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*

Abstract

The blast fungus *Magnaporthe oryzae* (Syn. *Pyricularia oryzae*) is a destructive plant pathogen that can infect about 50 species of both wild and cultivated grasses, including important crops such as rice and wheat. *M. oryzae* is composed of genetically differentiated lineages that tend to infect specific host genera. To date, most studies of *M. oryzae* effectors have focused on the rice-infecting lineage. We describe a clone resource of 195 effectors of *M. spp* predicted from all the major host-specific lineages. 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. We hope that this resource will encourage studies of *M. oryzae* effectors on diverse host species.

Plant pathogens secrete effectors that play central roles in subjugating plants for colonization. Fungal 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 *M. oryzae* is composed of genetically differentiated lineages which, in some cases, still exhibit a measurable degree of

gene flow (Gladieux *et al.,* 2018; Langner 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 *M. 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). One of these lineages of *M. oryzae* causes the wheat blast disease and has recently emerged as a serious threat to food security in South East Asia (Islam *et al*., 2016; Islam *et al.*, 2019). As part of the response to the wheat blast epidemic, we have committed to generating and sharing open source data and resources (Kamoun *et al.*, 2019). This project is part of this wider OpenWheatBlast initiative.

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 expanded in *M. oryzae* as MAX effectors account for 5-10% of the repertoire of secreted proteins and for ~50% of the already cloned effectors of *M. oryzae* (de Guillen *et al.,* 2015). Indeed, the identification of structural motifs has enabled 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. To develop this resource, we initially analysed secretomes of 13 *M. oryzae* isolates infecting different hosts such as rice, wheat, finger millet, foxtail millet and oat, and of 2 isolates of *M. grisea* infecting species of the *Digitaria* genus. The secretomes used in this study have been obtained from the genome resources associated with the studies of Chiapello et al., (2015) and Yoshida et al., (2016). We searched for effector candidates in those secretomes 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. Those 21 effectors are AVR-PikD, AVR-Pia and AVR-Pii (Yoshida et al., 2009), AVR1-CO39 (Cesari et al., 2013), AVR-Piz-t (Li et al., 2009), AVR-Pita1 (Orbach et al., 2000), AVR-Pita3 (Khang et al., 2008), AVR-Pib (Zhang et al., 2015), MGG_17227-t26_1 (Zhong et al., 2018),

AVR-Pi9 (Wu et al., 2015), AVR-Rmg8 (Anh et al., 2018), PWT3 and PWT4 (Inoue et al., 2017), Pwl2 (Sweigard et al., 1995), Pwl1, Pwl3 and Pwl4 (Kang et al., 1995), APikL1, APikL2, APikl3 and ms119 (unpublished). The BLAST search was performed with a blosum 62 substitution matrix, a gap opening penalty of 11 and a cut-off E-value of 10. 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) and the HMM profile generated by de Guillen *et al.,* (2015) with a cut-off E-value of 1e-4. We performed HMMsearch analyses with a cut-off E-value of 10. 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 manually removed candidates with poor gene model predictions and 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 from the *Lolium perenne* isolate PGKY to this library bringing the total number to 195 predicted *M. oryzae* effectors.

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. spp* 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).

We synthesized the coding sequences corresponding to the mature proteins (without the signal peptide) of the 195 candidates in the Golden Gate compatible vector pUC57-Kan. Each coding sequence was flanked by *BsaI* restriction enzyme sites and relevant overhang sequences for Golden Gate cloning 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 1 summarizes the different candidates identified, their host specificity and whether they were selected based on sequence or structural similarities. Supplementary File 1 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. Supplementary Figure 1 shows an overview of the Golden Gate system and a generic schematic representation of the pUC57-Kan vector insertion site. The protein sequences and codon-optimized sequences of the selected genes as well as additional information including the plasmid maps of the Level 0 constructs are provided through the data repository Zenodo as indicated below in the Recommended Resources.

Level 0 modules are available via Addgene. Catalog numbers are provided in Supplementary Table 1.

Author recommended resources

Kamoun Lab Darwin Database http://doi.org/10.5281/zenodo.3699564
Nucleotide and protein sequences of the candidate effectors in pUC57-Kan vector and plasmids maps. Zenodo http://doi.org/10.5281/zenodo.3268775
OpenWheatBlast http://wheatblast.net
Plasmids at Addgene https://www.addgene.org/Sophien_Kamoun/

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.

Literature Cited

Anh, V.L., Inoue, Y., Asuke, S., Vy, T.T.P., Anh, N.T., Wang, S., Chuma, I., and Tosa, Y. 2018. Rmg8 and Rmg7, wheat genes for resistance to the wheat blast fungus, recognize the same avirulence gene *AVR-Rmg8*. Mol. Plant Pathol. 19, 1252–1256.

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., et al. 2013. The Rice Resistance Protein Pair RGA4/RGA5 Recognizes the *Magnaporthe oryzae Effectors* AVR-Pia and AVR1-CO39 by Direct Binding. Plant Cell Online tpc.112.107201.

Chiapello, H., Mallet, L., Guérin, C., Aguileta, G., Amselem, J., Kroj, T., Ortega-Abboud, E., Lebrun, M.-H., Henrissat, B., Gendrault, A., Rodolphe, F., Tharreau, D., Fournier, E., 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., 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., Read, N.D., Lee, Y.-H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E., Birren, B.W., 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434, 980–986.

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

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M.J., Pennington, H.G., Kamoun, S., Banfield, M.J., 2017. Effectors of Filamentous Plant Pathogens: Commonalities amid Diversity. Microbiol. Mol. Biol. Rev. 81.

Gladieux, P., Ravel, S., Rieux, A., Cros-Arteil, S., Adreit, H., Milazzo, J., Thierry, M., Fournier, E., Terauchi, R., Tharreau, D., 2018. Coexistence of Multiple Endemic and Pandemic Lineages of the Rice Blast Pathogen. mBio 9, e01806-17.

Guillen, K. de, Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., 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., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Inoue, Y., Vy, T.T.P., Yoshida, K., Asano, H., Mitsuoka, C., Asuke, S., Anh, V.L., Cumagun, C.J.R., Chuma, I., Terauchi, R., et al. 2017. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. Science 357, 80–83.

Islam, M.T., Croll, D., Gladieux, P., Soanes, D.M., Persoons, A., Bhattacharjee, P., Hossain, S., Gupta, D., Rahman, Md.M., Mahboob, M.G., Cook, N., Salam, M., Surovy, M.Z., Bueno Sancho, V., Maciel, J.N., Nani, A., Castroagudin, V., de Assis Reges, J.T., Ceresini, P., Ravel, S., Kellner, R., Fournier, E., Tharreau, D., Lebrun, M.-H., McDonald, B., Stitt, T., Swan, D., Talbot, N., Saunders, D., Win, J., Kamoun, S. 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. BMC Biol. 14,84.

Islam, M.T., Kim, K.H., Choi, J. 2019. Wheat blast in Bangladesh: the current situation and future impacts. Plant Pathol J. 35,1–10.

Kamoun, S., Talbot, N.J., Islam, M.T. 2019. Plant health emergencies demand open science: Tackling a cereal killer on the run. PLOS Biology, 17,e3000302.

Kang, S., Sweigard, J.A., and Valent, B. 1995. The *PWL* host specificity gene family in the blast fungus *Magnaporthe grisea*. Mol. Plant-Microbe Interact. MPMI 8, 939–948.

Khang, C.H., Park, S.-Y., Lee, Y.-H., Valent, B., and Kang, S. 2008. Genome Organization and Evolution of the *AVR-Pita* Avirulence Gene Family in the *Magnaporthe grisea* Species Complex. Mol. Plant. Microbe Interact. 21, 658–670.

Langner, T., Białas, A., Kamoun, S., 2018. The Blast Fungus Decoded: Genomes in Flux. mBio 9, e00571-18.

Li, W., Wang, B., Wu, J., Lu, G., Hu, Y., Zhang, X., Zhang, Z., Zhao, Q., Feng, Q., Zhang, H., et al. 2009. The *Magnaporthe oryzae* Avirulence Gene AvrPiz-t Encodes a Predicted Secreted Protein That Triggers the Immunity in Rice Mediated by the Blast Resistance Gene Piz-t. Mol. Plant. Microbe Interact. 22, 411–420.

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

Orbach, M.J., Farrall, L., Sweigard, J.A., Chumley, F.G., and Valent, B. 2000. A Telomeric Avirulence Gene Determines Efficacy for the Rice Blast Resistance Gene *Pi-ta*. Plant Cell 12, 2019–2032.

Patron, N.J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., Smith, A., Hibberd, J., Webb, A.A.R., Locke, J., Schornack, S., Ajioka, J., Baulcombe, D.C., Zipfel, C., Kamoun, S., Jones, J.D.G., Kuhn, H., Robatzek, S., Esse, H.P.V., Sanders, D., Oldroyd, G., Martin, C., Field, R., O'Connor, S., Fox, S., Wulff, B., Miller, B., Breakspear, A., Radhakrishnan, G., Delaux, P.-M., Loqué, D., Granell, A., Tissier, A., Shih, P., Brutnell, T.P., Quick, W.P., Rischer, H., Fraser, P.D., Aharoni, A., Raines, C., South, P.F., Ané, J.-M., Hamberger, B.R., Langdale, J., Stougaard, J., Bouwmeester, H., Udvardi, M., Murray, J.A.H., Ntoukakis, V., Schäfer, P., Denby, K., Edwards, K.J., Osbourn, A., Haseloff, J., 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., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Sweigard, J.A., Carroll, A.M., Kang, S., Farrall, L., Chumley, F.G., and Valent, B. 1995. Identification, cloning, and characterization of *PWL2*, a gene for host species specificity in the rice blast fungus. Plant Cell 7, 1221–1233.

Wu J, Kou Y, Bao J, Li Y, Tang M, Zhu X, Ponaya A, Xiao G, Li J, Li C, *et al.* 2015. Comparative genomics identifies the *Magnaporthe oryzae* avirulence effector AvrPi9 that triggers Pi9-mediated blast resistance in rice. New Phytol. 206, 1463–1475.

Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., et al. 2009. Association Genetics Reveals Three Novel Avirulence Genes from the Rice Blast Fungal Pathogen *Magnaporthe oryzae*. Plant Cell 21, 1573–1591.

Yoshida, K., Saunders, D.G.O., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., Cano, L.M., Kamoun, S., Terauchi, R., 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.

Zhang, S., Wang, L., Wu, W., He, L., Yang, X., and Pan, Q. 2015. Function and evolution of *Magnaporthe oryzae* avirulence gene *AvrPib* responding to the rice blast resistance gene Pib. Sci. Rep. 5, 11642.

Zhong, Z., Chen, M., Lin, L., Han, Y., Bao, J., Tang, W., Lin, L., Lin, Y., Somai, R., Lu, L., et al. 2018. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. ISME J. 1.

Table 1. Features of the 195 *Magnaporthe oryzae* effectors. Number of candidate effectors selected based on their sequence identity to a known effector of *M. oryzae* or structural similarity with MAX effectors according to the HMM search. Multiple hosts: isolates that infect more than one host among the 7 hosts listed in this table.

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

Supplemental Material

Supplementary Figure 1. Synthetic Level 0 modules of candidate effector genes.

Supplementary file 1. 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 with highest sequence similarity to candidate. 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 encode this protein. Present_in: other isolates in which this effector gene can be found. Validated_effector: YES: this protein is an already-known *M. oryzae* effector. Proteins for which the ID is underlined have been identified as MAX effectors by De Guillen *et al,* (2015).