Genome assemblies of three strains of Pyricularia oryzae, representative of the variability of the population on Italian rice

Genome assemblies of three strains of *Pyricularia oryzae*, representative of the variability of the population on Italian rice

Marco Garello[1], Simone Bosco[1], Pamela Abbruscato[3], Davide Spadaro[1,2]

[1]Department of Agricultural, Forestry and Food Sciences (DISAFA), University of Turin, 10095 Grugliasco (TO), Italy.
[2]Centre of Competence for the Innovation in the Agroenvironmental Sector (AGROINNOVA), University of Torino, 10095 Grugliasco (TO), Italy.
[3]Parco Tecnologico Padano, 26900 Lodi (LO), Italy. Current address: Nuova Genetica Italiana, 22079 Villa Guardia (CO), Italy.

ABSTRACT

Despite recent population studies on Italian populations of *Pyricularia oryzae*, little information is available regarding their genomic structure and effector gene content. Here, we compared three representative Italian *P. oryzae* strains (ITC28, ITC35 and ITC82) with reference strain 70-15. This led to the identification of 1-2 Mb of new contigs with no match to the reference strain. Gene prediction performed on the contigs returned 894 genes, of which 19 encode proteins with effector characteristics. Among these candidates, protein FUN_000690-T1 matched the sequence of effector IUG6, which is associated with high virulence of strain 98-06. These data could be useful to better understand the molecular mechanism of pathogenicity in Italian strains and to provide targets for diagnostic tool development.

INTRODUCTION

Rice is one of the most important crops worldwide, as it is the main source of calories for almost half of Earth's population (Wei and Huang, 2019). Italy is the main rice producer in Europe (Tenni et al., 2021). One of the biggest threats for Italian rice cultivation is rice blast disease, caused by *Pyricularia oryzae* Cavara (Kraehmer et al., 2017). The most efficient integrated management strategies against this pathogen are chemical fungicides and resistant varieties (Srivastava et al., 2017). The first study that characterised European *P. oryzae* populations (Roumen et al., 1997) gave crucial information for rice breeding in Europe, allowing to understand which resistance genes are more efficient in disease tolerance based on the physiologic races that can be actually found in Europe. Since then, only one other research group explored the structure of Italian *P. oryzae* populations with the aim of

defining the physiologic races composition in Italy (Abbruscato et al., unpublished). They molecularly characterised 150 Italian *P. oryzae* isolates using SSR markers that allowed to identify three lineages. From these main groups, three strains (ITC28, ITC35 and ITC82), one for each lineage, were sequenced and used for artificial infection assays to find resistant rice cultivars. In this announcement, we report the results of our analyses for three *P. oryzae* strains (ITC28, ITC35 and ITC82) isolated in northern Italy. In particular, we focus on the identification and characterization of new effectors compared to the reference strain 70-15.

RESULTS

UNMAPPED REGION ASSEMBLY

Assembly statistics for the unmapped reads of the three strains are shown in **table 1**. Gene prediction on the strain-specific contigs revealed 894 genes, of which 276 on ITC28 contigs, 226 on ITC35 contigs and 392 on ITC82 contigs.

Table 1: assembly statistics for unaligned sequences of stains ITC28, ITC35 and ITC82.

Strain	N°contigs	N°contigs (>1000 bp)	Biggest contig	Total assembly length (Mb)	% GC	N50	Ns per 100 kbp
ITC28	1,060	464	34,987	1.46	44.97	1,649	0.68
ITC35	940	348	34,987	1.14	43.99	1,308	0.87
ITC82	1,175	620	34,987	2.07	45.23	2,446	0.48

EFFECTORS

Analyses indicated the presence of 19 effector-associated genes, whose products formed 16 protein clusters (**table 2**). Comparison of these protein sequences with those of known *P. oryzae* effectors in Phybase returns a single match, with protein FUN_000690-T1 of strain ITC82 emerging as a putative homologue to effector IUG6. This effector was first found in highly virulent *P. oryzae* strain 98—06 and its role in host infection was demonstrated through knock-out and complementation experiments coupled with infection assays and transcriptomics (Dong et al., 2015). Four additional proteins (FUN_000439-T1, FUN_000687-T1, FUN_000896-T1, FUN_000926-T1) present characteristics which match those of small secreted cysteine-rich proteins (Krijger et al., 2014) and thus represent prime

candidates for further analyses. Another protein (FUN_000331-T1) is suggested to act as a transcription factor, which is compatible with pathogenic mechanisms previously elucidated in *P. oryzae* (Kim et al., 2020).

Table 2*:* list of new putative effectors found in Italian strains of P. oryzae, compared to reference strain 70-15. Each protein is assigned to a cluster, based on all-vs-all sequence comparison, and is assigned a putative site of action based on EffectorP analysis results, as well as gene onthology terms, if found. Highlighted rows indicate small secreted cysteine-rich proteins (SSCPs).

Protein (strain)	**Cluster**	**Putative action site**	**GO terms**
FUN_000185-T1 (ITC28)	8	Cytoplasm	
FUN_000203-T1 (ITC28)	2	Apoplast	
FUN_000209-T1 (ITC28)	11	Apoplast	
FUN_000255-T1 (ITC28)	9	Cytoplasm	
FUN_000272-T1 (ITC28)	6	Cytoplasm	
FUN_000331-T1 (ITC35)	1	Cytoplasm	**GO:0000786** (nucleosome), **GO:0003677** (DNA binding), **GO:0030527** (structural constituent of chromatin), **GO:0046982** (protein heterodimerization activity)
FUN_000370-T1 (ITC35)	4	apoplast/ cytoplasm	**GO:0016491** (oxidoreductase activity)
FUN_000406-T1 (ITC35)	14	Cytoplasm	
FUN_000416-T1 (ITC35)	11	Apoplast	
FUN_000439-T1 (ITC35)	10	Apoplast	
FUN_000457-T1 (ITC35)	5	apoplast/ cytoplasm	
FUN_000687-T1 (ITC82)	12	Apoplast	
FUN_000690-T1 (ITC82)	15	Cytoplasm	
FUN_000831-T1 (ITC82)	4	Apoplast	**GO:0016491** (oxidoreductase activity)
FUN_000837-T1 (ITC82)	7	Cytoplasm	
FUN_000848-T1 (ITC82)	3	Cytoplasm	**GO:0016491** (oxidoreductase activity)
FUN_000896-T1 (ITC82)	16	apoplast/ cytoplasm	
FUN_000906-T1 (ITC82)	5	apoplast/ cytoplasm	
FUN_000926-T1 (ITC82)	13	Apoplast	

CONCLUSIONS

In this announcement, we presented new putative effectors associated with Italian *P. oryzae* strains. These data can be used to guide the selection of targets for resistance breeding in Italian rice varieties, as well as to develop screening tests in order to better quantify the presence of strain of interest.

MATERIALS AND METHODS

Pyricularia strains ITC28, ITC35 and ITC82 were sequenced at 10x coverage by Parco Tecnologico Padano using a next generation Illumina MiSeq sequencer. For each strain, a paired end library was generated using the Nextera XT DNA preparation kit (Illumina, San Diego, California, United States). Libraries were purified by AMPure XP beads and normalized to ensure equal library representation in the pools. Equal volumes of libraries were diluted in the hybridization buffer, heat denatured and sequenced. Standard phi X control library (Illumina) was spiked into the denatured HCT 116 library. The libraries and phi X mixture were finally loaded into a MiSeq 250 and MiSeq 300-Cycle v2 Reagent Kit (Illumina). Base calling was performed using the Illumina pipeline software. Demultiplexing was done using an Illumina provided software. Sequencing adapter contamination and reads shorter than 30 bp were removed with Trimmomatic (Bolger et al., 2014), then remaining reads were mapped to the 70-15 strain reference genome using Bowtie2 (Langmead and Salzberg, 2012). In order to investigate the presence of new effectors, unmapped sequences for each strain were extracted from the alignment files and assembled with SPAdes (Bankevich et al., 2012). Resulting contigs were masked using RepeatMasked (Smit et al, 2015a) and RepeatModeler (Smit and Hubley, 2015), then annotated with Funannotate (Palmer and Stajich, 2020). Predicted proteins were filtered first with SignalP (Teufel et al., 2022) to remove those without secretion signal, then with DeepTMHMM (Hallgren et al., 2022) to remove those with a non-terminal transmembrane domain. Another program, called TargetP (Almagro Armenteros et al., 2019), excluded plastid- and mitochondria-targeted proteins, while program PredGPI (Pierleoni et al., 2008) removed proteins with cell wall localisation. Finally, EffectorP (Sperschneider et al., 2016) was used to predict effector function and putative action site. Presence of homologues was investigated by clustering putative effectors using Hisat2 (Kim et al., 2019), with a sequence conservation cutoff value of 60% and a length conservation cutoff value of 90%. Recovered sequences were compared to those in PHI-base (Urban et al., 2020) in order to identify already characterized effectors.

Additional annotations, and in particular GO terms, were added using Interproscan v5.56-89.0 (Jones et al., 2014).

DATA AVAILABILITY

Raw sequencing data are available for strain ITC28 (accession number: ERR10176137), ITC35 (accession number: ERR10176471) and ITC82 (accession number: ERR10176472).

ACKNOWLEDGEMENTS

This project has received funding from AGER Foundation under Grant no. 2010–2369 ("RISINNOVA - Integrated genetic and genomic approaches for new Italian rice breeding strategies").

REFERENCES

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., &Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Dong, Y., Li, Y., Zhao, M., Jing, M., Liu, X., Liu, M., Guo, X., Zhang, X., Chen, Y., Liu, Y., Liu, Y., Ye, W., Zhang, H., Wang, Y., Zheng, X., Wang, P., & Zhang, Z. (2015). Global genome and transcriptome analyses of Magnaporthe oryzae epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and lose dynamics in genome evolution. *PLoS pathogens*, *11*(4), e1004801. https://doi.org/10.1371/journal.ppat.1004801

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Almagro Armenteros, J. J., Marcantili, P., Nielsen, H., Krogh, A., Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv 2022.04.08.487609; https://doi.org/10.1101/2022.04.08.487609

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kim, D., Paggi, J. M., Park, C., Bennett, C., &Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, S., Kim, C. Y., Park, S. Y., Kim, K. T., Jeon, J., Chung, H., Choi, G., Kwon, S., Choi, J., Jeon, J., Jeon, J. S., Khang, C. H., Kang, S., & Lee, Y. H. (2020). Two nuclear effectors of the rice blast fungus modulate host immunity via transcriptional reprogramming. Nature communications, 11(1), 5845. https://doi.org/10.1038/s41467-020-19624-w

Kraehmer H., Thomas C. &Vidotto F. (2017) Rice Production in Europe. In: Chauhan B, Jabran K and Mahajan G (eds) Rice Production Worldwide. Springer, Cham. https://doi.org/10.1007/978-3-319-47516-5_1

Krijger, J. J., Thon, M. R., Deising, H. B., &Wirsel, S. G. (2014). Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation.*BMC genomics*, *15*(1), 722. https://doi.org/10.1186/1471-2164-15-722

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Tenni D., Sinetti A., Waldner M., Torriani S. F., & Romani M. (2021). First report of QoI resistance in Italian population of Pyricularia oryzae. *Journal of Plant Diseases and Protection*, *128*(6), 1705-1709. https://doi.org/10.1007/s41348-021-00494-3

Palmer, J. M. & Stajich, J. (2020). Funannotate v1.8.1: Eukaryotic genome annotation (v1.8.1). Zenodo. https://doi.org/10.5281/zenodo.4054262

Pierleoni, A., Martelli, P. L., & Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC bioinformatics*, *9*, 392. https://doi.org/10.1186/1471-2105-9-392

Roumen E., Levy M., &Nottéghem J. L. (1997). Characterisation of the European pathogen population of Magnaporthe grisea by DNA fingerprinting and pathotype analysis. *European Journal of Plant Pathology*, *103*(4), 363-371. https://doi.org/10.1023/A:1008697728788

Smit, A. F. A., Hubley, R & Green, P. *RepeatMasker Open-4.0*. 2013-2015. https://www.repeatmasker.org

Smit, A. F. A. and Hubley, R. *RepeatModeler Open-1.0*. 2008-2015. https://www.repeatmasker.org

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., & Taylor, J. M. (2016). EffectorP: predicting fungal effector proteins from secretomes using machine learning. *The New phytologist*, *210*(2), 743–761. https://doi.org/10.1111/nph.13794

Srivastava D., Shamim M. D., Kumar M., Mishra A., Pandey P., Kumar D. & Singh, K. N. (2017). Current status of conventional and molecular interventions for blast resistance in rice. *Rice Science*, *24*(6), 299-321. https://doi.org/10.1016/j.rsci.2017.08.001

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature biotechnology*, *10*. https://doi.org/10.1038/s41587-021-01156-3

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., De Silva, N., Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K., & Hammond-Kosack, K. E. (2020). PHI-base: the pathogen-host interactions database. *Nucleic acids research*, *48*(D1), D613–D620. https://doi.org/10.1093/nar/gkz904

Wei X. & Huang X. (2019) Origin, taxonomy, and phylogenetics of rice. In: Bao J (ed) Rice (Fourth Edition). AAC International Press. https://doi.org/10.1016/B978-0-12-811508-4.00001-0