

# SCIENTIFIC REPORTS



OPEN

## Rapid and efficient genome-wide characterization of *Xanthomonas* TAL effector genes

Yan-Hua Yu\*, Ye Lu\*, Yong-Qiang He, Sheng Huang &amp; Ji-Liang Tang

Received: 17 March 2015

Accepted: 07 July 2015

Published: 14 August 2015

*Xanthomonas* TALE transcriptional activators act as virulence or avirulence factors by activating host disease susceptibility or resistance genes. Their specificity is determined by a tandem repeat domain. Some *Xanthomonas* pathogens contain 10–30 TALEs per strain. Although TALEs play critical roles in pathogenesis, their studies have so far been limited to a few examples, due to their highly repetitive gene structure and extreme similarity among different members, which constrict sequencing and assembling. To facilitate TALE studies, we developed an efficient and rapid pipeline for genome-wide cloning of *tal* genes as many as possible from a strain. Here, we report the pipeline and its use to identify all 18 *tal* genes from a newly isolated strain of the rice pathogen *Xanthomonas oryzae*. Target prediction revealed a number of potential rice targets including several notable genes such as genes encoding SWEET, WRKY, Hen1, and BAK1 proteins, which provide candidates for further experimental functional analysis of the TALEs.

*Xanthomonas* is a large genus of Gram-negative bacteria that cause disease in hundreds of plant hosts, including many economically important crops<sup>1</sup>. These pathogens use the type III secretion system to inject effector proteins into the plant cell, where the effectors promote disease symptoms or elicit disease resistance by diverse mechanisms<sup>2,3</sup>. The type III effectors can be grouped into different families based on sequence similarity and biochemical activity. The largest family is the TAL (transcription activator-like) (or AvrBs3/PthA) family which functions as transcriptional activators of plant genes<sup>4</sup>. A TAL effector (TALE) contains an N-terminal type III secretion signal, a central DNA-binding region, a nuclear localization signal, and a C-terminal transcriptional activation domain (Supplementary Fig. 1A). The DNA-binding region comprises variable numbers (from 1.5 up to over 30) of tandem 33–34 amino acid repeats that specify target nucleotide sequence. The repeats are almost identical, with variation occurring primarily at amino acids 12 and 13 in each repeat, which are referred as repeat variable di-residue (RVD). The number, sequence, and order of the RVDs across the whole repeat region of a TALE define its recognition and binding of distinctive DNA target<sup>4</sup>. Diverse TALEs with different RVDs may target various plant genes<sup>4–7</sup>. In addition to the importance to pathogen virulence, TALE target DNA-binding domain has been used to develop the so-called TALENs technology for high-throughput targeted genome editing<sup>8–11</sup>, which was named “Method of the Year” for 2011 by Nature Methods<sup>12</sup>.

The number of TALEs in *Xanthomonas* varies in different species, pathovars, and strains. The important rice pathogen *Xanthomonas oryzae* harbors more than 20 TALEs in some strains<sup>13</sup>. Up to date, about 100 TALE-encoding genes (*tal* genes) have been identified. Studies on the identified TALEs have revealed several target genes that confer disease resistance and/or host susceptibility<sup>6</sup>. Recently, a few algorithms have been developed for *in silico* TALE target predictions<sup>14–16</sup>. Even though such predictions are not flawless and the targets predicted for a given TALE could vary from script to script, the candidate targets obtained are of great value to the in-depth functional analysis of TALEs<sup>17</sup>. However, as a consequence of copious TALEs in a genome and their highly repetitive gene structures, TALE studies have been limited

State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, The Key Laboratory of Ministry of Education for Microbial and Plant Genetic Engineering, and College of Life Science and Technology, Guangxi University, 100 Daxue Road, Nanning, Guangxi 530004, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.-L.T. (email: jltang@gxu.edu.cn)

to a few examples over a long period of time. A large number of *tal* genes in *Xanthomonas* remain to be cloned and characterized. The cloning of *tal* genes from a strain could hardly be done by PCR<sup>18</sup>, it generally requires constructing a genomic library followed by dot blot/Southern blot-based or PCR screening<sup>19,20</sup>, which are laborious and inefficient. Furthermore, due to the tandem quasi-identical repeat units, *tal* gene sequencing has long been the bottle neck for large scale *tal* gene analysis. Unlike sequencing non-repetitive DNA molecules, sequencing the tandem repeats in *tal* genes could hardly be done neither by the conventional chromosome walking method nor using new high throughput sequencing technologies<sup>21</sup>. To facilitate TALE studies, we developed an efficient and rapid pipeline for genome-wide cloning of *tal* genes as many as possible from a strain, which include entire target DNA-binding repeat regions. Here, we report the pipeline and its use to identify most, if not all, of *tal* genes in a strain of *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), the causal agent of rice bacterial blight, which is a destructive rice disease worldwide. Subsequent *in silico* target prediction revealed a number of potential TALE targets in rice genome, which provide candidates for further experimentally functional analysis of the TALEs.

## Results

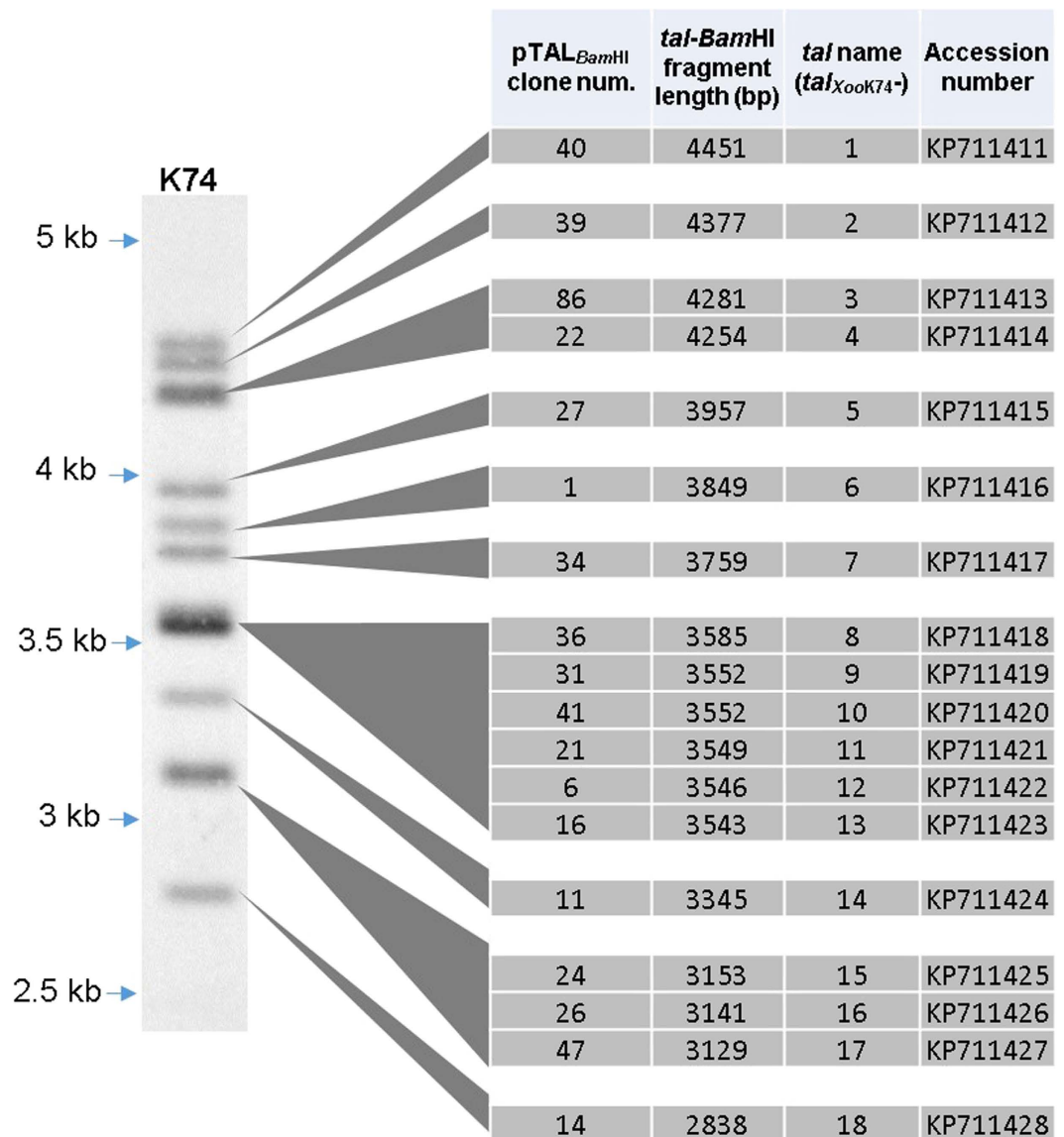
**Estimation of the number of *tal* genes in *Xoo* strain K74 by RFLP analysis.** Strain K74 is a spontaneous streptomycin-resistant mutant of *Xoo* strain 1074 collected from southern coastline region of China with strong virulence to a number of rice cultivars<sup>22</sup> (Supplementary Table 1). To approximate the number of *tal* genes in the strain, we performed a RFLP analysis. Since typical *tal* genes contain two conserved *Bam*HI restriction sites (Supplementary Fig. 1A), *Bam*HI was used to digest the genomic DNA of strain K74 followed by agarose gel electrophoresis. After transferred to a membrane from the gel, the DNA fragments were hybridized with labeled 500 bp N-terminal fragment of the previously identified *tal* gene *talC*<sup>23</sup> (Supplementary Fig. 1A). The hybridization result showed 10 clear, visible bands with sizes ranging from 2.8–4.5 kb. Of which, seven bands displayed similar hybridization signal intensity, while three exhibited significantly stronger signal intensity (Fig. 1), indicating more than one *tal* gene copy within each of the three bands. The RFLP result suggests that the genome of strain K74 may contain more than 13 *tal* gene copies.

**Cloning of *tal-Bam*HI fragments containing entire target DNA-binding motif.** As shown in Supplementary Figure 1A, the *Bam*HI fragment of a *tal* gene comprises a very large portion of the gene, including the entire tandem repeat region that encodes the essence for the functional specificity of TALEs. In addition, previous studies demonstrated that the *Bam*HI fragments of *tal* genes can be adapted with extraneous C-terminal transcriptional activation domain without alternating their original activities<sup>24</sup>. Therefore, we focused our study on the *Bam*HI fragments of *tal* genes in order to facilitate not only the subcloning and sequencing of the tandem repeats, but also to maintain the adaptation for further functional analysis. To this end, we envisioned a simple strategy with advantages in cloning massively *Bam*HI fragments of *tal* genes from strain K74 and screening unique types of *tal-Bam*HI fragments.

To increase cloning efficiency and facilitate the following downstream screening, the portion of *Bam*HI-digested total DNA of strain K74 with sizes ranging from 1.5 to 7.5 kb were recovered from agarose gel and cloned into the vector pUC19, followed by transformed into *E. coli* strain Trans5 $\alpha$ . About 3,000 individual transformant colonies were collected for dot blotting with the same probe as used for RFLP analysis. In total, 115 positive clones were obtained, which were further confirmed by Southern blot (Supplementary Fig. 2). These positive clones were denoted as pTAL<sub>*Bam*HI</sub> clones. Restriction analysis by *Bam*HI digestion revealed that 99 of the clones contained only one fragment, while 16 contained more than one, in addition to the vector fragment (data not shown). The Southern blot result showed that the sizes of hybridized *Bam*HI fragments were consistent with the RFLP pattern of *tal* genes in strain K74.

It is unlikely that the obtained 115 pTAL<sub>*Bam*HI</sub> clones derive from the same *tal-Bam*HI fragment. To isolate all unique *tal-Bam*HI fragments from the 115 pTAL<sub>*Bam*HI</sub> clones, we determined the DNA sequences of both 5' and 3' ends of the tandem repeat domain and their adjacent regions in all of the clones using the primers P1F and P1R (Supplementary Fig. 1B and Supplementary Table 2). Pairwise alignments were systematically performed with the experimental sequences to generate the identity matrix of Supplementary Tables 3 and 4. The clones containing *Bam*HI fragments with the same size and identical DNA sequences in both 5' and 3' portions of the repeat region were considered to derive from the same *tal* gene in strain K74. Based on these criteria, all pTAL<sub>*Bam*HI</sub> clones were re-classified into 18 groups which represent 18 unique *tal-Bam*HI fragments. Each of the 18 unique *tal-Bam*HI fragments was represented by at least four pTAL<sub>*Bam*HI</sub> clones, suggesting that the 18 groups may cover all of the *tal* genes in strain K74. These data indicate that the 10 RFLP bands contain 18 different types of *tal* genes; in other words, strain K74 carries 18 different *tal* genes. Eighteen pTAL<sub>*Bam*HI</sub> clones which represent the 18 unique *tal-Bam*HI fragments (Fig. 1) were chosen for further studies.

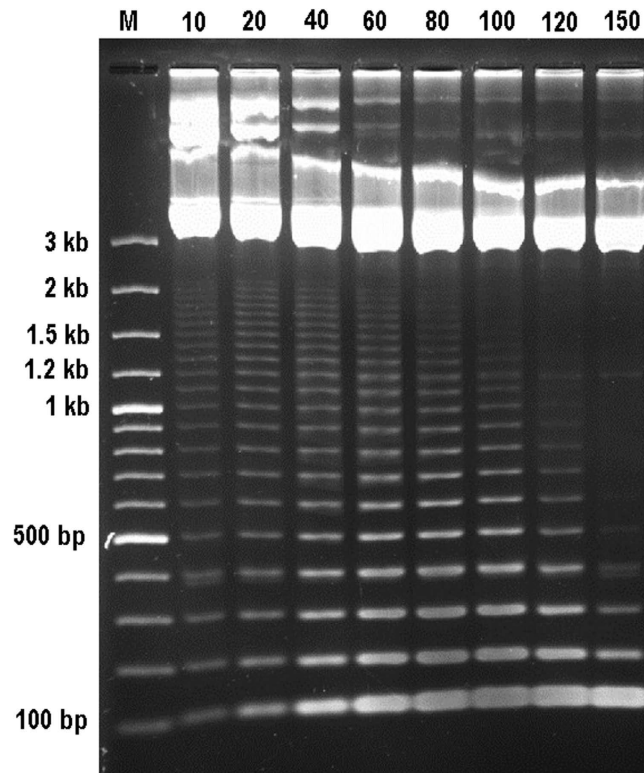
**Screen of a proper restriction enzyme for sequencing the tandem repeat regions.** As described above, the tandem repeat regions of many *tal* genes consist of a large number of quasi identical repeat units. If using short-read technologies to determine the sequence of the repeat regions, the sequence reads would be inaccessible to be assembled. We therefore used Sanger method to sequence the 18 unique *tal-Bam*HI fragments. For this purpose, we screened a proper restriction enzyme that has, if at all possible, only one repeat-exclusive digestion site in each repeat unit for partial digestion and



**Figure 1.** RFLP analysis of the *tal* genes in strain K74 and the correspondence between the banding pattern and the *tal*<sub>XooK74</sub> genes. The genomic DNA was digested with *Bam*HI and a portion of 5' sequence of *talC* was used as probe.

subcloning. DNA sequences of 87 *tal* genes reported previously were analyzed *in silico* and two such candidate enzymes (*Msc*I and *Bsm*BI) were found (Supplementary Table 5). Ideally, partial digestion of a tandem repeat region containing *n* repeat units with an enzyme should yield a ladder-like banding pattern with a range of sizes from one repeat unit to *n*-1 repeat units after agarose gel electrophoresis. Moreover, every band includes not only identical fragments but also fragments consisting of various repeat units. Therefore, cloning and sequencing the mixed populations of one or two bands could yield sufficient data to assemble the entire tandem repeat region. The number of recognition site of *Msc*I is in coherence with the number of repeat units of almost all of the *tal* genes tested (Supplementary Table 5). In addition, *Msc*I site is blunt-ended and absent in the vector pUC19. We thus chose *Msc*I to digest partially the *tal*-*Bam*HI fragments for subsequent sequence determination.

To test if the partial digestion is functional, the *Bam*HI fragment of *talC*<sup>23</sup> was cloned into pUC19 and partially digested with *Msc*I. The digestion time varied from 10 minutes to 150 minutes while the initial quantity of plasmid DNA (2 μg) and *Msc*I (5 units) was maintained at the same level in all digestions. Agarose gel electrophoresis revealed ladder-like banding patterns as expected. As shown in Fig. 2, apart from the bands around 3 kb, which correspond to the vector and its derivatives, a total of 21 cascading bands were visible, which are consistent with the number of repeat units of *talC*. The intensity of the 100-bp band increased along the digestion course, which is consistent with the fact that the final digestion products of the repeat region are the fragments with only one repeat unit. Similarly, partial



**Figure 2.** *MscI* partial digestion of *talC-BamHI* fragment cloned in pUC19. 2  $\mu$ g plasmid DNA was digested with 5 u *MscI* enzyme. The digestion times are given in minutes above the gel. M, 100 bp DNA ladder makers.

*MscI* digestion of the 18 unique pTAL<sub>*BamHI*</sub> clones all yielded expected ladder-like banding patterns after agarose gel electrophoresis (data not shown).

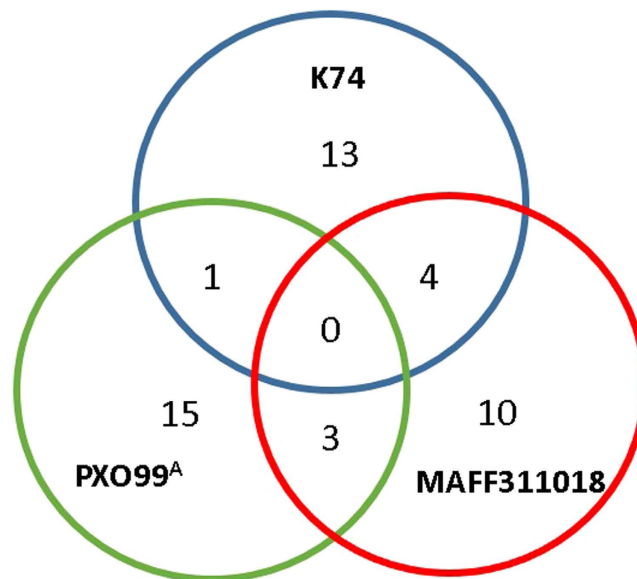
**Sequencing of all 18 unique *tal-BamHI* fragments of strain K74.** Three additional steps were required to sequence all of the 18 unique *tal-BamHI* fragments. First, both 5' and 3' non-repeat regions of the *tal-BamHI* fragments cloned in pUC19 were sequenced using the primers M13F/M13R, P2F/P2R and P3F/P3R (Supplementary Fig. 1B and Supplementary Table 2). Second, the mixed populations of the 1-kb band generated by partial *MscI* digestion and agarose gel electrophoresis were cloned into the vector pCC2FOS and sequenced using the primer pCC2-seqF (Supplementary Table 2) complementary to vector DNA. The number of the clones needed to be sequenced depends on the number of repeat units in the *tal-BamHI* fragment. The more repeat units in the fragment, the more clones need to be sequenced. The sequencing data of the 18 unique *tal-BamHI* fragments were listed in Supplementary Data 1. Finally, the *tal-BamHI* fragments were *de novo* assembled by using the sequence assembly module of Geneious 7 to analyze the sequence data of the repeat region together with the DNA sequences of both 5' and 3' ends of the tandem repeat domain and their adjacent non-repeat regions obtained above. The scheme for DNA sequence assembling of the 18 unique *tal-BamHI* fragments and the resulting DNA sequences are presented in Supplementary Figure 3 and Supplementary Data 2, respectively. The assembled DNA sequence lengths of all *tal-BamHI* fragments are consistent with the sizes revealed by RFLP analysis (Fig. 1). Notably, the biggest *tal-BamHI* fragment with 4,451 bp in length (No. 40) (Fig. 1) contains only 16.5 repeats, and the gene is interrupted by an *IS* element at its 3' end. To complete the *IS* element sequence determination, chromosome walking technology was employed. *talC* was taken as a control and the assembled result showed that its *BamHI* fragment consists of 3678 bp with 21.5 repeat units (Supplementary Fig. 4, Supplementary Data 2 and Supplementary Table 6), which is completely consistent with its published sequence<sup>23</sup>.

According to the sizes of the *tal-BamHI* fragments, the *tal* genes of *Xoo* strain K74 and their corresponding protein products were denoted as *tal*<sub>*XooK74*</sub>-1 to *tal*<sub>*XooK74*</sub>-18 (Fig. 1) and Tal<sub>*XooK74*</sub>-1 to Tal<sub>*XooK74*</sub>-18 (Fig. 3), respectively. The number of repeat units in the TALEs varies from 12.5 to 26.5 (Fig. 3) and the majority of the repeat units encode 34 amino acids. Similar to all known TALEs, the last unit in the tandem repeat motifs of all 18 Tal<sub>*XooK74*</sub> effectors is only a half repeat (Supplementary Table 6).

**Comparison of the Tal<sub>*XooK74*</sub> effectors with other *Xoo* TALEs and *in silico* target prediction.** As described above, DNA-binding specificity of TALEs to plant genes is determined by their RVD strings

Repeat num.	RVDs																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	NS	HD	NG	NG	NG	NG	HD	HD	HD	HD	NN	HD	HD	HD	HD	NN	H*											
2	NI	HG	NI	HG	NI	NI	NI	HD	NN	HD	HD	HD	NG	HD	N*	NI	HD	HD	NN	NI	NN	NN	NG	NN	HD	NG	NS	N*
3	NI	HG	NI	NI	HG	HD	NN	HD	HD	HD	NI	NI	NN	NI	HD	HD	HD	HG	NN	NN	HD	NS	NN	HD	NG	NS	N*	
4	HD	HD	NN	NN	NI	NG	HD	S*	HG	HD	NG	N*	NG	HD	HD	N*	NI	NI	NN	HD	HI	ND	HD	NG	NN	HG	N*	
5	NN	HD	NS	NG	HD	NN	N*	NI	HD	NS	HD	NN	HD	NN	HD	NN	NN	NN	NN	HD	NN	NN	HD	NG				
6	NI	NN	NI	NN	NN	NN	NN	HD	NI	NS	HG	HD	NI	N*	NS	NI	NI	HD	HD	N*	NS	N*						
7	NI	HG	NI	NN	NN	NI	NN	HD	NS	HD	NS	NS	NS	HD	NN	HD	NG	HD	HD	HD	NG	NG						
8	NI	HG	NI	NG	HG	HD	NS	NG	HD	NN	NG	HG	NG	HD	HG	HD	HD	NI	NN	NG								
9	NI	NI	HG	NI	NI	NS	HD	NN	HD	NS	NG	SS	HD	NI	NI	NN	NI	NN	NI	NG								
10	NI	HG	NI	NI	NI	NN	HD	NS	NN	NS	NN	HD	NN	NI	HD	NN	NI	NG	HD	NG								
11	NI	NG	NN	NG	NK	NG	NI	NN	NI	NN	NI	NN	NS	NG	NS	NN	NI	N*	NS	NG								
12	HD	HD	HD	NG	N*	NN	HD	HD	N*	NI	NI	NN	HD	HI	ND	HD	NI	HD	NG	NG								
13	NI	N*	NI	NS	NN	NG	NN	NS	N*	NS	NN	NS	N*	HD	HG	HD	NI	HD	HD	NG								
14	NI	NS	HD	NG	NS	NN	HD	N*	NN	NN	NI	NN	HD	HG	HD	HD	NN	NG										
15	NS	HD	NG	NG	NG	NG	HD	HD	HD	HD	HD	HD	NG	HD	NI	HD	NN	N*										
16	NI	NN	N*	NG	NS	NN	NN	NN	NI	NN	NI	NG	HD	HD	NI	NG												
17	NI	NN	NN	NI	NI	NI	HD	NS	HG	NN	NN	NN	NI	NI	NG	HD												
18	NI	NN	NI	HG	HG	HD	NG	HD	HG	HD	HD	HD	NG															

**Figure 3.** Repeat numbers and RVDs of the 18  $Tal_{XooK74}$  effectors. The last half repeat is counted as one repeat.



**Figure 4.** A Venn diagram illustrating the number of TALEs sharing identical RVD strings among *Xoo* strains K74, PXO99<sup>A</sup> and MAFF311018.

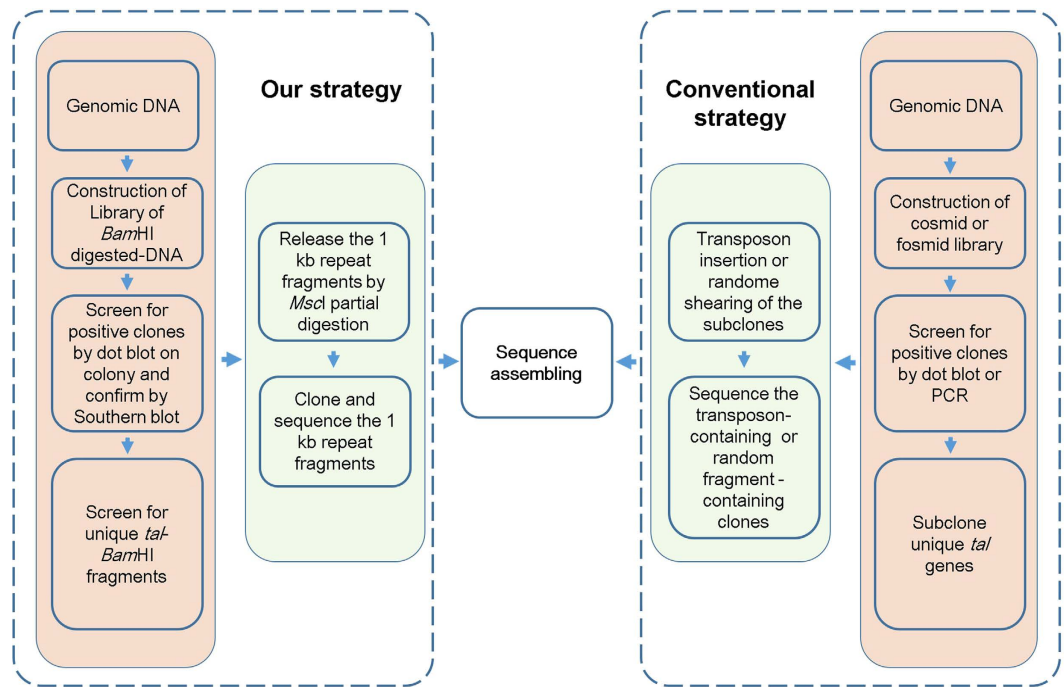
in the repeat motifs. To gain some general knowledge on the composition features and potential functions of the RVDs in the  $Tal_{XooK74}$  effectors, we compared them with those in the TALEs identified from the Philippine strain PXO99<sup>A</sup> and the Japanese strain MAFF311018<sup>25,26</sup>. The result showed that the RVDs of  $Tal_{XooK74}$ -1 and  $Tal_{XooK74}$ -15 are similar to each other, and the RVDs of one ( $Tal_{XooK74}$ -12) and four ( $Tal_{XooK74}$ -3,  $Tal_{XooK74}$ -11,  $Tal_{XooK74}$ -13, and  $Tal_{XooK74}$ -17) K74 TALEs are identical to those of PXO99<sup>A</sup> and MAFF311018 TALEs, respectively, and no TALE is identical among the three strains (Fig. 4; Supplementary Table 7). In total, there are 13 TALEs which are similar among the three strains and may compose a set of core TALEs of the *Xoo* strains (Supplementary Table 7). Interestingly, the RVD strings of  $Tal_{XooK74}$ -4 and  $Tal_{XooK74}$ -16 are very similar to those of the TALEs AvrXa23 and AvrXa27 identified in PXO99<sup>A</sup>, respectively (Supplementary Table 7), which are avirulence determinants conferring incompatibility on rice cultivars harboring the *R* genes *Xa23* and *Xa27*, respectively<sup>22</sup>.

As mentioned above, several algorithms have been developed for *in silico* TALE target prediction. To gain an insight into potential roles of the 18  $Tal_{XooK74}$  effectors in the interaction between strain K74 and its rice host, we employed one of such algorithms, TALgetter 1.0<sup>15</sup> to search rice promoterome (1 kb sequences upstream all the translation start sites) for latent targets of the TALEs. Table 1 lists some of

Tal <sub>XooK74</sub>	Target ID	Annotation	Target sequences	Score	p-value	E-value
1	LOC_Os06g18040.1	hypothetical protein	TCCTTTTCCCCTCCCCGC	-9.1854	2.1537E-08	1.4
2 <sup>a</sup>	LOC_Os11g31190.1	nodulin MtN3 family protein	TATATAAACCCCTCCAACCAGGTGCTAA	-16.1988	0	0
3 <sup>a</sup>	LOC_Os08g30910.1	YDG/SRA domain containing protein	TCTCCTCGCCAAAACCCCCCAACTCC	-22.1250	3.1084E-09	0.2
	LOC_Os01g61080.1	WRKY24	TCTATCCGCCAAAACCCCTACCCGCTCC	-22.7700	1.2434E-08	0.8
4	LOC_Os12g38380.1	tetratricopeptide repeat containing protein	TCCGAATCCGCTCTCCCAATCCTTCT	-21.8771	3.1084E-09	0.2
	LOC_Os12g18110.1	PQ loop repeat domain containing protein	CCCGGTTCGTCCCTCCAAGCCCCAACCC	-23.1760	1.7096E-08	1.1
5	LOC_Os07g09384.1	expressed protein	AGCCTCGCACGCGCGGAGGCCG	-16.0279	4.6387E-09	0.3
6 <sup>b</sup>	LOC_Os07g05400.1	retrotransposon protein	TAGAAGGGACAATCGCTAACCTCC	-18.3096	4.9530E-08	3.2
7 <sup>a</sup>	LOC_Os12g28330.1	retrotransposon protein	TATAAACCCCAACCCCTCCCCT	-16.6017	2.1647E-08	1.4
8	LOC_Os01g40290.1	expressed protein	TATATACCTCGTTTCTCCAGG	-11.4833	0	0
	LOC_Os12g30590.1	disease resistance protein RGA2	TATATTCATCATTTTACCAGT	-12.1436	1.5431E-09	0.1
9	LOC_Os06g29530.1	retrotransposon protein	TAATAGACGCATCCACAAGAT	-13.4391	2.0060E-08	1.3
10	LOC_Os08g07760.1	brassinosteroid insensitive 1-associated receptor kinase 1 precursor	TATAAAGCGAGGCGACGAACT	-12.4998	3.0861E-09	0.2
11	LOC_Os01g47650.1	vacuolar protein sorting-associated protein 16	TATGTCTAGAGAAATAGAAGT	-13.3070	1.8517E-08	1.2
12	LOC_Os07g06970.1	HEN1, putative	TCCCTTCCCTAAACCCCACTT	-10.4036	0	0
13	LOC_Os06g29790.1	phosphate transporter 1	TATAAGTGACAGCCCTCCCT	-11.2732	1.5431E-09	0.1
14	LOC_Os08g03550.1	expressed protein	TAGTCGCCGAGCACC	-12.6372	5.2400E-08	3.4
15	LOC_Os10g28990.1	MBTB42 - Bric-a-Brac, Tramtrack, Broad Complex BTB domain with Meprin and TRAF Homology MATH domain	TCCTTTTCCACGCGCAGC	-10.9153	2.4638E-08	1.6
16	LOC_Os06g32590.1	expressed protein	TACCTAGGCAGATCCAT	-8.5410	1.5368E-08	1
17	LOC_Os09g25200.1	protein binding protein	TAGCAAACATGCGAATC	-8.2442	1.2294E-08	0.8
18	LOC_Os08g05910.1	peptide transporter PTR2	TAGATTCTCTCCCT	-3.3726	7.6604E-09	0.5

**Table 1.** List of the top predicted targets for each of the 18 Tal<sub>XooK74</sub> effectors. <sup>a</sup>Counterpart presents in MAFF311018, but absent in PXO99<sup>A</sup>. <sup>b</sup>Counterpart presents in PXO99<sup>A</sup> but absent in MAFF311018.

the top score candidate targets predicted for the Tal<sub>XooK74</sub> effectors. The top ten high score predicted targets for each Tal<sub>XooK74</sub> effector were collected and listed in Supplementary Table 8. The predicted targets include several remarkable genes such as *Os11N3* (*Os11g31190*), *Os01g61080*, *Os12g18110*, *Os08g07760*, and *Os07g06970*, which were predicted to be potential targets for Tal<sub>XooK74</sub>-2, Tal<sub>XooK74</sub>-3, Tal<sub>XooK74</sub>-4, Tal<sub>XooK74</sub>-10, and Tal<sub>XooK74</sub>-12, respectively. *Os11N3* encodes a member of the SWEET sucrose-efflux transporter family. It has been demonstrated that *Xoo* activates the gene by the TALEs AvrXa7 and TalC, and thus diverts sugars from the plant cell so as to gratify the pathogen's nutritional needs and enhances its virulence<sup>24,27</sup>. The predicted recognition site of Tal<sub>XooK74</sub>-2 is 232 bp upstream of the start codon of *Os11N3*, which is different but overlapping with those recognized by AvrXa7 and its variants. The RVD string of Tal<sub>XooK74</sub>-2 differs mainly in additional "NI" and "HG" at the N-terminal from AvrXa7 and its variants (Supplementary Table 7). *Os01g61080* encodes the rice WRKY transcription factor OsWRKY24. Plant WRKY transcription factors are known to be involved in response to bacterial and environmental stresses via interactions with plant hormones. Interestingly, over expression of WRKY33 in *Arabidopsis*, which is an ortholog of rice OsWRKY24, can not only enhance resistance against necrotrophic fungal pathogens, but also increase susceptibility to the bacterial pathogen *Pseudomonas syringae*<sup>28</sup>. *Os08g07760* encodes a brassinosteroid insensitive 1-associated receptor kinase 1 (BAK1). BAK1 was originally discovered as a coreceptor for the brassinosteroid receptor, and later studies revealed that BAK1 is involved in recognizing the pathogen-associated molecular patterns (PAMPs) and thus leads to the induction of PAMP-triggered immunity (PTI)<sup>29</sup>. It has been demonstrated that several non-TAL effectors of *Pseudomonas* target BAK1, resulting in an abolishment of the pattern recognition receptor (PRR)-PAMPs recognition and signaling transduction<sup>30</sup>. *Os12g18110* which encodes a PQ-loop repeat domain-containing protein is in the top list of Tal<sub>XooK74</sub>-4 predicted targets. PQ-loop proteins may function as transporters of amino acids and other solutes across membranes<sup>31</sup>. *Os07g06970* encodes Hen1 which is a methyltransferase that modifies the 3' terminal nucleoside of small regulatory RNAs and increases their stabilities. Previous prediction using another algorithm showed that *hen1* is a potential target of the TALE Tal9a identified in PXO99<sup>A</sup> strain<sup>16</sup>. As the RVD strings of Tal<sub>XooK74</sub>-12 and Tal9a are identical (Supplementary Table 7), it is not surprising that they share a common target. Although further



**Figure 5.** A flow chart diagram illustrating the steps in the pipeline and in conventional methods.

experiments are needed to evaluate all the potential targets, the predictions provide us with useful clues for studies on the functions of the  $Tal_{XooK74}$  effectors.

## Discussion

Here, we report a simple and convenient method for genome-wide identification of *Xanthomonas tal* genes. To meet the challenges of highly conserved *tal* gene structures and their large tandem repeat segments, divers strategies have been used to determine accurately the DNA sequences of the repeat region; however, most of the strategies rely on the use of transposon-based or random DNA shearing sequencing technologies<sup>19,20</sup>, which are cumbersome and expensive. The pipeline described in this paper aimed to overcome these drawbacks with simple experiments and approaches. Most experiments involved in the pipeline are standard routine experiments in many laboratories. The method is efficient, rapid, and economical. In our laboratory, a well-trained technician can complete the pipeline to identify *tal* genes genome-widely from one or two *Xoo* strains in less than one month. A flow chart diagram illustrating the steps in the pipeline is shown in Fig. 5.

In the pipeline, there are two key steps for *tal* gene sequencing. The first one is the screening of a proper restriction enzyme for partial digestion of the *tal-Bam*HI fragments. An ideal enzyme should have only one repeat-exclusive restriction site within every repeat unit, and if not such perfect enzyme available, restriction enzymes with a single recognition site in each of the most repeat units are sufficient for subsequent sequencing and assembling. Use of an enzyme with too many recognition sites in the repeats will increase the sequencing and assembling-associated workload. As shown in Supplementary Table 5, in addition to *Msc*I, *Bsm*BI is also a good candidate. Another key step is the sequencing of the mixed populations of the 1-kb fragments generated by partial digestion with the restriction enzyme selected. Each repeat unit in a *tal* gene consists of about 100bp (encoding 33–34 aa) so that an 1-kb fragment should include about 10 repeat units. The mixed 1-kb fragments comprise all possible combinations of different 10-tandem-repeat-unit fragments. Therefore, sequencing and assembling of the mixed 1-kb fragment populations will reveal the entire tandem repeats in a *tal* gene. Since the normal credible read length obtained by Sanger sequencing method is around 1-kb, choosing the 1-kb fragments can minimize the sequencing cost while maintaining a high efficiency in assembling.

By the method, we were able to dissect genome-widely the *tal* genes of the newly isolated *Xoo* strain K74 without any prior genome sequence information. Since each of the 18 unique *tal-Bam*HI fragments was represented by at least four pTAL<sub>*Bam*HI</sub> clones and no solitary pTAL<sub>*Bam*HI</sub> clone was found, it is highly possible that the 18 *tal* genes identified in this study compose all the *tal* genes of strain K74. To the best of our knowledge, this is the first report on genome-wide identification of *tal* genes from a strain whose genome sequences are not available. As described above, TAL effectors play very important roles in *Xanthomonas*-host interactions. They function as virulence or avirulence factors which promote the expression of host genes associated with disease symptoms or trigger genes that confer disease resistance,

and their target specificity is determined by the RVDs. Identification of more *tal* genes from diverse *Xanthomonas* species, pathovars, or strains and analysis of their RVDs will facilitate greatly the investigation of disease resistance genes and susceptibility genes, and then enhance crop resistance by molecular plant breeding. In addition to *Xanthomonas*, TAL effectors have been also found in *Ralstonia solanacearum*<sup>4</sup>. The method may also apply to genome-wide identification of *Ralstonia tal* genes from diverse strains.

In addition to the variation in TALE numbers, some RVDs of the similar TALEs among strains K74, PXO99<sup>A</sup>, and MAFF311018 are diverse (Supplementary Table 7). Out of curiosity, we performed a target prediction for the counterparts of the Tal<sub>XooK74</sub> in PXO99<sup>A</sup> using TALgetter 1.0 with the same parameters. When comparing the results issued from the two sets of TALEs, we observed that the top targets of Tal<sub>XooK74</sub> strikingly depart from those of the TALEs in PXO99<sup>A</sup> (Supplementary Table 9), indicating that the polymorphism in the RVD sequences underlies diverging functions, which could be beneficial for maintaining virulence in different rice cultivars in different geographic regions.

*In silico* prediction revealed that the rice SWEET gene *Os11N3* is the top potential target of Tal<sub>XooK74</sub>-2 (Supplementary Table 8). As described above, some plant SWEET genes are susceptibility genes whose induction is essential for a pathogen's successful infection. This suggests that Tal<sub>XooK74</sub>-2 may be a major virulence determinant of strain K74, although further experiments are needed to validate the prediction. The prediction also revealed that the BAK1-encoding gene *Os08g07760* and the Hen1 gene *Os07g06970* are the most likely targets for Tal<sub>XooK74</sub>-10 and Tal<sub>XooK74</sub>-12, respectively, and the WRKY gene *Os01g61080* is the second most likely target for Tal<sub>XooK74</sub>-3 (Supplementary Table 8). Are these rice genes and other predicted targets real targets of the TALEs? If so, how does plant response to their induction and what will happen to the pathogen? These will be the topics that merit our further studies.

## Methods

**Bacterial strains, plasmids, and growth conditions.** *Xoo* and *E. coli* strains, as well as the plasmids used in this study are listed in Supplementary Table 1. *Xoo* was grown in PSA medium (10 g of peptone, 10 g of sucrose, 1 g of glutamic acid, 16 g of agar per liter of H<sub>2</sub>O) at 28 °C; *E. coli* was cultured in LB medium (10 g of tryptone, 5 g of yeast extract and 10 g of NaCl per liter of H<sub>2</sub>O) at 37 °C.

**RFLP analysis.** The RFLP analysis of *tal* genes of strain K74 was performed as follows: 5 µg of *Bam*HI completely digested genomic DNA of strain K74 were separated by electrophoresis on a 1% agarose gel in 1 × TAE (Tris-acetate-EDTA) buffer at 60 V for 38 h in a cold chamber; TAE buffer was refreshed every 13 h. The neutral transfer gel treatment procedure and the capillary blotting procedure were carried out according to the manufacture manual, and DNA fragments in the gel were transferred to an Amersham Hybond-N+ membrane (GE Healthcare). A 500-bp fragment of *talC* gene (from 144<sup>th</sup> bp to 644<sup>th</sup> bp) labeled with Roche DIG High Prime DNA Labeling and Detection Starter Kit II was used as probe for the hybridization. The 500-bp fragment of *talC* was amplified by PCR using *talC* DNA as template and the primer set *tal*-probe-F/*tal*-probe-R (Supplementary Table 2).

**Construction of the pTAL<sub>Bam</sub>HI library.** 20 µg of genomic DNA of strain K74 were completely digested with 50 u of the restriction enzyme *Bam*HI (Promega) and electrophoresised. The resulting *Bam*HI fragments with sizes from 1.5 kb to 7.5 kb were recovered from agarose gel using the QIAquick Gel Extraction Kit (Qiagen). The obtained *Bam*HI fragments were ligated into *Bam*HI-cleaved pUC19 vector as follows: 60 ng of purified DNA fragments, 100 ng of pUC19, 2 µl of ligation buffer, and 5 u of T4 ligase (Promega) were mixed; the final reaction volume was adjusted to 20 µl with water. The ligations were performed at 4 °C for 12 h, and the ligation products were desalted using MF<sup>TM</sup> membrane filters (0.025 µm, VSWP, Millipore) before being electroporated into *E. coli* strain Trans5α. The transformants were selected on ampicillin-containing LB medium (100 µg/ml) supplemented with IPTG (120 µg/ml) and X-gal (80 µg/ml), the plating density was around 300 colonies per plate (Ø = 14.4 cm). The dot blot on colony was performed following the manual provided with the Amersham Hybond-N+ Membranes (GE Healthcare). The probe used for the hybridization was the same as used in the RFLP analysis. The potential positive clones were confirmed by Southern blot using the same protocol as used in the RFLP analysis, except that the duration of the electrophoresis was 16 h.

**Construction of pCC2-*tal* clones for sequencing.** pTAL<sub>Bam</sub>HI DNA was partially digested with the restriction enzyme *Msc*I as follows: 2 µg pTAL<sub>Bam</sub>HI plasmid DNA, 5 u of *Msc*I (New England BioLabs), and 2 µl of digestion buffer were mixed; the final reaction volume was adjusted to 20 µl with water, and digestions were performed at 37 °C for a period of time and stopped by heating at 65 °C for 10 minutes and supplementing with 3 µl of 10% SDS-containing loading dye (Takara). The DNA fragments were separated by electrophoresis on agarose gel and the about 1-kb DNA band was sliced out from the gel and purified using QIAquick Gel Extraction Kit. The purified DNA fragments were vacuum-concentrated, and 20 ng of them was used for the ligation with 15 ng of pCC2FOS vector linearized with the restriction enzyme *Eco*72 I (Epicentre). The ligation products were electroporated into the *E. coli* strain EPI300-T1<sup>R</sup> (Epicentre) and the transformants were selected on chloramphenicol-containing LB medium supplemented with IPTG (120 µg/ml) and X-gal (80 µg/ml). The copy number induction of the pCC2FOS-*Msc*I clones were performed prior to plasmid extraction using CopyControl<sup>TM</sup> Induction Solution (Epicentre) according to the provided manual.



**DNA sequencing.** DNA sequencing was performed at Sangon Biotech, and all sequencing primers used in this study were listed in Supplementary Table 2. The sequences were extracted from .ab1 files using Geneious 7.0.6, and reading errors were corrected manually regarding the chromatograms.

**Bioinformatics analysis.** Multiple DNA sequence alignments were performed using the MUSCLE alignment module in Geneious 7. All parameters were set to default except that the output was grouped by similarity. The automatic *de novo* assembling of *tal-BamHI* fragments was carried out using the *De Novo Assemble Module* of Geneious 7. For the assembling of all *tal-BamHI* fragments, the minimum overlap value was set to 50 bp and 200 bp for the non-repeat and repeat regions, respectively; the minimum overlap identity was set to 100% in all cases. The online version of TALgetter (version 1.0)<sup>15</sup> was used to predict the potential direct targets of the Tal<sub>XooK74</sub> effectors in the promoterome of *Oryza sativa* cultivar Nipponbare (1 kb DNA sequences upstream of all start codons). The upstream and downstream offsets were set to 0, the fine-grained p-value mode was selected to compute the p-values, and 1e-06 was set as the maximum p-value to filter the hits generated by the program.

## References

- Ryan, R. P. *et al.* Pathogenomics of *Xanthomonas*: understanding bacterium-plant interactions. *Nat. Rev. Microbiol.* **9**, 344–355 (2011).
- Alfano, J. R. & Collmer, A. Type III secretion system effector proteins: double agents in bacterial disease and plant defense. *Annu. Rev. Phytopathol.* **42**, 385–414 (2004).
- Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M. & Dangl, J. L. Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.* **60**, 425–449 (2006).
- Boch, J. & Bonas, U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
- Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
- Doyle, E. L., Stoddard, B. L., Voytas, D. F. & Bogdanove, A. J. TAL effectors: highly adaptable phyto-bacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol.* **23**, 390–398 (2013).
- Mak, A. N., Bradley, P., Bogdanove, A. J. & Stoddard, B. L. TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* **23**, 93–99 (2013).
- Bogdanove, A. J. & Voytas, D. F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843–1846 (2011).
- Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
- Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2013).
- Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2011).
- Baker, M. Gene-editing nucleases. *Nat Methods* **9**, 23–26 (2012).
- Scholze, H. & Boch, J. TAL effectors are remote controls for gene activation. *Curr. Opin. Microbiol.* **14**, 47–53 (2011).
- Doyle, E. L. *et al.* TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* **40**, 117–122 (2012).
- Grau, J. *et al.* Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* **9**, e1002962 (2013).
- Perez-Quintero, A. L. *et al.* An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS ONE* **8**, e68464 (2013).
- Noël, L. D., Denancé, N. & Szurek, B. Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. *Front. Plant Sci.* **4**, 333 (2013).
- Hommelsheim, C. M., Frantzeskakis, L., Huang, M. & Ülker, B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Sci. Rep.* **4**, 5052 (2014).
- Yang, B. & White, F. F. Diverse members of the AvrBs3/PthA family of type III effectors are major virulence determinants in bacterial blight disease of rice. *Mol. Plant-Microbe Interact.* **11**, 1192–200 (2004).
- Cernadas, R. A. *et al.* Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* **10**, e1003972 (2014).
- Bart, R. *et al.* High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci.* **109**, 1972–1979 (2012).
- Yang, S. Q. *et al.* Molecular and pathogenic characterization of new *Xanthomonas oryzae* pv. *oryzae* strains from the coastline region of Fangchenggang city in China. *World J. Microbiol. Biotechnol.* **29**, 713–720 (2013).
- Yu, Y. *et al.* Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Mol. Plant-Microbe Interact.* **24**, 1102–1113 (2011).
- Zhu, W., Yang, B., Chittoor, J. M., Johnson, L. B. & White, F. F. AvrXa10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol. Plant-Microbe Interact.* **11**, 824–832 (1998).
- Salzberg, S. L. *et al.* Genome sequence and rapid evolution of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* PXO99<sup>A</sup>. *BMC Genomics* **9**, 204 (2008).
- Ochiai, H., Inoue, Y., Takeya, M., Sasaki, A. & Kaku, H. Genome sequence of *Xanthomonas oryzae* pv. *oryzae* suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. *Jpn. Agric. Res. Q.* **39**, 275–287 (2005).
- Antony, G. *et al.* Rice *xa13* recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene *Os-11N3*. *Plant Cell* **22**, 3864–3876 (2010).
- Zheng, Z., Qamar, S. A., Chen, Z. & Mengiste, T. *Arabidopsis* WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J.* **48**, 592–605 (2006).
- Monaghan, J. & Zipfel, C. Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant Biol.* **15**, 349–357 (2012).
- Macho, A. P. & Zipfel, C. Targeting of plant pattern recognition receptor-triggered immunity by bacterial type-III secretion system effectors. *Curr. Opin. Microbiol.* **23C**, 14–22 (2014).
- Jézégou, A. *et al.* Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. *Proc. Natl. Acad. Sci.* **109**, E3434–E3443 (2012).

## Acknowledgements

This work was supported by the 973 Program of the Ministry of Science and Technology of China (2012CB114003 and 2011CB100701) and the Ba Gui Scholar Program of Guangxi Zhuang Autonomous Region of China (2014A002).

## Author Contributions

J.L.T. and Y.H.Y. conceived of and organized the study. Y.H.Y. and Y.L. performed the experiments. Y.H.Y., J.L.T., Y.Q.H. and S.H. performed the data analyses. J.L.T. and Y.H.Y. wrote the paper. All authors have read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Yu, Y.-H. *et al.* Rapid and efficient genome-wide characterization of *Xanthomonas* TAL effector genes. *Sci. Rep.* **5**, 13162; doi: 10.1038/srep13162 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>