

The genome sequence of the rice blast fungus *Magnaporthe grisea*

Ralph A. Dean¹, Nicholas J. Talbot², Daniel J. Ebbole³, Mark L. Farman⁴, Thomas K. Mitchell¹, Marc J. Orbach⁵, Michael Thon³, Resham Kulkarni^{1*}, Jin-Rong Xu⁶, Huaqin Pan¹, Nick D. Read⁷, Yong-Hwan Lee⁸, Ignazio Carbone¹, Doug Brown¹, Yeon Yee Oh¹, Nicole Donofrio¹, Jun Seop Jeong¹, Darren M. Soanes², Slavica Djonovic³, Elena Kolomiets³, Cathryn Rehmeier⁴, Weixi Li⁴, Michael Harding⁵, Soonok Kim⁸, Marc-Henri Lebrun⁹, Heidi Bohnert⁹, Sean Coughlan¹⁰, Jonathan Butler¹¹, Sarah Calvo¹¹, Li-Jun Ma¹¹, Robert Nicol¹¹, Seth Purcell¹¹, Chad Nusbaum¹¹, James E. Galagan¹¹ & Bruce W. Birren¹¹

¹Center for Integrated Fungal Research, North Carolina State University, Raleigh, North Carolina 27695, USA

²School of Biological and Chemical Sciences, University of Exeter, Washington Singer Laboratories, Exeter EX4 4QG, UK

³Department of Plant Pathology and Microbiology, Texas A&M University, College Station, Texas 77843, USA

⁴Department of Plant Pathology, University of Kentucky, Lexington, Kentucky 40546, USA

⁵Department of Plant Pathology, University of Arizona, Tucson, Arizona 85721, USA

⁶Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907, USA

⁷Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK

⁸School of Agricultural Biotechnology, Seoul National University, Seoul 151-742, Korea

⁹FRE2579 CNRS-Bayer, Bayer CropScience, 69263 Lyon Cedex 09, France

¹⁰Agilent Technologies, Wilmington, Delaware 19808, USA

¹¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA

* Present address: RTI International, Research Triangle Park, North Carolina 27709, USA

***Magnaporthe grisea* is the most destructive pathogen of rice worldwide and the principal model organism for elucidating the molecular basis of fungal disease of plants. Here, we report the draft sequence of the *M. grisea* genome. Analysis of the gene set provides an insight into the adaptations required by a fungus to cause disease. The genome encodes a large and diverse set of secreted proteins, including those defined by unusual carbohydrate-binding domains. This fungus also possesses an expanded family of G-protein-coupled receptors, several new virulence-associated genes and large suites of enzymes involved in secondary metabolism. Consistent with a role in fungal pathogenesis, the expression of several of these genes is upregulated during the early stages of infection-related development. The *M. grisea* genome has been subject to invasion and proliferation of active transposable elements, reflecting the clonal nature of this fungus imposed by widespread rice cultivation.**

Outbreaks of rice blast disease are a serious and recurrent problem in all rice-growing regions of the world, and the disease is extremely difficult to control^{1,2}. Rice blast, caused by the fungus *Magnaporthe grisea*, is therefore a significant economic and humanitarian problem. It is estimated that each year enough rice is destroyed by rice blast disease to feed 60 million people³. The life cycle of the rice blast fungus is shown in Fig. 1. Infections occur when fungal spores land and attach themselves to leaves using a special adhesive released from the tip of each spore⁴. The germinating spore develops an appressorium—a specialized infection cell—which generates enormous turgor pressure (up to 8 MPa) that ruptures the leaf cuticle, allowing invasion of the underlying leaf tissue^{5,6}. Subsequent colonization of the leaf produces disease lesions from which the fungus sporulates and spreads to new plants. When rice blast infects young rice seedlings, whole plants often die, whereas spread of the disease to the stems, nodes or panicle of older plants results in nearly total loss of the rice grain². Different host-limited forms of *M. grisea* also infect a broad range of grass species including wheat, barley and millet. Recent reports have shown that the fungus has the capacity to infect plant roots⁷.

Here we present our preliminary analysis of the draft genome sequence of *M. grisea*, which has emerged as a model system for understanding plant–microbe interactions because of both its economic significance and genetic tractability^{1,2}.

Acquisition of the *M. grisea* genome sequence

The genome of a rice pathogenic strain of *M. grisea*, 70-15, was sequenced through a whole-genome shotgun approach. In all, greater than sevenfold sequence coverage was produced, and a

summary of the principal genome sequence data is provided in Table 1 and Supplementary Table S1. The draft genome sequence consists of 2,273 sequence contigs longer than 2 kilobases (kb), ordered and orientated within 159 scaffolds. The total length of all sequence contigs is 38.8 megabases (Mb), and the total length of the scaffolds, including estimated sizes for the gaps, is 40.3 Mb. The genome assembly has high sequence accuracy—96% of the bases have quality scores of greater than 40—and long-range continuity, with 50% of all bases residing in scaffolds longer than 1.6 Mb.

Reconstruction of the *M. grisea* genome was aided by the availability of genome maps^{8,9} (Supplementary Methods S1). Thirty-three scaffolds, representing 32.8 Mb or 85% of the draft assembly, were ordered on the genetic map and assigned to each of the seven chromosomes by virtue of containing an anchored genetic marker. In addition, 19 scaffolds (65% of genome assembly) contained more than one marker and could thus be oriented on the map. The ends of chromosomes were identified by the telomere repeat motif (TTAGGG)*n*. Thirteen telomeric sequences were placed at the ends of scaffolds, of which six could be placed at the ends of chromosomes, whereas the remainder were associated with unanchored scaffolds (Supplementary Table S2). Genome coverage was estimated by aligning 28,682 *M. grisea* expressed sequence tags (ESTs), representing genes expressed during a range of developmental stages and environmental conditions^{10,11}. Approximately 94% of the ESTs were aligned to the genome assembly, despite many of these ESTs being from different strains.

The gene content of a plant pathogenic fungus

Within the *M. grisea* genome, 11,109 genes were predicted with

protein products of longer than 100 amino acids (Supplementary Methods S2). These predicted genes comprise 48% of the assembly (1 gene per 3.5 kb). For comparison, 10,082 genes were predicted in *Neurospora crassa*, a related pyrenomyces, and 9,457 were predicted in the more distantly related plectomyces *Aspergillus nidulans* (also known as *Emericella nidulans*; <http://www.broad.mit.edu/annotation/fungi/aspergillus/>). Neither of these species causes plant disease. At the amino acid level, *M. grisea* orthologues in *N. crassa* and *A. nidulans* show an average identity of 47% and 46%, respectively.

Given that the *M. grisea* genome possesses more genes than the non-plant pathogens *N. crassa* and *A. nidulans*, and that these fungi are well-studied model organisms, we compared genes between these species to identify potential cases of gene family expansion that might be associated with the evolution of *M. grisea* as a plant pathogen.

Single linkage clustering of protein sequences resulted in 348 families containing five or more genes. Overall, more proteins from *M. grisea* (1,266, $P < 0.001$) and *A. nidulans* (1,424, $P < 0.001$) were classified into families as compared with those from *N. crassa* (950). An important factor in this comparison is that *N. crassa* has a

process called repeat-induced point mutation (RIP), a mechanism that eliminates duplicated genes during meiosis and therefore limits the ability of *N. crassa* to undergo paralogous gene duplication and consequent gene family expansion¹². Twenty-eight families were found with significant differences ($P < 0.05$) in gene content between the three species, nine of which were larger in *M. grisea* than they were in the genome of either *N. crassa* or *A. nidulans* (Table 2; see also Supplementary Table S3). Several gene families expanded in *M. grisea* exhibited sequence similarity to proteins that are involved in fungal pathogenicity¹³.

Phylogenetic trees (Supplementary Fig. S1) of the expanded gene families in *M. grisea* indicated that none of the families showed evidence for recent lineage-specific expansion. These gene families in *M. grisea* may therefore be the result of ancient gene duplication events followed by loss of gene family members in the *N. crassa* and *A. nidulans* lineages.

Genome architecture and co-linearity

Co-linearity of chromosome segments (synteny) is widely reported for plant and animal genomes¹⁴, but little is known about this for filamentous ascomycete fungi. Analysis of orthologous pairs of genes in *M. grisea* and *N. crassa* revealed no evidence for extensive regions of conserved synteny, although linkage group assignments were often conserved between the two species, notably chromosome 7 in *M. grisea* with linkage group I of *N. crassa* (Supplementary Table S4). Only 113 regions containing four or more genes were found to be co-linear between *M. grisea* and *N. crassa*. One example, also conserved in several other filamentous fungi, is the quinate/shikimate (Qa) metabolic pathway gene cluster (Supplementary Fig. S2). This seven-gene cluster, spanning ~20 kb, which participates in quinate use and aromatic amino acid catabolism, is found on chromosome 3 in *M. grisea*, and is syntenic in *N. crassa*, *A. nidulans* and other filamentous ascomycetes, but is not present in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* or other yeasts.

M. grisea has a family of novel G-protein-coupled receptors

To be a successful plant pathogen, a fungus must undergo a series of morphological and physiological programmes⁵. During these developmental transitions, the pathogen also overcomes (or suppresses) the plant's innate immune system and perturbs host metabolism and cell signalling to favour fungal growth. We have attempted to define the most important characteristics of the genome of *M. grisea* that are associated with its pathogenic lifestyle.

G-protein-coupled receptors with seven transmembrane helices (GPCRs) transduce environmental signals, by means of heterotrimeric G proteins, to activate secondary messengers and regulate gene expression^{15,16}. The *M. grisea* genome contains a large repertoire of GPCR-like genes, including 61 not previously described. Twelve of these genes form a subfamily (Supplementary Figs S3 and S4), and contain a conserved fungal-specific extracellular membrane-spanning domain (CFEM) at the amino terminus¹⁷ that resembles the epidermal growth factor (EGF) module present in certain human GPCRs¹⁵. Notably, one of the CFEM-GPCRs, Pth11,

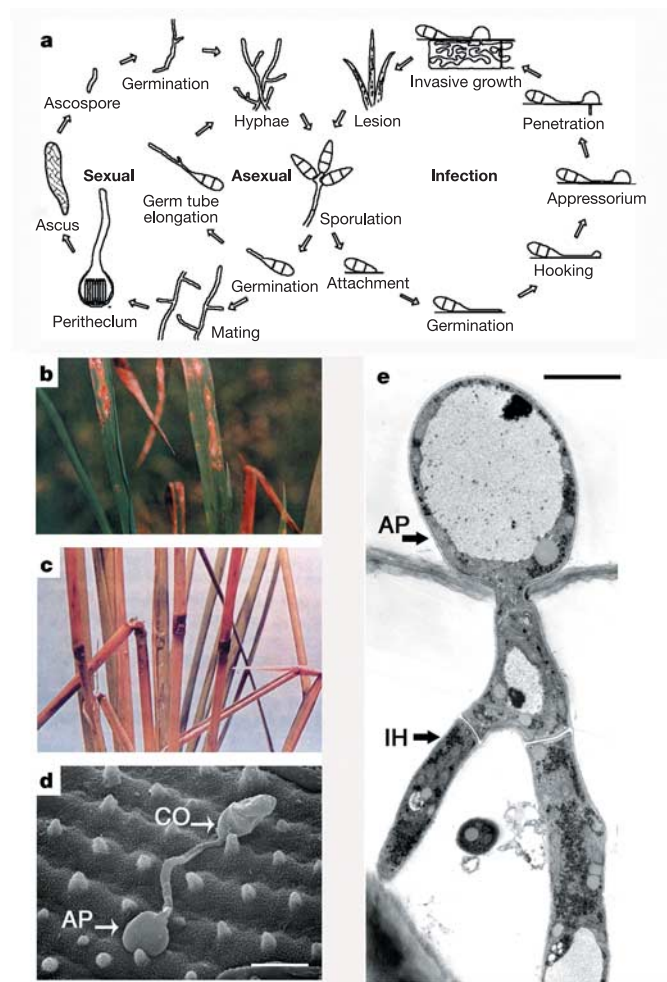


Figure 1 Life history of *Magnaporthe grisea*. **a**, Asexual spores called conidia germinate and develop a specialized infection structure, the appressorium. Invasive growth within and between cells culminates with sporulation and lesion formation. Sexual reproduction occurs when two strains of opposite mating type meet and form a peritheclum in which ascospores develop. Once released, ascospores can develop appressoria and infect host cells. **b**, Rice blast lesions on leaves. **c**, Node blast symptoms on rice, which cause lodging and yield loss. **d**, Scanning electron micrograph of an appressorium (AP) forming after germination of a conidium (CO) on the rice leaf cuticle. The appressorium generates enormous turgor to provide mechanical force to breach the leaf surface. Scale bar, 10 μ m. **e**, Transmission electron micrograph of an appressorium, showing a penetration hypha (IH) entering the leaf and developing invasive hyphae (IH). Scale bar, 5 μ m.

Table 1 *Magnaporthe grisea* assembly features

General genome features	Value
Size (bp)	37,878,070
Chromosomes	7
(G+C) percentage	51.6
Protein-coding genes	11,109
tRNA genes	316
Per cent coding	40.5
Average gene size (bp)	1,683
Average intergenic distance (bp)	1,503
Predicted protein-coding sequences	
Conserved hypothetical proteins	8,868 (79%)
Predicted proteins*	2,233 (20%)

*No similarity to known sequences.

is required for appressorium formation and pathogenesis^{2,18}, and *M. grisea* has the largest number of CFEM-GPCR proteins among sequenced fungi. In contrast, only one was detected in *N. crassa* and two in *A. nidulans*, and this type of GPCR is completely absent in the non-filamentous ascomycetes *S. cerevisiae*, *S. pombe*, *Candida albicans* and *Pneumocystis carinii*, or the basidiomycete fungi *Cryptococcus neoformans*, *Ustilago maydis* and *Phanerochaete chrysosporium* (Fig. 2). Moreover, whole-genome microarray analysis confirmed that the identified CFEM-GPCR-like proteins are expressed during infection-related development, and two CFEM-GPCR genes are specifically upregulated when the fungus is undergoing appressorium formation, as shown in Fig. 2. Together, these data suggest that *M. grisea* has greater flexibility to react to extracellular signals compared with saprobic fungi. The pathogenic lifestyle, which involves transitions from plant surface to colonization of distinct plant tissues, may therefore require the ability to respond to a wider range of physical and environmental stimuli.

Virulence-associated signalling pathways in *M. grisea*

The *M. grisea* genome contains three mitogen-activated protein kinase (MAPK) cascades that regulate appressorium development, penetration peg formation and adaptation to hyper-osmotic stress (Fig. 3). Although there are similar numbers of MAPK pathways in other fungi (*N. crassa* has three, *A. nidulans* four and *S. cerevisiae* five), the processes regulated by the pathways in different organisms are distinct. Two of the three MAPK pathways in *M. grisea* control virulence-associated development. The core of the Pmk1 MAPK (pathogenicity MAPK) pathway, which regulates appressorium formation in *M. grisea*, is clearly related to both the pheromone signalling and filamentation pathways from *S. cerevisiae*, and Pmk1 is able to function in place of either the Fus3 or Kss1 MAPK in yeast¹⁹. In *M. grisea*, however, Pmk1 pathway components such as the MAPK are involved in mating and pathogenesis, whereas Mst12, the Ste12-related transcription factor, is dispensable for mating altogether. These observations demonstrate the functional divergence of these signalling pathways. Furthermore, homologues of Pmk1 are required for fungal virulence in all plant pathogens in which they have been investigated²⁰. The Pmk1 signalling pathway revealed from the genome sequence is therefore likely to be of generic importance for fungal pathogenesis.

A principal difference in the operation of MAPK signalling in *M. grisea* compared with *S. cerevisiae* is the absence of a clear Ste5 homologue. A homologue also seems to be absent in other filamentous ascomycetes, including *A. nidulans* and *N. crassa*. In yeast, Ste5 is the scaffold protein that conditions specificity in MAPK signalling²¹. The absence of an identifiable scaffold implies that either a hitherto uncharacterized protein tethers the MAPK signalling module together in *M. grisea* and provides specificity in signal transmission, or that MAPK specificity is governed by a different mechanism in this fungus.

Cyclic AMP signalling is required for the induction of appressorium formation and for the turgor-driven process that leads to plant infection. Until now it has not been completely clear how this cAMP

signal is transmitted. A known cAMP-dependent protein kinase A (PKA) catalytic subunit (*CPKA*) is required for appressorium maturation, but is dispensable for early appressorium development^{19,22}. The genome has revealed a gene that putatively encodes a second PKA catalytic subunit (MG02832.4). Evaluation of this gene may shed light on how cAMP-mediated signalling operates both at initiation and later stages of appressorium formation. The Pth11 GPCR also operates upstream of the cAMP signalling pathway in *M. grisea*¹⁸, suggesting that the CFEM-GPCR family could provide a number of distinct and unforeseen inputs into this pathway.

Turgor-driven plant infection by *M. grisea*

Appressoria of *M. grisea* generate the enormous turgor pressure needed to breach the plant cuticle through accumulation of up to 3 M concentrations of glycerol. Analysis of the *M. grisea* genome suggests that germinating spores possess considerable versatility in their capacity to synthesize glycerol in the appressorium from storage products. Notably, in contrast to *S. cerevisiae*, where fatty acid β -oxidation occurs solely in peroxisomes, *M. grisea* has several genes that putatively encode acyl-CoA dehydrogenases, but does not appear to possess a gene encoding acyl-CoA oxidase. Thus, β -oxidation in *M. grisea* may occur both in mitochondria and in catalase-free glyoxysome-like bodies, as in *N. crassa*, which would allow use of a wide range of fatty acid substrates, including branched-chain fatty acids²³. *M. grisea* also seems to have the capacity to synthesize glycerol from the glycolytic intermediates dihydroxyacetone phosphate and dihydroxyacetone. Activity of both NADH-dependent glycerol-3-phosphate dehydrogenase and NADPH-dependent glycerol dehydrogenase has been reported in developing appressoria of *M. grisea*²⁴. Taken together, the apparent flexibility in lipid metabolism and ability to divert intermediates from glycolysis may be important for rapid glycerol accumulation during appressorium development.

***M. grisea* possesses a complex secreted proteome**

The secreted proteome is a crucial component of the ability of fungi to perceive and respond to the environment. We identified the presence of 739 proteins that are predicted to be secreted by *M. grisea*, approximately twice the corresponding number for *N. crassa*. Part of this difference reflects an expansion in protein families in *M. grisea*. For example, 163 putatively secreted proteins occur in families containing at least twice as many members as the corresponding family in *N. crassa* (Supplementary Table S5). Several of these expanded families are predicted to encode enzymes for degradation of the plant cell wall and cuticle. For example, eight genes in *M. grisea* putatively encode cutinases—methyl esterases that degrade cutin, the waxy polymer that forms the leaf cuticle. Several of these genes are significantly upregulated during infection-related development (Fig. 2). Previous experiments involving deletion of the cutinase *CUT1* (MG01943.4) led to the conclusion that enzymatic degradation of the cuticle was dispensable for plant infection²⁵. However, *CUT1* is not among those genes differentially regulated during appressorium formation. Any of the remaining seven

Table 2 Gene family content differences between *M. grisea*, *N. crassa* and *A. nidulans*

ID	Magnaporthe	Neurospora	Aspergillus	Annotation
78†	9	1*	3	Feruloyl esterase B precursor
247	6	1	0*	UDP-glucuronosyltransferase precursor, microsomal
134	8	4	0*	350–400 amino acid polypeptide
108†	8	0*	3	Cutinase precursor
4	34	11*	21	Isotrichodermin C-15 hydroxylase (cytochrome P450)
360	6	0*	0*	~180 amino acid cysteine-rich (6–8 cysteine residues) polypeptide
31†	15	1*	8	Cytochrome P450
96†	11	1*	0*	Subtilisin-like serine protease
180	10	0*	0*	~150 amino acid cysteine-rich (6–10 cysteine residues) polypeptide

Gene families larger in *M. grisea* than in *N. crassa* or *A. nidulans* (significant G-statistic; $P < 0.05$) are shown.

*Values are significantly smaller ($P < 0.05$) in pair-wise post-hoc G-tests of *M. grisea* versus *N. crassa* or *M. grisea* versus *A. nidulans*.

†Families implicated in pathogenicity¹³.

cutinases may be capable of complementing the activity of Cut1 in the *Δcut1* mutant. Coupled with the absence of cutinase-encoding genes in *N. crassa*—which colonizes dead plant tissues—these data strongly suggest that cutinases have a significant role in *M. grisea*.

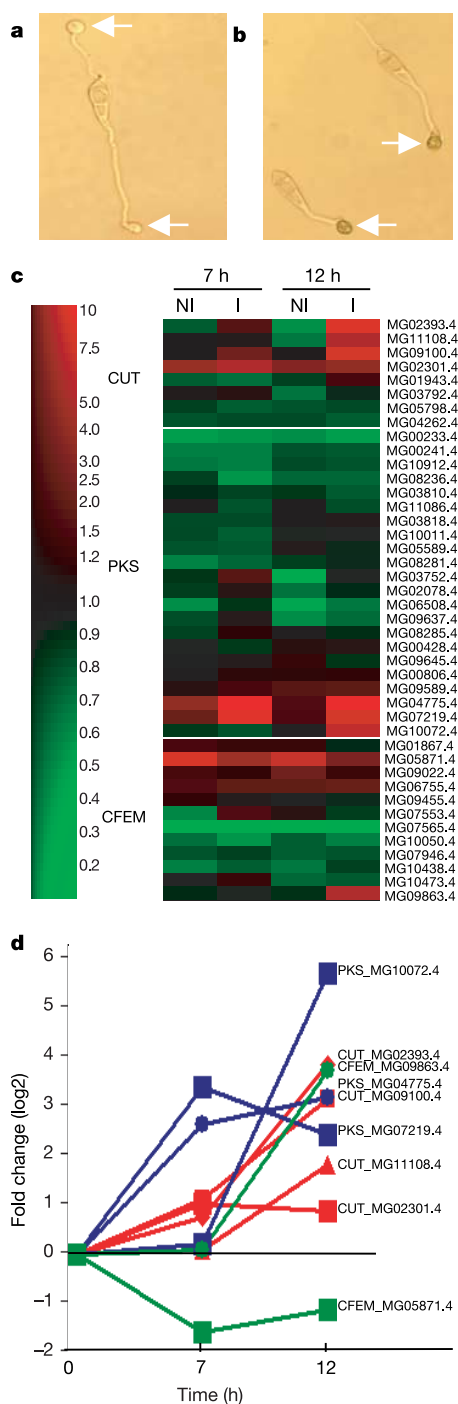


Figure 2 Differential expression of selected *M. grisea* genes during infection-related development. **a, b**, Appressorium formation (arrows) at 7 h (**a**; cessation of polar growth and tip hooking) and 12 h (**b**; tip swelling and melanization) after germination. **c**, Expression profiles of selected genes from conidia germinated on hydrophobic appressorium-inducing (I) and hydrophilic non-appressorium-inducing (NI) surfaces. The colour scale indicates transcript abundance relative to ungerminated conidia: red, increase in transcript abundance; green, decrease in transcript abundance. **d**, Fold change in transcript abundance (cutinases (CUT), red; polyketide synthases (PKS), blue; and Pth11 homologues with CFEM domain, green) at 7 and 12 h after germination on a hydrophobic surface compared with a hydrophilic surface. Only genes exhibiting a twofold or greater (≤ -1 or ≥ 1) change in expression are shown.

Among the secreted proteins predicted for *M. grisea*, many contained consensus carbohydrate substrate-binding domains, consistent with a role in attachment and colonization of plant tissue (Supplementary Table S6). Specifically, a role for chitin-binding proteins in plant–fungus interactions has been proposed from studies of the avirulence protein Avr4 of the tomato pathogen *Cladosporium fulvum*²⁶. Avr4 is a chitin-binding protein that may act to protect the fungal cell wall from chitinases produced by plants’ innate immune response. Inspection of proteins containing the cysteine pattern found in Avr4 and other variant motifs revealed a novel pattern that is highly abundant in *M. grisea*. The novel variant cysteine pattern CX₇CCX₅C is present in 36 copies of 21 predicted proteins. In contrast, this pattern occurs just eight times in *A. nidulans*, four times in *N. crassa*, and not at all in *S. cerevisiae* open reading frames (Supplementary Fig. S5). This motif probably represents a variant chitin-binding motif whose abundance may be diagnostic for plant-associated filamentous ascomycetes.

Fungal effectors and PAMPs

Pathogenic microorganisms of plants are known to secrete proteins directly into host plant cells to perturb host cell signalling or suppress the plant innate immune system^{27,28}. The plant adaptive immune system has, in turn, evolved to recognize pathogen effector proteins (often termed pathogen-associated molecular patterns, PAMPs). In the *M. grisea*–rice interaction, this immunity is governed by a gene-for-gene system (one gene in the host confers resistance to a pathogen effector protein encoded by a single pathogen gene).

Interrogation of the *M. grisea* genome for putative effector proteins revealed three families of putatively secreted, cysteine-rich polypeptides (clusters 180, 360 and 641) and a protein family with similarity to the necrosis-inducing peptide of *Phytophthora infestans* (NPP1, pfam05630)²⁹. As described above, cysteine-rich polypeptides are recognized as PAMPs, as exemplified by the *C. fulvum*–tomato interaction³⁰. In addition, a novel family of proteins with similarity to the N-terminal half (~150 amino acids) of the enterotoxin A chain (pfam01375) was noted. This region of the enterotoxin A chain contains ADP-ribosylation activity³¹, which suggests that the *M. grisea* protein may interact with rice GTP-binding proteins. The genome of the sequenced *M. grisea* strain 70-15 contains four known *M. grisea* avirulence genes: *AVR-Pita*, *ACE1*, *PWL2* and *PWL3*. No orthologues were found to *M. grisea* *PWL1*, *PWL4* or *AVR1-CO39*, or to well-characterized AVR genes from other pathogenic fungi including *Avr2*, *Avr4*, *Avr9*, *ECP2*, *ECP3* and *ECP5* from *C. fulvum*³⁰, and *NIP1* from *Rhynchosporium secalis*³². This highlights the diversity and lack of sequence similarity or conservation in the fungal avirulence gene products identified so far.

Secondary metabolic pathways of *M. grisea*

Filamentous fungi are well known producers of secondary metabolites, which in nature fulfil a variety of functions probably to allow for niche exploitation. Plant pathogenic fungi produce diverse secondary metabolites that aid in pathogenicity, such as host-selective toxins³³. The *M. grisea* genome displays a considerable capacity for production of secondary metabolites. There are 23 genes predicted to encode polyketide synthases (PKS), compared with seven PKS genes in *Neurospora*, and three of the PKS-encoding genes in particular are upregulated during infection-related development (Fig. 2). Beyond containing the ketosynthase domain, the structure of these proteins is highly divergent (Supplementary Fig. S6). Most of the 23 PKS genes appear to occur in gene clusters with neighbouring genes that are predicted to encode enzymes such as cytochrome P450 and mono-oxygenases, which typically modify or customize the polyketide backbone to form a functional secondary metabolite (Supplementary Table S7). The diversity of PKS genes in filamentous ascomycete fungi and the poor conservation of clearly orthologous PKS genes, even in related species³⁴, indicates

that considerable variability is likely to exist in the polyketide metabolites generated by pathogenic fungi.

Non-ribosomal peptide synthetases (NRPS), which catalyse production of cyclic peptides including numerous toxins, also seem to be well represented in the *M. grisea* genome. Two distinct subclasses exist: those that exist separately and those that are fused to a PKS (PKS–NRPS). Overall, there are six likely NRPS genes and eight PKS–NRPS genes in the *M. grisea* genome: six full-length PKS–NRPS genes and two PKS–NRPS genes with a truncated NRPS domain. This contrasts with two predicted NRPS genes and one NRPS-related gene reported in the *N. crassa* genome sequence. One of the hybrid PKS–NRPS proteins is encoded by the *ACE1* gene, which has recently been shown to act as an avirulence gene³⁵. The large number, and expression profile, of PKS and NRPS genes in *M. grisea* is consistent with the requirements of a fungal pathogen in adapting to diverse environments, perturbing host metabolism and ultimately causing plant cell death.

Repetitive DNA and repeat-induced point mutations

M. grisea exhibits a high degree of genetic variability, and novel pathogenic variants capable of infecting formerly resistant host plants arise with alarming frequency during rice cultivation³⁶. Such

gains of virulence are often associated with transposon-mediated inactivation or deletion of PAMP-encoding genes whose products trigger the plant adaptive immune system^{37,38}. Thus, an understanding of the natural history of repetitive elements in *M. grisea* not only provides an insight into their impact on genome evolution but also sheds light on mechanisms of pathogenic variation. Approximately 9.7% of the *M. grisea* genome assembly comprises repetitive DNA sequences longer than 200 base pairs (bp) and with greater than 65% similarity. Four previously unknown repeats were discovered, as were alternative forms for three previously described transposons. The genome sequence also revealed full-length sequences of two elements for which only incomplete sequences were previously available.

Most repetitive sequences in the assembly are retrotransposons comprising eight major families (Table 3). Five are retroelement families and three are DNA transposons. These repetitive elements are not uniformly distributed in the genome assembly, but form discrete clusters (Supplementary Fig. S7 and Supplementary Note S1)³⁹. Further examination revealed many examples of elements inserted into copies of themselves or other elements. Examination of integration events occurring within other transposons provides evidence for extensive past recombination in the genome

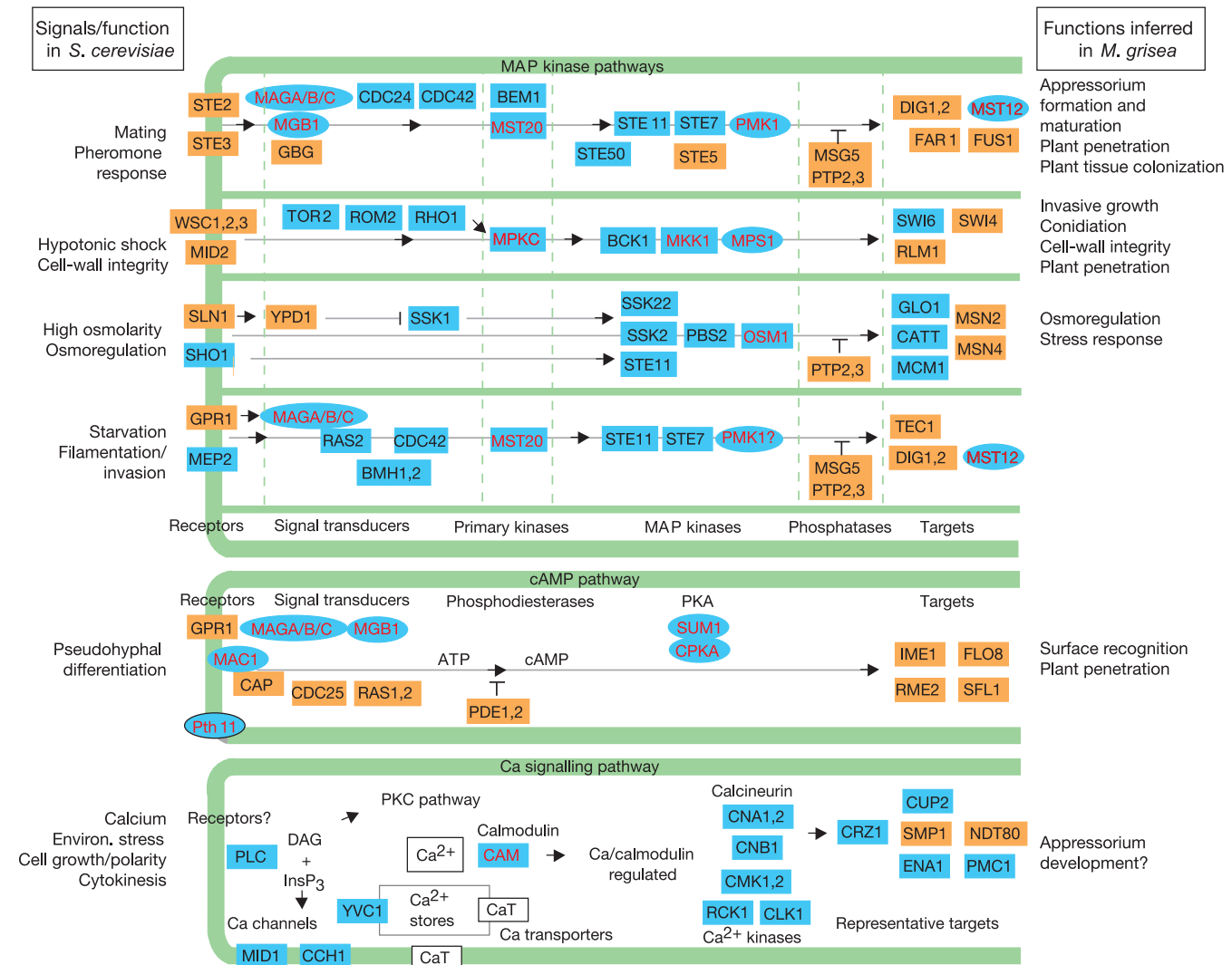


Figure 3 Comparison of major signalling pathways between *S. cerevisiae* and *M. grisea*. The core components of MAP kinase-, cAMP-dependent and calcium-signalling pathways are conserved; however, receptors and downstream targets are less conserved. Protein names in red indicate previously identified *M. grisea* homologues; names in black are *S. cerevisiae* protein names. The colour of the boxes around the protein names

indicates the degree of conservation between *M. grisea* and *S. cerevisiae* proteins based on BLASTP: blue, e -value $< 1 \times 10^{-10}$; orange, e -value $> 1 \times 10^{-10}$. Ovals indicate instances where the *M. grisea* homologue was shown to be required for pathogenicity. Ovals with black borders are *M. grisea* proteins not found in *S. cerevisiae* or not implicated in the *S. cerevisiae* pathway. InsP₃, inositol triphosphate; DAG, diacylglycerol.

(Supplementary Discussion S1). Given the prevalence of repetitive elements and their ability to participate in recombination, it is perhaps surprising that an organism could tolerate such substantial genomic change. However, in nature, rice pathogenic strains of *M. grisea* propagate asexually and, as such, genome organization is rarely, if ever, subject to the potentially catastrophic effects of meiotic recombination involving homologous chromosomes with radically different structures⁴⁰. Thus, rearrangements that would normally have been purged by meiosis appear to have been maintained in the absence of deleterious effects on vegetative fitness. Some rearrangements are expected to have positive fitness benefits, especially those that result in loss of genes whose products would normally trigger defence responses in potential hosts. Many host-specificity genes in *M. grisea* are situated in transposon-rich regions of the genome; this arrangement provides ample opportunity for host-range expansion through gene loss^{37,41}.

The prevalence of intact and essentially identical repeated DNA elements in the genome suggests that *M. grisea* has been unable to stop their proliferation. This is of interest, because RIP has previously been reported to occur in at least one strain (Br48) of *M. grisea*⁴². Inspection of *M. grisea* repeats reveals evidence for RIP (Supplementary Methods S3). However, for Pyret, the repeat family showing the most extensive RIP-like mutations, only three-quarters exhibit signs of RIP and show an average of only 11.4% sequence divergence from the reference sequence. Other elements such as Pot2 are present in over 100 apparently intact copies, displaying an average of 99.4% nucleotide identity. In contrast, all repeat elements in *N. crassa* show heavy RIP mutation, typically to greater than 20% divergence¹². The presence of many intact, highly similar elements despite evidence of RIP can be explained by a number of factors. First, it is possible that RIP was lost in the sequenced strain.

Table 3 Characterization of repeat elements in the *M. grisea* genome

Repeat name	Length (bp)	Per cent of genome*	IR/LTR length (bp)	Evidence of RIP†
Inverted repeat DNA transposons				
Pot2/4		1.3§		–§
Pot2‡	1,857		45	
Pot4	1,858		42	
Pot3	1,860	0.31	42	–
Occan	2,686	0.33	74	++
LTR retrotransposons				
MAGGY	5,638	0.86	253	+
MGLR-3	6,304	0.10	213	++
Pyret‡		1.47§		+++§
Pyret-1¶	7,407		450	
Pyret-2	7,441		474	
Pyret-3	7,407		476	
RETRO5	7,623	0.48	537	–
RETRO6‡		0.37§		+++§
RETRO6-1	5,871		193	
RETRO6-2	5,905		194	
RETRO7‡		0.49§		+++§
RETRO7-1	6,210		216	
RETRO7-2	6,134		208	
Non-LTR retrotransposons (LINEs and SINEs)				
MGL	5,977	1.43	–	++
Mg-SINE-A	472	0.16	–	++
REPBUF	1,111	0.01	–	–
Miscellaneous				
SR2	175	0.01	–	+
SR3	88	0.01	–	–
5S rRNA	70	0.01	–	–

Shown are elements that have a copy number of ten or greater and for which at least one full-length copy exists in the genome. IR, inverted repeat; LINE, long interspersed repeat element; LTR, long terminal repeat; SINE, short interspersed repeat element.

*Calculated using repeats identified with RepeatMasker (A. F. Smit, unpublished data; see <http://www.repeatmasker.org>). Some subfamilies cannot be distinguished using this method and were collapsed into their parent family.

†The number of plus signs indicates the approximate percentage of element insertions that exhibit RIP-like mutations (transition/transversion ratio >10): +++, >50%; ++, >10%; +, <10%; and –, no evidence of RIP.

‡Elements with different subclasses.

§The presence of truncated copies made it difficult to distinguish between subclasses.

||New elements identified in this study.

¶Previously reported sequence⁵⁰.

However, within the genome sequence there are numerous repetitive elements that show only transition mutations. This, combined with the presence of a gene in *M. grisea* predicted to encode a DNA methyltransferase homologous to the *RID* gene of *N. crassa*, which is required for RIP, indicates recent and possibly ongoing RIP⁴³. Second, the experimental evidence for RIP in *M. grisea* also demonstrated that the process is both less efficient at recognizing repeats and less severe in the number of mutations induced than in *N. crassa*⁴². Therefore, it is possible that many transposable elements in *M. grisea* simply escape damage by RIP. Finally, because RIP is only known to operate during the sexual cycle, the lack of RIP-like mutations in most repetitive elements may reflect the predominance of vegetative propagation during the recent evolution of rice pathogenic strains of *M. grisea*.

Discussion

This is the first analysis of the genome of a plant pathogenic fungus. Our analysis of the *M. grisea* genome has allowed a much greater appreciation of the likely attributes required by a plant pathogenic fungus to invade and colonize a living host plant. *M. grisea* has a considerably diverse set of proteins involved in extracellular perception and signal transduction, an extensive array of secreted proteins and secondary metabolites, specifically adapted regulatory pathways controlling infection-related development, and a genome capable of generating considerable genetic variation even in the absence of sexual reproduction. New opportunities for disease control and novel targets—such as unique families of CFEM-GPCR surface receptors and secreted cysteine-rich proteins—for development of durable fungicides are already apparent from our analyses. Acquisition of the genome sequence of *M. grisea* and that of its host, rice^{44,45}, coupled with the genetic and experimental tractability of this pathosystem, will enable a systems biology approach to the study of a plant–fungal interaction. Large-scale, insertional mutagenesis projects, transcriptional profiling and proteomic analysis are already in progress and offer the possibility of an enhanced understanding of the processes by which a fungus causes plant disease.

The genome sequence of *M. grisea* illustrates an extraordinary feature of the fungal kingdom, namely the extent of sequence diversity between what are thought of as closely related species from a taxonomic perspective. For example, although *M. grisea* and *N. crassa* are both pyrenomycetes and are often studied using similar methods in the same laboratory, they exhibit a degree of sequence diversity similar to that found between human and *Xenopus*⁴⁶. □

Methods

Strain and growth conditions

Rice-infecting *M. grisea* strain 70-15 (ref. 47) (Fungal Genetics Stock Center 8958) was grown in liquid complete medium, and DNA was extracted as previously described⁴⁸.

Sequencing and assembly

Plasmid (4-kb inserts) and fosmid (40-kb inserts) libraries were generated and end-sequenced as described (<http://www.broad.mit.edu/annotation/fungi/magnaporthe/assembly.html#clones>). Bacterial artificial chromosome (BAC) libraries were generated and end-sequenced as described⁴⁹. The draft genome sequence was assembled using Arachne (<http://www.broad.mit.edu/wga/>). Mapped genetic markers were associated with sequence scaffolds through hybridization to end-sequenced BACs (Supplementary Methods S1). ESTs were aligned to the genome using a BLAST *e*-value cutoff of $\leq 10^{-20}$. Assembly version 2 was used for all subsequent analyses.

Annotation and analysis

Automated gene predictions and annotations were performed using Calhoun (http://www.broad.mit.edu/annotation/fungi/magnaporthe/gene_finding.html). Gene predictions were performed using FGENESH/FGESH1+ trained on *M. grisea* sequences (SoftBerry) and GENESIS (Sanger Center) and validated against 65 characterized *M. grisea* genes. Additional information and gene identification numbers for genes featured in this article are presented in Supplementary Table S8.

To perform co-linearity analyses, amino acid identity between *M. grisea* and *N. crassa* was first determined by comparing the predicted proteins from each fungus using BLASTP. Homologues with the best hit were aligned using ClustalW and the amino acid per cent identity for each pair was calculated. *M. grisea* and *N. crassa* genes were considered

to belong to a conserved cluster if there was less than 10 kb between any two genes in the cluster. Homologous genes in clusters between species were accepted if alignments spanned $\geq 60\%$ of both genes and the alignment score was within 80% of the top score for either of the pair of genes. In this analysis, a gene may be placed in more than one cluster. No attempt was made to identify or resolve these cases.

Blastclust (<http://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/>) was used to cluster predicted peptide sequences from *M. grisea*, *N. crassa* and *A. nidulans* into families using threshold limits of 30% identity and 80% length overlap. Functional annotations were manually curated by examining blast searches to the GenBank NR protein database, SwissProt and a database of *M. grisea* repetitive sequences. A heterogeneity G-test was performed to identify families with significant content differences between species. Pair-wise post-hoc tests were performed using the Bonferroni correction for multiple comparisons. Families with significant similarity to proteins encoded by transposable elements were removed.

To identify G-protein-coupled receptor-like proteins, known GPCR sequences, including ones present in GPCRDB (www.gpcr.org/7tm/), were blasted (e -value limit of $\leq 10^{-9}$) against the *M. grisea* predicted proteins. These, in addition to candidates identified from an Interpro scan, were confirmed to have seven transmembrane spans by TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html), Phobius (<http://phobius.cgb.ki.se/>) and TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>). Default settings were used. *N. crassa* and *A. nidulans* CFEM-containing GPCRs were identified by BLASTP.

Putative polyketide synthases and non-ribosomal peptide synthetases were identified by comparison of the genome sequence to known sequences in GenBank from *A. nidulans*, *Leptosphaeria maculans*, *Nectria haematococca*, *Gibberella moniliformis* and *Cochliobolus heterostrophus*.

Microarray experiments were performed using conidia germinated in water on either the hydrophobic or hydrophilic side of GelBond. At 7 and 12 h after germination, samples were flash frozen with liquid nitrogen, scraped from the support, ground and RNA extracted using Sigma Trizol reagent. RNA was similarly extracted from ungerminated conidia. RNA from two biological replications of each treatment was pooled in equal amounts and labelled with Cy3 and Cy5 dyes using Agilent Technologies low input linear amplification kit. Samples were hybridized to the Agilent *M. grisea* whole genome oligo array (product G4137A) using manufacturer protocols and reagents. Hybridizations were performed in an interlaced loop where each treatment was paired with every other. A total of ten hybridizations were performed, each treatment was used in four hybridizations (two with Cy3 and two with Cy5). Spot fluorescence was normalized using Lowess within and between slides, and gene expression profiles were analysed in GeneSpring. Microarray data may be accessed through NCBI GEO (<http://www.ncbi.nlm.nih.gov/projects/geo/>) accession GSE1945.

Additional information concerning genome sequencing and analysis can be found at <http://www.broad.mit.edu/annotation/fungi/magnaporthe/>.

Received 4 November 2004; accepted 7 February 2005; doi:10.1038/nature03449.

1. Valent, B. & Chumley, F. G. Molecular genetic analysis of the rice blast fungus *Magnaporthe grisea*. *Annu. Rev. Phytopathol.* **29**, 443–467 (1991).
2. Talbot, N. J. On the trail of a cereal killer: investigating the biology of *Magnaporthe grisea*. *Annu. Rev. Microbiol.* **57**, 177–202 (2003).
3. Zeigler, R. S., Leong, S. A. & Teeng, P. S. *Rice Blast Disease* (CAB International, Wallingford, 1994).
4. Hamer, J. E., Howard, R. J., Chumley, F. G. & Valent, B. A mechanism for surface attachment in spores of a plant pathogenic fungus. *Science* **239**, 288–290 (1988).
5. Dean, R. A. Signal pathways and appressorium morphogenesis. *Annu. Rev. Phytopathol.* **35**, 211–234 (1997).
6. de Jong, J. C., McCormack, B. J., Smirnov, N. & Talbot, N. J. Glycerol generates turgor in rice blast. *Nature* **389**, 244–245 (1997).
7. Sesma, A. & Osbourn, A. E. The rice blast pathogen undergoes developmental processes typical of root-infecting fungi. *Nature* **431**, 582–586 (2004).
8. Sweigard, J. A. *et al.* in *Genetic Maps* (ed. O'Brien, S. J.) 3.112 (Cold Spring Harbor, New York, 1993).
9. Nitta, N., Farman, M. L. & Leong, S. A. Genome organization of *Magnaporthe grisea*: integration of genetic maps, clustering of transposable elements and identification of genome duplications and rearrangements. *Theor. Appl. Genet.* **95**, 20–32 (1997).
10. Takano, Y., Choi, W. B., Mitchell, T. K., Okuno, T. & Dean, R. A. Large scale parallel analysis of gene expression during infection-related morphogenesis of *Magnaporthe grisea*. *Mol. Plant Pathol.* **4**, 337–346 (2003).
11. Ebbole, D. J. *et al.* Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: Analysis of expressed sequence tags. *Mol. Plant Microbe Interact.* **17**, 1337–1347 (2004).
12. Galagan, J. E. *et al.* The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 859–868 (2003).
13. Idnurm, A. & Howlett, B. J. Pathogenicity genes of phytopathogenic fungi. *Mol. Plant Pathol.* **2**, 241–255 (2001).
14. Schmidt, R. Synteny: recent advances and future prospects. *Curr. Opin. Plant Biol.* **3**, 97–102 (2000).
15. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
16. Bolker, M. Sex and crime: heterotrimeric G proteins in fungal mating and pathogenesis. *Fungal Genet. Biol.* **25**, 143–156 (1998).
17. Kulkarni, R. D., Kelkar, H. S. & Dean, R. A. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Biochem. Sci.* **28**, 118–121 (2003).
18. DeZwaan, T. M., Carrol, A. M., Valent, B. & Sweigard, J. A. *Magnaporthe grisea* Pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive surface cues. *Plant Cell* **11**, 2012–2030 (1999).
19. Xu, J. R. & Hamer, J. E. MAP kinase and cAMP signalling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*. *Genes Dev.* **10**, 2696–2706 (1996).
20. Xu, J. R. MAP kinases in fungal pathogens. *Fungal Genet. Biol.* **31**, 137–152 (2000).

21. Ptashne, M. & Gann, A. Imposing specificity on kinases. *Science* **299**, 1025–1027 (2003).
22. Mitchell, T. K. & Dean, R. A. The cAMP-dependent protein kinase catalytic subunit is required for appressorium formation and pathogenesis by the rice blast pathogen *Magnaporthe grisea*. *Plant Cell* **7**, 1869–1878 (1995).
23. Theiringer, R. & Kunau, W. H. The β -oxidation system in catalase-free microbodies of the filamentous fungus *Neurospora crassa*. Purification of a multifunctional protein possessing 2-enoyl-CoA hydratase, L-3-hydroxyacyl-CoA dehydrogenase, and 3-hydroxyacyl-CoA epimerase activities. *J. Biol. Chem.* **266**, 13110–13117 (1991).
24. Thines, E., Weber, R. W. S. & Talbot, N. J. MAP kinase and protein kinase A-dependent mobilization of triacylglycerol and glycogen during appressorium turgor generation by *Magnaporthe grisea*. *Plant Cell* **12**, 1703–1718 (2000).
25. Sweigard, J. A., Chumley, F. G. & Valent, B. Disruption of a *Magnaporthe grisea* cutinase gene. *Mol. Gen. Genet.* **232**, 183–190 (1992).
26. Van den Berg, H. A. *et al.* Natural disulfide bond-disrupted mutants of AVR4 of the tomato pathogen *Cladosporium fulvum* are sensitive to proteolysis, circumvent Cf-4-mediated resistance, but retain their chitin-binding ability. *J. Biol. Chem.* **278**, 27340–27346 (2003).
27. Nimchuk, Z., Eulgem, T., Holt, B. F. III & Dangl, J. L. Recognition and response in the plant immune system. *Annu. Rev. Genet.* **37**, 579–609 (2003).
28. Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P. & Valent, B. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**, 4004–4014 (2000).
29. Felbrich, G. *et al.* NPP1, a Phytophthora-associated trigger of plant defense in parsley and *Arabidopsis*. *Plant J.* **32**, 375–390 (2002).
30. Joosten, M. H. A. J., Cozijnsen, T. J. & de Wit, P. J. G. M. Host resistance to the fungal tomato pathogen lost by a single base-pair change in the avirulence gene. *Nature* **367**, 384–386 (1994).
31. Moss, J., Stanley, S. J., Vaughan, M. & Tsuji, T. Interaction of ADP-ribosylation factor with *Escherichia coli* enterotoxin that contains an inactivating lysine 112 substitution. *J. Biol. Chem.* **268**, 6383–6387 (1993).
32. Steiner-Lange, S. *et al.* Differential defense reactions in leaf tissues of barley in response to infection by *Rhynchosporium secalis* and to treatment with a fungal avirulence gene product. *Mol. Plant Microbe Interact.* **16**, 893–902 (2003).
33. Wolpert, T. J., Dunkle, L. D. & Ciuffetti, L. M. Host-selective toxins and avirulence determinants: what's in a name? *Annu. Rev. Phytopathol.* **40**, 251–285 (2002).
34. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl Acad. Sci. USA* **100**, 15670–15675 (2003).
35. Bohnert, H. U. *et al.* A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. *Plant Cell* **16**, 2499–2513 (2004).
36. Bonman, J. M. Durable resistance to rice blast disease—environmental influences. *Euphytica* **63**, 115–123 (1992).
37. Farman, M. L. *et al.* Analysis of the structure of the AVR1-CO39 avirulence locus in virulent rice-infecting isolates of *Magnaporthe grisea*. *Mol. Plant Microbe Interact.* **15**, 6–16 (2002).
38. Kang, S., Lebrun, M.-H., Farrall, L. & Valent, B. Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. *Mol. Plant Microbe Interact.* **14**, 671–674 (2001).
39. Thon, M. R., Martin, S. L., Goff, S., Wing, R. A. & Dean, R. A. BAC end sequences and a physical map reveal transposable element content and clustering patterns in the genome of *Magnaporthe grisea*. *Fungal Genet. Biol.* **41**, 657–666 (2004).
40. Farman, M. L. Meiotic deletion at the *BUF1* locus of the fungus *Magnaporthe grisea* is controlled by interaction with the homologous chromosome. *Genetics* **160**, 137–148 (2002).
41. Kang, S., Sweigard, J. A. & Valent, B. The PWL host specificity gene family in the rice blast fungus, *Magnaporthe grisea*. *Mol. Plant Microbe Interact.* **8**, 939–948 (1995).
42. Ikeda, K. *et al.* Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. *Mol. Microbiol.* **45**, 1355–1364 (2002).
43. Freitag, M., Williams, R. L., Kothe, G. O. & Selker, E. U. A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **99**, 8802–8807 (2002).
44. Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
45. Goff, S. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
46. Hedges, S. B. The origin and evolution of model organisms. *Nature* **3**, 838–849 (2002).
47. Chao, C. & Ellingboe, A. Selection for mating competence in *Magnaporthe grisea* pathogenic to rice. *Can. J. Bot.* **69**, 2130–2134 (1991).
48. Weiland, J. J. Rapid procedure for extraction of DNA from fungal spores and mycelium. *Fungal Genet. Newslett.* **44**, 60–63 (1997).
49. Zhu, H., Blackmon, B. P., Sasinowski, M. & Dean, R. A. Physical map and organization of chromosome 7 in the rice blast fungus, *Magnaporthe grisea*. *Genome Res.* **9**, 739–750 (1999).
50. Nakayashiki, H. *et al.* Pyret, a Ty3/Gypsy retrotransposon in *Magnaporthe grisea* contains an extra domain between the nucleocapsid and protease domains. *Nucleic Acids Res.* **29**, 4106–4113 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors acknowledge the USDA-CSRES and the National Science Foundation for funding this work. We thank all the members of the Dean Laboratory at NCSU, and members of the Broad Institute Sequencing Platform, Assembly and Annotation teams, and members at each collaborating laboratory. We also thank the rice blast research community at large. We acknowledge other fungal research communities, in particular the *A. nidulans* community, for making it possible to have access to genome sequence information before publication.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.D. (Ralph_Dean@ncsu.edu). The whole-genome shotgun data has been deposited at GenBank/EMBL/DDJB under the project accession number AACU00000000.