Rice blast fungus sequenced

Subhankar Roy-Barman and Bharat B. Chattoo

Rice (*Oryza sativa* L.) is a major staple food crop for half of the world's population. The International Rice Research Institute, Philippines, estimates that in order to feed the growing global population, rice production must increase by another one-third by the year 2020. *Magnaporthe grisea* (anamorph, *Pyricularia oryzae*) causes rice blast disease (Figure 1) and is the most destructive pathogen of rice worldwide; around 50% of production may be lost in a field moderately affected by infection. Each year the fungus destroys rice enough to feed an estimated 60 million people. The disease is currently managed using resistant cultivars, fungicides and cultural practices. Strains of this fungus have also been reported to infect wheat, barley and turf grass. Most of the rice cultivars are susceptible to some strain of this fungus and since the pathogen is highly variable, breeding for durable resistance to blast remains a major challenge.

The rice blast fungus has emerged as a model system for the study of plant–pathogen interactions[1]. This is largely due to the efforts of a small community of researchers worldwide, who have developed the necessary tools for molecular biological investigations in this fungus. The mechanism of plant infection by *M. grisea* is currently better understood than for any other cereal disease. Infection by the rice blast fungus starts when the three-celled conidia from conidiophores land on a host leaf and anchor the leaf cuticle with the help of spore-tip mucilage. Germination proceeds with the extension of germ tube, which undergoes hooking and swelling at its tip and then differentiates into a melanized infection structure called 'appressorium'. The formation of this infection structure on the host surface marks the onset of the disease (Figure 2).

Publication of the draft sequence of this fungus[2] being the first such sequence from a fungal pathogen of an economically important crop plant is a welcome addition to the list of sequenced genomes. The genome was sequenced by International Rice Blast Genome Consortium with participating laboratories from USA, UK, France and Korea. Sadly, there was no Indian participation in this effort, even though rice blast remains a major challenge for this country. The availability of the genome sequence will provide an enormous impetus to better understanding the molecular basis of fungal disease development in plants.

The genome of *M. grisea* is around 40 Mb, distributed among seven chromosomes. The draft sequence predicts 11,109 genes, with a frequency of 1 gene every 3.5 kb in the rice blast genome. This information will facilitate a systematic characterization of pathogen-specific genes and the study of the biology of this fungus. The genome of *M. grisea* possesses more genes compared to the non-pathogenic *Neurospora crassa* and *Aspergillus niger*,
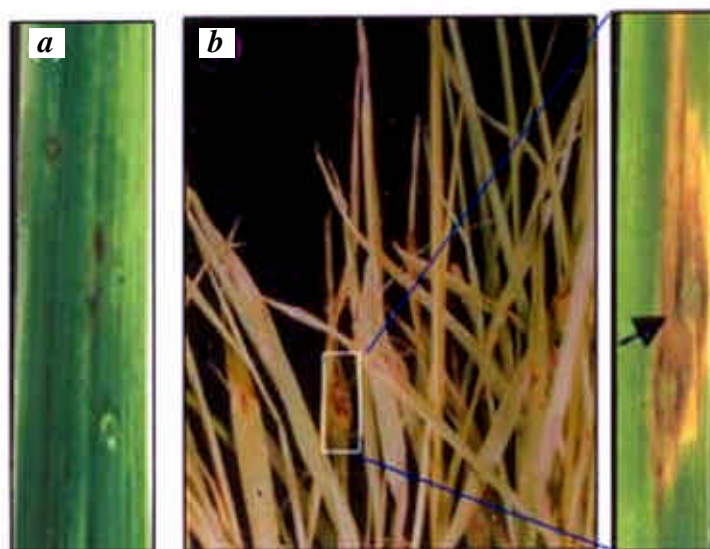

Figure 1. Rice leaves showing (*a*) incompatible and (*b*) compatible interactions. In the inset typical lesion caused by the fungus is shown.
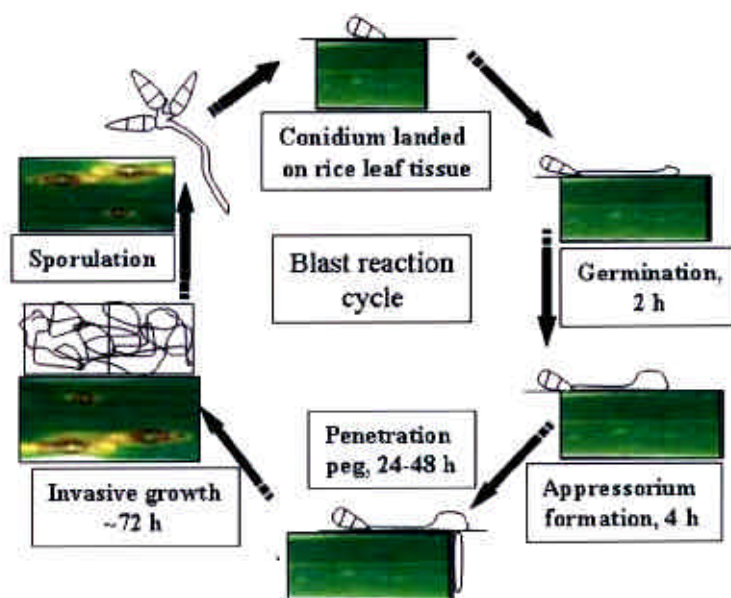

Figure 2. Life cycle of the rice blast fungus.

which harbour an estimated 10,082 and 9,457 genes, respectively. Neither of these species is known to cause any plant disease. The marginally large number of genes may code for some specific functions relevant to pathogenesis or recognition and interaction with the host. Several gene families from *M. grisea* exhibit sequence similarity to proteins that are known to be involved in the process of fungal pathogenicity.

The orthologues of *M. grisea* in *N. crassa* and *A. niger* show an average identity of ~ 46.5%. Family-classified proteins are more in *M. grisea* (1266) and *A. niger* (1424) compared to *N. crassa* (950), which is explained by the existence of a process called repeat-induced point mutation (RIP) in *N. crassa*. RIP eliminates paralogous gene duplication and consequent gene-family expansion that might occur during meiosis and consequent gene-family expansion. There are several gene families observed to be expanded in *M. grisea*, which exhibit sequence similarity to proteins involved in fungal pathogenicity.

Involvement of G-protein-coupled receptors (GPCRs), mitogen-activated protein kinases (MAPKs) and cAMP-dependent protein kinase A (CPKA) has been implicated in the development of the 'appressorium' and disease progression. The GPCRs are involved in the transduction of environmental signals, an important feature associated with the pathogenic lifestyle of the blast fungus. The genome contains the largest repertoire of GPCR-like genes among the sequenced fungi so far. Remarkably, one of the GPCRs, Pth11 is required for the formation of appresorium and pathogenesis[2]. The genome also contains three genes that code for MAPKs, which are known to regulate development of the 'appressorium', formation of the penetration peg and adaptation to osmotic stress. Existence of these genes could possibly be associated with virulence-associated signalling pathways. It is not yet clear how precisely CPKA-mediated signalling is required for the induction of 'appressorium' formation. Interestingly, a second PKA catalytic subunit domain has also been unravelled in the genome. Evaluation of the function of this gene will further our understanding the development of the infection structure.

The 'secretome' of *M. grisea* is reported to include 739 proteins, which is twice that of *N. crassa*. Several of these secreted proteins are predicted to code for enzymes which can help the pathogen pass through the first line of plant defence. For instance,

eight genes encode cutinases – methyl esterases required to degrade cutin that forms leaf cuticle. Many members of the secretome family contain consensus carbohydrate substrate-binding domains, consistent with their role in attachment and colonization of the plant tissue. The role of chitin-binding proteins has been proposed in plant–fungus interaction from studies of avirulence protein Avr4, of the tomato pathogen *Cladosporium fulvum*, which probably protects the fungus from chitinases produced by the plant. A novel variant of the cysteine pattern of Avr4 was found in *M. grisea*. It is likely to represent the chitin-binding motif, which is a characteristic feature of the plant pathogenic filamentous ascomycetes. In addition, genes encoding the effector proteins required by the plant pathogens to perturb the host cell signalling or suppress the host immune system were also predicted from the sequence. There are three families of putatively secreted, cysteine-rich polypeptides and a protein family with similarity to the necrosis-inducing peptide of *Phytophthora infestans* (NPP1). Furthermore, the genome of the sequenced blast strain contains four known avirulence genes: *AVR-Pita*, *ACE1*, *PWL2* and *PWL3*. Although a near homologue of *AVR1–CO39* was found to occur in the genome, no orthologue for any other well-characterized *AVR* genes from other pathogenic fungi could be identified. This indicates lack of sequence conservation in the fungal avirulence genes identified so far. The plant pathogenic fungi are well known to produce secondary metabolites as well. Interestingly, there are 23 genes that have been predicted to encode polyketide synthases (PKS) involved in the synthesis of secondary metabolites. This number is significantly higher than that of *N. crassa* and the functions of majority of the genes remain to be elucidated in detail.

Isolates of *M. grisea* are known to generate new pathogenic variants at a high frequency. The postulated mechanisms by which these variants arise include heterokinesis, parasexual cycle, aneuploidy and transposons. An inverted repeat transposon, *Pot2*, was cloned and characterized in the blast fungus[3]. Subsequently, several other transposons and retrotransposons have been identified[1,4]. With the genome sequence now available, 9.7% of the *M. grisea* genome appears to consist of repetitive DNA sequences longer than 200 bp. Four previously unknown repeats have now been discovered. The genome sequence also revealed full-length sequences

of the two elements for which only incomplete sequences were previously available. Most of the repetitive sequences are retrotransposons comprising eight major families. However, these repeat elements are not uniformly distributed in the genome, but occur in discrete clusters. Interestingly, many of the host-specificity genes in the blast fungus appear to be located in the transposon-rich regions of the genome. Contribution of these repeat elements to promoting genetic diversity in this fungus remains an area of active interest.

It is crucial to identify the weapons used by the fungus, to breach the host defence mechanisms, for strategic control of the disease. A significant number of the predicted genes may be related to the process of pathogen ingress and establishment in the host system. Comparative genomic analysis between the non-pathogenic and pathogenic fungi might help identify the components that determine pathogenicity. It is already apparent that there are many gene products constituting the pathogen arsenal for establishment and development of the disease. Initiatives for large-scale insertional mutagenesis and expression profiling should lead to the generation of valuable data on the biology of this fungus and towards a better understanding of the plant fungal diseases.

Sequence information from the blast fungus promises to kick-start designing of experiments for the identification of novel targets, and the development of novel fungicides. Availability of the rice genome sequence is now complemented by sequence information of its most important pathogen, thus paving the way for carrying out some insightful experiments on fungal biology as well as molecular host–pathogen interactions.

1. Talbot, N. J., *Annu. Rev. Microbiol.*, 2003, **57**, 177–202.
2. Dean, R. A. *et al.*, *Nature*, 2005, **434**, 980–986.
3. Kachroo, P., Leong, S. A. and Chattoo, B. B., *Mol. Gen. Genet.*, 1994, **245**, 339–348.
4. Kachroo, P., Leong, S. A. and Chattoo, B. B., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 11125–11129.

Subhankar Roy-Barman and Bharat B. Chattoo* are in the Department of Microbiology and Biotechnology Centre, Faculty of Science, M.S. University of Baroda, Vadodara 390 002, India.
*e-mail: chattoo@bcmsu.ac.in