daTALbase: A Database for Genomic and Transcriptomic Data Related to TAL Effectors

Alvaro L. Pérez-Quintero,^{1,2} Léo Lamy,¹ Carlos A. Zarate,¹ Sébastien Cunnac,¹ Erin Doyle,³ Adam Bogdanove,^{3,4} Boris Szurek,¹ and Alexis Dereeper^{1,*}

¹IRD, Cirad, Université Montpellier, IPME, Montpellier (34000), France; ²Institut de Biologie de l'Ecole Normale Supérieure, Ecole Normale Supérieure, CNRS, INSERM, PSL Research University, 75005 Paris, France; ³Department of Biology, Doane University, 1014 Boswell Avenue, Crete, NE 68333, U.S.A.; and ⁴Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, 334 Plant Science Building, Ithaca, NY 14853, U.S.A.

Accepted 15 November 2017.

Transcription activator-like effectors (TALEs) are proteins found in the genus *Xanthomonas* of phytopathogenic bacteria. These proteins enter the nucleus of cells in the host plant and can induce the expression of susceptibility genes (*S* genes), triggering disease. TALEs bind the promoter region of *S* genes following a specific code, which allows the prediction of binding sites based on TALEs amino acid sequences. New candidate *S* genes can then be discovered by finding the intersection between genes induced in the presence of TALEs and genes containing predicted effector binding elements. By contrasting differential expression data and binding site predictions across different datasets, patterns of TALE diversification or convergence may be unveiled, but this requires the seamless integration of different genomic and transcriptomic data. With this in mind, we present daTALbase, a curated relational database that integrates TALE-related data including bacterial TALE sequences, plant promoter sequences, predicted TALE binding sites, transcriptomic data of host plants in response to TALE-harboring bacteria, and other associated data. The database can be explored to uncover new candidate *S* genes as well as to study variation in TALE repertoires and their corresponding targets. The first version of the database here presented includes data for *Oryza* sp.–*Xanthomonas* pv. *oryzae* interactions. Future versions of the database will incorporate information for other pathosystems involving TALEs.

The plant-pathogenic bacteria of genus *Xanthomonas* cause devastating diseases on a wide range of hosts and impact the yield of important crops such as rice, cassava, cotton, wheat, banana, mango, citrus, and cabbage, both quantitatively and qualitatively (Hayward 1993). In rice, the two closely related pathovars *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas oryzae* pv. *oryzicola* are responsible for two diseases, bacterial leaf blight and bacterial leaf streak, respectively. *X. oryzae* pv. *oryzae* is a vascular pathogen that enters leaves via hydathodes and colonizes the xylem parenchyma, while *X. oryzae* pv. *oryzicola* is an intercellular pathogen that enters through stomata and colonizes the mesophyll

apoplast (Niño-Liu et al. 2006; White and Yang 2009). Yield losses caused by these pathogens can amount up to 50% for *X. oryzae* pv. *oryzae* and 30% for *X. oryzae* pv. *oryzicola*. These diseases are therefore important constraints for rice production worldwide (Niño-Liu et al. 2006).

To colonize their host, *Xanthomonas* species, like other bacteria, rely on a type III secretion system specialized in the injection of virulence factors (also called type III effectors [T3Es]) into the host cell. They notably rely on a family of T3Es known as transcription activator-like effectors (TALEs), which act as plant transcription factors to reprogram the host transcriptome upon translocation into the plant cell and localization to the nucleus (Boch and Bonas 2010). To induce host genes, TALEs are able to directly bind DNA through their central repeat region according to the so-called TALE code (Boch et al. 2009; Moscou and Bogdanove 2009). Each repeat forms a hairpin structure made by two α -helices connected by a loop. Upon binding to DNA, the repeats form a superhelix wrapped around the DNA major groove with the loops from each repeat on the inner side of the helix, directly interacting with the DNA (Deng et al. 2012; Mak et al. 2013). The specificity of interaction with DNA is determined by amino acids located within the loop of each repeat at positions 12 and 13, which are usually highly variable and are, thus, designated RVD, for repeat variable di-residues. Within the RVD, amino acid 12 helps stabilize the loop, while amino acid 13 can interact directly or indirectly with the nitrogenous bases through hydrogen bonds and van der Waals forces (Deng et al. 2012; Mak et al. 2012).

TALE-mediated induction of a subset of genes, referred to as susceptibility genes (*S* genes), can promote host colonization and disease. To date, several of them have been described (particularly in rice), and *S* genes with similar function are often targeted by multiple TALEs in a redundant and convergent manner (Boch et al. 2014; Pérez-Quintero et al. 2013). *S* genes targeted by *X. oryzae* pv. *oryzae* TALEs include sugar transporters of the *SWEET* family (Boch et al. 2014; Chen et al. 2010) as well as multiple types of transcription factors (Sugio et al. 2007). In contrast, OsSULTR3;6, a putative sulfate transporter, is, so far, the only *S* gene identified as target for *X. oryzae* pv. *oryzicola* TALEs (Cernadas et al. 2014). Proposed common targets for *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* include the small RNA 2'-*O*-methyltransferase *HEN1* and a flavanone 3-hydroxylase (F3H), which have been shown to be induced by TALEs from both pathovars (Moscou and Bogdanove 2009; Pérez-Quintero et al. 2013), but no phenotype has yet been associated to their induction. Importantly, plants have evolved different mechanisms to detect or neutralize TALEs. They

*Corresponding author: Alexis Dereeper: E-mail: alexis.dereeper@ird.fr

Funding: This work was funded by the Fondation Agropolis grant number 1403-073, the Agence Nationale de la Recherche grant number ANR-14-CE19-443-0002, and the United States National Science Foundation, grant number IOS-1444511.

include loss-of-susceptibility alleles such as *xa13*, *xa25*, and *xa41*, in which TALE binding to the promoters of *S* genes is precluded by target sequence polymorphism (Chu et al. 2006; Hutin et al. 2015a and b; Liu et al. 2011). Other forms of resistance also entail direct recognition of TALE structures (potentially Xo1 and Xa1) and subsequent defense response activation (Ji et al. 2016; Read et al. 2016; Triplett et al. 2016) or so-called executor *E* gene (*Xa7*, *Xa10*, *Xa23*, *Xa27*) induction (Zhang et al. 2015).

Because the mechanism of action of TALEs is relatively well-understood, they have become an ideal probe to investigate physiological processes governing plant susceptibility to bacteria. Binding sites for TALEs can be predicted in the host genomes, using various available softwares (Doyle et al. 2012; Grau et al. 2013; Pérez-Quintero et al. 2013; Rogers et al. 2015), and these predictions can be contrasted with transcriptomic data to identify genes that are likely to be targets of TALEs, i.e., genes that contain a predicted binding site (effector-binding element [EBE]) in their promoters and that are shown to be induced in presence of a bacteria harboring the TALE (Noël et al. 2013). These candidate targets can then be tested experimentally for either a role in disease (Cernadas et al. 2014), resistance (Strauss et al. 2012), or both.

In recent years, genomic and transcriptomic resources for the rice-*X. oryzae* system have expanded considerably; transcriptomic profiles for plants infected with various *X. oryzae* pv. *oryzicola* and *X. oryzae* pv. *oryzae* strains are becoming increasingly available (Wilkins et al. 2015). SMRT (single molecule, real time) sequencing technologies now allow the straightforward assembly of TALE repetitive regions and several finished *X. oryzae* genomes with full TALome sequences (i.e., TAL effector repertoires) have also been released (Grau et al. 2016; Quibod et al. 2016; Wilkins et al. 2015). Likewise, recent sequencing projects have made available multiple sets of genomic sequences from rice, including fully assembled *de novo* genomes (Chen et al. 2013; Wang et al. 2014) and rich single nucleotide polymorphism (SNP) data encompassing more than 3,000 rice cultivars (The 3,000 Rice Genomes Project 2014). For TALE research, this data holds the promise of not only helping discover new *S* genes but, also, of bringing important insight into the coevolution of the interacting organisms.

While there are currently multiple tools available to predict TALE binding sites (Doyle et al. 2012; Grau et al. 2013; Pérez-Quintero et al. 2013; Rogers et al. 2015) as well as tools for analyzing genomic data (The 3,000 Rice Genomes Project 2014) and pathogen-specific transcriptomic data. (Dash et al. 2012), the type of data produced by these tools is often heterogeneous and comparisons among them are often burdensome and time-consuming. Seeing the need for an accessible way to interrogate these types of data, we here present daTALbase, a relational database that integrates publicly available TALE-related genomic and transcriptomic data. This database will easily allow users to explore TALE sequences from *X. oryzae*, their predicted targets in available *Oryza* sp. genomes, target expression in transcriptomic data, target genomic variation, and more. Future versions of the database will integrate data for other pathosystems.

RESULTS

Description of the database.

The database consists mainly of five types of information: i) TALE sequences, ii) predicted targets for these sequences in promoters of annotated genes in available genomes, iii) orthology relations among genes in the available genomes, iv) genetic variants in the predicted binding sites in promoters, and v) transcriptomic data.

daTALbase v.1 includes a total of 528 TALE sequences from two *X. oryzae* pathovars, *X. oryzae* pv. *oryzae* (30 strains, 270

effectors) and *X. oryzae* pv. *oryzicola* (10 strains, 258 effectors) (Fig. 1A). RVD sequences for these available TALEs were used to predict EBEs on available assembled and annotated *Oryza* genomes (13 genomes in total) (Fig. 1B). A total of 3,405,793 putative EBEs were incorporated into the database. Among them, 259,000 were predicted in the reference *O. sativa* Nipponbare genome, corresponding to 39,811 potential target genes. More precisely, we found 8,472 genes targeted by a single TALE and 9,872 possible “hub” genes that were predicted to be targeted by at least 10 TALEs.

Distribution of EBEs along the *O. sativa* Nipponbare genome is reported in Figure 2 and reveals that predicted EBEs are distributed continuously and homogeneously along the chromosomes for both strains *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* taken together. All information regarding the scoring and location of the EBEs was included in the database. Orthology relations among annotated genes in the available genomes were also predicted to facilitate comparisons between different species or cultivars. We identified 50,015 orthologous gene sets containing 392,183 genes, accounting for between 66% (21,194 of 32,037 genes of *O. brachyantha*) and 92.7% (35,447 of 38,245 genes of *O. sativa* DJ123) of all predicted proteins. Additionally, we identified SNPs and indels in the predicted EBEs from publicly available data. In total, 112,202 SNPs and 13,605 indels from the 3,000 Rice Genomes dataset, 2,280 SNPs from the high density rice array (HDRA) were incorporated into daTALbase.

Published RNA-seq and microarray experiments comparing rice plants inoculated with various *X. oryzae* strains and compared with control conditions have been integrated into daTALbase. These included nine microarray experiments and one RNA-seq experiment, and represented experimental treatments involving 14 of the *X. oryzae* strains included in the database (Fig. 1A).

Querying the web interface.

daTALbase has been made available online. The interface is organized in five main tabs representing the main types of data integrated in the database: TALE sequences, EBE predictions, orthology relations among genes, transcriptomic data, and SNP/indel data. An additional tab “My gene lists” allows the user to compare lists generated in the other tabs, mainly to contrast EBE predictions with transcriptomic data. daTALbase also provides links to external sources for further exploration of data, including Talvez (Pérez-Quintero et al. 2013), QueTAL (Pérez-Quintero et al. 2015), GEO datasets (Edgar et al. 2002), as well as the Rice genome browser (JBrowse) of the South Green bioinformatics platform. In each tab, the data can be filtered according to relevant fields (i.e., strain for TALE sequences) and the results can be exported as Excel files (.xlsx).

daTALbase can be used for multiple types of queries, depending on the interests of the researcher, including, for example: What are the genes predicted as targets for all TALEs from a certain strain and are these genes induced? Is a certain target conserved across different *Oryza* genomes?

The interface was organized to be as intuitive as possible, so that users can perform these types of queries. The different tabs are connected to each other and allow researchers to easily find relationships among the different types of data as depicted in Figure 3. For example, users can select TALEs of a strain of interest using the filters available in the “TAL effector” tab (Fig. 3A) and, from there, they could find the predicted targets available in the database for any desired genome (Fig. 3D) or use the external link to do their own predictions using Talvez (Fig. 3B). Users can also use the external link to the QueTAL suite to draw phylogenetic relationships among TALEs of interest (Fig. 3C).

From the “TAL targets in plants” tab (Fig. 3D), users can see results for TALEs chosen in the “TAL effector” tab or they can

search for EBEs predicted in any genes of interest. For the displayed set of predicted targets, users can then check whether there is expression data available (Fig. 3H), they can display the genomic region of the EBE in a genome browser (Fig. 3F), they

can save a list of predicted target genes to compare with selected experiments in the “RNA-seq/microarray” tab (Fig. 3I), or they can search for associated SNP data in the available datasets (Fig. 3E). Detailed information of genomic variation is

Fig. 1. A, Transcription activator-like effector (TALE) sequences and transcriptomic data included in daTALbase v1. On the left, transcriptomic experimental data included in the database associated to *Xanthomonas oryzae* strains. Each treatment represents a transcriptomic set from rice inoculated with the corresponding strain under a unique set of conditions (e.g., time postinoculation, rice variety). On the right, number of TALEs included in the database for each *X. oryzae* strain. Only strains with fully sequenced genomes are shown. Individual TALE sequences not coming from full genomes were added up as “other”. Bars are shaded according to the country of origin of each strain. Xoo = *X. oryzae* pv. *oryzae*, Xoc = *X. oryzae* pv. *oryzicola*. **B**, *Oryza* species phylogenetic tree adapted from Timetree (Hedges and Kumar 2009). Stars indicate independently assembled sequenced genomes, both draft and complete, for which predicted effector binding element data and orthology relations are available in daTALbase. The triangle indicates cultivars with available single nucleotide polymorphism data.

shown in the “SNPs/Indels” tab, these EBE variants can be of particular interest when looking for loss-of-susceptibility alleles. To assess the predicted impact of EBE variants on TALE binding, users can choose the option “Re-evaluate mutated EBEs prediction using Talvez”, which allows running TALE binding predictions on the different variants and compare their prediction scores.

In the “orthologs” tab (Fig. 3G), users can look for genes similar to any gene of interest in the available genomes and, then, look for predicted EBEs in these orthologs. Finally, the RNA-seq/microarray tab (Fig. 3H) allows users, in addition to obtaining data for previously selected genes, to explore differentially expressed genes in any of the available experiments. Users can, for example, select experiments showing genes induced in the presence of their strain of interest, save this list, and then compare it to predicted EBEs for TALEs from a said strain, using the previously described tab. For any set of genes, this tab also displays bar plots showing expression values in the relevant experiments, one gene at a time. Other possible interactions with the data are displayed in Figure 3.

Examples of usage and analysis of results from the database.

If users are interested in a specific strain of *X. oryzae*, they can use the database to identify candidate targets for all TALEs from this strain. For example, we can study the candidate targets for TALEs from the strain *X. oryzae* pv. *oryzicola* BLS256. Using daTALbase (“TAL effectors” tab, filtering according to strain), we can see that this strain has 28 TALEs, whose

predicted EBEs could be identified in the Nipponbare genome by using the link to the “TAL targets in plants” tab (2,722 genes in total, with rank less than 100). We can then explore all the experimental data available for this strain (six treatments) and identify 2,525 differentially expressed genes in the presence of this strain. Intersection between predictions and expression data represents 182 candidate target genes (Fig. 4A), which includes previously identified targets for TALEs from this strain (Cernadas et al. 2014).

This analysis can be made for each of the strains for which there is available experimental data. This reveals 747 candidate target genes for 315 TALEs from 14 strains. A hierarchical clustering based on induction of target genes reveals that strains in the database can be grouped into three main groups: i) Asian *X. oryzae* pv. *oryzae*, ii) African and Indian *X. oryzae* pv. *oryzicola*, and iii) east Asian *X. oryzae* pv. *oryzicola* (Fig. 4B), suggesting that strains from related populations have similar TALE repertoires and activate similar sets of genes, as has been previously suggested for *X. oryzae* pv. *oryzicola* (Wilkins et al. 2015) and *X. oryzae* pv. *oryzae* (Quibod et al. 2016).

Notably, some genes were identified as targets of multiple strains including both *X. oryzae* pv. *oryzicola* and *X. oryzae* pv. *oryzae*. These included “LOC_Os01g40290”, an expressed protein with unknown function, predicted as a target for TALEs from 12 of the 14 strains analyzed and differentially expressed in 41 conditions in the available transcriptome data. Other common targets included *OsSULTR3;6* (LOC_Os01g52130), a *S* gene involved in sulfate transport previously reported as a

Fig. 2. Distribution of genomic features along the rice genome *O. sativa* cv. Nipponbare MSU7 (200-kb sliding windows). 1) Gene density, 2) total effector binding elements (EBEs) predicted with rank less than 500, 3) target genes for EBEs predicted with rank less than 10 for *Xanthomonas oryzae* pv. *oryzicola*, and 4) *X. oryzae* pv. *oryzae* transcription activator-like effectors, 5) single nucleotide polymorphisms and indels from The 3,000 Rice Genome Project (2014) located in predicted EBEs.

common target for *X. oryzae* pv. *oryzicola* (Cernadas et al. 2014), and *OsHEN1* (LOC_Os07g06970), a common target for both *X. oryzae* pv. *oryzicola* and *X. oryzae* pv. *oryzae*, involved in the stability of small RNAs but with a yet-unknown function in the rice-*X. oryzae* interaction (Moscou and Bogdanove 2009).

Users can also use the database to explore commonalities of target genes. For instance, it has been recently reported that TALEs can induce gene expression bidirectionally (Streubel et al. 2017; Wang et al. 2017), that is, binding to either strand in the promoter region of a gene can drive transcription of the downstream gene. With this in mind, we can look at the

Fig. 3. Navigation process and links using the daTALbase interface, shown in screenshots of different tabs or links accessible from the web interface. Thick arrows indicate links between different tabs, bidirectional arrows indicate that queries can be made in both directions between the linked tabs. **A**, Transcription activator-like (TAL) effector tab, **B**, link to Talvez prediction, **C**, link to QueTAL phylogeny, **D**, TAL targets in plants effector binding elements (EBEs) tab, **E**, single nucleotide polymorphisms (SNPs)/Indels tab, **F**, link to Jbrowse displaying EBEs and SNPs, **G**, orthologs tab, **H**, RNA-Seq/microarray tab, and **I**, My gene lists tab.

frequency at which candidate target genes contain EBEs in the forward (same orientation as the gene) or reverse strand of the promoter. This suggests that binding in the forward strand is more common (almost twice) than binding in the reverse

strand but that, nonetheless, a large number of targets might be induced through “antisense” transcription (Fig. 4B). It’s also possible that this is the result of unknown biases in the target predictions.

Fig. 4. A, Venn diagram showing the intersection between genes containing predicted effector binding elements (EBEs) for transcription activator-like effectors (TALEs) from *Xanthomonas oryzae* pv. *oryzicola* BLS256 and genes differentially expressed in transcriptomic data comparing plants inoculated with *X. oryzae* pv. *oryzicola* BLS256 against a control, as identified using daTALbase. A gene is considered as differentially expressed if it is identified as such in any of the six experimental treatments evaluated. **B**, Target genes were identified for TALEs from 14 *X. oryzae* strains, i.e., genes containing predicted EBEs for the corresponding strain and induced by said strain in transcriptomic data as shown for BLS256 in A. The heatmap shows the highest log fold change for each gene in the treatments evaluated. On the left, hierarchical clustering showing grouping of target genes (top) and hierarchical clustering showing grouping of strains used (bottom); bottom right, country of origin of each strain is shown. Three genes shown to be common targets for various strains are highlighted. **C**, For the target genes identified in B, the bar graph shows the frequency at which EBEs were identified in the forward strand (same orientation as the gene) or the reverse strand in the promoter of each gene. **D**, Screenshot of a JBrowse session showing the 5' untranslated region (UTR) of *OsHEN1* in the *O. sativa* cv. Nipponbare genome. The region targeted by several TALEs from *X. oryzae* pv. *oryzicola* or *X. oryzae* pv. *oryzae* is highlighted. Tracks on the top indicate single nucleotide polymorphisms (SNPs) and indels detected in The 3,000 Rice Genome Project dataset, and tracks at the bottom indicate predicted EBEs.

Finally, a user can also query the database to look further into genomic variation in the predicted EBEs for these genes. For example, we can look for possible orthologs of *HEN1*, thus identifying 11 orthologs in the 13 *Oryza* genomes included in the database (no orthologs were identified in *O. punctata* or *O. sativa* cv. *kassalath* under the parameters used). When looking for predicted EBEs for these orthologs, it can be seen that EBEs for TALEs of both *X. oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* are greatly conserved across the different *Oryza* species, with some variation in the *O. glaberrima* and *O. barthii* genomes (Table 1). Likewise, we can look at the available genetic variants for this region, which reveals three SNPs and three insertion or deletion events identified in the 3,000 accessions. A researcher could then perform wet-lab experiments to associate the variation found in orthologous EBEs with possible phenotypes in the presence of strains harboring *HEN1*-inducing TALEs. This could help in the search for loss-of-susceptibility alleles as a source of resistance against *Xanthomonas* spp.

DISCUSSION

Data curation and future improvement.

daTALbase is conceived to be a constantly expanding and curated database for TALE-related data. The current version only integrates data related to the rice-*X. oryzae* system because a wealth of transcriptomic and genomic resources is available for this system. We are currently in the process of integrating additional rice transcriptomic data and TALE sequences generated in our laboratory related to African strains of *X. oryzae* pv. *oryzae* that await publication (T. T. Tran, A. L. Perez-Quintero, M. Hutin, and B. Szurek *in preparation*) and adding recently released rice genomes (Li et al. 2017), and we plan to add more data as it becomes available. New data can also be integrated upon request.

Future versions of the database will incorporate data related to *Xanthomonas* pathogens of beans, cabbage, citrus, wheat, and cassava that are currently being generated in collaboration with partners from the CropTAL project and the International Center for Tropical Agriculture (CIAT) cassava website. The working version of the cassava database integrates publicly available data corresponding to seven TALEs sequences from *Xanthomonas axnopodis* pv. *manihotis* (Bart et al. 2012; Castiblanco et al.

2013), their predicted targets on the cassava genome (v 6.1) (Bredeson et al. 2016), and two sets of RNA-seq data (Cohn et al. 2016, 2014; Muñoz-Bodnar et al. 2014). We expect to expand this database to include newly sequenced TALEs upon their release. Integrating other hosts will be of special interest to study convergence and evolution of targets, considering how some targets like the *SWEET* family of genes are being found to be important for different pathosystems (Cohn et al. 2014; Cox et al. 2017; Hu et al. 2014).

We also envision improving on the methods used for curating the data, including the possibility of adding EBE predictions using other available software (Doyle et al. 2012; Grau et al. 2013; Rogers et al. 2015) and improving the existing predictions using different sets of parameters. Likewise, we plan on improving the strategy to identify orthologs to make sure it is suitable for the inclusion of phylogenetically distant genomes. Finally, we hope daTALbase will constitute both a reference and an analysis tool for the community of TALE researchers and we encourage feedback for its improvement and curation.

MATERIALS AND METHODS

Data collection, contents, and features.

TALE sequences. TALE sequences have been retrieved from the National Center for Biotechnology Information protein databases from the two *X. oryzae* pathovars *X. oryzae* pv. *oryzae* (30 strains, 270 effectors) and *X. oryzae* pv. *oryzicola* (10 strains, 258 effectors) (Fig. 1A). Of these sequences, 487 were extracted from complete genome sequences. Each TALE was assigned an identifying number for the database in the format TBv1_001 (TBv1 indicates daTALbase version 1). For each TALE, associated information was registered including: published identifiers (e.g., PthXo1, Tal2g), gene bank database identifier of the TALE nucleotide sequence or the corresponding genome sequence, RVD sequence, the *X. oryzae* strain in which it was found, and its country of origin. TALEs with identical sequences found in different strains are considered as different entries in the database. RVD sequences were extracted using in-house perl scripts. TALEs were also assigned to groups according to similarities in their repeat sequences, as determined using the program DisTAL (Pérez-Quintero et al. 2013).

Table 1. Predicted effector binding elements (EBEs) for two transcription activator-like effectors (TALEs) from *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* in the promoters of orthologs of *HEN1* as identified using daTALbase

TALE name/genome	Target gene	Talvez rank	Talvez score	EBE sequence
XocBLS256_Tal1c				
<i>Oryza barthii</i>	OBART07G04060	1	15.51	TCCCCCTCGCTTCCCTT
<i>O. glaberrima</i> CG14	ORGLA07G0034600	1	15.522	TCCCCCTCGCTTCCCTT
<i>O. glumaepatula</i>	OGLUM07G03470	1	15.471	TCCCCCTCGCTTCCCTT
<i>O. meridionalis</i>	OMERI07G02610	1	15.406	TCCCCCTCGCTTCCCTT
<i>O. nivara</i>	ONIVA07G02640	1	15.493	TCCCCCTCGCTTCCCTT
<i>O. rufipogon</i>	ORUFI07G03760	1	15.462	TCCCCCTCGCTTCCCTT
<i>O. sativa</i> DJ123	Maker-scaffold_131-snap-gene-2.32	1	15.533	TCCCCCTCGCTTCCCTT
<i>O. sativa indica</i> 93-11	BGIOSGA024768	1	15.471	TCCCCCTCGCTTCCCTT
<i>O. sativa</i> IR65	Maker-scaffold_17-snap-gene-8.38	1	17.377	TCCCCCTCGCTTCCCTT
<i>O. sativa</i> Nipponbare	LOC_Os07g06970	1	15.434	TCCCCCTCGCTTCCCTT
XooPXO99A_Tal9a				
<i>O. barthii</i>	OBART07G04060	144	11.893	TCCCTTCCCTAATCCCCAACT
<i>O. glaberrima</i> CG14	ORGLA07G0034600	123	11.899	TCCCTTCCCTAATCCCCAACT
<i>O. glumaepatula</i>	OGLUM07G03470	1	17.395	TCCCTTCCCTAATCCCCAACT
<i>O. meridionalis</i>	OMERI07G02610	1	17.364	TCCCTTCCCTAATCCCCAACT
<i>O. nivara</i>	ONIVA07G02640	1	17.406	TCCCTTCCCTAATCCCCAACT
<i>O. rufipogon</i>	ORUFI07G03760	5	14.858	TCCCTTCCCTAATCCCCAACT
<i>O. sativa</i> DJ123	Maker-scaffold_131-snap-gene-2.32	1	17.426	TCCCTTCCCTAATCCCCAACT
<i>O. sativa indica</i> 93-11	BGIOSGA024768	1	17.395	TCCCTTCCCTAATCCCCAACT
<i>O. sativa</i> IR64	Maker-scaffold_17-snap-gene-8.38	1	15.493	TCCCTTCCCTAATCCCCAACT
<i>O. sativa</i> Nipponbare	LOC_Os07g06970	1	17.377	TCCCTTCCCTAATCCCCAACT

TALE targets (EBEs) in different *Oryza* genomes. Genomes included in the database are the reference *O. sativa* cv. Nipponbare (assembly and annotation version MSU7) from the Rice Genome Annotation Project (Kawahara et al. 2013). Ten rice genomes were obtained from the Ensembl genome database release 35: *O. barthii* (ABRL000000000), *O. brachyantha* (v1.4b), *O. glaberrima* cv. CG14 (AGH1.1), *O. glumaepatula* (*O. glumipatula*) (ALNW000000000), *O. meridionalis* (ALNW000000000), *O. nivara* (AWHD000000000), *O. punctata* (AVCL000000000), *O. rufipogon* (PRJEB4137), and *O. sativa* cv. 93-11 (ASM465v1). *O. sativa* cvs. DJ123 and IR64 (versions CSHL 1.0) were obtained from the Schatz lab (Schatz et al. 2014) and the *O. sativa* cv. Kasalath genome (v. NIAS-RAP-1.0) was obtained from rap-db (Ohyanagi et al. 2006).

Predictions were made using the Talvez software (Pérez-Quintero et al. 2013). This prediction tool uses the TALE-DNA code to convert the RVD sequence in a positional weight matrix. Then, the program uses the matrix to scan all the possible EBEs in the host genome sequence and gives a rank and a score for each putative EBE. For each of the genomes used, promoter sequences (1,000 bp upstream) were extracted from all annotated genes and Talvez was used to find EBEs on both strands of the promoter to reflect their bidirectional binding, allowing 500 hits per TALE, a minimum score of 7, and using updated RVD-DNA specificities that reflect recent experimental data for TALE-binding, including predictions for all possible RVD combinations (Yang et al. 2014) and the contribution of strong versus weak RVDs (Streubel et al. 2012) as employed in the program FuncTAL (Pérez-Quintero et al. 2015).

Transcriptomic data. Published RNA-seq and microarray experiments comparing rice plants inoculated with various *X. oryzae* strains and compared with control conditions have been integrated into daTALbase. Nine microarray experiments

(GSE16793, GSE19239, GSE19844, GSE33411, GSE34192, GSE36093, GSE36272, GSE43050, GSE8216) and one RNA-seq experiment (GSE67588) were used to feed the database. Microarray data were obtained from the PlexDB database, mean MAS-normalized values were downloaded from the database and differential expression was assessed using the limma package, as described by Pérez-Quintero et al. (2013) and Smyth (2005). RNA-seq data were obtained from GEO datasets (Edgar et al. 2002) and were processed as reported in (Wilkins et al. 2015).

Annotation of probes and RNA-seq mappings were based on the reference genome sequence of Nipponbare (MSU7 annotation). For all experiments, only genes considered as significantly induced or repressed, when the *P* value was <0.05, were kept and stored in the database. In total, 104,346 entries are recorded in a dedicated table for expression information, representing 14,071 differentially expressed genes potentially involved in the molecular basis of diseases caused by *X. oryzae*.

Orthology information. To allow comparisons between the available GFF-file *Oryza* genomes, the annotated proteome for each species and cultivar was obtained from the corresponding assembly, and the reconstruction of orthology groups was based on the commonly used approach combining an “all against all” BLASTP of whole proteomes and the clustering of blast results by the OrthoMCL suite (Li et al. 2003) (default parameters). Future versions of the database will include orthology with available genomes from other genera to allow the study of TALE target convergence in a wide scale.

Genetic variants (SNPs and indels) in predicted EBEs. The 3,000 Rice Genomes Project (2014) provides a considerable genetic resource recording millions of SNPs and indels. Another important resource with genotype information for 700,000 SNPs from a diverse set of rice accessions is the HDRA (McCouch

Fig. 5. daTALbase architecture and process for constructing the daTALbase database. Each square represents a table, number of entries in version 1 of daTALbase are also shown.

et al. 2016) with more than 1,600 genotyped accessions. We used this data to search for variations within the predicted EBEs in the *O. sativa* Nipponbare reference.

Variants overlapping a predicted EBE were extracted from these resources, using PLINK 1.9 (Chang et al. 2015) and EBE coordinates. In addition, EBEs and associated polymorphisms were also integrated as specific new tracks into the Rice genome browser (JBrowse) of the South Green bioinformatics platform.

System architecture and implementation.

The daTALbase database system combines both a relational MySQL database and JSON flat files. The web interface is implemented in Perl CGI scripts running on an Apache web server. The interactivity and smooth navigation is allowed by JavaScript and various libraries, such as jQuery user interface and Highcharts application programming interface, to manage graphical layouts, JVenn (Bardou et al. 2014) to handle Venn diagram representation, or DataTables plugin for jQuery to facilitate the manipulation of output tables. To manage the access to private data, the application is also equipped with login authentication to keep private entries password protected.

The database is normalized and consists mainly of nine tables, which approximately correspond either to the information reported by the different tabs of the web interface (TALS, EBEsInPromoters, OrthologGroups, GeneExpDiffData, SnpInfo) derived from genome wide analyses or to sparser information that can be applied for filtering (Bacteria, Host, HostGeneInfo, RnaseqCondition), plus two additional association tables. Tables and processes associated with the data are summarized in Figure 5. Finally, the application also includes a series of Perl scripts facilitating the extraction, conversion, and integration of new data.

Data availability.

The instance for rice hosted at the French Research Institute for Development (IRD) may be accessed at the daTALbase website. The source code of the application, including Perl, CGI, as well as SQL scripts for populating the database, is available for download and installation at GitHub South Green. Full portable copies of the current release of the database including the data presented here is available upon request.

ACKNOWLEDGMENTS

We thank the South Green bioinformatics platform and the French Research Institute for Development (IRD) bioinformatics “i-trop” for hosting the database and providing computational resources. A. Pérez-Quintero was supported by doctoral fellowship awarded by the Erasmus Mundus Action 2 PANACEA, PRECIOUS program of the European Community. C. A. Zarate is supported by the Allocations de recherche pour une thèse au Sud (ARTS) program (IRD). This project was supported by a grant from Agence Nationale de la Recherche (ANR-14-CE19-443-0002) and from Fondation Agropolis (number 1403-073) and from the United States National Science Foundation (IOS-1444511) to A. Bogdanove and E. Doyle. We also acknowledge L.-A. Becerra and A. Gkanogiannis for their collaboration in the deployment of the cassava instance of daTALbase at the International Center for Tropical Agriculture.

LITERATURE CITED

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. 2014. jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics* 15:293.
 Bart, R., Cohn, M., Kassen, A., McCallum, E. J., Shybut, M., Petriello, A., Krasileva, K., Dahlbeck, D., Medina, C., Alicai, T., Kumar, L., Moreira, L. M., Rodrigues Neto, J., Verdier, V., Santana, M. A., Kositcharoenkul, N., Vanderschuren, H., Gruissem, W., Bernal, A., and Staskawicz, B. J. 2012. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. U.S.A.* 109:E1972-E1979.

Boch, J., and Bonas, U. 2010. *Xanthomonas* AvrBs3 family-type III effectors: Discovery and function. *Annu. Rev. Phytopathol.* 48:419-436.
 Boch, J., Bonas, U., and Lahaye, T. 2014. TAL effectors—Pathogen strategies and plant resistance engineering. *New Phytol.* 204:823-832.
 Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326:1509-1512.
 Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I. Y., Egesi, C., Nauluvula, P., Lebot, V., Ndunguru, J., Mkamilo, G., Bart, R. S., Setter, T. L., Gleadow, R. M., Kulakow, P., Ferguson, M. E., Rounsley, S., and Rokhsar, D. S. 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat. Biotechnol.* 34:562-570.
 Castiblanco, L. F., Gil, J., Rojas, A., Osorio, D., Gutiérrez, S., Muñoz-Bodnar, A., Perez-Quintero, A. L., Koebnik, R., Szurek, B., López, C., Restrepo, S., Verdier, V., and Bernal, A. J. 2013. TALE1 from *Xanthomonas axonopodis* pv. *manihotis* acts as a transcriptional activator in plant cells and is important for pathogenicity in cassava plants. *Mol. Plant Pathol.* 14:84-95.
 Cernadas, R. A., Doyle, E. L., Niño-Liu, D. O., Wilkins, K. E., Bancroft, T., Wang, L., Schmidt, C. L., Caldo, R., Yang, B., White, F. F., Nettleton, D., Wise, R. P., and Bogdanove, A. J. 2014. Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* 10:e1003972.
 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
 Chen, J., Huang, Q., Gao, D., Wang, J., Lang, Y., Liu, T., Li, B., Bai, Z., Luis Goicoechea, J., Liang, C., Chen, C., Zhang, W., Sun, S., Liao, Y., Zhang, X., Yang, L., Song, C., Wang, M., Shi, J., Liu, G., Liu, J., Zhou, H., Zhou, W., Yu, Q., An, N., Chen, Y., Cai, Q., Wang, B., Liu, B., Min, J., Huang, Y., Wu, H., Li, Z., Zhang, Y., Yin, Y., Song, W., Jiang, J., Jackson, S. A., Wing, R. A., Wang, J., and Chen, M. 2013. Whole-genome sequencing of *O. brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* 4:1595.
 Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X.-Q., Guo, W.-J., Kim, J.-G., Underwood, W., Chaudhuri, B., Chermak, D., Antony, G., White, F. F., Somerville, S. C., Mudgett, M. B., and Frommer, W. B. 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* 468:527-532.
 Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B., Xu, C., Li, X., Fu, B., Li, Z., Benetzen, J. L., Zhang, Q., and Wang, S. 2006. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 20:1250-1255.
 Cohn, M., Bart, R. S., Shybut, M., Dahlbeck, D., Gomez, M., Morbitzer, R., Hou, B.-H., Frommer, W. B., Lahaye, T., and Staskawicz, B. J. 2014. *Xanthomonas axonopodis* virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant-Microbe Interact* 27:1186-1198.
 Cohn, M., Morbitzer, R., Lahaye, T., and Staskawicz, B. J. 2016. Comparison of gene activation by two TAL effectors from *Xanthomonas axonopodis* pv. *manihotis* reveals candidate host susceptibility genes in cassava. *Mol. Plant Pathol.* 17:875-889.
 Cox, K. L., Meng, F., Wilkins, K. E., Li, F., Wang, P., Booher, N. J., Carpenter, S. C. D., Chen, L.-Q., Zheng, H., Gao, X., Zheng, Y., Fei, Z., Yu, J. Z., Isakeit, T., Wheeler, T., Frommer, W. B., He, P., Bogdanove, A. J., Shan, L., 2017. TAL effector driven induction of a SWEET gene confers susceptibility to bacterial blight of cotton. *Nat. Commun.* 8:15588.
 Dash, S., Van Hemert, J., Hong, L., Wise, R. P., and Dickerson, J. A. 2012. PLEXdb: Gene expression resources for plants and plant pathogens. *Nucleic Acids Res.* 40:D1194-D1201.
 Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., and Yan, N. 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335:720-723.
 Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Vandyk, J. K., and Bogdanove, A. J. 2012. TAL effector-nucleotide targeter (TALE-NT) 2.0: Tools for TAL effector design and target prediction. *Nucleic Acids Res.* 40:W117-W122.
 Edgar, R., Domrachev, M., and Lash, A. E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30:207-210.
 Grau, J., Reschke, M., Erkes, A., Streubel, J., Morgan, R. D., Wilson, G. G., Koebnik, R., and Boch, J. 2016. AnnoTALE: Bioinformatics tools for identification, annotation, and nomenclature of TALEs from *Xanthomonas* genomic sequences. *Sci. Rep.* 6:21077.
 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.

- Hayward, A. C. 1993. The hosts of *Xanthomonas*. Pages 1-119 in: *Xanthomonas*. J. G. Swings, and E. L. Civerolo, eds. Springer, Dordrecht, The Netherlands.
- Hedges, S. B., and Kumar, S. 2009. The Timetree of Life. Oxford University Press, Oxford.
- Hu, Y., Zhang, J., Jia, H., Sosso, D., Li, T., Frommer, W. B., Yang, B., White, F. F., Wang, N., and Jones, J. B. 2014. *Lateral organ boundaries 1* is a disease susceptibility gene for citrus bacterial canker disease. *Proc. Natl. Acad. Sci. U.S.A.* 111:E521-E529.
- Hutin, M., Pérez-Quintero, A. L., Lopez, C., and Szurek, B. 2015a. MorTAL Kombar: The story of defense against TAL effectors through loss-of-susceptibility. *Front. Plant Sci.* 6:535.
- Hutin, M., Sabot, F., Ghesquière, A., Koebnik, R., and Szurek, B. 2015b. A knowledge-based molecular screen uncovers a broad-spectrum *OsSWEET14* resistance allele to bacterial blight from wild rice. *Plant J.* 84:694-703.
- Ji, Z., Ji, C., Liu, B., Zou, L., Chen, G., and Yang, B. 2016. Interfering TAL effectors of *Xanthomonas oryzae* neutralize R-gene-mediated plant disease resistance. *Nat. Commun.* 7:13435.
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., Itoh, T., Buell, C. R., and Matsumoto, T. 2013. Improvement of the *O. sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4.
- Li, L., Stoeckert, C. J., Jr., and Roos, D. S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13:2178-2189.
- Li, L.-F., Li, Y.-L., Jia, Y., Caicedo, A. L., and Olsen, K. M. 2017. Signatures of adaptation in the weedy rice genome. *Nat. Genet.* 49:811-814.
- Liu, Q., Yuan, M., Zhou, Y., Li, X., Xiao, J., and Wang, S. 2011. A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. *Plant Cell Environ.* 34:1958-1969.
- Mak, A. N.-S., Bradley, P., Bogdanove, A. J., and Stoddard, B. L. 2013. TAL effectors: Function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* 23:93-99.
- Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335:716-719.
- McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., Singh, N., DeClerck, G., Agosto-Perez, F., Korniliev, P., Greenberg, A. J., Naredo, M. E. B., Mercado, S. M. Q., Harrington, S. E., Shi, Y., Branchini, D. A., Kuser-Falcão, P. R., Leung, H., Ebana, K., Yano, M., Eizenga, G., McClung, A., Mezey, J., 2016. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7:10532.
- Moscou, M. J., and Bogdanove, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* 326:1501-1501.
- Muñoz-Bodnar, A., Perez-Quintero, A. L., Gomez-Cano, F., Gil, J., Michelmore, R., Bernal, A., Szurek, B., and Lopez, C. 2014. RNAseq analysis of cassava reveals similar plant responses upon infection with pathogenic and non-pathogenic strains of *Xanthomonas axonopodis* pv. *manihotis*. *Plant Cell Rep.* 33:1901-1912.
- Niño-Liu, D. O., Ronald, P. C., and Bogdanove, A. J. 2006. *Xanthomonas oryzae* pathovars: Model pathogens of a model crop. *Mol. Plant Pathol.* 7:303-324.
- Noël, L. D., Denancé, N., and Szurek, B. 2013. Predicting promoters targeted by TAL effectors in plant genomes: From dream to reality. *Front. Plant Sci.* 4:333.
- Ohyanagi, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T., Fujii, Y., Antonio, B. A., Nagamura, Y., Imanishi, T., Ikeo, K., Itoh, T., Gojobori, T., and Sasaki, T. 2006. The Rice Annotation Project Database (RAP-DB): Hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.* 34:D741-D744.
- Pérez-Quintero, A. L., Lamy, L., Gordon, J. L., Escalon, A., Cunnac, S., Szurek, B., and Gagnevin, L. 2015. QueTAL: A suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Front. Plant Sci.* 6:545.
- Pérez-Quintero, A. L., Rodriguez-R, L. M., Dereeper, A., López, C., Koebnik, R., Szurek, B., and 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.
- Quibod, I. L., Perez-Quintero, A., Booher, N. J., Dossa, G. S., Grande, G., Szurek, B., Vera Cruz, C., Bogdanove, A. J., and Oliva, R. 2016. Effector diversification contributes to *Xanthomonas oryzae* pv. *oryzae* phenotypic adaptation in a semi-isolated environment. *Sci. Rep.* 6:34137.
- Read, A. C., Rinaldi, F. C., Hutin, M., He, Y.-Q., Triplett, L. R., and Bogdanove, A. J. 2016. Suppression of *Xo1*-mediated disease resistance in rice by a truncated, non-DNA-binding TAL effector of *Xanthomonas oryzae*. *Front. Plant Sci.* 7:1516.
- Rogers, J. M., Barrera, L. A., Reyon, D., Sander, J. D., Kellis, M., Joung, J. K., and Bulky, M. L. 2015. Context influences on TALE-DNA binding revealed by quantitative profiling. *Nat. Commun.* 6:7440.
- Schatz, M. C., Maron, L. G., Stein, J. C., Hernandez Wences, A., Gurtowski, J., Biggers, E., Lee, H., Kramer, M., Antoniou, E., Ghiban, E., Wright, M. H., Chia, J., Ware, D., McCouch, S. R., and McCombie, W. R., 2014. Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica*. *Genome Biol.* 15:506.
- Smyth, G. K. 2005. LIMMA: Linear models for microarray data. Pages 397-420 in: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Statistics for Biology and Health. R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, S. Dudoit, eds. Springer, New York.
- Strauss, T., van Poecke, R. M. P., Strauss, A., Römer, P., Minsavage, G. V., Singh, S., Wolf, C., Strauss, A., Kim, S., Lee, H.-A., Yeom, S.-I., Parniske, M., Stall, R. E., Jones, J. B., Choi, D., Prins, M., and Lahaye, T. 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.
- Streubel, J., Baum, H., Grau, J., Stuttmann, J., and 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., Blücher, C., Landgraf, A., and Boch, J. 2012. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 30:593-595.
- 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 *OSTFIIAyl* and *OSTFXI* during bacterial blight of rice. *Proc. Natl. Acad. Sci. U.S.A.* 104:10720-10725.
- The 3,000 Rice Genomes Project. 2014. The 3,000 rice genomes project. *GigaScience* 3:7.
- Triplett, L. R., Cohen, S. P., Heffelfinger, C., Schmidt, C. L., Huerta, A. I., Tekete, C., Verdier, V., Bogdanove, A. J., and Leach, J. E. 2016. A resistance locus in the American heirloom rice variety Carolina Gold Select is triggered by TAL effectors with diverse predicted targets and is effective against African strains of *Xanthomonas oryzae* pv. *oryzicola*. *Plant J.* 87:472-483.
- Wang, L., Rinaldi, F. C., Singh, P., Doyle, E. L., Dubrow, Z. E., Tran, T. T., Pérez-Quintero, A. L., Szurek, B., and Bogdanove, A. J. 2017. TAL effectors drive transcription bidirectionally in plants. *Mol. Plant* 10:285-296.
- Wang, M., Yu, Y., Haberer, G., Marri, P. R., Fan, C., Goicoechea, J. L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J. S. S., Cossu, R. M., Maldonado, C., Chen, J., Lee, S., Sisneros, N., de Baynast, K., Golser, W., Wissotski, M., Kim, W., Sanchez, P., Ndjioudjop, M.-N., Sanni, K., Long, M., Carney, J., Panaud, O., Wicker, T., Machado, C. A., Chen, M., Mayer, K. F. X., Rounsley, S., and Wing, R. A. 2014. The genome sequence of African rice (*O. glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46:982-988.
- White, F. F., and Yang, B. 2009. Host and pathogen factors controlling the rice-*Xanthomonas oryzae* interaction. *Plant Physiol.* 150:1677-1686.
- Wilkins, K. E., Booher, N. J., Wang, L., and Bogdanove, A. J. 2015. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. *Front. Plant Sci.* 6:536.
- Yang, J., Zhang, Y., Yuan, P., Zhou, Y., Cai, C., Ren, Q., Wen, D., Chu, C., Qi, H., and Wei, W. 2014. Complete decoding of TAL effectors for DNA recognition. *Cell Res.* 24:628-631.
- Zhang, J., Yin, Z., and White, F. 2015. TAL effectors and the executor R genes. *Front. Plant Sci.* 6:641.

AUTHOR-RECOMMENDED INTERNET RESOURCES

ComputerScience and Quantitative Biology Schatz Lab Cold Spring Harbor Laboratory and Johns Hopkins University:
<http://schatzlab.cshl.edu/data/rice>
 CropTAL project: <https://umr-pvbm.cirad.fr/principaux-projets/croptal>
 daTALbase homepage: <http://bioinfo-web.mpl.ird.fr/cgi-bin2/datalbase/home.cgi>
 daTALbase for Rice: <http://bioinfo-web.mpl.ird.fr/cgi-bin2/datalbase/index.cgi>
 Ensembl genome database: <http://www.ensembl.org>
 GitHub South Green bioinformatics platform:
<https://github.com/SouthGreenPlatform/daTALbase>
 Highsoft AS Highcharts application programming interface:
<http://api.highcharts.com/highcharts>
 The International Center for Tropical Agriculture (CIAT)cassava website:
<http://ciat.cgiar.org/what-we-do/breeding-better-crops/rooting-for-cassava>
 jBrowse: http://jbrowse.southgreen.fr/?data=oryza_sativa_japonica_v7
 jQuery user interface: <https://jqueryui.com>
 DataTables plugin for jQuery: <https://datatables.net>
 PlexDB (the plant expression database): <http://www.plexdb.org>
 The Rice Annotation Project database rap-db:
<http://rapdb.dna.affrc.go.jp/download/irgsp1.html>
 South Green bioinformatics platform: <http://www.southgreen.fr>