-tttggtctcAAGGTTGCCAGAGGCGACACACGAAGCGA-3′ and the reverse primers 5′-tttggtctcACTTGCTTTTTCACTGCATCCAGCGCAGGA-3′ or 5′-tttggtctcACTTGCTTTCACTGCATTCAGCGCAGAG-3′. PCR amplicons were cloned by GoldenGate cloning using *Bsa*I into the GoldenGate-compatible pK18 derivative pΔ13 (Guy *et al.*, 2013). Sequence analyses including alignments were performed using Geneious software (Kearse *et al.*, 2012). At least two clones from independent PCR reactions were obtained for each *tal* gene.

Xcc genomic DNA extraction and genome sequencing using SMRT technology

Genomic DNA of *Xc* pv *campestris* (*Xcc*) was extracted from 15 ml of overnight culture in MOKA medium essentially as described in Chen & Kuo (1993). RNAse A treatment was performed during the lysis step. A final chloroform extraction was performed. The quality and quantity of gDNA was evaluated on agarose gel and nanodrop. Library preparation (3–20 kb) and sequencing (one single molecule real time (SMRT) cell per strain) was performed at the Yale Center for Genome Analysis (CT, USA). Sequences were assembled using HGAP3 with default parameters and an estimated genome size of 5 Mb. Genome sequences were deposited at GenBank under accession numbers CP017308–CP017309 (CN03), CP017310 (CN12), CP017317–CP017318 (CN14), CP017323–CP017325 (CN15), CP017389–CP017401 (CN16), CP017307 (CN17) and CP017319–CP017322 (CN18).

Plant material, growth conditions and infection tests

Brassica oleracea var. *botrytis* (*Bob*) cultivar Clovis, *Brassica rapa* ssp *trilocularis* (*Brt*) line R-o-18, *B. rapa* ssp *pekinensis* (*Brp*) accession Chiifu-401-42 were grown under glasshouse conditions. For virulence assays with the wild-type (WT) strain Xca5 and its *tal* mutant derivatives (Kay *et al.*, 2005), the central veins of *Bob* and *Brt* leaves were wounded using a needle dipped in a bacterial suspension (10^8 colony forming units (CFU) ml^{-1}). The first and second true leaf of at least six plants (4-wk-old) was inoculated for each strain. Bacterial multiplication and disease index were determined at 0, 4 and 6 d post-inoculation (dpi), respectively, as described previously (Xu *et al.*, 2008). At least three independent experiments were performed. For quantitative reverse transcription polymerase chain reaction (qRT-PCR) analyses, leaves of *Brp* were infiltrated using a needleless syringe with suspensions (10^7 CFU ml^{-1}) of *Xc* strains containing *tal* genes or *Xcc* CFBP 4955, which contains none, as a negative control. Two strains per leaf were inoculated, and tests were done on three leaves per plant, with a total of 16 plants. After inoculation, plants were placed for 24 h in miniature glasshouses with a clear plastic cover and kept at 100% relative humidity. Statistical significance was assessed as describes (Methods S1).

TALE target validation assays by qRT-PCR

Infiltrated areas of *Brp* leaves were harvested at 24 h post-inoculation. Total RNA was extracted with the NucleoSpin® RNA Plant kit (Macherey-Nagel, Düren, Germany). Specific primers were designed using default parameters of the OligoArchitect Online v.2.0 software (Sigma-Aldrich; Table S2). One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega) and used to generate cDNA using the SuperScript® III First-Strand Synthesis System (Life Technologies, Carlsbad, CA, USA). Real-time PCR was performed as described previously (Hutin *et al.*, 2015b). The specificity of the primer pairs was checked by a melting curve analysis and an agarose gel electrophoresis. The amplification efficiency for each primer pair was analyzed using the LinRegPCR analysis program (Ramakers *et al.*, 2003). Expression levels were normalized with *BrpUBC* constitutive control (Qi *et al.*, 2010), and then with expression in the mock sample. The ratios of normalized values for TALE-containing *Xc* vs strain devoid of TALEs (CFBP 4955) were calculated. qRT-PCR primer sequences are provided in Table S2. The statistical significance of the results was assessed using a *t*-test.

Results

Host species-dependent, cooperative contribution to virulence by two TALEs of *Xcc* strain Xca5

No *tal* genes are present in the four *Xc* genome sequences completed to date (*Xcc* strains 8004, B100 and ATCC 33913, and

Xcr strain 756C) (da Silva *et al.*, 2002; Qian *et al.*, 2005; Vorhölter *et al.*, 2008; Bogdanove *et al.*, 2011). A first hint of the presence of *tal* genes in *Xcc* came from studies on strain Xca5, initially wrongly assigned to pathovar *armoraciae* and now corrected to pathovar *campestris* (Bolot *et al.*, 2013). Indeed, it had been shown that strain Xca5 carries the *tal* genes named *hax2*, *hax3* and *hax4* (Kay *et al.*, 2005), suggesting that more *Xc* strains might express TALEs. To be consistent with a more common nomenclatural convention for *tal* genes, we refer to those genes herein by the number of repeats followed by a letter and the subscripted strain name so that *hax2*, *hax3* and *hax4* are called $tal22a_{Xca5}$, $tal12a_{Xca5}$ and $tal15a_{Xca5}$, respectively (Table S1). The letter was used to be able to distinguish genes with the same number of repeats but different encoded RVD sequences. The order (a, b, c, . . .) reflects the order of identification. Interestingly, a previous study (Kay *et al.*, 2005) using the triple knockout strain Xca5Δ*tal* (lacking *tal22a*, *tal12a* and *tal15a*) implicated $tal22a_{Xca5}$, $tal12a_{Xca5}$ and $tal15a_{Xca5}$ in the development of necrotic symptoms when force-infiltrated into the mesophyll of radish leaves. Because *Xcc* strains are known vascular pathogens, we decided to inoculate Xca5 bacteria by wounding the main vein of *B. oleracea* var. *botrytis* cv Clovis leaves. Xca5 strain caused spreading necrotic lesions typical of a virulent *Xcc* strain (Fig. 1a) and multiplied strongly in tissues indicating that cauliflower cv Clovis is susceptible to infection by Xca5 strain. We also observed that Xca5Δ*tal* caused reduced symptoms (Fig. 1a,b), but not on susceptible *B. rapa* ssp *trilocularis* line R-o-18 (Fig. S1). This phenotype was most striking at 6 dpi. Tested in pairs, only *tal12a* and *tal15a* together rescued the virulence defect (Fig. 1a,b). Bacterial multiplication was also measured at 4 and 6 dpi in *B. oleracea* (Fig. 1c). Bacterial growth of the strain Xca5Δ*tal* was not significantly different from that of the WT, increasing *c.* 100-fold by 4 dpi (Fig. 1c). These data demonstrate that *Xcc* TALEs $Tal12a_{Xca5}$ and $Tal15a_{Xca5}$ synergistically contribute to disease development on *B. oleracea*.

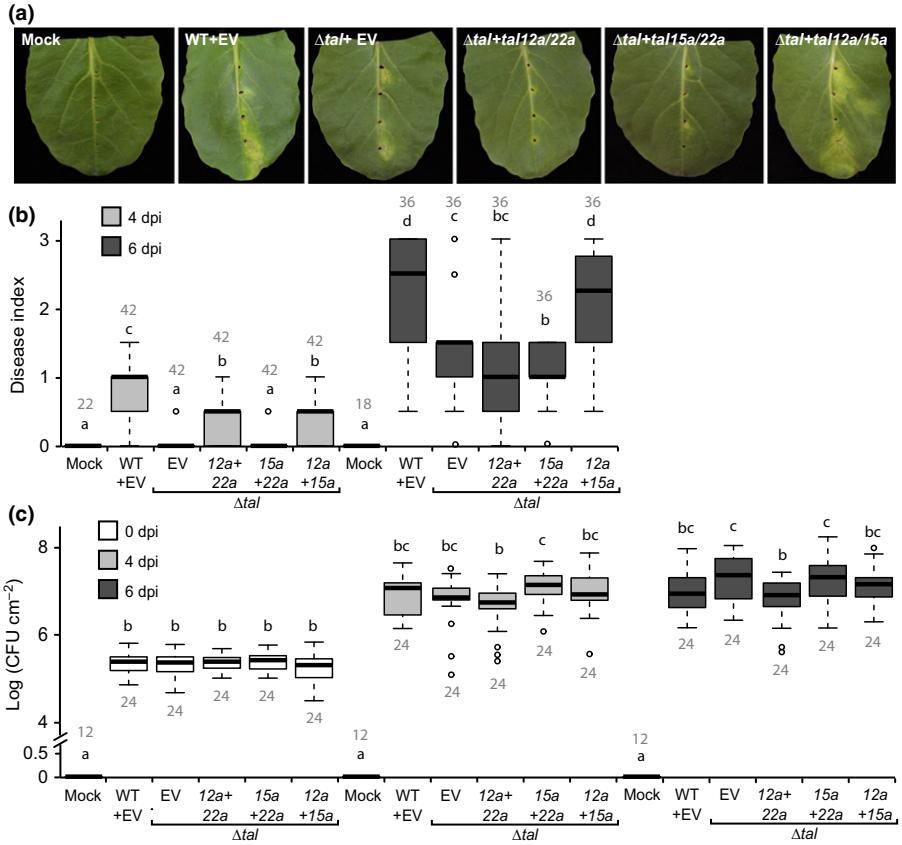

Fig. 1 *Xanthomonas campestris* pv *campestris* (*Xcc*) strain Xca5Δ*tal* (Δ*tal*) is reduced in virulence on *Brassica oleracea* var. *botrytis* (cv Clovis) compared to the wild-type (WT) strain. The Xca5Δ*tal* strain lacks *tal12a*, *tal15a* and *tal22a* (Kay *et al.*, 2005). WT strains carrying an empty plasmid vector (EV) and Xca5Δ*tal* carrying EV or plasmids carrying two of the *tal* genes were inoculated by wounding the central vein of the leaf with a needle dipped in a bacterial suspension at 10^8 colony forming units (CFU) ml^{-1}. (a) Disease symptoms observed at 6 d post-inoculation (dpi). Leaves representative of the median disease index found in (b) are shown. (b) Boxplot representation of the disease symptoms caused by *Xcc* at 4 or 6 dpi. The disease index was as follows: 0, no symptoms; 1, weak chlorosis around wounded sites; 2, strong chlorosis; 3, first necrotic symptoms; 4, large necrotic lesions. (c) Bacterial populations of *Xcc* in leaves were measured at 0, 4 and 6 dpi. (b, c) Three independent experiments were performed. Statistical groups were determined using a nonparametric Kruskal–Wallis test ($P < 0.01$) and are indicated by different letters. Boxplot representations are as follows: middle bar, median; box limit, upper and lower quartiles; and extremes, minimum and maximum values. Dots indicate outliers. Numbers in gray indicate the number of data points from the three combined experiments.

Identification of a large repertoire of TALEs in *X. campestris*

In order to determine the prevalence of TALEs in *Xc*, 49 strains (38 *Xcc*, three nonpathogenic, five *Xci* and three *Xcr*) were selected to represent the genomic and geographical diversity of the species with a particular focus on pathovar *campestris* (CIRM-CFBP, Angers, France and CN; He *et al.*, 2007; Table S1). A phylogenetic tree of these 49 strains was obtained based on AFLP analysis (Fig. 2), showing seven clades (A–G) of *Xcc* as described previously (Guy *et al.*, 2013) and grouping the *XcNP*, *Xcr* and *Xci* strains distinctly within a separate clade (Roux *et al.*, 2015). These clades also were supported by average nucleotide identities (ANI) of 21 strains

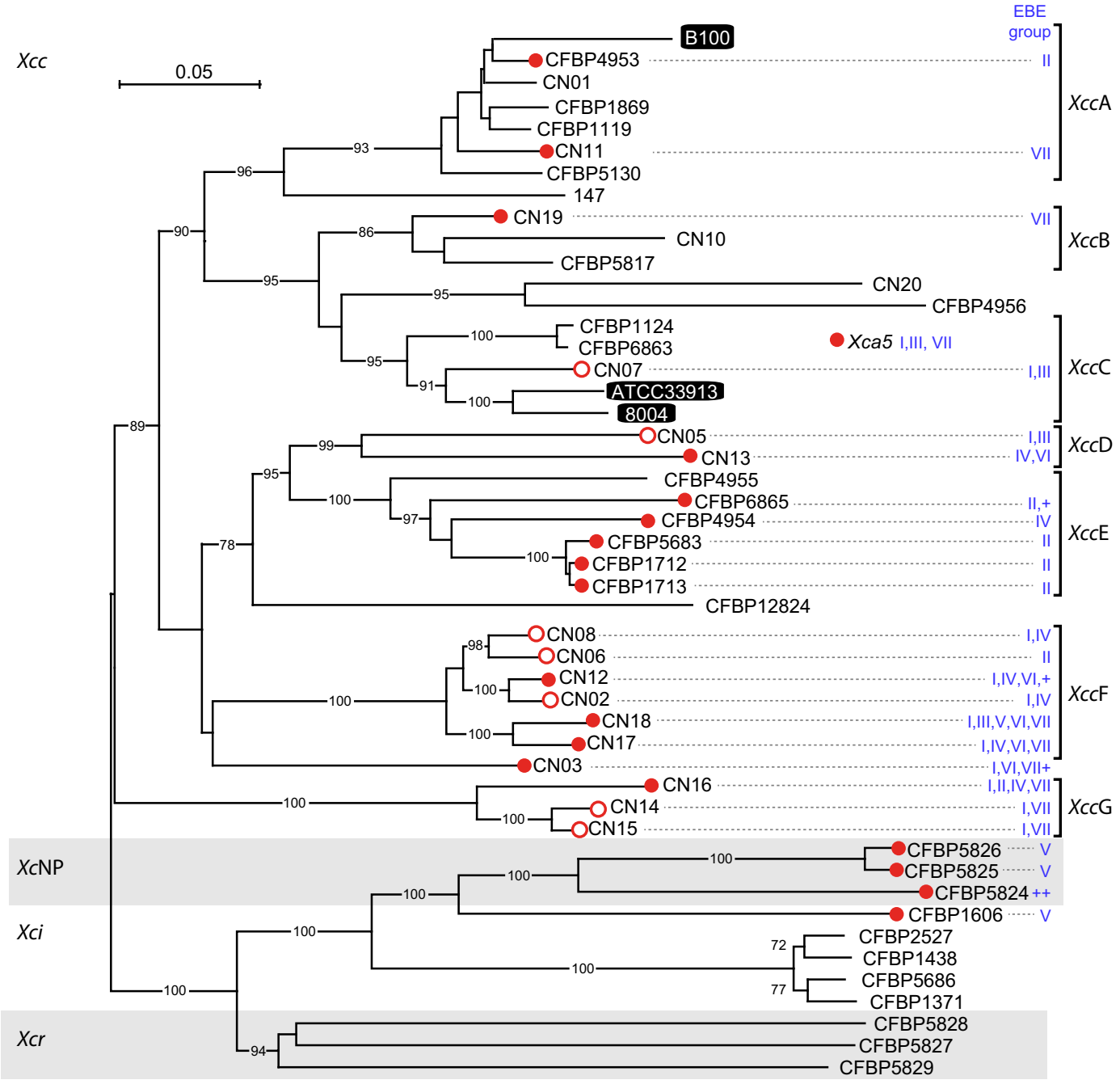

Fig. 2 Neighbor-joining tree based on Dice similarity indices (1942 amplified fragment length polymorphism (AFLP) markers) showing the relationships among *Xanthomonas campestris* (*Xc*) strains. Bootstrap values higher than 70% are indicated. Bar, substitutions per site. *Xcc* (*Xc* pv *campestris*) strains for which full genome sequence is available are boxed in black. *Xc* strains with complete or partial *tal* gene repertoires as described in this study are indicated with a filled or empty red dot, respectively. On the right, blue roman numerals indicate which effector-binding element (EBE) groups (I–VII) or orphan members ('+') are present for each strain (see Fig. 4 for details). Brackets indicate AFLP- and MLST-based subgroups of *Xc* (Fargier *et al.*, 2011; Guy *et al.*, 2013). Pathovars are indicated to the left with *Xcc*, *Xc*NP, *Xci* and *Xcr*, respectively, corresponding to *X. campestris* pv *campestris*, *X. campestris* nonpathogenic strains, *X. campestris* pv *incanae* and *X. campestris* pv *raphani*.

for which genomic sequences were available (Table S3). Most of the *Xcc* strains were isolated from diseased plant material, including *B. rapa* (10/38) (Table S1).

The 49 *Xc* strains were screened for the presence of *tal* genes by PCR based on Xca5 *hax* sequences. Subsequent Southern blot analysis of genomic DNA using the conserved *Bam*HI sites within *tal* genes revealed that strains carry from zero to at least four *tal* genes (Fig. 3a; Table S1). The minimal *Xc* TALome size (across strains) is at least 61 genes. Overall 26 of 49 *Xc* strains were found to harbor *tal* genes: 22 of 38 *Xcc* strains, each of the three *XcNP*, one out of the five *Xci* strains, and none of the *Xcr* strains. *tal* gene distribution seems to relate to *Xc* phylogeny: most strains in clades *Xcc*D-G contain *tal* genes, whereas two-thirds of the strains in clades *Xcc*A-C do not (Fig. 2). Western blot analysis confirmed that all of the *tal* genes are expressed, yielding proteins of the expected sizes except Tal15g$_{CFBP4954}$, Tal15g$_{CN13}$ and Tal16a$_{CFBP6865}$ whose electrophoretic mobility was faster than would be expected for complete TALEs with those numbers of repeats, and Tal16b$_{CFBP1606}$ which we could not reproducibly detect (Fig. 3b). The results for these three gene products are likely to be the result of variation outside the (sequenced) repeat region, such as an early stop codon, an alternative start codon, or an internal deletion(s), or in the case of Tal16b$_{CFBP1606}$, a destabilizing missense mutation, an early frameshift mutation, or a promoter polymorphism.

By combining PCR-based cloning and sequencing of repeat regions and full genome sequencing of seven *Xcc* genomes using the SMRT technology (Pacific Biosciences, Menlo Park, CA), a total of 53 *tal* gene sequences was obtained in 26 *Xc* strains (Methods S2). Based on cross-referencing with the Southern blot of Fig. 3(a), nine *tal* gene sequences are still to be determined (Table S1). The number of repeats ranges from 11 to 22 (including the final, truncated repeat; Fig. 4). Twenty-five proteins have repeats of 34 aa, 13 have repeats of 35 aa, and 15 have repeats of both lengths (Figs 4, 5a,b). The 34- and 35-aa length repeats generate two distinct consensus sequences (Fig. 5a): repeats of the 34-amino acid type end with 'HG' whereas those of 35 amino acids finish with 'PHD/C'. *tal* genes were named using the nomenclature described in the previous section. A correspondence between the *tal* and AnnoTALE nomenclatures (Grau *et al.*, 2016) is provided for full-length *tal* genes (Table S1).

The 53 *tal* genes represent 21 unique RVD sequences (Fig. 4). Among the 322 total repeats, only seven RVDs are found (Fig. 5c). The five most represented RVDs (HD, NI, NG, NN and NS) also are the most abundant in 113 TALEs from various *Xanthomonas* species (Boch & Bonas, 2010). As for the two less common RVDs, IG is a particularly rare RVD found in 3% of the *Xc* TALEs whereas NT is found exclusively in Tal14a$_{CFBP4953}$, all Tal14b, and Tal14c$_{CN16}$). Interestingly, this RVD has been reported so far only in TAL-like effectors of *Ralstonia solanacearum*, *Burkholderia* species and unknown marine bacteria (Streubel *et al.*, 2012; de Lange *et al.*, 2013, 2014, 2015). The 34-aa and 35-aa repeats exhibit differences in the RVDs they contain. RVD NS is specific to 34-aa repeats, whereas RVDs NN, IG and NT are found exclusively in 35-aa repeats (Fig. 5c). RVDs HD, NI and NG occur in both repeat types.

In order to further investigate the relationships among *Xc* TALEs, we used FuncTAL in the QueTAL suite, which classifies TALEs targeting similar EBEs (Pérez-Quintero *et al.*, 2015). This divided the TALome into seven EBE groups (I–VII) and five orphans each with distinct DNA-binding specificities (Fig. 5d). The seven EBE groups encompass 90% of the cloned *tal* genes revealing a limited complexity in the TALome. To further investigate the relationships among *Xcc* TALEs, we used the program DisTAL, which groups TALEs based directly on RVD sequence similarity (Pérez-Quintero *et al.*, 2015), on the full-length *tal* sequences obtained by SMRT sequencing (Fig. 5e). This analysis

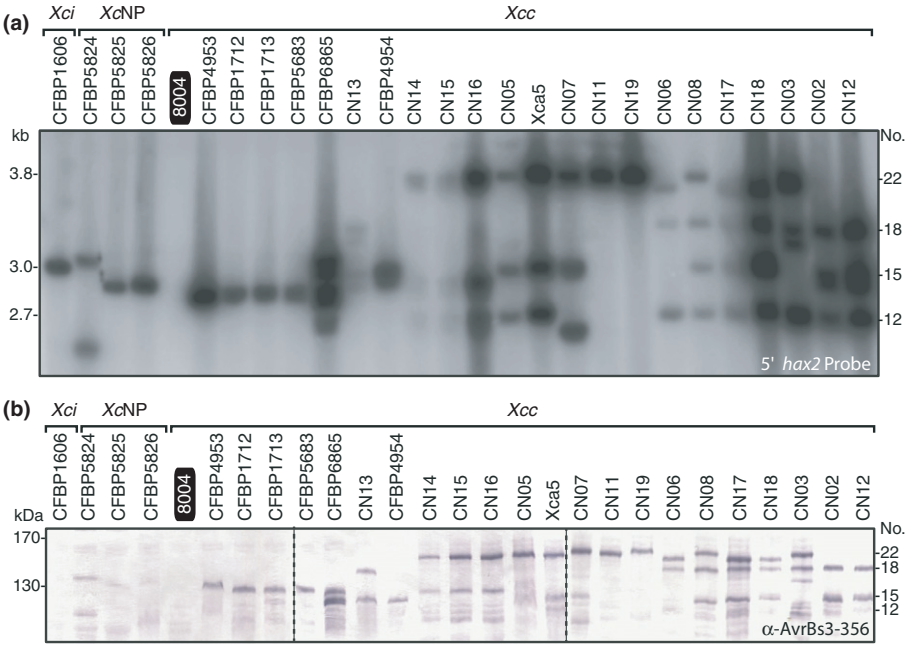

Fig. 3 Detection of transcription activator-like effector (TALE)-encoded *tal* genes in *Xanthomonas campestris* (*Xc*) strains. (a) Southern analysis of *Xc* genomic DNA digested by *Bam*HI using a 5′ *tal22a* (*hax2*) sequence as probe. The 865-bp probe was amplified by PCR from *Xcc* strain Xca5 genomic DNA. DNA sizes are indicated (kb). (b) Western-blot analysis of *Xc* total protein extracts resolved by SDS-PAGE using an anti-AvrBs3-356 antibody. *Xcc* strain 8004, which does not contain *tal* genes, serves as negative control. Note the presence of weak background bands in all strains which likely correspond to proteins cross-reacting with the antiserum (AvrBs3-356). Molecular weight is indicated (kDa). Expected position of a *tal* gene or TALE with given number of repeats is shown on the right.

essentially recapitulated the FuncTAL results, indicating that each EBE group is made up of phylogenetically related TALEs with no hints of convergent evolution. It also highlights the fact that a certain degree of polymorphism can be detected within EBE groups that can be used to trace the evolution of this gene family.

Numerous traces of DNA rearrangements in *Xc tal* genes

Notably, phylogenetically closely related strains usually harbor similar TALomes (Fig. 2) as described for other type III effectors (Guy *et al.*, 2013). This indicates that *tal* genes are predominantly vertically inherited. Yet, a detailed analysis of the *tal* genes suggests that recombination events shaped the current *Xc* TALome. Evidence of such events is often masked by the extreme sequence conservation of *tal* genes. We drew trees based on alignments of the 5′ and 3′ sequences of *Xcc tal* CDS (Fig. 6c,d; Notes S1, S2). The two trees were incongruent, suggesting that recombination events might have occurred among *tal* genes. A greater diversity was observed for the 3′ sequences. Furthermore, although group VII members target essentially the same EBE, their 3′ and 5′ sequences are not monophyletic. To highlight polymorphisms in 5′ and 3′ *tal* gene sequences, IPS (informative polymorphic sites; polymorphic positions in alignments that are common to at least two sequences) were identified to define DNA signatures for the 5′ and 3′ ends. Interestingly, every *tal* gene sequence reflected a combination of only two IPS signatures (Figs 6a, S2). This observation suggests that present *tal* gene sequences are the result of DNA recombination events between two ancestral genes. In support of this observation, all *Xc* TALEs start with a 34-aa repeat with the RVD NI or a 35-aa repeat with the RVD NN, whereas all but four end with a truncated (20-aa) repeat with the RVD NG. These end repeats are likely to reflect the ancestral sequences, because one would expect end repeats to remain constant relative to more central repeats, which would be expected to be more often affected by recombination. As mentioned earlier, 11 of the 21 repeat regions are composed of a mixture of 34-aa and 35-aa repeats (Fig. 4). This duality affords an opportunity to trace recombination within the repeat region. For instance, although encoding the same RVD, the third repeat of *tal18a* can encode either 34 or 35 amino acids, depending on the strain. This suggests a recombination event between the RVD and the end of the repeat, at which the length polymorphism occurs. Other examples of possible recombination include the first eight RVDs of Tal14d$_{CN12}$ which are identical to those of group VII members or the last five RVDs which are identical between Tal15e$_{CN08}$ (group IV) and Tal17d$_{CN03}$ (orphan). Altogether, the observations discussed above reveal that gene duplications and complex sequence rearrangements likely occurred to shape the present *Xc* TALome from two ancestral *tal* sequences.

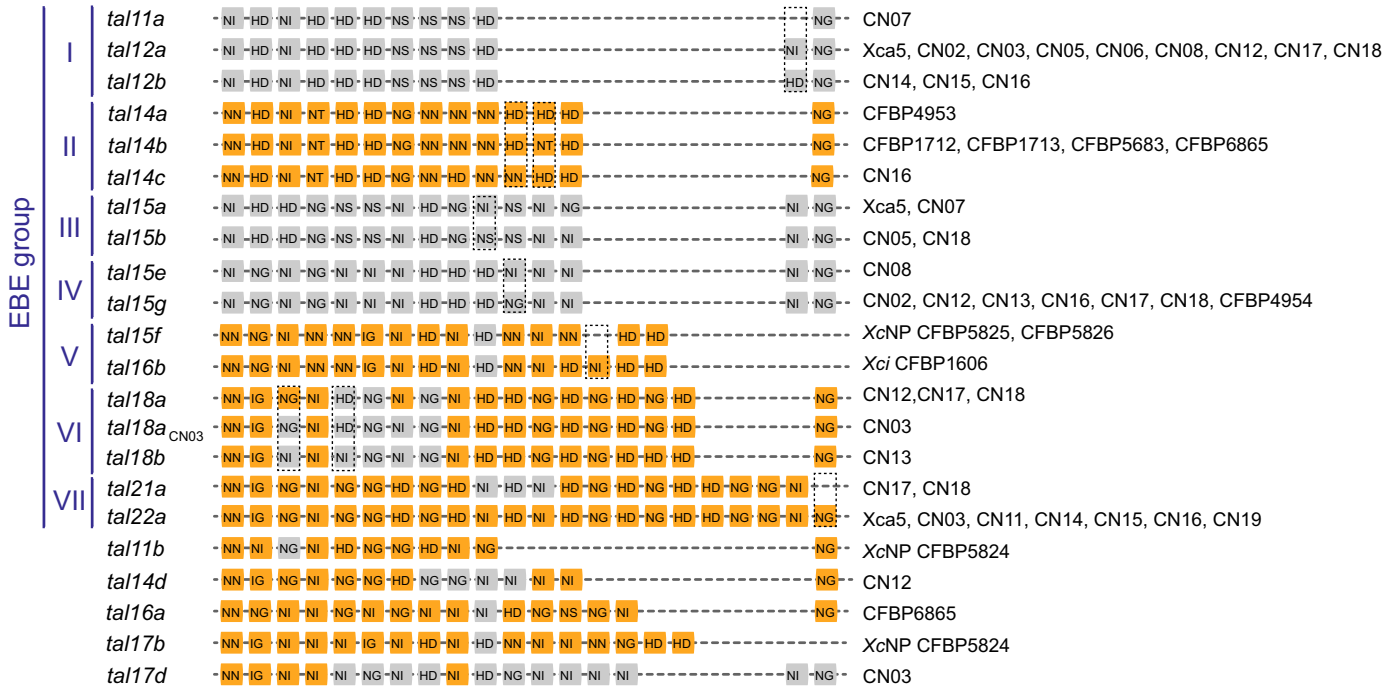

Fig. 4 Properties of central repeat regions of *Xanthomonas campestris* (*Xc*) *tal* (transcription activator-like effector (TALE)-encoded) genes. Properties of the *tal* gene repeats identified either by PCR amplification or single molecule real time (SMRT) genome sequencing are represented. Repeats of 102- or 105-bp (coding for 34 or 35 amino acids) are indicated by gray or orange boxes, respectively. The sequence of repeat variable di-residues (RVDs) in each protein is represented by the paired single letter amino acid codes in the boxes. The repeat regions were organized in seven effector-binding element (EBE) groups (I–VII; shown in blue on the left) using FuncTAL based on the similarity of the corresponding predicted EBEs (Fig. 5d). Alignment of the repeats was performed manually to optimize comparisons within and between EBE groups. Dashed boxes indicate polymorphic RVDs within each EBE groups. *Xc* strains in which the corresponding TALE repeats were identified are indicated on the right.

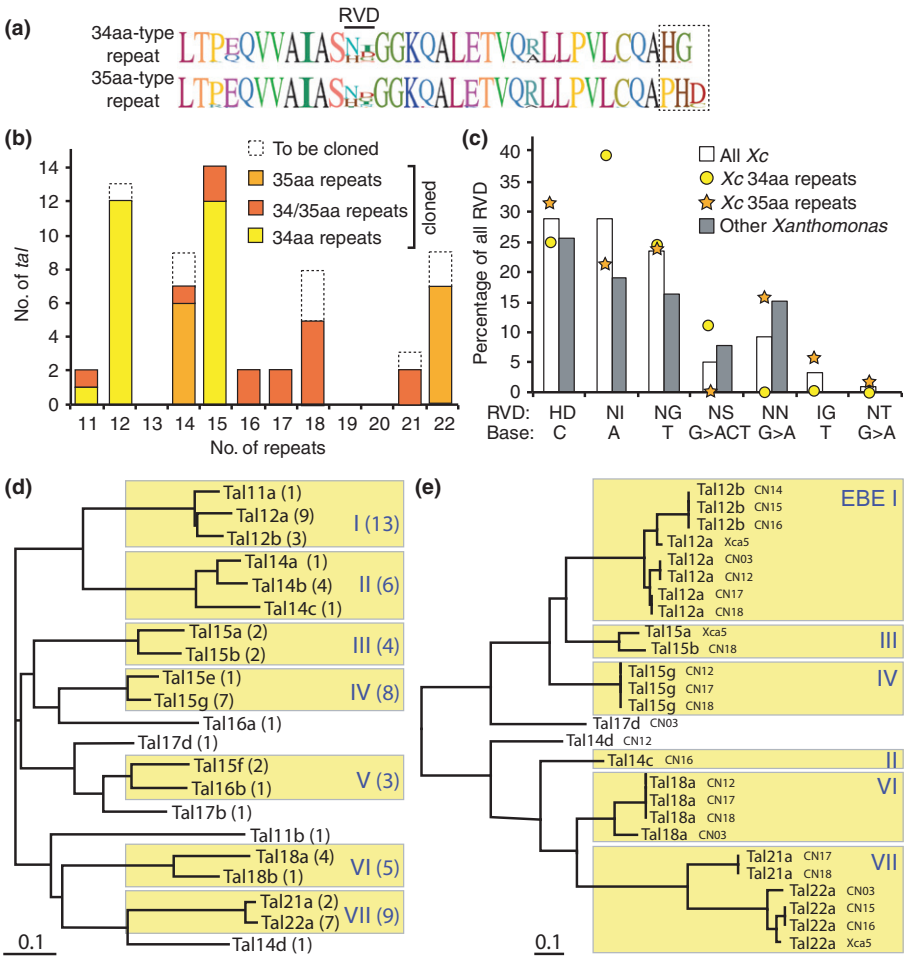

Fig. 5 Properties of the *Xanthomonas campestris* (*Xc*) TALome. (a) Sequence logo representation of 34- and 35-aa type repeats extracted from *Xc* transcription activator-like effectors (TALEs) as described in Fig. 4. Repeat variable di-residue (RVD) position is indicated. Residues that were used to define repeats of 34 or 35 residues are boxed in a dashed line. (b) Distribution of repeat number and length of the encoded repeat (34 or 35 aa) in the TALE proteins encoded by the 53 sequenced *Xc tal* genes (out of 61 detected). (c) RVD usage in *Xc* TALEs compared to other *Xanthomonas* TALEs described in Boch & Bonas (2010). The preferred DNA base(s) for each RVD is given at bottom as described (Doyle *et al.*, 2012; Streubel *et al.*, 2012). (d) Tree obtained upon analysis of the 21 RVD combinations represented in the *Xc* TALome (Fig. 4) highlighting the functional relatedness of TALE members within seven effector-binding element (EBE) groups (I–VII, written in blue on yellow background). RVD combinations with no close (less than two RVDs different) homologues are defined as orphans. Numbers between brackets refer to the number of members in each RVD combination or EBE group. (e) Phylogeny of *Xcc tal* genes based on the DisTAL program. Only *Xcc tal* (TALE-encoded) genes for which full length sequence was available were included in this analysis. EBE groups as defined in (d) are indicated by a roman numeral and a yellow box.

Genomic environment of *Xc tal* genes

SMRT sequencing of seven *Xcc* genomes enabled examination of the immediate genomic environment of *tal* genes (Fig. S3). A first remarkable feature is that *Xcc tal* genes can be plasmid- or chromosome-borne as reported for Xca5 (Kay *et al.*, 2005). Although strains CN12 and CN17 are devoid of plasmids, other strains contain one or two (Guy *et al.*, 2013). Seven of 26 *tal* genes in these strains were in contigs that match the plasmid sizes observed previously (Guy *et al.*, 2013), and we therefore conclude that these are plasmid-borne (Fig. S3). *tal* genes with the same RVD sequence can be found in the chromosome (e.g. $tal12a_{CN17}$) or in a plasmid (e.g. $tal12a_{CN03}$) depending on the strain. Second, we did not detect any *tal* genes in clusters as is observed in *Xo* (Booher *et al.*, 2015) and *Xcm* (Cox *et al.*, 2017); instead, the closest *tal* genes were > 10 kb apart in the same (e.g. $tal18a_{CN12}$-$tal15g_{CN12}$) or inverted orientation (e.g. $tal12a_{CN17}$-$tal21a_{CN17}$) (Fig. S3). Third, using MAUVE, we detected collinear blocks at these *tal* loci (Fig. S3). Some of these blocks of synteny are shared between plasmids and the chromosome (e.g. $tal12a_{CN03}$-$tal22a_{CN03}$ vs $tal12a_{CN12}$-$tal14d_{CN12}$). Some show evidence of exchange or recombination. For instance, *tal14d* might have been the result of rearrangement in a former *tal22a*-containing block or vice-versa. This supposition is supported by the IPS signatures in the 5′ and 3′ end sequences of the *tal* genes (Fig. 6a). Traces of large DNA insertion/deletion or inversion within blocks also are observed.

Last, sequences related to the Tn*Xax1* of the Tn*3* family are in close vicinity to *tal* genes as observed in *Xac* (Ferreira *et al.*, 2015). Characteristic of these mobile elements is a pair of 62-bp inverted repeats with a 'GAGGG' tip motif (Fig. 6b). Surprisingly, this motif was found intact upstream of only eight *tal*

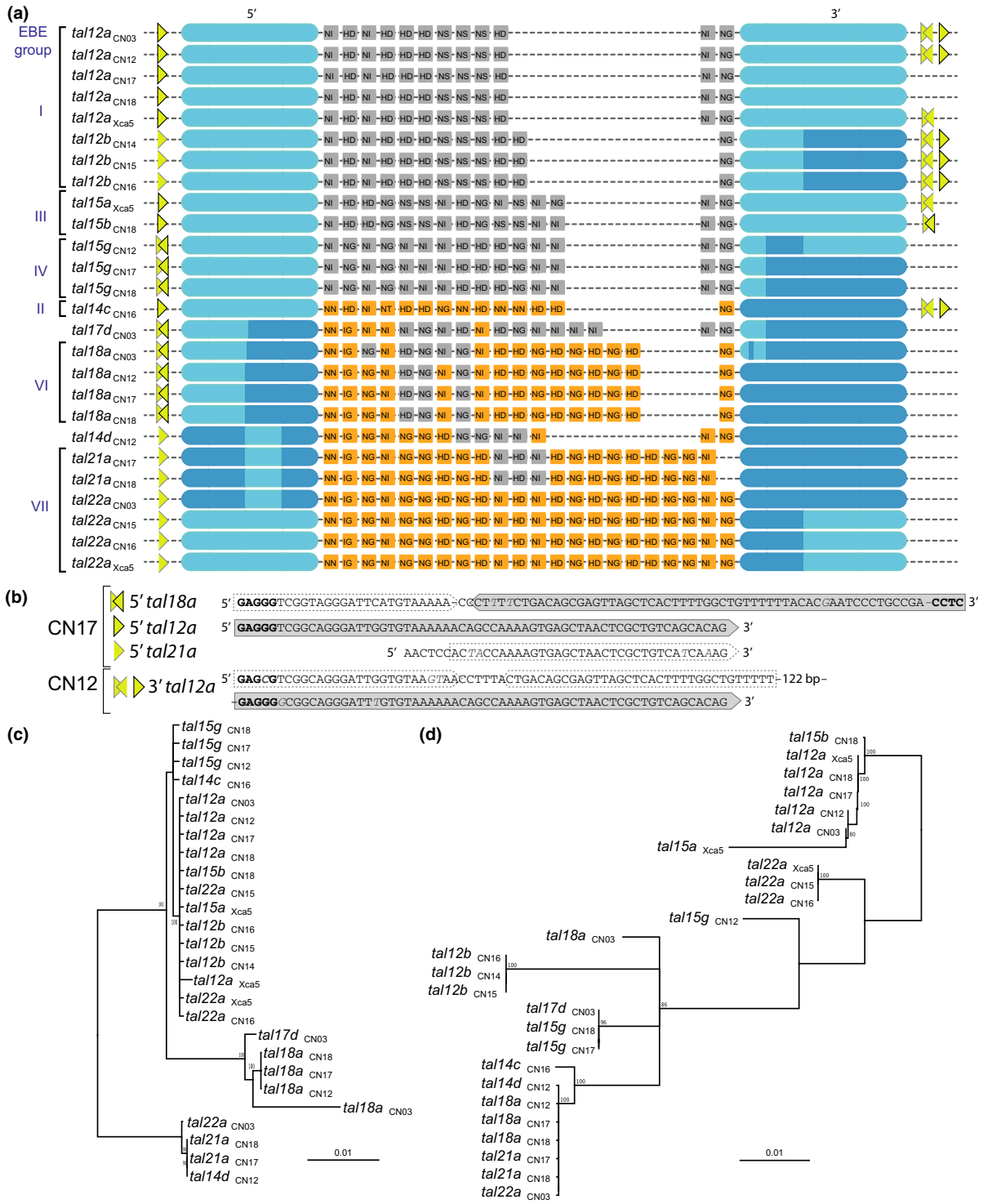

Fig. 6 Diversity, phylogeny and apparent recombination patterns of *Xanthomonas campestris* pv *campestris* (*Xcc*) *tal* (transcription activator-like effector (TALE)-encoded) genes. Twenty-six *Xcc tal* genes for which full-length sequences were available (databases or our single molecule real time (SMRT) sequencing) were included in this analysis. (a) Schematic representation of *tal* gene loci including 5′ and 3′ regions. Effector-binding element (EBE) groups are indicated by a roman numeral as defined in Fig. 5(d). Yellow triangles represent sequence motif variants usually flanking *Xanthomonas tal* genes as inverted repeats. Examples are presented in (b). Identical blue colours in the 5′ or 3′ *tal* gene sequence indicate that the sequences of the corresponding informative polymophic sites (IPS) are identical. Repeats of 102- or 105-bp (coding for 34 or 35 aa) are indicated by gray or orange boxes, respectively. The two-letter code in each repeat corresponds to the encoded repeat variable di-residue (RVD) as in Fig. 4. (b) Representative complete or rearranged sequence motifs identified upstream or downstream of *tal* genes of *Xcc* strains CN12 and CN17. Gray arrows indicate full length sequence motifs and their direction, whereas dashed lines delimit truncated ones. Bases that do not match the consensus inverted repeat of Tn*Xax1* transposon as defined (Ferreira *et al.*, 2015) are indicated in gray and italic. The 'GAGGG' tip motif typical of Tn*3*-like transposons is indicated in bold. (c, d) Unrooted phylogenetic trees of (c) 5′ and (d) 3′ regions of *tal* genes were constructed using Geneious Alignment and Tree Builder tools with default parameters.

genes, yet it was truncated or scrambled at the other 18 genes (Fig. 6a). Downstream, we only detected the corresponding inverted repeat at *tal15b*$_{CN18}$. We observed two distinct repeat rearrangements downstream of eight *tal* genes (Fig. 6a): one including scrabbled repeats plus a full direct repeat (at six genes; e.g. *tal12a*$_{CN03}$) or only the scrabbled repeats (at two genes; e.g. *tal15a*$_{Xca5}$). However, 17 *tal* genes (e.g. *tal12a*$_{CN17}$) lacked any sign of downstream direct or inverted repeat (Fig. 6a).

These observations together further reinforce the conclusion that the *tal* loci and the *tal* genes themselves are highly plastic, and have undergone major rearrangements during their evolution.

Prediction of *Xc* TALE EBEs in the promoterome of *B. rapa* ssp *pekinensis*

We sought to identify putative *S* genes targeted by *Xc* TALEs in a natural host. *Brassica rapa* ssp *pekinensis* accession Chiifu-401-42 was selected because a draft genomic sequence is available (Wang *et al.*, 2011) and the accession is susceptible to the five *Xcc* strains tested. The *Brp* promoterome was searched for EBEs using Target Finder from the TALE-NT suite (Doyle *et al.*, 2012). For each of the 21 RVD combinations, a maximum of 45 EBEs were considered for further analyses, resulting in a total of 517 pairs of TALE and *Brp* EBEs (Methods S3; Table S4). To test predictions, expression of 12 *Brp* putative targets (Table 1) was quantified by qRT-PCR at 24 h post-inoculation with strains carrying the appropriate *tal* gene. Candidates were selected that showed similarity to TALE targets already identified in other plants (*Bra030336* and *Bra029914* which encode nodulins from the MtN21/UMAMIT and SWEET families, respectively) or to immune regulators (*Bra001424*, an U-Box type ubiquitin ligase and *Bra003239*, a WRKY family transcription factor), or because EBEs were conserved in both *Brp* and Arabidopsis (Bra002616, a protease; Bra024326, a root meristem growth factor; and Bra039705 and Bra039823, two myrosinases). Some targets also were selected at random (*Bra013242*, a CLE peptide; *Bra018741*, an AGL family transcription factor; *Bra025053*, a protein kinase and *Bra035278*, an unknown protein). In nine of the 17 *Brp* gene/*Xc* strain combinations tested, the expression of the gene was significantly higher following inoculation of the strain than upon treatment with *Xcc* control strain CFBP 4955, which is devoid of TALEs (P-value < 0.01) (Fig. S4), validating five of the 12 putative targets (Table 1). Among these validated target genes, Bra002616 encodes an FtsH9 protease, Bra024326 encode a root meristem growth factor 9 protein, and Bra039823 and Bra039705 code for TGG1 myrosinases (Table 1).

We tested whether TALEs belonging to the same group were predicted to target the same set of *Brp* promoters, and found that, except for EBE group VI, the different TALEs within each EBE group are predicted to target at least one promoter in common (on identical, overlapping or distant EBEs) (Fig. 7a; Table S4). Interestingly, three *Brp* genes which encode a ribosomal protein, a protein with a DOF zinc finger domain and an ERF transcription factor, are targeted by multiple TALEs binding to unrelated nonoverlapping EBEs (Fig. 7c; Table 2). These putative TALE target combinations represent potential cases of functional convergence and compelling *S* gene candidates, for future examination.

Discussion

Xca5 TALEs contribute to virulence on cauliflower

Numerous transcription activator-like effector (TALE)-encoding (*tal*) genes have been reported in *Xanthomonas* pathogenic on rice, pepper, tomato, citrus, cotton or cassava (Schornack *et al.*, 2013). *Xanthomonas campestris* (*Xc*) pv *campestris* (*Xcc*) genomes were long thought to be devoid of *tal* genes. The recent reclassification of *Xc* pv *armoraciae* strain Xca5 as a *bona fide Xcc* suggested the presence of *tal* genes at least in some strains of this pathovar (Bolot *et al.*, 2013) (Fig. 2). Indeed, *tal22a*$_{Xca5}$ (*hax2*), *tal12a*$_{Xca5}$ (*hax3*) and *tal15a*$_{Xca5}$ (*hax4*) genes were found in seven, nine and two strains of *Xcc*, respectively, from a collection of 38 strains representative of the overall *Xcc* biodiversity. Importantly, *tal22a*$_{Xca5}$, *tal12a*$_{Xca5}$ and *tal15a*$_{Xca5}$ were collectively implicated in the development of necrotic lesions on radish using mesophyll infiltration of a triple knockout mutant Xca5Δ*tal* (Kay *et al.*, 2005). Whether this *tal*-dependent necrosis results in the promotion of virulence by *tal* genes or by the recognition of some of those *tal* genes in radish cannot be determined unambiguously based on the experiments presented by Kay *et al.* (2005). When wound-inoculated in the vasculature of susceptible *Brassica oleracea* var. *botrytis* cv Clovis, strain Xca5Δ*tal* causes reduced disease symptoms. The combination of *tal12a*$_{Xca5}$ and *tal15a*$_{Xca5}$, and neither of the other two pairwise combinations of the three *tal* genes, were able to complement the reduced virulence phenotype (Fig. 1). By contrast, Xca5Δ*tal* was unchanged in virulence on *B. rapa* ssp *trilocularis* line R-o-18 (Fig. S1). The contribution of individual *Xc* TALEs to virulence thus can depend on the host plant, which may be due to allelic differences in corresponding susceptibility genes. In cauliflower, we observed no significant difference in multiplication of Xca5Δ*tal* compared to wild-type or of Xca5Δ*tal* compared to derivatives carrying each two-gene combination of the three *tal* genes. Additional experiments including transcript profiling with EBE prediction should help identify the biologically relevant target of each of Xca5 TALE in cauliflower.

The *Xc* TALome comprises seven distinct recognition specificities

A repertoire composed of at least 61 *tal* genes was defined in 26 of the 49 *Xc* strains. In *Xcc*, we identified a total of 48 *tal* genes, including the known *tal12a*, *tal15a* and *tal22a* genes, in 22 of the 38 strains analyzed. These *tal* genes encode 21 unique repeat variable di-residue (RVD) combinations (DNA binding specificities) and can be sorted into seven EBE groups based on relatedness of the sequences predicted to be recognized by their RVD sequences (Figs 4, 5c). The repeat regions of a substantial proportion of these TALEs are made up of 35-aa repeats.

Table 1 Features and properties of *Brassica rapa* ssp *pekinensis* (*Brp*) genes whose transcription activator-like effector (TALE)-dependent expression was tested experimentally (Supporting Information Fig. S4)

TALE	*Brp* Gene	Gene product	Aligned EBE (top) and RVD (bottom) sequences[b]	Score[c]	Ratio[d]	EBE to TLS[e]	Experimental validation[f]
Tal22a	Bra013242	CLE41 peptide	T c a T A T T C T C A C A C T C T t C T T c T NN IG NG NI NG NG HD NG HD NI HD NI HD NG HD NG HD HD NG NG NI NG	13.72	2.55	–71	No
Tal11b	Bra001424	PUB24 U-BOX ubiquitin-protein ligase	T G A T A C a T C A T T NN NI NG NI HD NG NG HD NI NG NG	5.05	1.66	–63	No
Tal11b	Bra030336	AtUMAMIT40; nodulin MtN21 family	T A A c A C T T C A T T NN NI NG NI HD NG NG HD NI NG NG	5.25	1.72	–31	No
Tal11a	Bra003239	WRKY70 transcription factor	T A C A C a C A c A C T NI HD NI HD HD HD NS NS NS HD NG	6.24	2.23	–93	No
Tal15e	**Bra002616**[a]	**FtsH protease 9**	**T A g A T A A A C C C A A A A a** **NI NG NI NG NI NI NI HD HD HD NI NI NI NI NG**	**7.86**	**2.75**	**–73**	**Yes**
Tal15e	**Bra024326**[a]	**Root Meristem Growth Factor 9**	**T A T A T A A A C a C A A A A a** **NI NG NI NG NI NI NI HD HD HD NI NI NI NI NG**	**7.18**	**2.52**	**–76**	**Yes**
Tal17b	Bra018741	AGL53 transcription factor	T A T A A A T A a A C A c A A c C C NN IG NI NI NI IG NI HD NI HD NN NI NI NN NG HD HD	12.12	2.35	–88	No
Tal17b	Bra029914	AtSWEET11; nodulin MtN3 family	T A T A t A T A C A C G A A G T C C NN IG NI NI NI IG NI HD NI HD NN NI NI NN NG HD HD	8.87	1.72	–143	No
Tal12b	Bra025053	MEE62 protein kinase	T A C A a C C A A A C C T NI HD NI HD HD HD NS NS NS HD HD NG	5.28	1.79	–96	No
Tal16a	**Bra035278**	**Unknown protein**	**T A a A A T A T A A A C T A T A T** **NN NG NI NI NG NI NG NI NI NI HD NG NS NG NI NG**	**6.54**	**1.51**	**–68**	**Yes**
Tal15g	**Bra039705**[a]	**TGG1 myrosinase**	**T A T A T A A A C C C T A g A T** **NI NG NI NG NI NI NI HD HD HD NG NI NI NI NG**	**6.47**	**2.19**	**–63**	**Yes**
Tal15g	**Bra039823**[a]	**TGG1 myrosinase**	**T A T A T A A A C C C T A g A c** **NI NG NI NG NI NI NI HD HD HD NG NI NI NI NG**	**8.48**	**2.86**	**–63**	**Yes**

[a]The corresponding *Arabidopsis thaliana* orthologues were also predicted *in silico* as TALE targets.
[b]Repeat variable di-residue (RVD) sequence shown at bottom, effector-binding element (EBE) at top. Mismatches are indicated by lower case in the EBE. Matches are defined as (with X equal to any amino acid) XD to C, XI to A, XG to T, XN to G or A, NS to G or A, and N* and other, rare RVDs to any base. Distinct colors were attributed to each RVD.
[c]Score in *B. rapa* promoters is according to Doyle *et al.* (2012).
[d]Ratio of EBE score to best possible score for the TAL effector in each *B. rapa* or *A. thaliana* promoters (Doyle *et al.*, 2012).
[e]Distance in bases from the 5′ end of the EBE to the translational start site (TLS) of the target locus; a negative value indicates an EBE location upstream of the TLS.
[f]Experimental validation of the target predictions by quantitative reverse transcription polymerase chain reaction in *B. rapa*. Yes, validated (true target, in bold); No, not validated (false positive).

Sequence rearrangements between two ancestral *tal* genes shaped the *Xc* TALome

Although recombination leading to variation in *tal* repeat numbers can be observed *in vitro* (Yang & Gabriel, 1995b; Yang *et al.*, 2005; Lau *et al.*, 2014), little is known about how *tal* gene diversity is generated and maintained in *Xanthomonas*. A major reason for this lack of knowledge is the very high sequence conservation of *tal* genes and their flanking regions, which can mask recombination events. Interestingly, *Xc tal* genes are composed of a combination of two major sequence types (Fig. 6a). This includes both the 5′ and 3′ portions of the genes as well as the repeat region, which encodes two types of repeats, 34 or 35 aa long (Fig. 5a). This mosaic structure suggests that two ancestral *tal* genes gave rise to the current *Xc* TALome through duplication and recombination (Fig. 6a). The alternative hypothesis that the TALome derived from a single TALE ancestor with both 34- and 35-aa repeats cannot be totally excluded. Interestingly, *tal21a* and *tal22a* differ by repeats 10–12. These repeats harbor the same RVDs but are different in length. The 34-aa repeats 10–12 of *tal21a* are identical to repeats 1–3 of *tal12a*. These observations suggest that repeats 10–12 of *tal21a* might have been acquired through gene conversion by *tal12a* present in the same strains (Fig. S5a). Comparable 'silent' sequence exchanges that do not affect RVD composition also are observed. For instance, in $tal18a_{CN03}$, the position of a sequence exchange can be pinpointed to the sequence from the middle of repeat 2 until the middle of repeat 4 that does not impact upon RVD sequence and DNA binding specificity (Fig. S5b). Point mutations also appear to be a source of

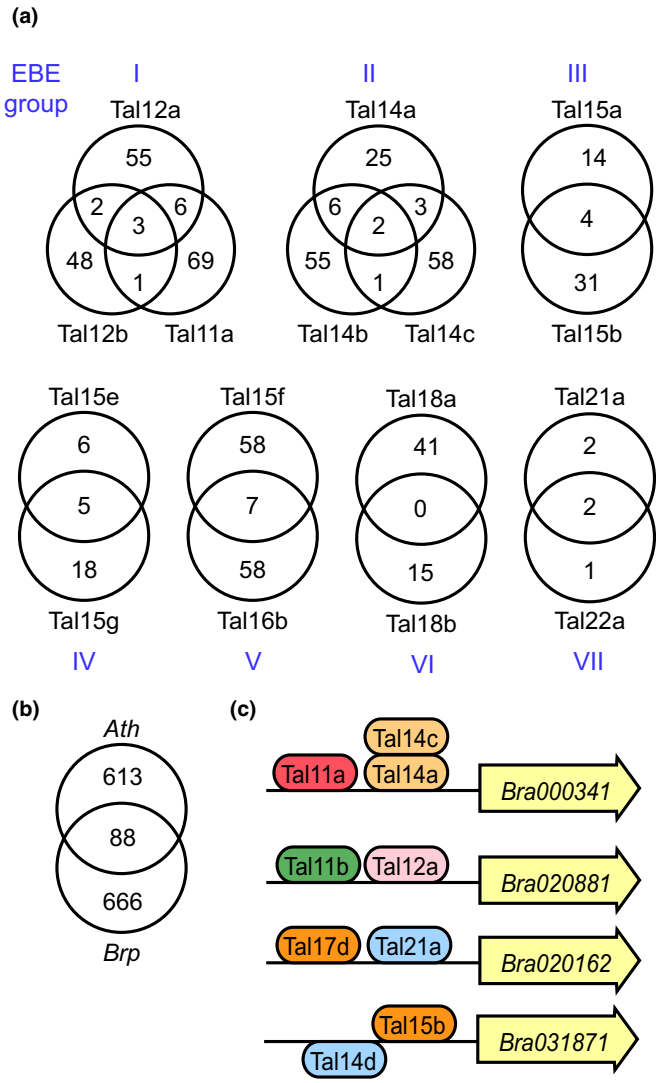

Fig. 7 Results of *in silico* prediction of transcription activator-like effector (TALE) effector-binding elements (EBEs) in *Brassica rapa* ssp *pekinensis* (*Brp*). (a) Overlap between *Brp* predicted targets of TALEs within the seven I–VII TALE EBE groups. (b) Comparative analysis of TALE targets in *B. rapa* ssp *pekinensis* and *Arabidopsis thaliana* predicted by the Target Finder software. A list of *Arabidopsis* orthologs of the targets predicted in *B. rapa* ssp *pekinensis* (*Brp*) was compared to the list of targets predicted in *Arabidopsis* (*Ath*; Col-0 ecotype) to identify overlaps between both datasets. (c) Example of three *Brp* genes predicted to bear several unrelated EBEs in their promoter sequences highlighting potential functional convergence among TALEs.

sequence variation both at the DNA and the protein level. For instance, 11 $tal18a_{CN03}$-specific DNA polymorphisms are found in the 5′ region of this gene, five of which are nonsynonymous when compared to other *tal18a* sequences (data not shown). Finally, variants generated by loss or gain of single repeats are common (e.g. *tal11a*/*tal12a* and *tal21a*/*tal22a*; Fig. 4). However, internal duplication events are difficult to track. One possible illustration of such an event is repeats 6–9 and 14–17 in *tal22a*, which differ by only one nonsynonymous substitution (Fig. S5c). Although not detectable in our dataset, among the TAL-like proteins of *Ralstonia solanacearum*, evidence of loss, duplication or mutations of individual repeats also was observed, enabled by the relatively high repeat to repeat sequence diversity (de Lange *et al.*, 2013; Schandry *et al.*, 2016). In *Xac*, the repeat arrays of *tal* genes are highly variable in their encoded RVD composition and seem to have evolved by inter- and intragenic recombination (Yang *et al.*, 2005). *tal* gene shuffling in *Xac* strains may be mediated by plasmid copy number variations facilitated by Tn*3-like* transposons (Gochez *et al.*, 2018). Another recent study examined the large datasets available for the *Xoo* and *Xoc* TALomes and provided evidence that *tal* genes evolve by base substitution at RVD codon pairs, by recombination between sequences coding for N- or C-terminal TALE domains and by insertion/deletion of individual repeats (Erkes *et al.*, 2017). Yet, precise delineation of recombination breakpoints has remained difficult due to high sequence conservation. The dual origin of *Xc tal* genes offers an unique opportunity to infer the evolution of *tal* genes in *Xanthomonas* by processes proposed for *Ralstonia*, *Xoo*, *Xoc* and *Xac*. This being said, *Xc tal* sequences appear to be relatively evolutionarily stable. Certain *tal* genes are found over and over in chromosomes of phylogenetically distant strains (e.g. *tal12a* in Xca5 and CN12). Sequence drift and recombination would be expected to have yielded a greater sequence diversification and obscured evidence of specific events. Instead, for *Xc tal* genes, recombination can be effectively tracked and functional conservation of RVD arrangements observed. It is tempting to speculate that the relative conservation of *Xc* TALE sequences across diverse strains reflects a fitness advantage that they confer during infection.

Xcc tal genes are dispersed on chromosome and plasmids

SMRT sequencing allowed us to inventory and compare *tal* gene repertoires across strains and pathovars in their genomic context, providing further insight into *tal* gene evolution. Our data show that the *Xcc* TALome is distributed between chromosomal and plasmid locations. Such distribution is unusual: *tal* genes are strictly chromosomal in *Xo* (Ochiai *et al.*, 2005; Salzberg *et al.*, 2008; Bogdanove *et al.*, 2011; Wilkins *et al.*, 2015), and strictly or mainly plasmid-borne in strains of *Xac* (Swarup *et al.*, 1991), *Xe* (Bonas *et al.*, 1989, 1993), or bean-infecting *Xanthomonas* (Ruh *et al.*, 2017). Interestingly, in *Xc*, the same *tal* gene can be found located on plasmids or on the chromosome, hinting at shuffling mechanisms between these different genomic molecules (Fig. S3). Although it is well known that plasmid-borne DNA sequences can integrate into the chromosome, and to a lesser extent vice-versa, the mechanisms allowing such transfer of *tal* genes are unknown. A clue comes from the immediate flanking regions of *tal* genes in *Xc*, which exhibit 62-bp inverted repeats like those found in association with *tal* genes in other species (Bonas *et al.*, 1993; Ferreira *et al.*, 2015). As suggested in *Xac*, these genetic structures may form Tn*Xax1*-like mobile insertions cassettes, potentially leading to *tal* diversification during the transposition process. However, although such mobile elements may have been active in promoting *tal* gene movement during the evolutionary history of *Xc*, this does not seem to be the case

Table 2 *Brassica rapa* ssp *pekinensis* (*Brp*) promoters predicted to bear unrelated effector-binding elements (EBEs) for distinct *Xanthomonas campestris* transcription activator-like effectors (TALEs)

Gene_ID	Gene_Description[a]	TALE	EBE (plus strand)[b]	Ratio[c]	Start[d]
Bra000341	AT2G44120 (E = 4e-056) | 60S ribosomal protein L7 (RPL7C)	Tal11a	NI HD NI HD HD HD NS NS NS HD NG T A C A C C C t A A C T	2.15	–188
		Tal14a	NN HD NI NT HD HD NG NN NN NN HD HD HD NG	2.09	–164
		Tal14c	NN HD NI NT HD HD NG NN HD NN NN HD HD NG T G C A G C C a A C A C C a T		
Bra020881	AT4G21050 (E = 3 e-046) | Dof-type zinc finger domain-containing protein (AtDOF4.4)	Tal11b	NN NI NG NI HD NG NG HD NI NG NG T A A T A C a T C A T T	1.72	–168
		Tal12a	NI HD NI HD HD HD NS NS NS HD NI NG T A a A C C C A A A C A T	1.78	–125
Bra020162	AT5G21960 (E = 3e-050) | AP2 domain-containing transcription factor, putative (AtERF016)	Tal17d	NN IG NI NI NI NG NI HD NI HD NG NI NI NI NI NI NG T A T A A A T A C A C T A A c A A g	2.31	–150
		Tal21a	NN IG NG NI NG NG HD NG HD NI HD NI HD NG HD NG HD HD NG NG NI T t c T A T T C T t A C A C T C T a C T T A	2.89	–187
Bra031871	AT5G65210 (E = 2e-195) TGA1 | TGA1; DNA binding/calmodulin binding/transcription factor	Tal14d	NN IG NI NG NG HD NG NG NI NI NI NI NG T A T T c T T C T T A A A A T	1.59	–31
		Tal15b	NI HD HD NG NS NS NI HD NG NS NS NI NI NI NG T c C C T c A A C T G c A A A T	2.42	–4

[a]http://plants.ensembl.org/Brassica_rapa/Info/Index; EnsemblPlant release 28 (August 2015).
[b]Repeat variable di-residue (RVD) sequence at top, EBE at bottom. Matches are defined as (with X equal to any amino acid) XD to C, XI to A, XG to T, XN to G or A, NS to G or A, and N* and other, rare RVDs to any base. Mismatches are indicated by gray lower case in the EBE. Distinct colors were attributed to each RVD.
[c]Prediction ratio 'score/best possible score' (lower is better) according to Doyle *et al.* (2012).
[d]Distance in bases from the 5′ end of the EBE to the translational start site (TLS) of the target locus; a negative value indicates an EBE location upstream of the TLS.

anymore because only one *tal* gene flanked on both sides with intact inverted repeats could be detected. Along the same lines, no Tn*Xax1*-like sequences could be found associated with the *tal* genes in *Xoc*, which may explain their lower diversity across strains, as opposed to *tal* genes in *Xoo* (Wilkins *et al.*, 2015). Interestingly and in contrast to *Xo* TALE repertoires, which are often organized in clusters of *tal* genes moderately conserved across strains (Bogdanove *et al.*, 2011; Wilkins *et al.*, 2015), the *tal* genes in the seven new, complete *Xcc* genomes are not in clusters. This suggests the absence of local duplication of *tal* genes in *Xc*, as opposed to *Xo* pathovars where such a mechanism appears to be a major force for TALome expansion (Grau *et al.*, 2016). Additional SMRT-derived full genome sequences of *Xc* strains will be needed for a better understanding of the molecular mechanisms underlying the high plasticity of *Xc* TALomes.

Brassica rapa as a crop model for identifying *Xc* TALE targets

In our collection, most of the *Xc* strains containing *tal* genes were isolated from *B. oleracea* or *B. rapa* (Table S1). Because the genome of *Brp* was determined (http://plants.ensembl.org/Brassica_rapa, Wang *et al.*, 2011) and this accession was susceptible to the *Xcc* strains tested, we selected this host to search *in silico* for TALE EBEs. We also searched the Arabidopsis promoterome because *Xcc* strains infect Arabidopsis in nature (Buell, 2002) as well as under laboratory conditions (Meyer *et al.*, 2005; Guy *et al.*, 2013). Somewhat surprisingly, candidate TALE target genes based on EBE prediction were mostly distinct between Arabidopsis and *Brp*, probably because of the overall sequence divergence of the two genera since they split 13–17 Myr ago (Fig. 7b). This poor overlap between predictions in both genera could result from many of the host-specific EBEs being false positives or chance, 'collateral' targets inconsequential to disease. Indeed, one might speculate that the shared candidate targets are more likely to be important. Alternatively, the incongruity may indicate that the *Xc* TALome includes different TALEs for different hosts and that the virulence contribution of individual TALEs may not be conserved throughout the whole host range (Noël *et al.*, 2013). The different TALEs could have been selected to target different *S* genes in different hosts, or have adapted to target allelic variants of the same gene across different hosts. The latter hypothesis could explain why closely related TALE proteins are predicted to activate nearly nonoverlapping gene sets (Fig. 7a). Subtle changes in RVD composition of closely related *Xoc* TALEs, such as might occur during adaptation to different alleles of an *S* gene, have been observed to affect both target predictions and levels of transcriptional activation of targets in rice (Erkes *et al.*, 2017). These considerations are important for pathogen groups like *Xc* that infect a broad range of host plant species, but may also apply to more host-specific

pathogens that nonetheless are exposed to genotypic variability within the host species. Characterization of the virulence contributions of *Xc* TALEs in different hosts is an important first step toward identifying key targets and whether they are shared by or unique to each host.

Biological relevance of TALEs and their plant targets for brassicaceae susceptibility to black rot disease

Experimentally validated targets are often predicted by all the different available algorithms for TALE target prediction, yet their score and ranking by these algorithms can vary due to differences in the algorithms, all of which generate some number of false positives (Doyle *et al.*, 2012; Grau *et al.*, 2013; Noël *et al.*, 2013; Pérez-Quintero *et al.*, 2013). Experimental evidence of TALE-dependent induction of predicted target genes therefore remains essential (Cernadas *et al.*, 2014). Out of the 12 *Brp* TALE target genes selected for experimental validation by quantitative reverse transcription polymerase chain reaction (qRT-PCR), five were transcriptionally activated upon infection by strains carrying the corresponding TALE(s) (Fig. S4). Interestingly, in line with the speculation discussed above that targets predicted in more than one host could be more likely to be real, predicted targets in *Brp* that also were predicted in Arabidopsis were more often validated than those that were predicted only in *Brp* (Table 1). Although this dataset is too small to draw a firm conclusion, these observations suggest that conservation of EBEs in both plant species might be used as a criterion for prioritizing candidates for experimental validation. As indicated earlier, convergent targeting of a putative *S* gene by distinct TALEs might also be weighed positively in prioritizing candidates. Because the *Xc* TALome is composed of seven TALE specificity groups, we believe that future research should focus specifically on targets of representative members of those groups. The effect of mutations in those *tal* genes and ectopic manipulation of expression of the respective targets should be carried out to understand how *Xc* strains use TALEs to promote susceptibility in Brassicaceae.

Acknowledgments

We thank Feng Cheng and the Korea Brassica Genome Resource Bank for providing *Brp* chiifu seeds, the CIRM-CFBP for *Xc* strains, Ulla Bonas and Sabine Kay for anti-AvrBs3-356 antibody and the Xca5 strain derivatives, Ivanna Fuentes for technical support and Orlando De Lange, Alice Boulanger and Alvaro Pérez-Quintero for critically reading the manuscript. This work was supported by PhD grants from the French Ministry of National Education and Research to A.H. and E.G., a French Guyana grant to E.G., Agence Nationale de la Recherche grants (XANTHOMIX ANR-2010-GENM-013-02 and CROpTAL ANR-14-CE19-0002-01) to L.D.N., a NSF Plant Genome Research Program award to A.J.B. (IOS 1238189), an Agropolis Foundation grant (#1200-003) to B.S. and a French Laboratory of Excellence project (TULIP ANR-10-LABX-41; ANR-11-IDEX-0002-02). This work benefited from interactions promoted by COST Actions FA1208 and CA16107.

Author contributions

N.D., B.S., M.A., A.J.B., T.B., S.P. and L.D.N. planned and designed the research; N.D., B.S., E.L.D., L.F-B., S.C., E.L., E.G. A.H., L.D.N. performed the experiments; N.D., B.S., E.L.D., A.C., M.A., A.J.B. and L.D.N. analyzed the data; N.D., B.S. and L.D.N. wrote the manuscript, with assistance from E.L.D., S.C., A.H., S.P., M.A. and A.J.B.

ORCID

Nicolas Denancé http://orcid.org/0000-0003-0173-3970
Boris Szurek http://orcid.org/0000-0002-1808-7082
Adam J. Bogdanove http://orcid.org/0000-0003-1683-4117
Laurent D. Noël http://orcid.org/0000-0002-0110-1423

References

van den Ackerveken G, Marois E, Bonas U. 1996. Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* **87**: 1307–1316.

Ah-You N, Gagnevin L, Chiroleu F, Jouen E, Neto JR, Pruvost O. 2007. Pathological variations within *Xanthomonas campestris* pv. *mangiferaeindicae* support its separation into three distinct pathovars that can be distinguished by amplified fragment length polymorphism. *Phytopathology* **97**: 1568–1577.

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.

Blanvillain S, Meyer D, Boulanger A, Lautier M, Guynet C, Denance N, Vasse J, Lauber E, Arlat M. 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS ONE* **2**: e224.

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

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

Bogdanove AJ, Koebnik R, Lu H, Furutani A, Angiuoli SV, Patil PB, Van Sluys MA, Ryan RP, Meyer DF, Han SW *et al.* 2011. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *Journal of Bacteriology* **193**: 5450–5464.

Bolot S, Guy E, Carrère S, Barbe V, Arlat M, Noël LD. 2013. Genome Sequence of *Xanthomonas campestris* pv. *campestris* strain Xca5. *Genome Announcements* **1**: e00032-12.

Bonas U, Conrads-Strauch J, Balbo I. 1993. Resistance in tomato to *Xanthomonas campestris* pv *vesicatoria* is determined by alleles of the pepper-specific avirulence gene *avrBs3*. *Molecular & General Genetics* **238**: 261–269.

Bonas U, Stall RE, Staskawicz B. 1989. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Molecular & General Genetics* **218**: 127–136.

Booher NJ, Carpenter SC, Sebra RP, Wang L, Salzberg SL, Leach JE, Bogdanove AJ. 2015. Single molecule real-time sequencing of *Xanthomonas oryzae* genomes reveals a dynamic structure and complex *TAL* (transcription activator-like) effector gene relationships. *Microbial Genomics* **1**: doi: 10.1099/mgen.0.000032.

Buell CR. 2002. Interactions between *Xanthomonas* species and *Arabidopsis thaliana*. *Arabidopsis Book* **1**: e0031.

Büttner D, Bonas U. 2010. Regulation and secretion of *Xanthomonas* virulence factors. *FEMS Microbiology Reviews* **34**: 107–133.

Castiblanco LF, Gil J, Rojas A, Osorio D, Gutierrez S, Munoz-Bodnar A, Perez-Quintero AL, Koebnik R, Szurek B, Lopez C *et al.* 2013. TALE1 from *Xanthomonas axonopodis* pv. *manihotis* acts as a transcriptional activator in plant cells and is important for pathogenicity in cassava plants. *Molecular Plant Pathology* **14**: 84–95.

Cernadas RA, Doyle EL, Nino-Liu DO, Wilkins KE, Bancroft T, Wang L, Schmidt CL, Caldo R, Yang B, White FF *et al.* 2014. Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathogens* **10**: e1003972.

Chen WP, Kuo TT. 1993. A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic Acids Research* **21**: 2260.

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. *Theoretical and Applied Genetics* **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. *Molecular Plant–Microbe Interactions* **27**: 1186–1198.

Cox KL, Meng F, Wilkins KE, Li F, Wang P, Booher NJ, Carpenter SCD, Chen LQ, Zheng H, Gao X *et al.* 2017. TAL effector driven induction of a *SWEET* gene confers susceptibility to bacterial blight of cotton. *Nature Communications* **8**: 15 588.

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.

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 Research* **40**: W117–W122.

Erkes A, Reschke M, Boch J, Grau J. 2017. Evolution of transcription activator-like effectors in *Xanthomonas oryzae*. *Genome Biology and Evolution* **9**: 1599–1615.

Fargier E, Fischer-Le Saux M, Manceau C. 2011. A multilocus sequence analysis of *Xanthomonas campestris* reveals a complex structure within crucifer-attacking pathovars of this species. *Systematic and Applied Microbiology* **34**: 156–165.

Fargier E, Manceau C. 2007. Pathogenicity assays restrict the species *Xanthomonas campestris* into three pathovars and reveal nine races within *X. campestris* pv. *campestris*. *Plant Pathology* **56**: 805–818.

Ferreira RM, de Oliveira AC, Moreira LM, Belasque J Jr, Gourbeyre E, Siguier P, Ferro MI, Ferro JA, Chandler M, Varani AM. 2015. A TALE of transposition: Tn*3*-like transposons play a major role in the spread of pathogenicity determinants of *Xanthomonas citri* and other xanthomonads. *MBio* **6**: e02505-14.

Gochez AM, Huguet-Tapia JC, Minsavage GV, Shantaraj D, Jalan N, Strauss A, Lahaye T, Wang N, Canteros BI, Jones JB *et al.* 2018. Pacbio sequencing of copper-tolerant *Xanthomonas citri* reveals presence of a chimeric plasmid structure and provides insights into reassortment and shuffling of transcription activator-like effectors among *X. citri* strains. *BMC Genomics* **19**: 16.

Grau J, Reschke M, Erkes A, Streubel J, Morgan RD, Wilson GG, Koebnik R, Boch J. 2016. AnnoTALE: bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Scientific Reports* **6**: 21 077.

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.

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

Guy E, Genissel A, Hajri A, Chabannes M, David P, Carrère S, Lautier M, Roux B, Boureau T, Arlat M *et al.* 2013. Natural genetic variation of *Xanthomonas campestris* pv. campestris pathogenicity on arabidopsis revealed by association and reverse genetics. *MBio* **4**: e00538-12.

Hajri A, Brin C, Hunault G, Lardeux F, Lemaire C, Manceau C, Boureau T, Poussier S. 2009. A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* **4**: e6632.

He YQ, Zhang L, Jiang BL, Zhang ZC, Xu RQ, Tang DJ, Qin J, Jiang W, Zhang X, Liao J *et al.* 2007. Comparative and functional genomics reveals genetic diversity and determinants of host specificity among reference strains and a large collection of Chinese isolates of the phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Biology* **8**: R218.

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. *Proceedings of the National Academy of Sciences, USA* **111**: E521–E529.

Hutin M, Perez-Quintero AL, Lopez C, Szurek B. 2015a. MorTAL Kombat: the story of defense against TAL effectors through loss-of-susceptibility. *Frontiers in Plant Science* **6**: 535.

Hutin M, Sabot F, Ghesquiere A, Koebnik R, Szurek B. 2015b. A knowledge-based molecular screen uncovers a broad-spectrum *OsSWEET14* resistance allele to bacterial blight from wild rice. *Plant Journal* **84**: 694–703.

Jacques MA, Arlat M, Boulanger A, Boureau T, Carrere S, Cesbron S, Chen NW, Cociancich S, Darrasse A, Denance N *et al.* 2016. Using ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annual Review of Phytopathology* **54**: 163–187.

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. *Molecular Plant–Microbe Interactions* **18**: 838–848.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

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 Phytologist* **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 Research* **42**: 7436–7449.

de Lange O, Wolf C, Thiel P, Kruger 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 Research* **43**: 10065–10080.

Lau CH, Zhu H, Tay JC, Li Z, Tay FC, Chen C, Tan WK, Du S, Sia VK, Phang RZ *et al.* 2014. Genetic rearrangements of variable di-residue (RVD)-containing repeat arrays in a baculoviral TALEN system. *Molecular Therapy – Methods & Clinical Development* **1**: 14 050.

Lee BM, Park YJ, Park DS, Kang HW, Kim JG, Song ES, Park IC, Yoon UH, Hahn JH, Koo BS *et al.* 2005. The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. *Nucleic Acids Research* **33**: 577–586.

Mak AN, Bradley P, Bogdanove AJ, Stoddard BL. 2013. TAL effectors: function, structure, engineering and applications. *Current Opinion in Structural Biology* **23**: 93–99.

Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA *et al.* 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology* **13**: 614–629.

Meyer D, Lauber E, Roby D, Arlat M, Kroj T. 2005. Optimization of pathogenicity assays to study the *Arabidopsis thaliana-Xanthomonas campestris* pv. *campestris* pathosystem. *Molecular Plant Pathology* **6**: 327–333.

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

Noël LD, Denancé N, Szurek B. 2013. Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. *Frontiers in Plant Science* **4**: 333.

Ochiai H, Inoue V, Takeya M, Sasaki A, Kaku H. 2005. Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. *JARQ – Japan Agricultural Research Quarterly* **39**: 275–287.

Pérez-Quintero AL, Lamy L, Gordon JL, Escalon A, Cunnac S, Szurek B, Gagnevin L. 2015. QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Frontiers in Plant Science* **6**: 545.

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.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, Couloux A, Darrasse A, Gouzy J, Jacques MA *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.

Qi J, Yu S, Zhang F, Shen X, Zhao X, Yu Y, Zhang D. 2010. Reference gene selection for real-time quantitative polymerase chain reaction of mRNA transcript levels in Chinese cabbage (*Brassica rapa* L. subsp. *pekinensis*). *Plant Molecular Biology Reporter* **28**: 597–604.

Qian W, Jia Y, Ren SX, He YQ, Feng JX, Lu LF, Sun Q, Ying G, Tang DJ, Tang H *et al.* 2005. Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Research* **15**: 757–767.

Ramakers C, Ruijter JM, Deprez RH, Moorman AF. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**: 62–66.

Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences, USA* **106**: 19126–19131.

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.

Roux B, Bolot S, Guy E, Denance N, Lautier M, Jardinaud MF, Fischer-Le Saux M, Portier P, Jacques MA, Gagnevin L *et al.* 2015. Genomics and transcriptomics of *Xanthomonas campestris* species challenge the concept of core type III effectome. *BMC Genomics* **16**: 975.

Ruh M, Briand M, Bonneau S, Jacques MA, Chen NWG. 2017. *Xanthomonas* adaptation to common bean is associated with horizontal transfers of genes encoding TAL effectors. *BMC Genomics* **18**: 670.

Ryan RP, Dow JM. 2011. Communication with a growing family: diffusible signal factor (DSF) signaling in bacteria. *Trends in Microbiology* **19**: 145–152.

Salzberg SL, Sommer DD, Schatz MC, Phillippy AM, Rabinowicz PD, Tsuge S, Furutani A, Ochiai H, Delcher AL, Kelley D *et al.* 2008. Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99A. *BMC Genomics* **9**: 204.

Schandry N, de Lange O, Prior P, Lahaye T. 2016. TALE-like effectors are an ancestral feature of the *Ralstonia solanacearu*m species complex and converge in DNA targeting specificity. *Frontiers in Plant Science* **7**: 1225.

Schornack S, Moscou MJ, Ward ER, Horvath DM. 2013. Engineering plant disease resistance based on TAL effectors. *Annual Review of Phytopathology* **51**: 383–406.

Schwartz AR, Morbitzer R, Lahaye T, Staskawicz BJ. 2017. TALE-induced bHLH transcription factors that activate a pectate lyase contribute to water soaking in bacterial spot of tomato. *Proceedings of the National Academy of Sciences, USA* **114**: E897–E903.

da Silva AC, Ferro JA, Reinach FC, Farah CS, Furlan LR, Quaggio RB, Monteiro-Vitorello CB, Van Sluys MA, Almeida NF, Alves LM *et al.* 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* **417**: 459–463.

Strauss T, van Poecke RM, Strauss A, Romer P, Minsavage GV, Singh S, Wolf C, Kim S, Lee HA, Yeom SI *et al.* 2012. RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. *Proceedings of the National Academy of Sciences, USA* **109**: 19480–19485.

Streubel J, Baum H, Grau J, Stuttman J, Boch J. 2017. Dissection of TALE-dependent gene activation reveals that they induce transcription cooperatively and in both orientations. *PLoS ONE* **12**: e0173580.

Streubel J, Blucher C, Landgraf A, Boch J. 2012. TAL effector RVD specificities and efficiencies. *Nature Biotechnology* **30**: 593–595.

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 Phytologist* **200**: 808–819.

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 *OsTFIIAgamma1* and *OsTFX1* during bacterial blight of rice. *Proceedings of the National Academy of Sciences, USA* **104**: 10720–10725.

Swarup S, De Feyter R, Brlansky RH, Gabriel DW. 1991. A pathogenicity locus from *Xanthomonas citri* enables strains from several pathovars of *X. campestris* to elecit cankerlike lesions on citrus. *Phytopathology* **81**: 802–809.

Tao F, Wang X, Ma C, Yang C, Tang H, Gai Z, Xu P. 2012. Genome sequence of *Xanthomonas campestris* JX, an industrially productive strain for Xanthan gum. *Journal of Bacteriology* **194**: 4755–4756.

Thieme F, Koebnik R, Bekel T, Berger C, Boch J, Buttner D, Caldana C, Gaigalat L, Goesmann A, Kay S *et al.* 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. *Journal of Bacteriology* **187**: 7254–7266.

Tian D, Wang J, Zeng X, Gu K, Qiu C, Yang X, Zhou Z, Goh M, Luo Y, Murata-Hori M *et al.* 2014. The rice TAL effector-dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *Plant Cell* **26**: 497–515.

Vera Cruz CM, Bai J, Ona I, Leung H, Nelson RJ, Mew TW, Leach JE. 2000. Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proceedings of the National Academy of Sciences, USA* **97**: 13500–13505.

Vorhölter FJ, Schneiker S, Goesmann A, Krause L, Bekel T, Kaiser O, Linke B, Patschkowski T, Ruckert C, Schmid J *et al.* 2008. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *Journal of Biotechnology* **134**: 33–45.

Wang L, Rinaldi FC, Singh P, Doyle EL, Dubrow ZE, Tran TT, Perez-Quintero AL, Szurek B, Bogdanove AJ. 2017. TAL effectors drive transcription bidirectionally in plants. *Molecular Plant* **10**: 285–296.

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F *et al.* 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**: 1035–1039.

Wang C, Zhang X, Fan Y, Gao Y, Zhu Q, Zheng C, Qin T, Li Y, Che J, Zhang M *et al.* 2015. XA23 is an executor R protein and confers broad-spectrum disease resistance in rice. *Molecular Plant* **8**: 290–302.

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

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. *Frontiers Plant Science* **6**: 536.

Xu RQ, Blanvillain S, Feng JX, Jiang BL, Li XZ, Wei HY, Kroj T, Lauber E, Roby D, Chen B *et al.* 2008. AvrAC(Xcc8004), a type III effector with a leucine-rich repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. *Journal of Bacteriology* **190**: 343–355.

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*. *Molecular Plant–Microbe Interactions* **18**: 142–149.

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

Yang Y, De Feyter R, Gabriel DW. 1994. Host-specific symptoms and increased release of *Xanthomonas citri* and *X. campestris* pv. *malvacearum* from leaves are determined by the 102-bp tandem repeats of *pthA* and *avrb6*, respectively. *Molecular Plant–Microbe Interactions* **7**: 345–355.

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

Yang Y, Gabriel DW. 1995b. *Xanthomonas* avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. *Molecular Plant–Microbe Interactions* **8**: 627–631.

Yang Y, Yuan Q, Gabriel DW. 1996. Watersoaking functlon(s) of XcmH1005 are redundantly encoded by members of the *Xanthomonas avr/pth* gene family. *Molecular Plant–Microbe Interactions* **9**: 105–113.

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 J, Huguet-Tapia JC, Hu Y, Jones J, Wang N, Liu S, White FF. 2017. Homologues of *CsLOB1* in citrus function as disease susceptibility genes in citrus canker. *Molecular Plant Pathology* **18**: 798–810.

Zhang J, Yin Z, White F. 2015. TAL effectors and the executor *R* genes. *Frontiers in Plant Science* **6**: 641.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 *Xcc* strain Xca5Δ*tal (Δtal)* is unchanged in virulence on *Brassica rapa* cv R-o-18 compared to WT strain.

Fig. S2 5′ and 3′ domains of *Xcc tal* genes are mosaics composed of two sequence types.

Fig. S3 Comparison of the genomic environments of *tal* genes in *Xcc*.

Fig. S4 qRT-PCR validation of TALE candidate gene targets in *Brp*.

Fig. S5 Hypothetical scenarios of sequence rearrangements between or within *tal* genes of *Xanthomonas campestris*.

Table S1 List of the 49 WT *Xanthomonas campestris* strains used in this study

Table S2 Primers used for qRT-PCR validation of TALE target candidates

Table S3 ANI values between publicly available genome sequences of *Xanthomonas campestris* strains studied in this manuscript

Table S4 Results of prediction of *Brassica rapa* ssp *pekinensis* (*Brp*) promoter gene targets of *Xanthomonas campestris* TALEs using TARGET FINDER

Methods S1 Statistical analyses.

Methods S2 Annotation and analyses of *tal* sequences in *Xcc* genomes.

Methods S3 TALE target predictions in *B. rapa* ssp *pekinensis*.

Notes S1 DNA alignment of the 5′ region of the ORF of *Xcc tal* genes (ATG to first repeat).

Notes S2 DNA alignment of the 3′ region of the ORF of *Xcc tal* genes (last repeat to stop codon).

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.