



# Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality

Laurent D. Noël<sup>1,2</sup>, Nicolas Denancé<sup>1,2</sup> and Boris Szurek<sup>3\*</sup>

<sup>1</sup> CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR 2594, Castanet-Tolosan, France

<sup>2</sup> INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR 441, Castanet-Tolosan, France

<sup>3</sup> Institut de Recherche pour le Développement, UMR RPB IRD-CIRAD-UM2, Montpellier Cedex 5, France

\*Correspondence: [boris.szurek@ird.fr](mailto:boris.szurek@ird.fr)

## Edited by:

Laurent Deslandes, Centre National de la Recherche Scientifique, France

## Reviewed by:

Adam Bogdanove, Cornell University, USA

Sven-Erik Behrens, Martin Luther University Halle-Wittenberg, Germany

**Keywords:** TAL effectors, targets, prediction, *Xanthomonas*, susceptibility genes

## INTRODUCTION

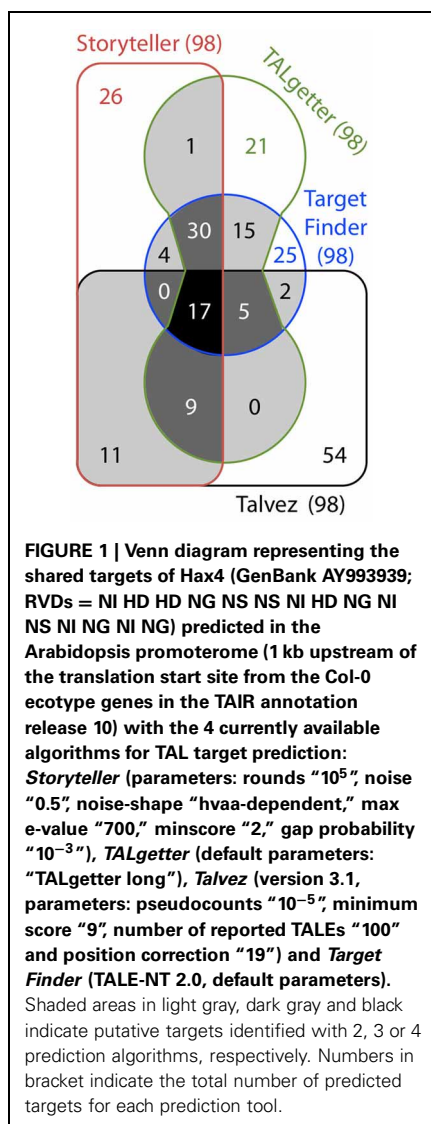
Transcription Activator-Like (TAL) effectors from the plant pathogenic bacteria of the genus *Xanthomonas* are molecular weapons injected into eukaryotic cells to modulate the host transcriptome. Upon delivery, TAL effectors localize into the host cell nucleus and bind to the promoter of plant susceptibility (S) genes to activate their expression and thereby facilitate bacterial multiplication (Boch and Bonas, 2010; Schornack et al., 2013). In resistant plants, a few TAL effectors have been shown to bind to promoters of executor resistance (R) genes, resulting in localized cell death and preventing pathogen spread (reviewed in Doyle et al., 2013). Remarkably, TAL effectors harbor a novel type of DNA-binding domain with a unique modular architecture composed of 1.5–33.5 almost identical tandem repeats of 33–35 amino acids. Each repeat type specifies one or more bases through direct interaction with the second amino acid in a centrally located “Repeat Variable Di-residue” (RVD). The number and sequence of the RVDs across the whole repeat region of the TAL protein defines the DNA target. The code of DNA-binding specificity of *Xanthomonas* TAL effectors was inferred from experimental, computational and later on structural approaches (Boch et al., 2009; Moscou and Bogdanove, 2009; Deng et al., 2012; Mak et al., 2012). This new paradigm for protein-DNA interaction is now revolutionizing our perspectives for the understanding of TAL effectors roles during plant disease and defense since the identification of their plant targets is largely facilitated. A few algorithms are now available to predict *in silico* candidate

genes of a given TAL effector. This *Opinion* gives an overview of the current tools and strategies that may be applied for finding targets of TAL effectors. We also raise limitations and pitfalls and emphasize what may be improved to gain in prediction accuracy. Finally, we also highlight several perspectives offered by these new tools.

## *In silico* PREDICTION OF TAL EFFECTORS TARGETS

One major output of the modular TAL effector–DNA recognition code discovery is the possibility to predict through computer programs, the DNA binding sites of a TAL effector within a whole plant genome or promoterome (i.e., the sequences immediately upstream of the transcriptional start sites) of any sequenced organism. Four bio-informatic tools are currently available and enable to scan genomes for TAL effectors binding sites, rapidly providing users with lists of potential S or R targets. *Target Finder* from the TALE-NT 2.0 suite (<https://tale-nt.cac.cornell.edu/>, Doyle et al., 2012), *Talvez* (<http://bioinfo.mpl.ird.fr/cgi-bin/talvez/talvez.cgi>, Pérez-Quintero et al., 2013) and *Storyteller* (<http://bioinfo-prod.mpl.ird.fr/xantho/tales>, Pérez-Quintero et al., 2013) algorithms are available as web interface and/or standalone software. For these three examples, predictions rely on the use of a RVD-nucleotide association matrix based on known TAL effector–target pairs, to convert a sequence of RVDs of a given TAL effector into a positional weight matrix (PWM). These PWM are regularly updated based on novel experimental insights into TAL-DNA binding or the availability of experimentally confirmed TAL target sequences. *Target Finder*

and *Talvez* both use the PWM to scan and score all possible binding sites in a promoter region with a log-likelihood function. In contrast, *Storyteller* uses this matrix to generate a set of possible binding sequences and takes the advantage of a faster pattern-search algorithm based on Hidden Markov models. Moreover, *Talvez* incorporates a position correction parameter, which enables to tolerate RVD-nucleotide mismatches toward the C-terminal end of RVD sequences and improves target sites prediction. Finally, *TALgetter* (<http://galaxy.informatik.uni-halle.de>; web interface or standalone) differs from the above-mentioned programs as it is based on a statistical model which parameters are estimated from training data computationally (Grau et al., 2013). Furthermore, *TALgetter* decodes the RVDs according to their binding specificity, but takes into account RVD “efficiency” or affinity, as reported by Streubel et al. (2012). Though these predictions yield a number of validated targets, we are still at early days. As an example, we used the Hax4 RVDs from *Xanthomonas campestris* pv. *campestris* strain Xca5 (Kay et al., 2005; Bolot et al., 2013) to mine for Arabidopsis targets in Col-0 promoterome using all 4 algorithms mentioned above. Among the 98 top targets identified by each algorithm, only 17 targets were predicted by all 4 algorithms and 51 by at least 3 algorithms (Figure 1). Although our knowledge on the efficiency (i.e., the percentage of validation of the predictions) of these bioinformatics tools remains poorly documented, significant differences exist between the four algorithms. Thus, combining predictions might help to reduce numbers of true



targets for subsequent experimental validation. Yet, false negatives appear as the greatest threats in such approach since true biological targets could be missed this way. Further experimental validation of TAL targets in different plant genomes are needed to improve the quality of the algorithms and move toward a higher confidence in the predictions.

## PREDICTIONS: CAVEATS AND POSSIBLE IMPROVEMENTS

Yet, our incomplete knowledge of the molecular mechanisms underlying the TAL-DNA interaction and subsequent transcriptional activation is a major limit of these predictions. Since the RVD-DNA code is somewhat degenerate, predictions for TALs with fewer repeats or rich in

unspecific RVDs will yield significant amounts of false positive/false negative target sites especially when scrutinizing large plant genomes. Besides, our understanding of the relative contribution of each individual RVDs to the general protein affinity is still very scarce. Recently, it was shown both by reporter gene expression-based *in vivo* assays and/or biochemical studies that RVDs display different affinity to their favored nucleotides (Christian et al., 2012; Streubel et al., 2012; Meckler et al., 2013). These pioneer studies clearly point to the necessity of systematically evaluating the affinity of each individual most frequent but also rare RVDs for a given nucleotide. Other uncertainties include our incapacity to predict the effect of neighboring RVDs over the binding of a particular RVD, as well as the influence of the binding-site direct environment and the status of epigenetic marks. Finally, another source of inaccuracy in DNA-binding sites prediction deals with our difficulty to estimate TAL effectors tolerance for imperfect pairings which may vary depending on the type, position and context of the mismatch (Doyle et al., 2013). In the same line of idea, Meckler et al. (2013) recently showed that N-terminal RVDs contribute more to the overall DNA affinity than C-terminal RVDs. This result is corroborated by the analysis of Pérez-Quintero et al. (2013), showing from a set of well-characterized RVD-DNA interactions that perfect RVD-nucleotide pairing in TAL effectors N-terminal region (first 15–19 RVDs) probably determines for the most part the target DNA recognition and activity. Thus, mismatches in the C-terminal end of the repeat region generally appear to be better tolerated than in the N-terminal end. Altogether, this illustrates the fact that additional systematic experiments of both the binding affinity and specificity of each RVDs for their preferred nucleotides are required to optimize current predictive models, which would also gain in accuracy if trained with additional experimentally validated pairs of TALs and targets.

Independently of DNA-binding itself, transcriptional activation was strongly enhanced for TAL target sites in the –300 to +200 region relative to the transcriptional start site (TSS, Grau et al., 2013). Thus, proper structural annotation of

genomes including RNAseq-based or EST based annotation of TSS should greatly enhance the quality of the predictions. Though not formally included in the current algorithms, filtering manually for putative target sites close to transcriptional or translational start sites is advisable.

## **In silico OR WET LAB?: PROBABLY BOTH!**

What comes up as an obvious and promising strategy is the use of experimental data to identify new targets. Recently, *Bs4C* executor target was identified solely based on a thoroughly designed RNAseq approach in pepper. The *X. axonopodis* pv. *vesicatoria* AvrBs4 TAL effector target was pinned down to a single promoter to which direct binding was demonstrated (Strauss et al., 2012). Yet, Q-RT-PCR can also be used to confirm predicted targets and may be a cheap shortcut when whole transcriptome profiling (micro-arrays or RNAseq) is not an option: prediction algorithms yield a number of true positives. For instance, 21 TAL targets predicted by *Target Finder* in the rice genome for 14 presumably *X. oryzae* TALs predicted could be verified experimentally (Doyle et al., 2012). As already shown in several studies, comparing the transcriptome of plants challenged with *Xanthomonas* strains carrying a TAL effector of interest vs. a strain defective for that particular *tal* gene or mock inoculation, produces lists of up-regulated genes which are enriched for direct *S* or *R* targets of the TAL effector under study (Yang et al., 2006; Sugio et al., 2007; Yu et al., 2011). Hence, one strategy for evaluating the validity of computationally predicted virulence targets is certainly to benchmark them against TAL effector-dependent profiling experiments, as successfully applied to assess the validity of *TALgetter* (Grau et al., 2013) and *Talvez* (Pérez-Quintero et al., 2013). Nevertheless, a main concern in the overall process is due to the difficulty of discriminating direct and biologically relevant TAL targets from direct and biologically irrelevant TAL targets or secondary/indirect targets. Indeed, off-targets can be found predicted and induced, inherently to the degeneracy of the TAL effector—DNA recognition code and as exemplified by the *X. oryzae* pv. *oryzae* TAL effector AvrXa7 which induces both the

expression of the well-characterized susceptibility gene *OsSWEET14* and a gene coding for a retrotransposon (Li et al., 2012). Secondary indirect targets might be induced independently of the presence of the *tal* gene or as a result of the induction of a TAL direct target. The use of a cycloheximide treatment can help to identify genes which expression does not directly result from TAL activity. One alternative way to counter select off-targets may be to favor candidate targets subjected to functional convergence events, as illustrated for the rice susceptibility gene *OsSWEET14*, which was found to be activated by 4 different TAL effectors originating from 4 different strains of *Xoo* and binding to 3 different target sites in the *OsSWEET14* promoter. Upon the analysis of the *Xoo* TAL repertoire for which targets were predicted and compared to publicly available expression data, several instances of functional convergence between different strains could be demonstrated (Pérez-Quintero et al., 2013).

### PERSPECTIVES: TAL S TARGETS AS NEW TOOLS TO DECIPHER HOST SPECIALIZATION OF *Xanthomonas* SPECIES?

Despite recent breakthrough in TAL effectors biology, the contribution of TAL targets in promoting susceptibility is yet poorly understood. This is particularly true considering the diversity of the *Xanthomonas* genus (27 species and more than 100 pathovars), of the diseases caused on more than 400 different host plants and of the corresponding TAL repertoire (none to 26 TAL copies per strain). The discovery of the *Xanthomonas* TALome is a major task which is seriously hindered by the fact that current sequencing technologies and genome assembly pipelines cannot properly assemble the highly repetitive TAL DNA sequences from whole genome shotgun sequencing data. Also, our knowledge about the relative contribution of TAL effectors to pathogenicity in strains containing multiple *tal* genes is limited to a “happy few” pathosystems such as *X. axonopodis* pv. *vesicatoria*, *X. citri* pv. *citri*, *X. oryzae* pv. *oryzae*, *X. axonopodis* pv. *malvacearum* and *X. axonopodis* pv. *manihotis*. Revealing the susceptibility genes involved in these processes will be key to deciphering as many potentially unique disease

scenarios and represent unprecedented means to access a wealth of information and dissecting the molecular executors of susceptibility. In fact, identifying major virulence TAL effectors of well-studied and more exotic *Xanthomonas* pathovars and fishing their targets offers a unique strategy to understand what may drive host specialization in a species level.

### CONCLUDING REMARKS

As time passes, experimental data will accumulate and help to refine the prediction algorithms. Yet, the most challenging aspect remains the biology of the *Xanthomonas*/plant interaction. During the co-evolution process, bacteria have selected a TAL repertoire to adapt to the diversity of natural hosts and the selection of novel crop species by humans. The latter might be the reason why some strains of the rice pathogens *Xoo* and *Xoc* have so many TALs (up to 26) (Schornack et al., 2013). Therefore, the choice of the right host plant genotype to find the genuine TAL targets is critical. One will always find a target for a TAL in any plant or even animal genome. The experimental validation of target gene induction or direct TAL-binding to the promoter still does not indicate that the right biological system was studied. If the *tal* gene studied contributes significantly to the pathogenicity on the selected plant genotype, one has the chance to find important S genes. Yet, in nature, the contribution of many TAL effectors to disease development will be subtle and dependent on the plant genotype. This means that in the future, *Xanthomonas* and diseased hosts should be sampled together in epidemics to advance in the identification of genuine TAL targets and in our understanding of *Xanthomonas* virulence strategies. Combining pathosystems isolated from natural epidemics with *in silico*, genomic and transcriptomic approaches are certainly the way to go in the next decade. These approaches should yield a large number of targets which contribute quantitatively to susceptibility and resistance for marker-assisted breeding in important crop species.

### ACKNOWLEDGMENTS

We are grateful to A. Pérez-Quintero for providing us with the predicted targets of

Hax4 obtained with *Talvez* and *Storyteller* and E. Doyle for providing us with Hax4 targets with *Target Finder*. This work was in part supported by the LABEX TULIP (ANR-10-LABX-41) and a Jeunes Chercheurs grant from the Agence Nationale de la Recherche (Xopaque ANR-10-JCJC-1703-01) to Laurent D. Noël.

### REFERENCES

- Boch, J., and Bonas, U. (2010). *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* 48, 419–436. doi: 10.1146/annurev-phyto-080508-081936
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., et al. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509–1512. doi: 10.1126/science.1178811
- Bolot, S., Guy, E., Carrere, S., Barbe, V., Arlat, M., and Noël, L. D. (2013). Genome Sequence of *Xanthomonas campestris* pv. *campestris* Strain Xca5. *Genome Announc.* 1, e00032–e00012.
- Christian, M. L., Demorest, Z. L., Starker, C. G., Osborn, M. J., Nyquist, M. D., Zhang, Y., et al. (2012). Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS ONE* 7:e45383. doi: 10.1371/journal.pone.0045383
- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J. K., et al. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335, 720–723. doi: 10.1126/science.1215670
- Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Vandyk, J. K., et al. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40, W117–W122. doi: 10.1093/nar/gks608
- Doyle, E. L., Stoddard, B. L., Voytas, D.F., and Bogdanove, A. J. (2013). TAL effectors: highly adaptable phyto-bacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol.* 23, 390–398. doi: 10.1016/j.tcb.2013.04.003
- Grau, J., Wolf, A., Reschke, M., Bonas, U., Posch, S., and Boch, J. (2013). Computational predictions provide insights into the biology of TAL effector target sites. *PLoS Comput. Biol.* 9:e1002962. doi: 10.1371/journal.pcbi.1002962
- Kay, S., Boch, J., and Bonas, U. (2005). Characterization of AvrBs3-like effectors from a Brassicaceae pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol. Plant-Microbe Interact.* 18, 838–848. doi: 10.1094/MPMI-18-0838
- Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30, 390–392. doi: 10.1038/nbt.2199
- Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335, 716–719. doi: 10.1126/science.1216211

- Meckler, J. F., Bhakta, M. S., Kim, M. S., Ovadia, R., Habrian, C. H., Zykovich, A., et al. (2013). Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.* 41, 4118–4128. doi: 10.1093/nar/gkt085
- Moscou, M. J., and Bogdanove, A. J., (2009). A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501. doi: 10.1126/science.1178817
- Pérez-Quintero, A., Rodríguez-R, L., Dereeper, A., Lopez, C., Koebnik, R., Szurek, B., et al. (2013). An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS ONE* 8:68464. doi: 10.1371/journal.pone.0068464
- Schornack, S., Moscou, M. J., Ward, E. R., and Horvath, D. M., (2013). Engineering plant disease resistance based on TAL effectors. *Annu. Rev. Phytopathol.* 51, 383–406. doi: 10.1146/annurev-phyto-082712-102255
- Strauss, T., Van Poecke, R. M., Strauss, A., Romer, P., Minsavage, G. V., Singh, S., et al. (2012). RNA-seq pinpoints a *Xanthomonas* TAL-effector activated resistance gene in a large-crop genome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19480–19485. doi: 10.1073/pnas.1212415109
- Streubel, J., Blucher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 30, 593–595. doi: 10.1038/nbt.2304
- Sugio, A., Yang, B., Zhu, T., and White, F. F. (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. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10720–10725. doi: 10.1073/pnas.0701742104
- Yang, B., Sugio, A., and White, F. F. (2006). *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10503–10508. doi: 10.1073/pnas.0604088103
- Yu, Y., Streubel, J., Balzergue, S., Champion, A., Boch, J., Koebnik, R., et al. (2011). Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 *Os11N3* gene. *Mol. Plant-Microbe Interact.* 24, 1102–1113. doi: 10.1094/MPMI-11-10-0254

Received: 12 June 2013; accepted: 08 August 2013; published online: 03 September 2013.

Citation: Noël LD, Denancé N and Szurek B (2013) Predicting promoters targeted by TAL effectors in plant genomes: from dream to reality. *Front. Plant Sci.* 4:333. doi: 10.3389/fpls.2013.00333

This article was submitted to *Plant-Microbe Interaction*, a section of the journal *Frontiers in Plant Science*.

Copyright © 2013 Noël, Denancé and Szurek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.