A clone resource of *Magnaporthe oryzae* effectors that share sequence and structural similarities across host-specific lineages

Yohann Petit-Houdenot[1-2], Thorsten Langner[1], Adeline Harant[1], Joe Win[1*], Sophien Kamoun[1*]

[1]*The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, NR4 7UK, United Kingdom.*
[2]*UMR BIOGER, INRA, AgroParisTech, Université Paris-Saclay, Avenue Lucien Brétignières, BP 01, F-78850 Thiverval-Grignon, France. * for correspondence: joe.win@tsl.ac.uk; sophien.kamoun@tsl.ac.uk*

We describe a clone resource of 195 effectors of the blast fungus *Magnaporthe oryzae*. These clones are freely available as Golden Gate compatible entry plasmids. Our aim is to provide the community with an open source effector clone library to be used in a variety of functional studies.

Introduction

Plant pathogens secrete effectors that play central roles in subjugating plants for colonization. Effectors typically have signal peptides, and occasionally carry conserved folds and motifs (Lo Presti *et al.*, 2015; Franceschetti *et al.*, 2017). *Magnaporthe oryzae* (Syn. *Pyricularia oryzae*) is an important plant pathogen that is able to infect around 50 species of both wild and cultivated grasses including important cereals of the Poaceae family. *M. oryzae* is mostly known to cause rice blast but can also cause disease on other crops such as barley, wheat, foxtail millet, and finger millet. The global population of *Magnaporthe* is composed of genetically differentiated lineages which, in some cases, still exhibit a measurable degree of gene flow (Gladieux *et al.*, 2018). Fungal isolates from each of those lineages show a preference for a specific host and also encode distinct repertoires of effector genes (Yoshida *et al.*, 2016).

The first genomic sequence of *Magnaporthe oryzae* was released in 2005 for the lab strain 70-15 and allowed to predict a large set of secreted proteins such as enzymes involved in secondary metabolism and virulence-associated factors including putative effectors (Dean *et al.*, 2005). Recently an increasing number of genome sequences of isolates from different lineages have become available, allowing the research community to perform comparative genomic studies (Chiapello *et al.*, 2015; Yoshida *et al.*, 2016).

Many of the validated effectors of *M. oryzae* are known as the MAX (***M****agnaporthe* **A**VRs and To**x**B like) effectors. These effectors, while showing little primary sequence similarity, share a conserved structural fold composed of 6 β-sheets alternating in an anti-parallel manner (de Guillen *et al.*, 2015). The MAX family has been largely expanded in *Magnaporthe* as those effectors account for 5-10% of the effector repertoire and for 50% of the already cloned effectors of *Magnaporthe* (de Guillen *et al.*, 2015). Indeed, the identification of structural motifs enables more sensitive predictions of effectors from pathogen genomes compared to sequence similarity searches (Franceschetti *et al.*, 2017).

The aim of this project was to computationally identify a set of *M. oryzae* effectors from the main host-specific lineages and develop an open access clone resource for functional analyses.

Results and Discussion

Bioinformatics. Here, we describe a bioresource of 195 predicted *M. oryzae* effectors that we assembled from available genome sequences. To develop this resource, we initially analysed secretomes of *M. oryzae* isolates infecting different hosts such as rice, wheat, finger millet, foxtail millet, oat and some species of the *Digitaria* genus using the following computational pipeline. First, we identified proteins showing sequence similarity to known *M. oryzae* effectors with an avirulence activity (i.e. detected by plant immune receptors), using BLAST and a library of 21 previously characterized effectors. Second, we performed a search for remote relationships using a Hidden Markov Model (HMM) trained with multiple, structure-based alignments of *M. oryzae* MAX effectors (de Guillen *et al.,* 2015) using the HMMER program (Finn *et al.*, 2011). Next, we grouped candidates obtained by both approaches to form a non-redundant list of putative effectors and functionally annotated them using two methods. The first is a classical Gene Ontology (GO) search using the Blast2GO program (Conesa *et al.,* 2005). The second is a BLASTP similarity search against a custom-made database (the Darwin database) which contains more than 2,600,000 predicted proteins of 137 different eukaryotic organisms. We removed candidates with functional domain annotation as most *M. oryzae* effectors are proteins with no predicted function. Finally, we also removed candidates with poor gene model predictions and very low sequence similarity to *M. oryzae* effectors. The combination of these approaches resulted in 194 effector candidates that we selected for gene synthesis as a clone resource. We also added a homolog of the well-studied *M. oryzae* effector AVR-Pik of the *Lolium perenne* isolate PGKY to this library.

Golden Gate system. The Golden Gate cloning system enables rapid and high throughput assembly of multiple sequence modules, such as promotor, terminator or tags into a common vector (Patron *et al.,* 2015). This cloning strategy is ideal for wide dissemination of cloning material by using a universal code for cloning. We synthesized the 195 *M. oryzae* effectors in a Golden Gate compatible fashion to enable the transfer of these genes into a variety of vectors for various applications such as yeast two-hybrid assays, heterologous protein expression, and fungal transformation (see for example the fungal transformation vectors described by Pennington *et al*., 2017, 2018).

Effector clone resource. We synthesized the coding sequences corresponding to the mature proteins (without the signal peptide) of 195 candidates in the Golden Gate compatible vector pUC57-Kan. Each coding sequence was flanked by *BsaI* restriction enzyme sites and relevant overhang sequences to ensure optimal Golden Gate reactions. The coding sequences were manually codon-optimised for expression in *Saccharomyces cerevisiae*, *Nicotiana benthamiana* and *Escherichia coli* (when it was not possible to optimise for all three organisms, priority was given to *S. cerevisiae* and *N. benthamiana*). Table S1 describes the 195 candidates and includes information such as sequence similarity to *M. oryzae* effectors, results of the HMM search, SignalP scores, functional annotation and host specificity of the strain in which the effector candidate was identified. Table 1 summarizes the different candidates identified, their host specificity and the rationale behind their selection whether it was based on sequence or structural similarities. Fig 1 shows an overview of the Golden Gate system and a generic schematic representation of the pUC57-Kan vector insertion site. Additional supplementary files include the DNA and amino acid sequences of the selected genes.

Clone distribution. Level 0 modules are available via Addgene (addgene.com). Catalog numbers are provided in Table S1.

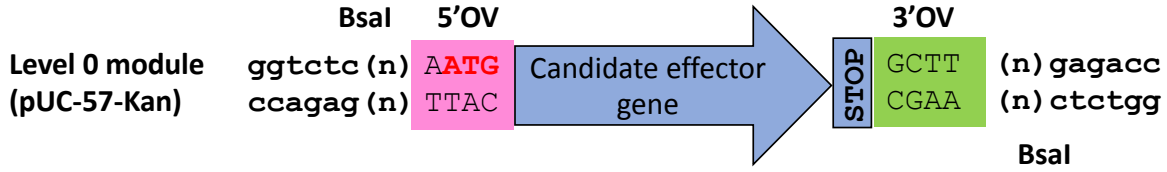

Figure 1. Synthetic Level 0 modules of candidate effector genes. Candidate effectors were synthesized in the pUC57-Kan vector. The sequence encoding the signal peptide and the start codon were removed (the ATG is part of the Golden Gate 5' overhang sequence). Effector sequences contain a stop codon at the 3' end. 5'OV: 5' overhang. 3'OV: 3' overhang.

Table 1. Features of the 195 *Magnaporthe oryzae* effectors. The number of candidates selected on the base of their sequence identity to a known effector of *M. oryzae*; and to similarity with MAX effectors according to the HMM search.

Host of *M. oryzae* strain	N with similarity to AVR effectors	N with similarity to MAX effectors	Total
Rice	23	27	50
Wheat	6	5	11
Setaria	3	3	6
Eleusine	6	2	8
Avena	3	0	3
Digitaria	20	19	39
Lolium	1	0	1
Multiple hosts	38	39	77
	100	95	195

Acknowledgements

This research was funded by The Gatsby Charitable Foundation and European Research Council BLASTOFF project. Y.P.-H. was supported by a Young Scientist Contract from INRA.

References

Chiapello H, Mallet L, Guérin C, Aguileta G, Amselem J, Kroj T, Ortega-Abboud E, Lebrun M.-H, Henrissat B, Gendrault A, *et al.* 2015. Deciphering Genome Content and Evolutionary Relationships of Isolates from the Fungus *Magnaporthe oryzae* Attacking Different Host Plants. *Genome Biol. Evol.* 7, 2896–2912.

Conesa A, Götz S, García-Gómez J.M, Terol J, Talón M, and Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676.

Dean R.A, Talbot N.J, Ebbole D.J, Farman M.L, Mitchell T.K, Orbach M.J, Thon M, Kulkarni R, Xu,J.-R, Pan H, *et al*. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980–986.

Finn R.D, Clements J, and Eddy S.R. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39, W29–W37.

Franceschetti, M., Maqbool, A., Jimenez-Dalmaroni, M.J., Pennington, H.G., Kamoun, S., and Banfield, M.J. 2017. Effectors of filamentous plant pathogens: commonalities amid diversity. *Microbiology Molecular Biology Reviews*, 81, e00066-16.

Gladieux P, Condon B, Ravel S, Soanes D, Maciel J.L.N, Nhani A, Chen L, Terauchi R, Lebrun M.-H, Tharreau D, *et al*. 2018. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus *Magnaporthe oryzae. MBio* 9, e01219-17.

De Guillen K, Ortiz-Vallejo D, Gracy J, Fournier E, Kroj T, and Padilla A. 2015. Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLOS Pathog*. 11, e1005228.

Krogh A, Larsson B, von Heijne G, and Sonnhammer E.L.L. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.

Patron N.J, Orzaez D, Marillonnet S, Warzecha H, Matthewman C, Youles M, Raitskin O, Leveau A, Farré G, Rogers C, *et al*. 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytol. 208, 13–19*.

Pennington, H.G., Youles, M., and Kamoun, S. 2017. Golden-Gate compatible *Magnaporthe oryzae* protoplast transformation vectors. Figshare. https://doi.org/10.6084/m9.figshare.5375428.v1

Pennington, H.G., Youles, M., and Kamoun, S. 2018. Golden-Gate compatible *Magnaporthe oryzae* Agrobacterium transformation vectors. Figshare. https://doi.org/10.6084/m9.figshare.5808348.v1

Petersen T.N, Brunak S, von Heijne G, and Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786.

Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, and Kahmann R. 2015. Fungal Effectors and Plant Susceptibility. *Annu. Rev. Plant Biol*. 66, 513–545.

Yoshida K, Saunders D.G.O, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano L.M, *et al*. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics* 17, 370.

Supplementary material

Table S1. Characteristics of the selected *Magnaporthe oryzae* candidate effectors. Addgene catalog number: number under which each effector candidate can be found in the Addgene catalog. match_blast_AVR: *M. oryzae* effector against which the best blast hit is found. Evalue_hmm: evalue found with the HMM search. score_hmm: score found with the HMM search. Score_signalP: score from signalP search (Petersen *et al.*, 2011). signalP_HMM_score: combined score from signalP and tmhmm searches (Krogh *et al.*, 2001). hit_blast_Darwin_without_magnaporthe: best blast hit against the Darwin database after removing the proteins from *Magnaporthe spp*. GO: Gene Ontology attributed by the Blast2GO program. Host: host of the isolate from which the gene sequence originates. Specificity: hosts of all the isolates that express this protein. Present_in: other isolates in which this protein can be found. Validated_effector: YES: this protein is an already-known *M. oryzae* effector.

Supplementary file 1. Protein sequences of the selected *Magnaporthe oryzae* candidate effectors in FASTA format.

Supplementary file 2. Codon-optimised nucleotide sequences of the *Magnaporthe oryzae* candidate effectors (with overhang sequences) in FASTA format.

Supplementary folder 1. Nucleotide sequences of the *Magnaporthe oryzae* candidate effectors in the pUC57-Kan vector (Genbank format).

Supplementary folder 2. Plasmid maps.