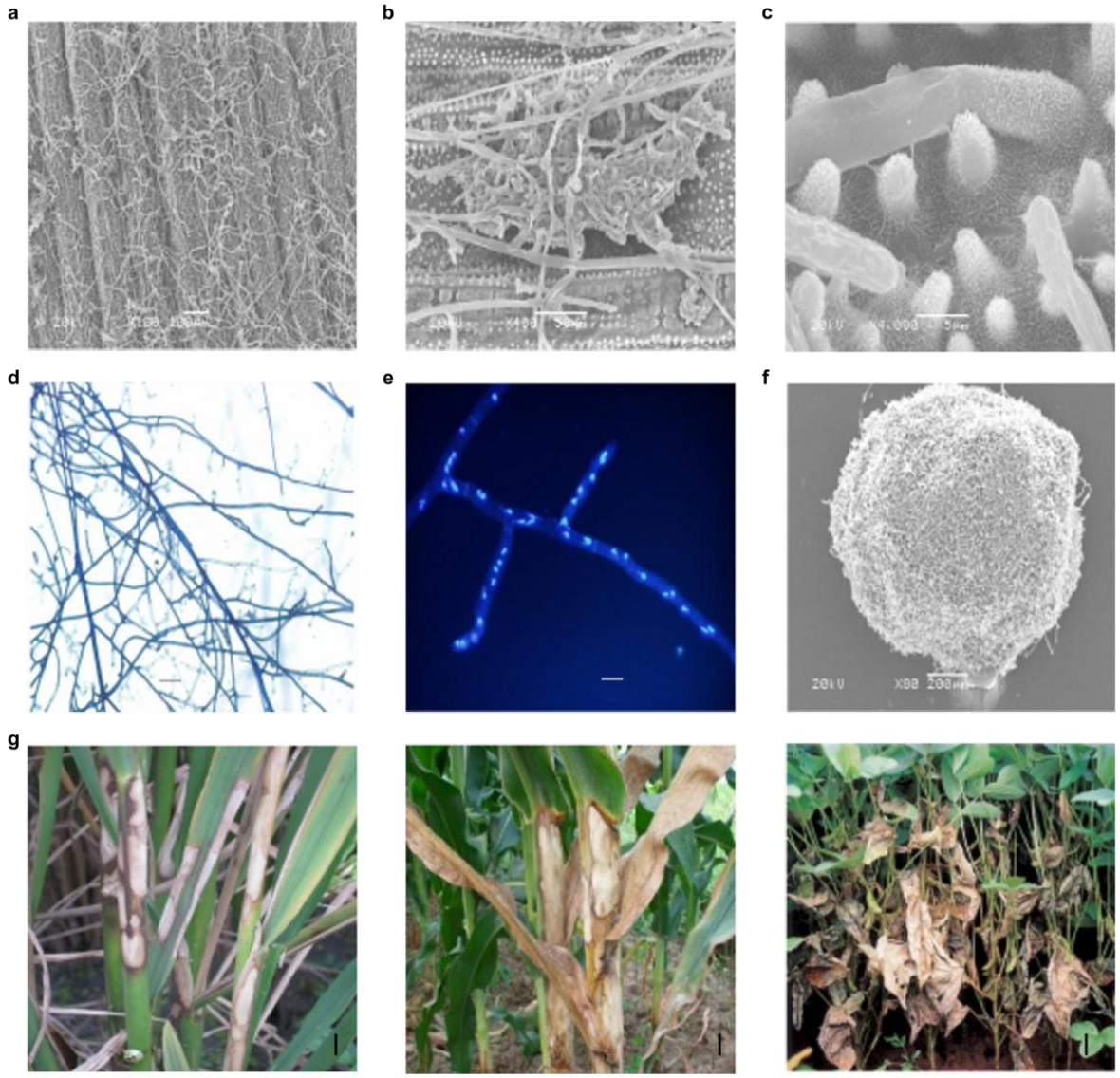
# Supplementary Information

## The Evolution and Pathogenic Mechanisms of the Rice Sheath Blight Pathogen

Aiping Zheng[1,2,3,]*, Runmao Lin[1,]*, Danhua Zhang[1], Peigang Qin[1], Lizhi Xu[1], Peng Ai[1], Lei Ding[3], Yanran Wang[1], Yao Chen[1],Yao Liu[1], Zhigang Sun[1], Haitao Feng[1], Xiaoxing Liang[1], Rongtao Fu[1], Changqing Tang[1], Qiao Li[1], Jin Zhang[1], Zelin Xie[3], Qiming Deng[1,2,3], Shuangcheng Li[1,2,3], Shiquan Wang[1,2,3], Jun Zhu[1,2,3], Linxia Wang[2], Huainian Liu[2] & Ping Li[1,2,3]

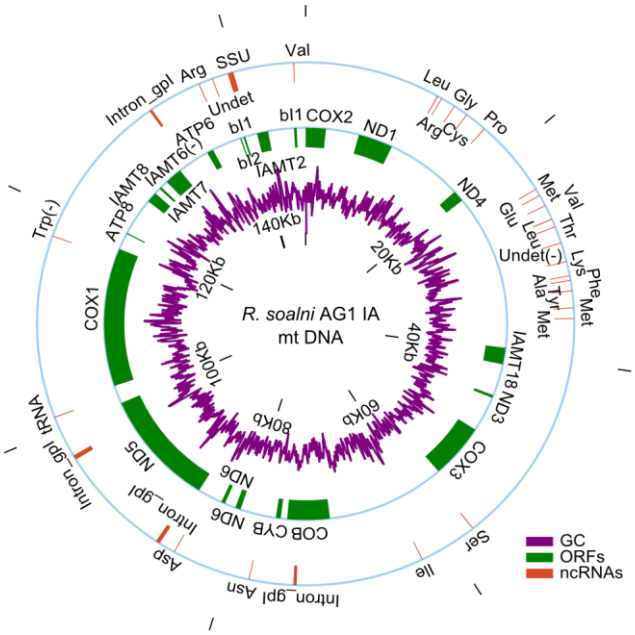
[1] State Key Laboratory of Hybrid Rice, Sichuan Agricultural University, Chengdu 611130, China. [2] Rice Research Institute of Sichuan Agricultural University, Chengdu 611130, China. [3] Key Laboratory of Southwest Crop Gene Resource and Genetic Improvement of Ministry of Education, Sichuan Agricultural University, Ya'an 625014, China. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.Z. or P.L. (email: apzh0602@gmail.com or liping6575@163.com).

## SUPPLEMENTARY FIGURES



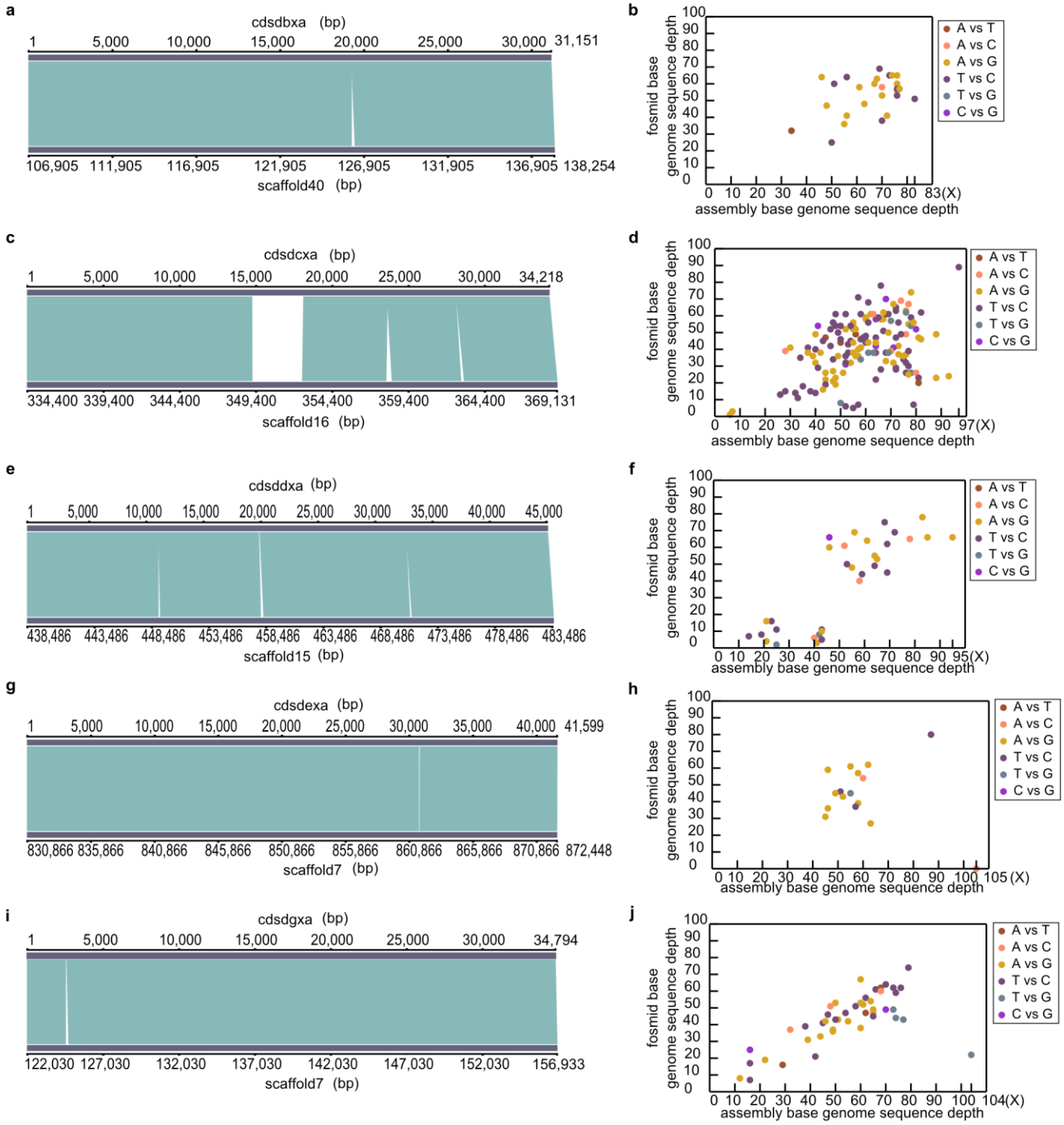

**Supplementary Figure S1. Characteristics of *R. solani* AG1 IA strain.**

a The strain growth on the rice 9311 leaf at 24 hours post inoculation observed using a scanning electron microscope (SEM). Scale bars , 100 μm. b The infection cushion formed at 18 hours post inoculation observed by an SEM. Scale bars, 50 μm. c The haphae penetrating into the rice tissue observed using an SEM. Scale bars, 5 μm. d The vegetative mycelium dyed with lactophenol cotton blue cultured on potato dextrose agar (PDA) medium for 24 hours. Scale bars, 50 μm. e The multiple nuclear mycelia dyed with 4', 6-diamidino-2-phenylindole (DAPI) after culture on PDA medium for 24 hours, observed using a fluorescence microscope. Scale bars,10 μm. f The sclerotia cultured on PDA medium for 72 hours, scanned by an SEM. Scale bars, 100 μm. g The infected rice, corn and soybean plants in the field with sheath blight disease (SBD) symptoms. Scale bars (rice), 4cm; Scale bars (corn), 9cm; Scale bars (soybean), 6cm.



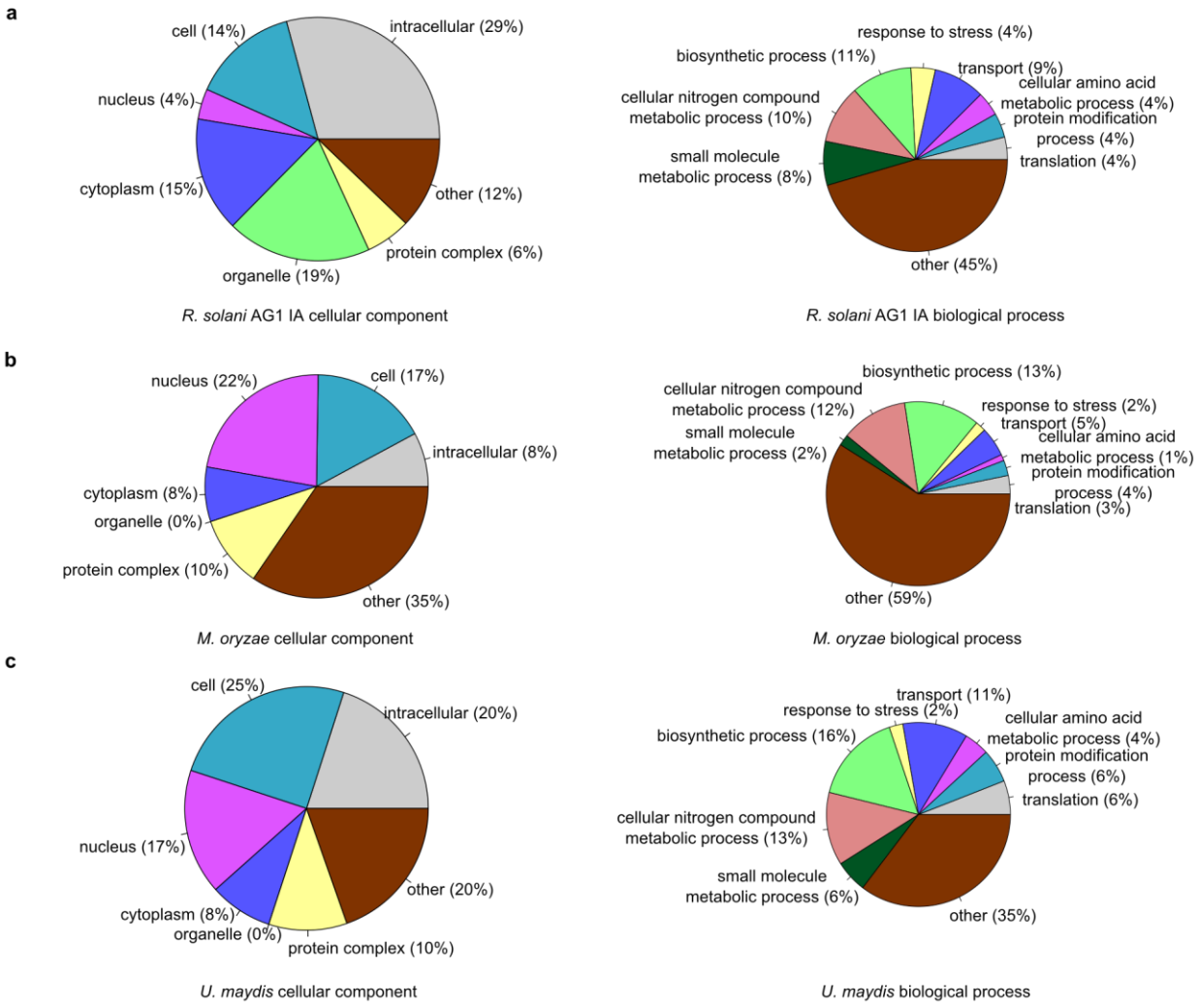

**Supplementary Figure S2. The Map of the *R. solani* AG1 IA mitochondrial genome.**

The inner circle presents the GC content. The ORFs are in green on the middle circle. The ncRNAs are presented as orange blocks on the outer circle. The ORFs are named according to the conserved genes of Swiss-Prot or 'IAMT+num', the names of ncRNA given by Rfam families' id or tRNA type from tRNAscan-SE. ORFs (green), tRNAs, one rRNA and ribozymes (orange) were predicted.



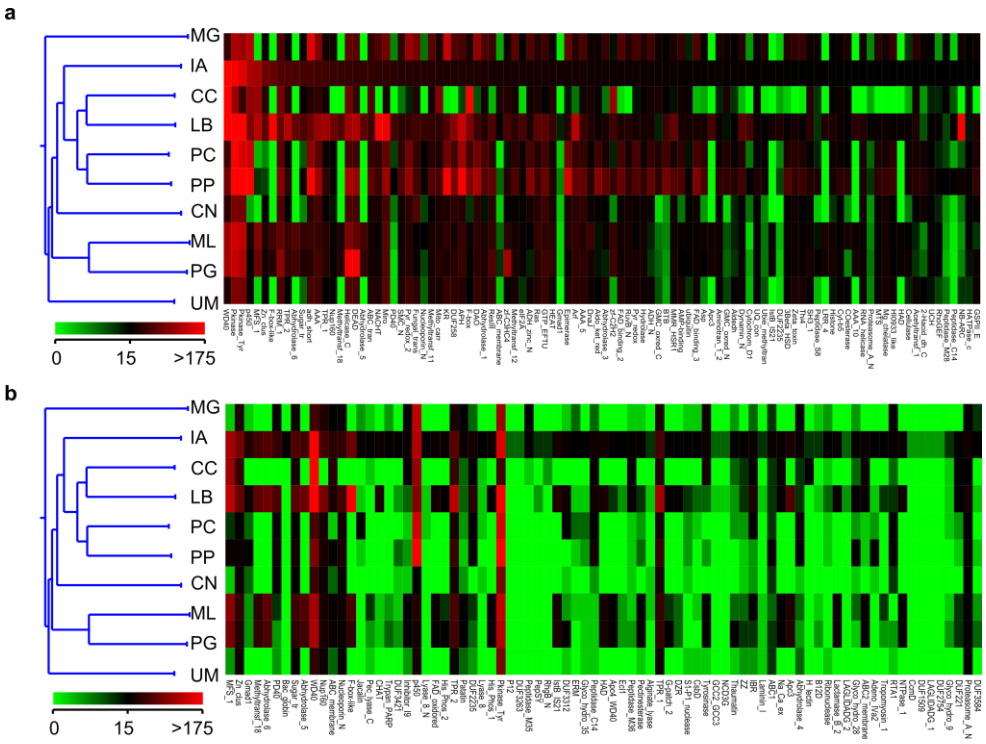

**Supplementary Figure S3. Comparison of fosmid and genome assembly.**

Distribution of the genomic coverage and paired reads coverage of five fosmids from a to j. The sequence depth was calculated by mapping the Illumina Genome Analyzer short reads with an insert size of 173 bp to the genome assembly. For mismatches of the fosmid bases and genome assembly bases, the sequence depth is shown and higher coverage of genome assembly bases is identified. a, b cdsdbxa vs scaffold40; c, d cdsdcxa vs scaffold16; e, f cdsddxa vs scaffold15; g, h cdsdexa vs scaffold7; i, j cdsdgxa vs scaffold7.



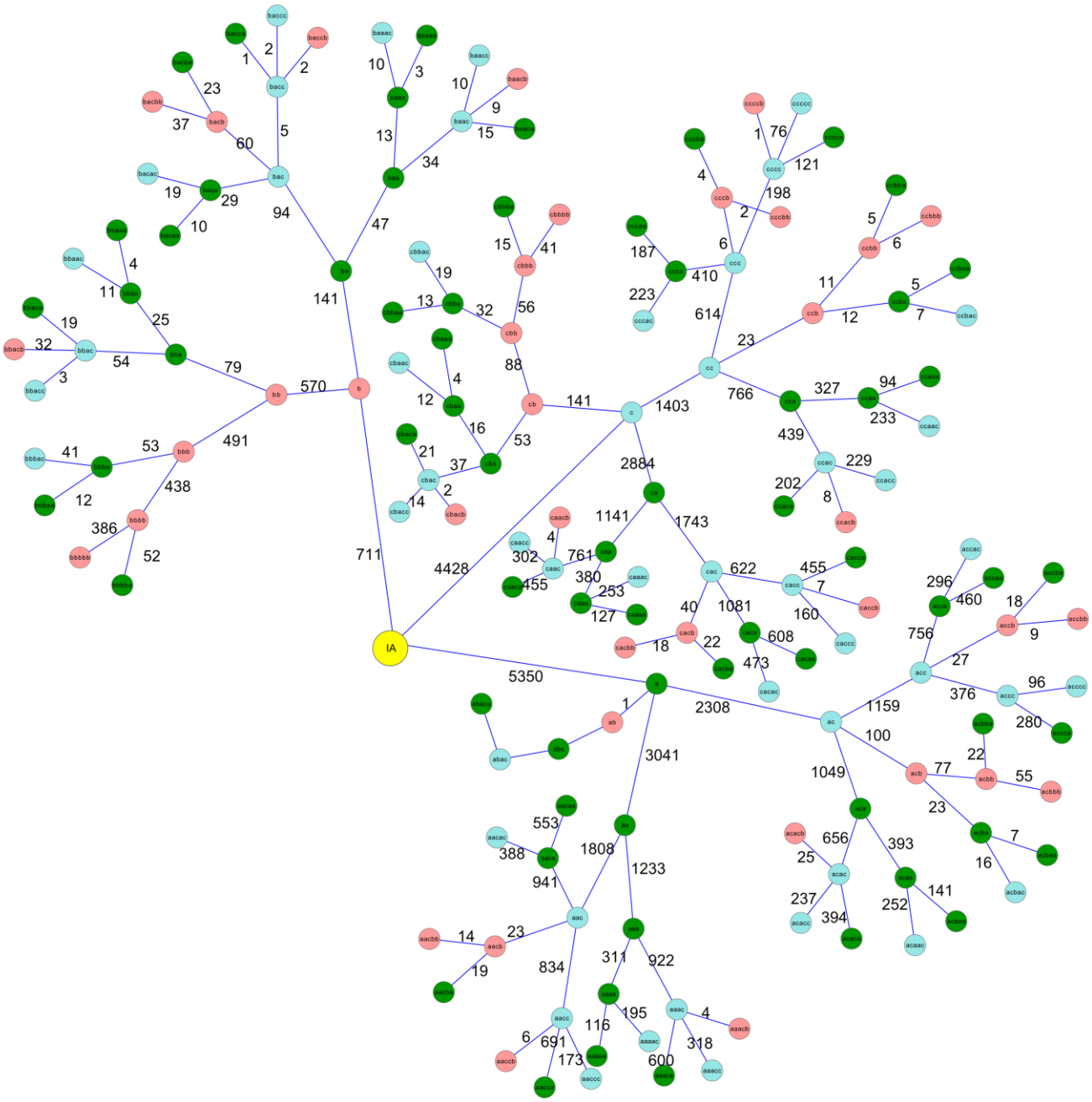

**Supplementary Figure S4. Gene Ontology Slim analysis of *R. solani* AG1 IA.**

Using Gene Ontology Slim, representatives of three cereal pathogens, *R. solani*, *M. oryzae*, and *U. maydis* were compared. The comparison shows different GO values in cellular components and biological processes among the three fungi. *R. solan*i AG1 IA had more genes with products in the cytoplasm (blue), organelle (green) and intracellular (gray) components as well as in small molecule metabolic processes (dark green) of biological progress compared with *M. oryzae* and *U. maydis*.



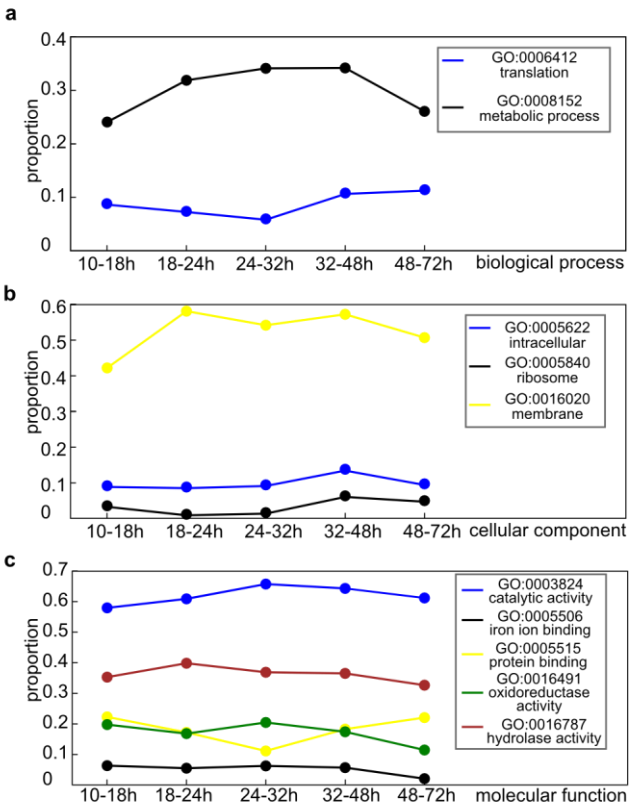

**Supplementary Figure S5. Enriched PFAM families of *R. solani* AG1 IA.**

a Functional comparison of enriched PFAM families of *R. solani* AG1 IA with those of other fungi. The top 100 PFAM domains found in *R. solani* AG1 IA were selected and compared with 9 other fungi. The frequencies in each domain were used to measure the relationship of enrichment (red) and depletion (green). b Enriched PFAM domains involved in the diverse biological functions of *R. solani* AG1 IA were detected in the Basidiomycetes. Abbreviations: IA: *R. solani* AG1 IA, CC: *Coprinus cinereus*, LB: *Laccaria bicolor*, PC: *Phanerochaete chrysosporium*, PP: *Postia placenta*, CN: *Cryptococcus neoformans*, ML: *Melampsora laricis-populina*, PG: *Puccinia graminis*, UM: *Ustilago maydis*, MG: *Magnaporthe oryzae*.




**Supplementary Figure S6. Total number of genes differentially expressed during infection by *R. sonali* AG1 IA at 10 hours, 18 hours, 24 hours, 32 hours, 48 hours and 72 hours after inoculation on rice.**

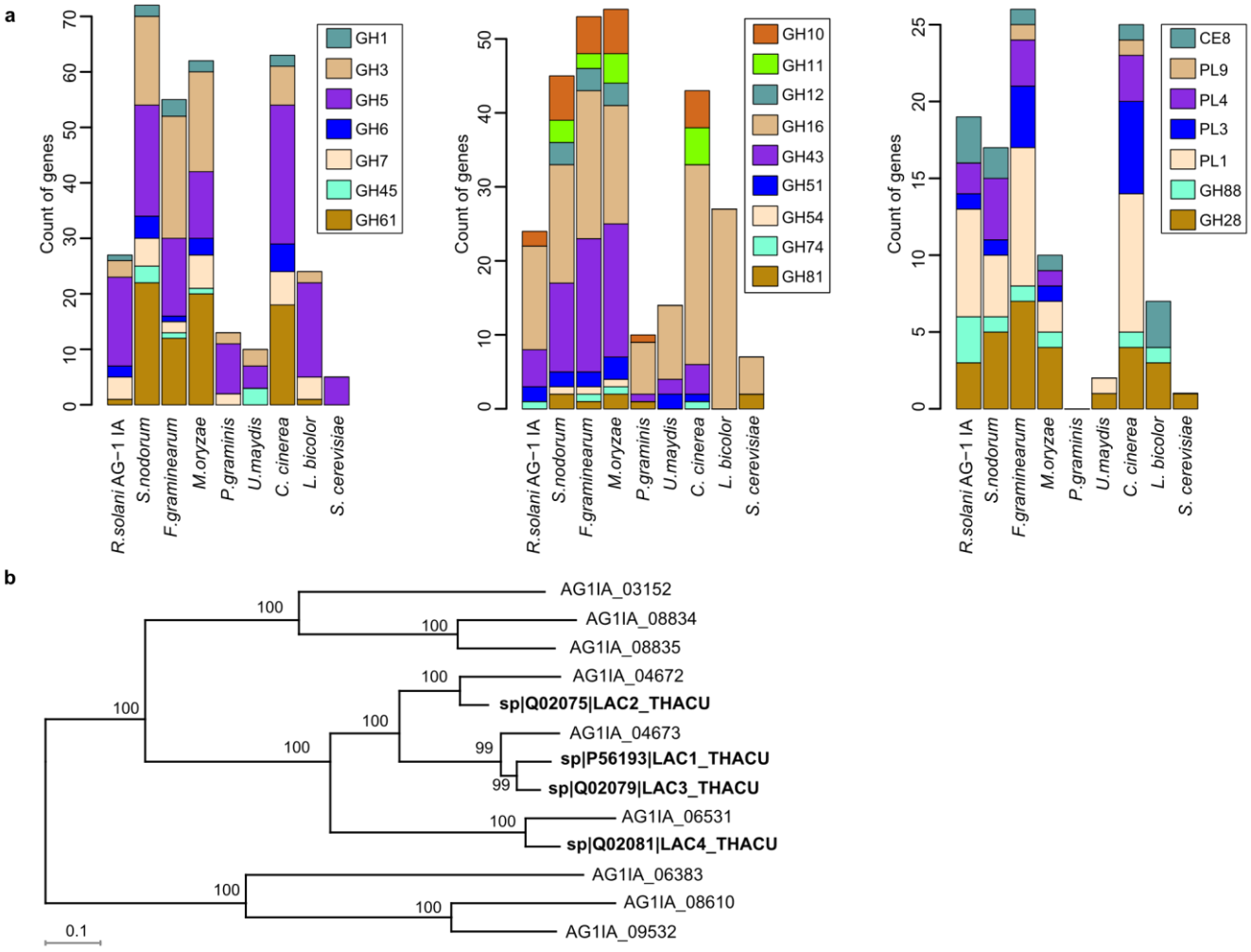
Green circle and 'a' mean up-regulated, bluish and 'c' mean down-regulated, peach and 'b' mean FPKMs did not change, 'aaacc' means up-regulated in four stages from 10h to 32h, and down-regulated in 3 stages from 32h to 72h, and other letters explain similar meanings.




**Supplementary Figure S7. Distribution of genes that were up-regulated during the host infection.**

The up-regulated genes are displayed based on the Gene Ontology (GO) annotation. The proportion was calculated for the up-regulated genes of a GO term/all genes of the GO term.

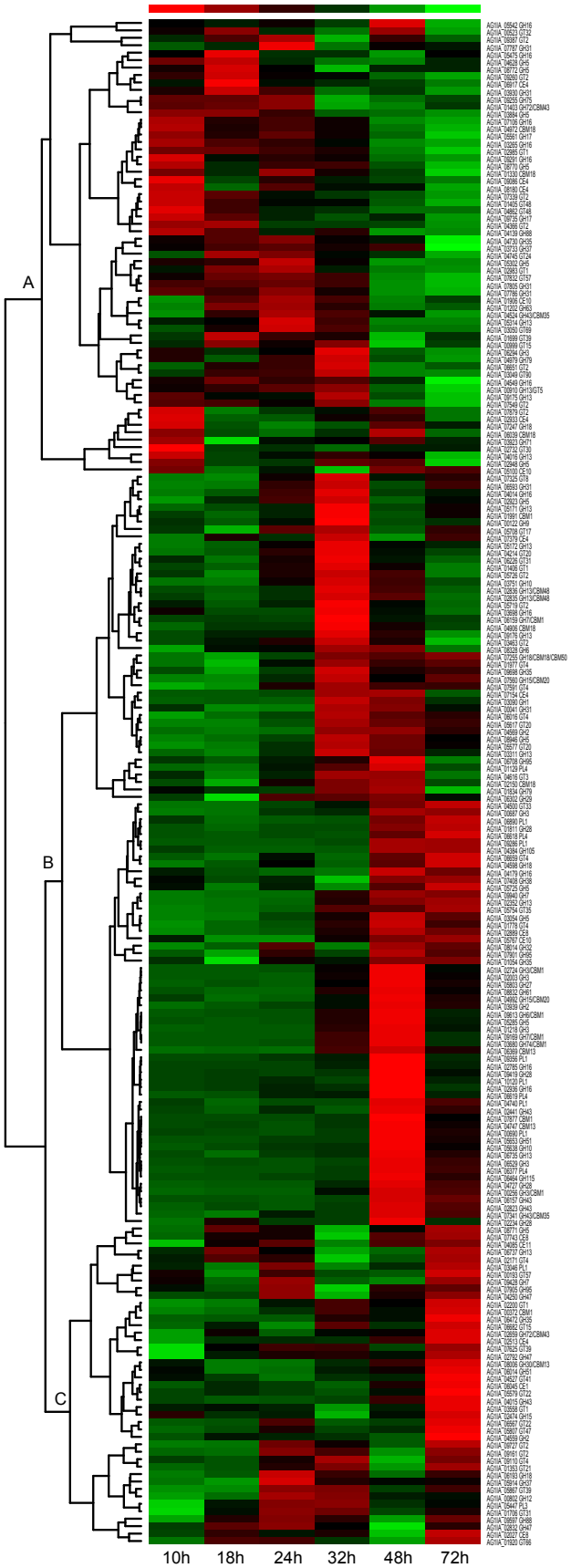
a Distribution based on the GO biological process class. b Distribution based on the GO cellular component class. c Distribution based on the GO molecular function class.




**Supplementary Figure S8. The prediction of cell wall degrading enzymes prediction.**

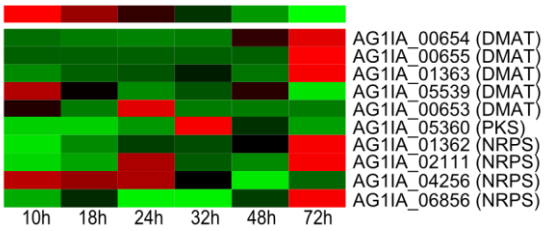
a The distributions of cellulose degrading enzymes (left), hemi-cellulose degrading enzymes (middle) and pectin degrading enzymes (right) in fungi.

b The phylogenetic tree of laccase proteins predicted from *R. solani* AG1 IA. Nine laccase genes were predicted in the *R. solani* AG1 IA genome by alignment to other fungi laccases using BLASTP with an E-value < 1e-50. The protein alignments of nine laccase genes of *R. solani* AG1 IA and four reported laccase genes of *R. solani* AG6 were performed using MUSCLE, and the phylogenetics were analyzed using TREEBEST with the neighbor-joining method and the JTT model.


A
B
C
10h
18h
24h
32h
48h
72h

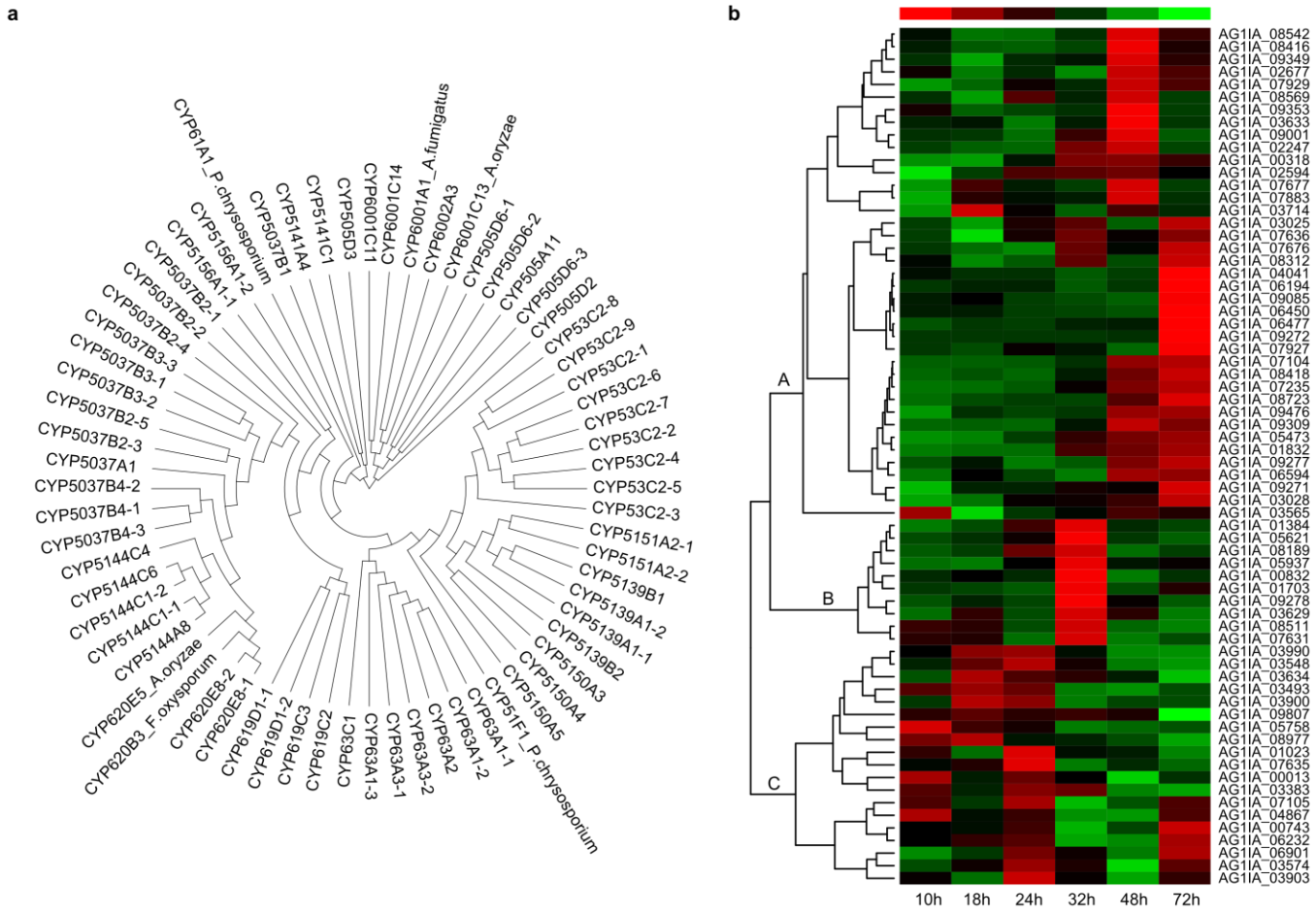
**Supplementary Figure S9. Carbohydrate active enzyme family genes expressed in *R. solani* AG1 IA during the host infection.**

A total of 205 genes showed a specific expression pattern in six infection stages, and these genes were clustered. CAZymes were predicted using BLASTP with an E-value less than 1e-50 in the CAZymes Analysis Toolkit (CAT) database. For the expression values, a Pearson correlation was used for distance calculations, and the average linkage method was chosen for clustering.




**Supplementary Figure S10. The gene expression of secondary metabolite genes clusters of *R. solani* AG1 IA during the host infection.**

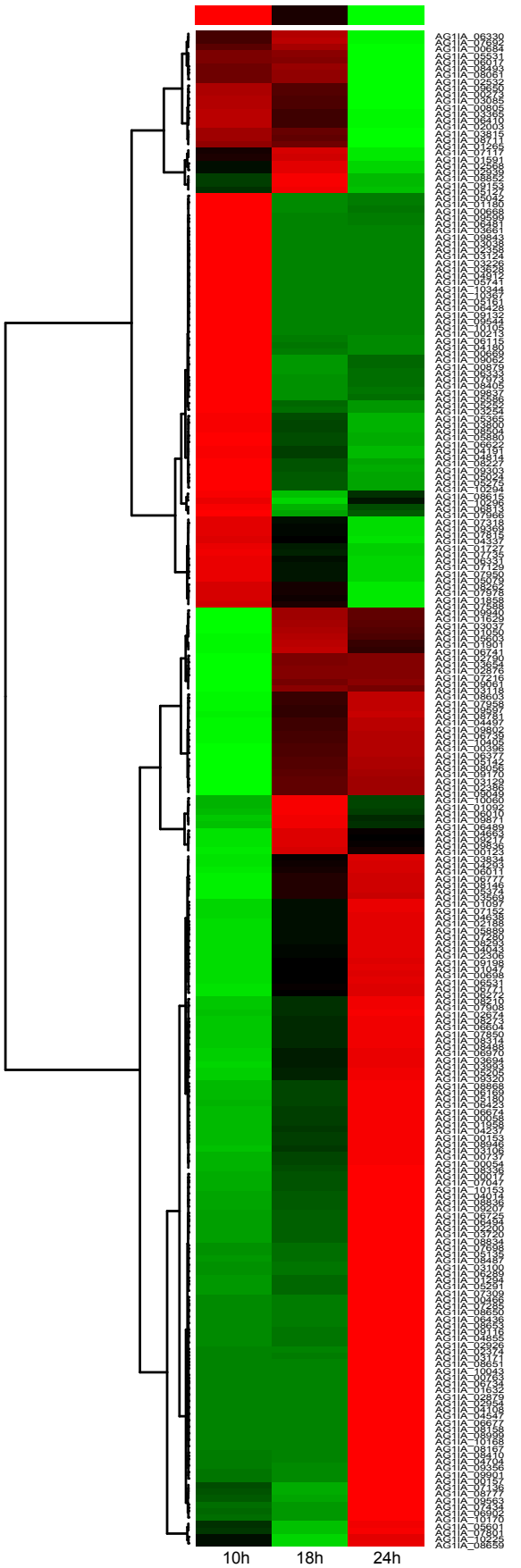
The genes encoding nonribosomal peptide synthases (NRPSs) (4), prenyltransferases (DMATSs) (5), polyketide synthases (PKSs) (1) were predicted using SMURF and and NRPSpredictor.



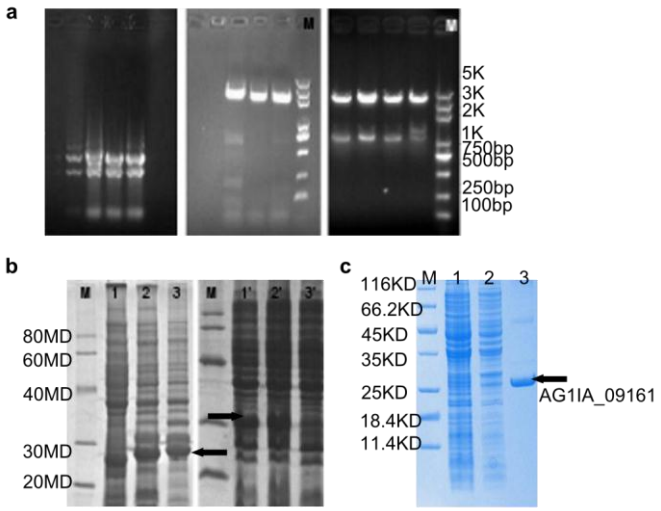

**Supplementary Figure S11. Phylogenetic relationship between the P450 protein sequence and their gene expression during the host infection of *R. solani* AG1 IA.**

a The phylogenetic analysis of P450 families. Putative cytochrome P450 CYPs were identified by assigning proteins of *R. solani* AG1 IA to the Cytochrome P450 Homepage http://drnelson.uthsc.edu/CytochromeP450.html with a BLASTP E-value less than 1e-50. The alignment of proteins was performed using ClustalW v2.1. The protein distance was calculated using the PHYLIP http://evolution.genetics.washington.edu/phylip/ Protdist program with the JTT model. The tree was constructed using the PHYLIP UPGMA program.

b The heatmap of P450 genes: the gene expression patterns of each P450 family member allowed them to be clustered during the host infection stages. The clustering of expressed genes was based on the Pearson correlation and the average linkage method.


10h
18h
24h

**Supplementary Figure S12. Differential expression of selected *R. solani* AG1 IA genes encoding secreted proteins during infection.** The secreted proteins of *R. solani* AG1 IA were analyzed by the prediction algorithms, and they exhibited a two-fold change in expression between 10 hours and 24 hours post inoculation.

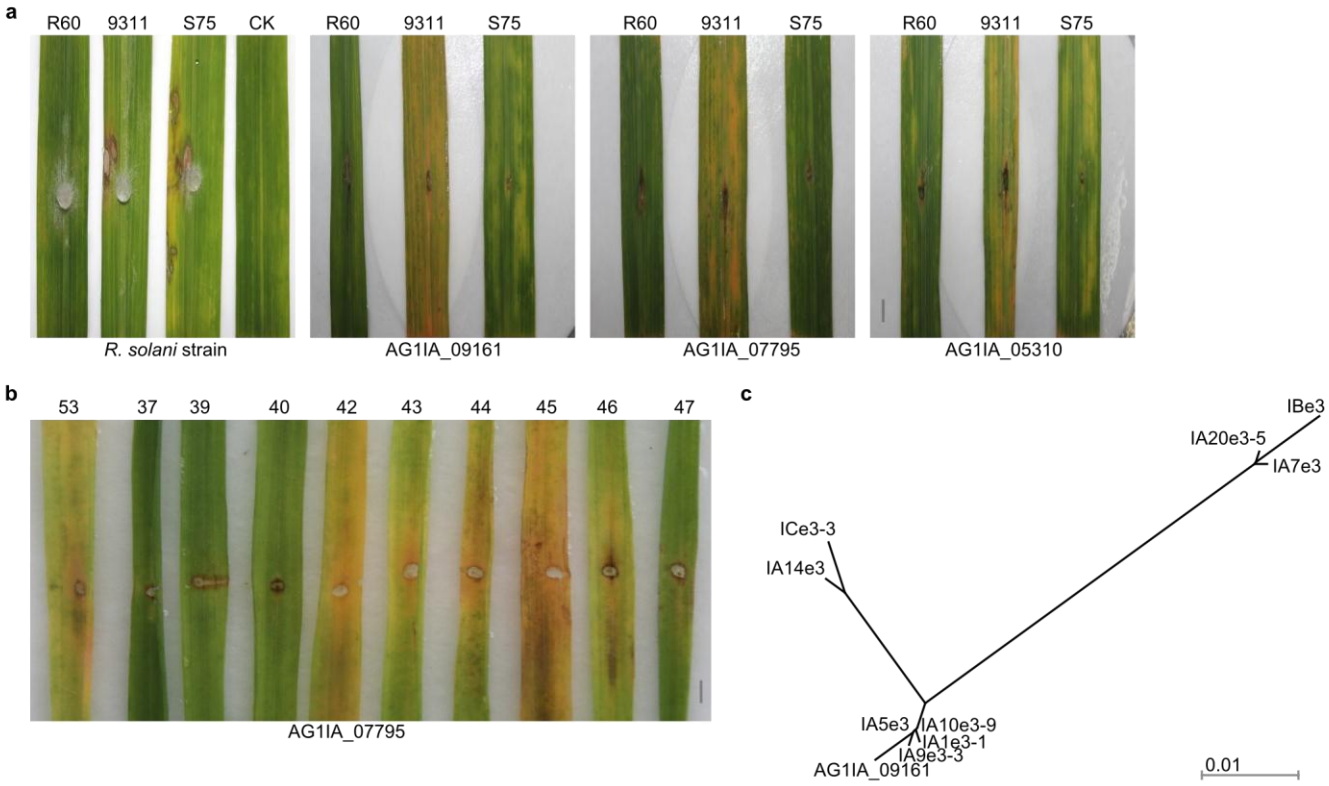



**Supplementary Figure S13. Gene coding for cloning and expression of predicted effectors in *E.coli*.**

a The total mRNA and cDNA of the rice plants and the 1,144 bp genes of AG1IA_07795 when the plasmids were digested by BamHI and NotI (from left to right).

b The fusion proteins encoded by AG1IA_09161 and AG1IA_05310 were separated using SDS-PAGE (arrow). Lane M: Protein Marker; Lane 1: CK1 proteins; Lane 2: CK2 proteins; Lane 3: induced proteins harboring AG1IA_09161with IPTG 0.5Mm/L. Lane1' and Lane2': induced proteins harboring AG1IA_05310 with IPTG 0.5Mm/L; Lane3': non-induced proteins harboring AG1IA_05310 with IPTG 0.5Mm/L.

c AG1IA_09161 expressed protein. M: Protein marker standards; 1: The total expressed proteins of *E.coli* harboring AG1IA_09161 without IPTG induction. 2: The total induced proteins of *E.coli* harboring AG1IA_09161. 3: The purified protein expressed from AG1IA_09161.

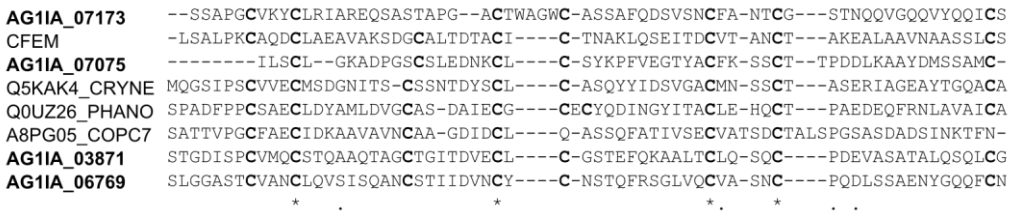



**Supplementary Figure S14. *R. solani* inoculation and infiltration of three effectors and phylogram of AG1IA_09161 coding regions of the isolates and intraspecific groups (ISG).**

a Comparisons of three rice cultivar symptom phenotypes for three candidate effectors. R60, resistant mutant; 9311, moderately sensitive cultivar; S75, sensitive mutant. Type 9311 is sensitive to the three effectors. R60 is sensitive to AG1IA_05310 and is not sensitive to AG1IA_07795 and AG1IA_09161. S75 is sensitive to AG1IA_09161 and AG1IA_05310 and is not sensitive to AG1IA_07795. Scale bars, 2cm.

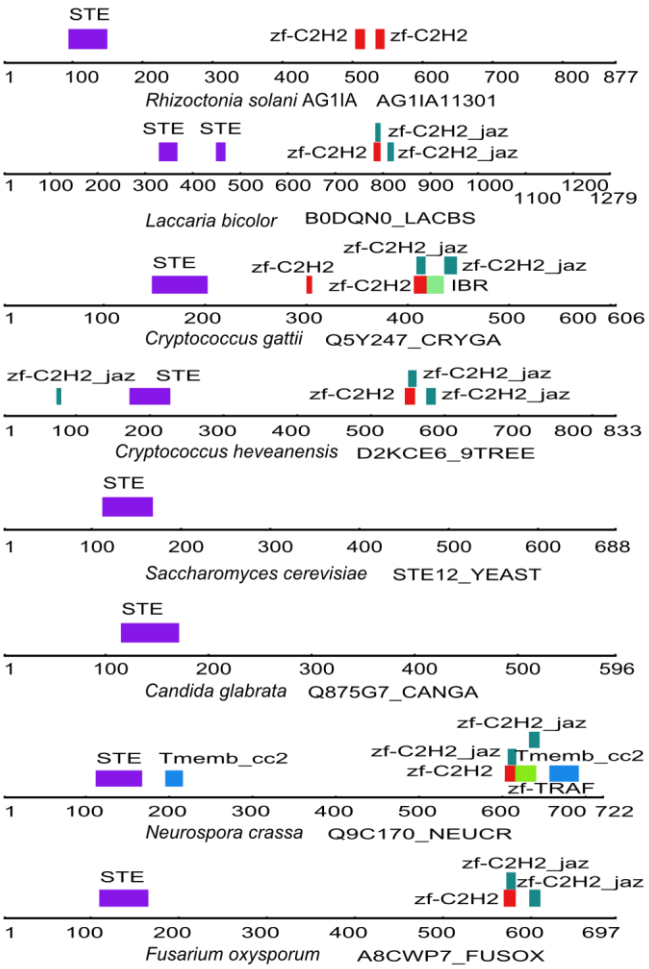
b Rice cultivars showed the different degrees of necrosis in the field for AG1IA_05310 (from the left to right: 53: 88B; 37: Longhuamaohu; 39: Zhonglou 1 hao; 40: Yelicanghua; 42: Chengdu ai 3 hao 43: Sankecun 44: Gongwei 73 45: Jiabala 46: Taishan ruo 47: Guichao 2 hao). Scale bars, 2cm.

c The phylogram of AG1IA_09161 coding regions of eight isolates and two intraspecific groups (ISG). IBe3 and ICe3-3 belong to intraspecific groups: *R. solani* AG1 IB and *R. solani* AG1 IC, respectively; others are the isolates from different regions.




**Supplementary Figure S15. The predicted CFEM domains.**

The AG1IA_07173, AG1IA_07075, AG1IA_03871, and AG1IA_06769 proteins had eight-cysteine-containing CFEM domains (letters in bold).




**Supplementary Figure S16. Domains of STE12 of the MAPK pathway.**

Ste12 in *R. solani* AG1 IA contains two C-terminal C2H2 zinc finger motifs, a Ste domain, and a novel protein structure when compared with *Laccaria bicolor*, *Cryptococcus gattii*, *Cryptococcus heveanensis*, *Saccharomyces cerevisiae*, *Candida glabrata*, *Neurospora crassa* and *Fusarium oxysporum*.

# SUPPLEMENTARY TABLES

**Supplementary Table S1. Summary of the high-quality sequenced data.**

| Insert size (bp) | 173 | 347 | 2200 | 5300 | 5600 |
|---|---|---|---|---|---|
| Reads length | 100, 100 | 75, 75 | 44, 44 | 44, 44 | 44, 44 |
| Rawdata length | 5,409,846,200 | 1,769,797,200 | 1,678,203,824 | 1,585,179,024 | 1,935,934,880 |
| Rawdata cov (X) | 138 | 45 | 43 | 40 | 49 |
| Filter length | 3,950,040,402 | 797,443,221 | 284,163,725 | 215,111,983 | 358,478,120 |

**Supplementary Table S2. Coverage of transcriptome reads of *R. solani* AG1 IA.**
The sum of the alignment reads include BWA and BLAT alignment reads.

| **Samples** | **#All reads (filter)** | **#BWA alignment reads** | **#BLAT alignment reads** | **#Sum of alignment reads** | **Coverage percentage (%)** |
|---|---|---|---|---|---|
| AG1 IA | 12,416,622 | 11,166,638 | 885,231 | 12,051,869 | 97.062 |
| 1-1 [a] | 9,538,188 | 8,636,112 | 657,719 | 9,293,831 | 97.438 |
| 1-2 | 8,941,304 | 8,096,604 | 619,593 | 8,716,197 | 97.482 |
| 1-3 | 9,252,916 | 8,356,449 | 663,441 | 9,019,890 | 97.482 |
| 1-4 | 9,441,970 | 8,339,580 | 666,382 | 9,005,962 | 95.382 |
| 1-5 | 9,276,342 | 8,052,490 | 712,139 | 8,764,629 | 94.484 |
| 1-6 | 8,787,298 | 7,133,795 | 655,027 | 7,788,822 | 88.637 |

**a** 1-1 to 1-6 are, respectively, the fungal mRNAs samples extracted at 10 hours, 18 hours, 24 hours, 32 hours, 48 hours and 72 hours during the rice infection.

**Supplementary Table S3. Statistical analysis of fosmids.**

| **Fosmid ID** | **Fosmid Length (bp)** | **Scaffold Number** | **Contig Number** | **GC content (%)** | **Align to Scaffold** | **Scaffold aligned BeginPos** | **Scaffold aligned EndPos** |
|---|---|---|---|---|---|---|---|
| cdsdaxa | 36,847 | 1 | 1 | 48.26 | scaffold17 | 230,761 | 269,062 |
| cdsdbxa | 31,151 | 1 | 1 | 48.08 | scaffold40 | 106,905 | 138,254 |
| cdsdcxa | 34,218 | 1 | 1 | 47.89 | scaffold16 | 334,400 | 369,131 |
| cdsddxa | 45,481 | 1 | 2 | 47.07 | scaffold15 | 438,486 | 484,471 |
| cdsdexa | 41,599 | 1 | 2 | 47.88 | scaffold7 | 830,866 | 872,448 |
| cdsdgxa | 34,794 | 1 | 1 | 48.66 | scaffold7 | 122,030 | 156,933 |

**Supplementary Table S4. Summary of transposable elements of *R. solani* AG1 IA.**
Abbreviations: LINE, long interspersed elements; LTR, long terminal repeat retrotransposons; SINE, short interspersed elements; RC, rolling circle replication.

| Class | | Family | Count | Length (bp) | Percentage (%) |
|---|---|---|---|---|---|
| Class II | DNA | *Chapaev* | 175 | 30,017 | 0.08665 |
| | | *Chapaev-Chap3* | 33 | 22,261 | 0.06426 |
| | | *En-Spm* | 6 | 1,079 | 0.00311 |
| | | *Maverick* | 1 | 88 | 0.00025 |
| | | *MuDR* | 103 | 11,779 | 0.034 |
| | | *Sola* | 19 | 3,699 | 0.01068 |
| | | *TcMar* | 55 | 10,131 | 0.02925 |
| | | *TcMar-Mariner* | 7 | 979 | 0.00283 |
| | | *TcMar-Tc1* | 65 | 19,016 | 0.0549 |
| | | *TcMar-Tigger* | 1 | 54 | 0. 00016 |
| | | *hAT-Ac* | 25 | 5,971 | 0.01724 |
| | | *hAT-Charlie* | 23 | 7,010 | 0.02024 |
| | | *hAT-Hobo* | 10 | 6,520 | 0.01882 |
| Class I | LINE | *CR1* | 53 | 8,247 | 0.02381 |
| | | *L1* | 332 | 39,196 | 0.11315 |
| | | *L1-Tx1* | 40 | 6,349 | 0.01833 |
| | | *L2* | 89 | 15,989 | 0.04616 |
| | | *Penelope* | 38 | 7,487 | 0.02161 |
| | | *R2* | 143 | 18,048 | 0.05210 |
| | | *RTE-BovB* | 135 | 38,953 | 0.11245 |
| | | *Tad1* | 36 | 8,177 | 0.02361 |
| | LTR | *Copia* | 603 | 208,455 | 0.60178 |
| | | *DIRS* | 125 | 18,483 | 0.05336 |
| | | *ERV* | 1 | 40 | 0.00012 |
| | | *ERV1* | 512 | 103,545 | 0.29892 |
| | | *ERVK* | 73 | 5,105 | 0.01474 |
| | | *Gypsy* | 2,228 | 1,189,261 | 3.43320 |
| | | *Lenti* | 114 | 29,225 | 0.08437 |
| | | *Ngaro* | 31 | 4,128 | 0.01192 |
| | | *Pao* | 11 | 2,409 | 0.00695 |
| | RC | *Helitron* | 1 | 202 | 0.00058 |
| | SINE | *MIR* | 3 | 163 | 0.00047 |
| Unknown | -- | *Unknown* | 44 | 2,573 | 0.00743 |
| Total | -- | -- | 5,135 | 1,824,639 | 5.26743 |

**Supplementary Table S5. Summary of gene family clustering of *R. solani* AG1 IA.**

The genes from the eight sequenced fungal genomes were collected and aligned to each other using BlastP. The pairs and their weights were used as the orthomcl graph for clustering by the MCL algorithm.

| Species | # of genes | # in families | Unique families | Unique genes | Single-gene family | Genes per family |
|---|---|---|---|---|---|---|
| *M. oryzae* | 11,054 | 9,547 | 4,829 | 5,750 | 4,325 | 1.16 |
| *U. maydis* | 6,522 | 6,153 | 1,755 | 1,931 | 1,651 | 1.06 |
| *M. laricis-populina* | 16,399 | 9,670 | 4,430 | 9,638 | 3,117 | 1.70 |
| *P. graminis* | 15,979 | 9,931 | 4,964 | 8,714 | 3,708 | 1.61 |
| *C. gattii* | 6,210 | 5,913 | 1,458 | 1,546 | 1,397 | 1.05 |
| *R. solani* AG1 IA | 10,489 | 9,324 | 4,863 | 5,530 | 4,538 | 1.12 |
| *P. chrysosporium* | 10,048 | 7,886 | 1,966 | 2,864 | 1,592 | 1.27 |
| *P. placenta* | 17,173 | 8,629 | 2,850 | 6,040 | 1,748 | 1.99 |
| *C. cinereus* | 13,342 | 10,467 | 3,586 | 5,209 | 2,952 | 1.27 |
| *L. bicolor* | 23,127 | 13,908 | 6,833 | 13,424 | 4,688 | 1.66 |
| Total | 130,343 | 91,428 | 37,534 | 60,646 | 29,716 | 1.39 |

# in families: family numbers; genes per family, calculated by # of genes / # in families.

**Supplementary Table S6. Characterization of carbohydrate-active enzymes in *R. solani* AG1 IA.**

The codes indicating the enzyme classes are those defined by the CAZyme database (http://www.cazy.org/). GH: glycoside hydrolase, GT: glycosyltransferase, PL: polysaccharide lyase, CE: carbohydrate esterase, CBM: carbohydrate-binding module.

| Fungus | Carbohydrate active enzymes | | | | | |
|---|---|---|---|---|---|---|
| | GH | GT | CE | PL | CBM | Total |
| *M. oryzae* | 257 | 105 | 53 | 5 | 79 | 499 |
| *F. graminearum* | 248 | 98 | 45 | 18 | 56 | 465 |
| *S. nodorum* | 261 | 92 | 48 | 9 | 82 | 492 |
| *R. solani* AG1 IA | 114 | 55 | 15 | 12 | 27 | 223 |
| *P. graminis* | 108 | 69 | 20 | 2 | 26 | 225 |
| *M. laricis-populina* | 126 | 72 | 30 | 3 | 31 | 262 |
| *U. maydis* | 62 | 37 | 16 | 1 | 10 | 126 |
| *L. bicolor* | 129 | 71 | 19 | 2 | 22 | 243 |
| *C. cinerea* | 192 | 70 | 44 | 10 | 86 | 402 |

**Supplementary Table S7. Enriched families of carbohydrate active enzymes of *R. solani* AG1 IA.**

| Family | *R. solani* AG1 IA | *S. nodorum* | *F. graminearum* | *M. oryzae* | *P. graminis* | *M. laricis-populina* | *U. maydis* | *L. bicolor* | *C. cinerea* |
|---|---|---|---|---|---|---|---|---|---|
| CBM35 | 2 | 3 | 2 | 3 | 1 | 0 | 0 | 0 | 0 |
| CBM43 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 |
| CE8 | 3 | 2 | 1 | 1 | 3 | 2 | 1 | 4 | 0 |
| GH13 | 13 | 6 | 8 | 10 | 5 | 7 | 4 | 10 | 9 |
| GH35 | 4 | 4 | 3 | 0 | 1 | 1 | 0 | 1 | 0 |
| GH5 | 12 | 20 | 14 | 12 | 16 | 17 | 4 | 17 | 25 |
| GH7 | 4 | 5 | 2 | 6 | 8 | 8 | 0 | 0 | 6 |
| GH79 | 2 | 1 | 0 | 2 | 1 | 2 | 0 | 8 | 5 |
| GH88 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| GH95 | 3 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 0 |
| GT4 | 7 | 7 | 6 | 5 | 5 | 5 | 3 | 6 | 6 |
| GT48 | 2 | 1 | 1 | 1 | 1 | 6 | 1 | 2 | 2 |
| PL1 | 7 | 4 | 9 | 2 | 2 | 3 | 1 | 0 | 1 |
| PL4 | 4 | 4 | 3 | 1 | 0 | 0 | 0 | 0 | 2 |

**Supplementary Table S8. The different counts and families of up-regulated CAZymes involved in the five infection stages.**

| CAZY | 18h | 24h | 32h | 48h | 72h |
|---|---|---|---|---|---|
| CBM | 5 (14) | 8 (17) | 7 (22) | 6 (15) | 6 (8) |
| CE | 5 (9) | 4 (10) | 3 (6) | 5 (9) | 5 (8) |
| GH | 25 (61) | 34 (80) | 30 (69) | 32 (61) | 23 (40) |
| GT | 18 (33) | 19 (38) | 15 (28) | 10 (14) | 16 (26) |
| PL | 3 (5) | 3 (11) | 3 (8) | 2 (11) | 3 (4) |
| Total | 56 (122) | 68 (156) | 58 (133) | 55 (110) | 53 (86) |

18 h: at the 18-hour stage; 24 h: at the 24-hour stage; 32 h: at the 32-hour stage; 48 h: at the 48-hour stage; 72 h: at the 72-hour stage. The number means gene family, and the number in the bracket means gene count in the different infection stage.

**Supplementary Table S9. The different quantities of up-regulated cellulose, hemi-cellulose and pectin- degrading enzymes involved in the five infection stages.**

| **CAZymes** | | **18h** | **24h** | **32h** | **48h** | **72h** |
|---|---|---|---|---|---|---|
| Cellulose degrading enzymes | GH1 | 1 | 0 | 1 | 0 | 0 |
| | GH3 | 5 | 4 | 7 | 6 | 0 |
| | GH5 | 7 | 8 | 6 | 5 | 6 |
| | GH6 | 2 | 1 | 2 | 1 | 0 |
| | GH7 | 3 | 1 | 3 | 2 | 1 |
| | GH61 | 0 | 1 | 1 | 1 | 0 |
| | GH12 | 1 | 1 | 0 | 0 | 1 |
| | GH74 | 1 | 1 | 1 | 1 | 0 |
| **Total** | | 20 | 17 | **21** | 16 | 8 |
| Hemicellulose degrading enzymes | GH10 | 1 | 2 | 2 | 1 | 0 |
| | GH12 | 1 | 1 | 0 | 0 | 1 |
| | GH16 | 5 | 6 | 4 | 4 | 2 |
| | GH43 | 3 | 4 | 3 | 5 | 1 |
| | GH51 | 1 | 2 | 1 | 2 | 1 |
| | GH74 | 1 | 1 | 1 | 1 | 0 |
| | GH115 | 1 | 1 | 1 | 1 | 0 |
| | GH27 | 1 | 0 | 1 | 1 | 0 |
| | GH29 | 0 | 1 | 1 | 1 | 0 |
| | GH31 | 4 | 5 | 2 | 2 | 1 |
| | GH35 | 2 | 3 | 2 | 1 | 2 |
| **Total** | | 20 | **26** | 18 | 19 | 8 |
| Pectin degrading enzymes | CE8 | 3 | 1 | 2 | 2 | 2 |
| | PL4 | 2 | 4 | 3 | 4 | 1 |
| | PL3 | 1 | 1 | 1 | 0 | 1 |
| | PL1 | 2 | 6 | 4 | 7 | 2 |
| | GH88 | 1 | 1 | 1 | 0 | 2 |
| | GH28 | 3 | 2 | 4 | 4 | 1 |
| | GH43 | 3 | 4 | 3 | 5 | 1 |
| | GH105 | 0 | 0 | 1 | 1 | 0 |
| **Total** | | 15 | 19 | 19 | **23** | 10 |

**Supplementary Table S10. Prediction of the secondary metabolite biosynthesis gene clusters.**

| Gene ID | Gene cluster type | Software |
|---|---|---|
| AG1IA_05360 | PKS-like | SMURF |
| AG1IA_01362 | NRPS-like | SMURF |
| AG1IA_02111 | NRPS-like | SMURF |
| AG1IA_01363 | DMAT | SMURF |
| AG1IA_00653 | DMAT | SMURF |
| AG1IA_00654 | DMAT | SMURF |
| AG1IA_00655 | DMAT | SMURF |
| AG1IA_05539 | DMAT | SMURF |
| AG1IA_04256 | NRPS | NRPS predictor |
| AG1IA_06856 | NRPS | NRPS predictor |

SMURF (http://jcvi.org/smurf/run_smurf.php), NRPS predictor (Christian Rausch, 2005)

**Supplementary Table S11. Comparison of the secondary metabolite biosynthesis gene clusters.**

| Species | PKS | NRPS | DMATS | Hybrid |
|---|---|---|---|---|
| *R. solani* AG1 IA | 1 | 4 | 5 | 0 |
| *S. nodorum* | 9 | 16 | 2 | 0 |
| *F. graminearum* | 15 | 21 | 0 | 1 |
| *M. oryzae* | 15 | 11 | 3 | 3 |
| *U. maydis* | 5 | 9 | 0 | 0 |
| *C. neoformans* | 0 | 1 | 0 | 0 |
| *P. chrysosporium* | 1 | 8 | 0 | 0 |

**Supplementary Table S12. Prediction of secreted proteins of *R. solani* AG1 IA.**

| Species | Total Secreted (Per %) | GPI (Per %) | Small Cysteine-Rich (Per %) |
|---|---|---|---|
| *A. nidulans* | 1,040 (9.85) | 70 (0.66) | 48 (0.45) |
| *F. graminearum* | 1,405 (10.55) | 132 (0.99) | 109 (0.82) |
| *N. crassa* | 893 (9.01) | 91 (0.92) | 71 (0.72) |
| *M. oryzae* | 1,646 (14.89) | 38 (0.34) | 164 (1.48) |
| *R. solani* AG1 IA | 965 (9.20) | 20 (0.19) | 103 (0.98) |
| *C. cinereus* | 1,401 (10.50) | 91 (0.68) | 117 (0.88) |
| *L. bicolor* | 2,062 (8.92) | 82 (0.35) | 282 (1.22) |
| *P. chrysosporium* | 1,010 (10.05) | 63 (0.63) | 63 (0.63) |
| *P. placenta* | 1,606 (9.35) | 47 (0.27) | 41 (0.24) |
| *C. gattii* | 401 (6.46) | 36 (0.58) | 9 (0.14) |
| *M. laricis-populina* | 2,199 (13.41) | 62 (0.38) | 454 (2.77) |
| *P. graminis* | 2,242 (14.03) | 63 (0.39) | 411 (2.57) |
| *U. maydis* | 701 (10.75) | 32 (0.49) | 25 (0.38) |

# The percentage of genes divided by total genes.

**Supplementary Table S13. Sequence analysis, likelihood values and parameter estimates for positive selection of amino acid sites in the candidate effector AG1IA_09161.**

**a All sequences were analysed without signal peptide.**

IBe3 and ICe3-3 belong to intraspecific groups: *R. solani* AG1 IB and *R. solani* AG1 IC, respectively; others are the isolates from different regions. NSI: national standard isolate. N: no testing.

| Gene | Sequenced Isolate | Average disease score (Grade) | Location | CDS length (bp) | CDS GC content (%) |
|---|---|---|---|---|---|
| AG1IA_09161 homologous | IA1e3-1 | 6.69±2.27 | MiYi | 720 | 50.69 |
| | IA3e | 5.67±1.59 | NingNan | – | – |
| | IA5e3 | 2.27±2.41 | DeChang | 720 | 50.56 |
| | IA7e3 | 3.04±3.34 | YanBian | 726 | 50.96 |
| | IA8e3 | 3.40±1.96 | Pengshan | – | – |
| | IA9e3-3 | 5.87±1.98 | RenShou | 720 | 50.69 |
| | IA10e3-9 | 5.30±2.72 | DanLing | 720 | 50.56 |
| | IA11e3 | 3.00±2.68 | QingShen | – | – |
| | IA12e3 | 0.11±0.19 | HongYa | – | – |
| | IA13e3 | 3.07±0.74 | MeiShan | – | – |
| | IA14e3 | 3.98±1.19 | JiaJiang | 705 | 50.78 |
| | IA15e3 | 4.78±1.64 | JingYan | – | – |
| | IA20e3-5 | 4.47±1.82 | FuShun | 726 | 51.10 |
| | AG1IA_09161 | 5.73±1.82 | NSI | 825 | 50.67 |
| | IBe3 | N | NSI | 741 | 51.42 |
| | ICe3-3 | N | NSI | 744 | 50.94 |

**b Parameter estimates and likelihood scores under models of variable ω ratios among sites.**
The LRT statistic for comparing M7 (beta) and M8 (beta&ω) is 2Δl=2*(1752.28-1745.64) =13.28, with P=0.00130 compared with $\chi^2$ the distribution with d.f. =2.

| Gene | Model | lnL | Tree Length | dN/dS | Estimated Parameters | Chi-square test |
|---|---|---|---|---|---|---|
| AG1IA_09161 homologous | M0 | -1781.18 | 1.07339 | 0.1326 | $\omega = 0.1326$ | |
| | M1 | -1751.76 | 1.16462 | 0.1782 | $p_0 = 0.83773$ | |
| | M2 | -1745.60 | 3.31186 | 1.2115 | $p_0 = 0.91930$, $p_1 = 0.00000$, $\omega = 14.47135$ | P=0.00212 |
| | M7 | -1752.28 | 1.17879 | 0.2082 | p = 0.01613, q = 0.05670 | |
| | M8 | -1745.64 | 3.29066 | 1.2033 | $p_0 = 0.91973$, p = 5.06205, q = 99.00000, $p_1 = 0.08027$, $\omega = 14.43785$ | P=0.00130 |

**Supplementary Table S14. Annotation of G protein subunits in *R. solani* AG1 IA.**
*R. solani* AG1IA G protein subunits were identified by assignment to the Swiss-Prot database with an e-value ≤ 6e-14 (five genes less than 1e-50).

| Protein Class | *R. solani* AG1 IA | *U. maydis* | *N. crassa* | *M. Oryzae* | *F. graminearum* |
|---|---|---|---|---|---|
| G alpha subunit | AG1IA_03263 | GPA1 | GNA-1 | MGG_00365 | FGSG_05535 |
| G alpha subunit | AG1IA_02624 | GPA2 | GNA-2 | MGG_04204 | FGSG_09988 |
| G alpha subunit | – | – | GNA-2 | – | – |
| G alpha subunit | AG1IA_03961 | GPA3 | GNA-3 | MGG_01818 | FGSG_09614 |
| G alpha subunit | AG1IA_01027 | GPA4 | – | – | – |
| G alpha subunit | AG1IA_01029 | GPA4 | – | – | – |
| G alpha subunit | AG1IA_05917 | GPA4 | – | – | – |
| G beta subunit | AG1IA_05961 | BPP1 | GNB-1 | MGG_05201 | FGSG_04104 |
| G gamma subunit | AG1IA_00962 | – | GNG-1 | MGG_10193 | FGSG_07235 |
| RACK1 homologue | AG1IA_00847 | UM02408 | CPC-2 | MGG_04719 | FGSG_09870 |

**Supplementary Table S15. Predicted G protein coupled receptors (GPCRs) in *R. solani* AG1 IA.**

| GPCRs Classes | *R. solani* AG1 IA | *N. crassa* | *M. oryzae* | *A. nidulans* | *F. graminearum* |
|---|---|---|---|---|---|
| Pheromone receptors | 3 | 2 | 2 | 2 | 2 |
| cAMP Receptor-like proteins | 2 | 3 | 3 | 2 | 5 |
| Carbon Sensors | − | 1 | 2 | | 1 |
| Stm1-related proteins | 2 | 2 | 2 | 2 | 2 |
| Microbial opsins | − | 2 | 1 | 1 | 3 |
| Homologues of *Magnaporthe grisea* PTH11 | − | 25 | 60 | 70 | 98 |
| Homologues of *Homo sapiens* mPR-like GPCRs | 5 | 2 | 3 | 3 | 2 |
| Rat growth hormone releasing factor-related protein | − | 1 | 1 | 3 | 1 |
| GprK/AtRGS1-like GPCRs | − | 1 | 2 | 1 | 1 |
| FlbA | − | 1 | 1 | 1 | 2 |
| NCD3G | 1 | – | – | – | – |

The GPCR-like proteins were evaluated for 7 transmembrane (7-TM) helices by TMPRED (http://www.ch.embnet.org/software/TMPRED_form.html), Phobius (http://phobius.sbc.su.se/) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/).

## SUPPLEMENTARY NOTES

### 1. Selection of the *R. solani* AG1 IA strain to be sequenced

A classification based on a hyphal anastomosis grouping has been widely accepted over the last 35 years, although the taxonomical status of the different anastomosis groups (AGs) described for *Rhizoctonia* is controversial. 14 (AG1 to AG13 and AGBI) anastomosis groups (AGs) are considered to be an independent evolutionary entity or phylospecies[61] and are divided into different intraspecific groups (ISG) according to the characteristics of its culture medium and hosts, including physiological and biochemical properties. AG1 contains three main intraspecific groups (ISG): *R. solani* AG1 IA, IB and IC.

The sequenced *R. solani* AG1 IA strain was selected from a heavily infected rice plant from South China Agricultural University (SCAU) and was identified as the national standard isolate. This strain was subjected to the hyphal anastomosis test with different AG isolates from different ecological regions[62]. 55 *R. solani* AG1 IA isolates included in this study were isolated on agar plates without additional nutrition[5] from heavily infected rice plants in Sichuan Province. Based on the browning of the colony base, mycelial morphology and lactophenol cotton blue staining properties, they were selected by transferring the tip of a single mycelium. The purified strains were cultured at 28℃ and then identified via the hyphal anastomosis test with the national standard isolate. All of the isolates showed identical internal transcribed spacer (ITS) of ribosomal DNA and different virulence in rice plants[5,62]. We sequenced the national standard *R. solani* AG1 IA strain, which is a multinucleate diploid without the plasmids and hypovirulence. We observed the 13 chromosomes through the cytological observations performed in our lab，by applying a traditional squash technique and modified phenol-magenta staining after culture on potato dextrose agar (PDA) medium for 36 hours (unpublished). A total of 8-10 nuclei were observed using a fluorescence microscope following staining with 4', 6-diamidino-2-phenylindole (DAPI) after culture in PDA medium for 24 hours. It has been reported that the haploid of *R. solani* has 6 chromosomes[63], *R. solani* isolates have at least 11 chromosomes, which range in size from 0.6 to 6 Mb based on pulsed field gel electrophoresis (PFGE) analysis[64].

### 2. Genome sequencing and assembly

We assembled a 36.94Mb draft genome sequence using the high-quality sequenced data (Supplementary Table S1) and SOAPdenovo that was developed specifically for use with next-generation short-read sequences[10]. A total of 880 scaffolds with a length of 1,000 bp or longer constituted more than 96.67% of the assembly

sequences (35,708,106 bp) (Table 1, Figure 1). We detected 43,121 SNPs in the assembly, with 54.94% of the SNPs occurring within genes (including 37.39% located in CDS regions, 6,843 non-synonymous and 9,452 synonymous coding SNPs ) and 45.06% of the SNPs occurring within intergenic regions. We predicted the open reading frames (ORFs) of the *R. solani* AG1 IA genome using a combination of manual and automated annotation[11]. These coding DNA sequences contain 10,489 ORFs with 18,342,233 bp gene sequence lengths (exon sequence 12,900,156 bp). In total, 4,340 ORFs were annotated in the UniProt database, and the domains of 5,610 ORFs were found in the Pfam database; 4,305 genes were assigned to the EuKaryotic Orthologous Groups (KOG) database (E-value cut-off of 1e-5), and 5,258 genes that have homologues in the NCBI RefSeq fungi protein set were assigned using BLASTP (E-value cut-off of 1e-6). In total, 6,156 genes were annotated with integrated annotation from these databases, and, among these genes 257 were assigned to the Pathogen-Host Interaction (PHI) database[65] using BLASTP with an e-value cut-off of 1e-50.

We obtained the mitochondrial genome of *R. solani* AG1 IA, which was identified by aligning the sequence to the NT database using BLAST. The 146 Kb genome sequence was obtained including 29 gaps, which were filled using PCR methodology. It contains the typical genes encoding 12 inner mitochondrial membrane proteins involved in electron transport and coupled oxidative phosphorylation (nad1, 3, 4, 5, 6, cytb, cyb, cox1-3, atp6, apt8), the 5S ribosomal protein, five open reading frames ORFs (IAmt2, IAmt6, IAmt7, IAmt8, IAmt18) of unknown function. The 26 tRNAs can carry all 20 amino acids, among them, tRNA-Thrs (1), tRNA-Args (2), tRNA-mets (3) and tRNA-Undets (2) were predicted. All tRNA secondary structures showed the expected cloverleaf form.

To assess the large-scale and local assembly accuracy of the scaffolds, we aligned mRNA reads from different transcriptomes, including six infection stages, to the assembly and observed high coverage rates of 97.06%, 97.44%, 97.48%, 97.48%, 95.38%, 94.48% and 88.64% (Supplementary Table S2). We randomly selected six fosmids to be sequenced by Sanger technology and the sequences were all aligned to only one genome scaffold. On average, 98.0% of the total fosmid sequences were covered by the assembled contigs (Figure 1c, Supplementary Table S3 and Supplementary Figure S3). Additionally, all known *R. solani* proteins in the UniProt database were aligned to the genome based on BLASTP. We designed 28 primer pairs for the regions based on our assembly. Twenty-eight genomic regions were amplified and Sanger-sequenced from both ends. All of the PCR products had the predicted size, and the sequences showed an average of 95% identity to the genome assembly. Moreover, we used BLASTP searching homologous genes against 246 single-copy orthologous gene families of fungi genomes, 227 families (92.28%) were identified[66].

**3. The *R. solani* AG1 IA repeat content**

Repetitive elements of *R. solani* AG1 IA were identified de novo by identifying sequence elements. DNA transposons (Class I Transposable elements) and retrotransposons (Class II Transposable elements) are abundant in *R. solani* AG1 IA and comprised 1,824,639 bp. Furthermore, tandem repeats of approximately 55,417 bp were identified. In all, repeat elements comprising more than 1,880,056bp were identified, which was approximately 5.43% of the assembly. Except for the enriched *Gypsy LTR* retrotransposons, one percent of the genome could be annotated with four families: *L1*, *RTE-BovB*, *Copia*, and *ERV1*. Two larger DNA transposon families: *Chapaev* and *Chapaev-Chap3* were predicted. The genes associated with TEs in *R. solani* AG1 IA were also detected. Among these predicted genes, 25 annotated proteins from *R. solani* AG1 IA with coding potential for retroviral components and other elements were identified, including three *Tc5* transposase DNA-binding domain genes, one *hAT* family dimerisation domain gene, and five *RNase H* domain genes. The repeats of *R. solani* AG1 IA genome is compared to those of other pathogenic genomes including Ascomycetes and Basidiomycetes. The similarity of repeat contents among the sequenced necrotrophs may exist because of selection against the accumulation of large amounts of DNA sequences that are superfluous to the necrotroph's necessary functions[67.] The Ascomycete *Stagonospora nodorum*, for example, has an estimated genome size of 37 Mb and contains 4.52% repeats[68], while repeats account for 7% of the 42.8 Mb Oomycete *Pythium ultimum* genome[69].

**4. Evolution and comparative genomics**

Comparative analyses of the available Basidiomycete genomes show different gene numbers ranging from less than 7,000 genes in *C. neoformans* and *U. maydis* to more than 10,000 in the sequenced Agaricomycete fungi including *R. solani* AG1 IA[70]. To determine the mechanism underlying the observed gene increases and the evolution of *R. solani* AG1 IA in Basidiomycota, we constructed gene families that were based on sequence similarity using MCL and orthoMCL, and we determined which *R. solani* AG1 IA genes are orthologous or paralogous. We used eight Basidiomycete fungi and *M. oryzae* in *Sordariomycete* for this analysis. Among the ortholog and inparalog pairs, 272,336 ortholog pairs and 38,495 inparalog pairs were derived from orthoMCL. Additionally, 18,206 ortholog groups (100,627 genes) were clustered using the MCL algorithm, which presented approximately 5.5 gene counts per group. A total of 1,661 ortholog groups (19,079 genes) containing at least one gene from each fungus were identified. There were 608 ortholog groups with only a single gene from each fungus and 102 ortholog groups of nine fungi, but no *M. oryzae* groups were found. Among the 102 groups, only

a single gene from each fungus was included in the 37 ortholog groups. The genes in these 102 groups were assigned to 426 KOG groups. The conserved and divergent clusters between Basidiomycetes and Ascomycetes were revealed. On the basis of pairwise protein sequence similarities, we carried out a gene family clustering analysis on all genes in nine sequenced fungi, using *M. oryzae* as an outgroup. The families, unique gene families and single-gene families were compared among fungi (Supplementary Table S5). In eukaryotic genomes, a cornerstone of gene creation is extension of paralogous families by gene duplication[71]. This is reflected in the slow increase of new gene families with genome size, which does correlate with an increase in the size of the families[72]. The relative content of paralogous genes increases with genome size, largely due to the expansion of gene family size in large genomes. We calculated the tree branch length of 608 clusters of the 1:1 orthologs of ten fungi (Figure 2b). The tree branch lengths indicate different evolutionary rates for these ortholog groups. The distribution of branch length values was different from the normal distribution determined by the Shapiro-Wilk test ($p < 0.01$). A peak (111 ortholog groups, 18.26%) was distributed near the tree length of three. The tree length of 40 ortholog groups (6.58%) was estimated to be more than eight.

The Gene Ontology (GO) term analysis was performed for the very important cereal pathogens *U. maydis*, *M. oryzae* and *R. solani* AG1 IA. *R. solani* AG1 IA had unique enriched functional genes, notably an obviously expanded collection of potentially pathogenic GO classifications (e.g., GO: 0004497 protein kinase activity and GO: 0016787 hydrolase activity), and the GO classifications for zinc binding[73], ATP binding[74], and carbohydrate metabolic processes[68-70,75] were also far more abundant in *R. solani* AG1 IA. The gene numbers of the GO classification involved in pathogenesis was most abundant in *M. oryzae*. The three pathogens appeared to be very similar with regard to transcription factor activity[76]. Moreover, using Gene Ontology Slim, which provides an "overview of the ontology content" and summarises the results of the Gene Ontology annotation[17], we described the location of a gene product or its association with cellular components, its activity in biological processes and the molecular functions it performs in these fungal pathogens. The ortholog genes from the three pathogens showed different GO values for cellular components and biological processes.The largest values for genes in the cytoplasm, organelle and intracellular cellular components, and small molecule metabolic processes were detected in *R. solani* AG1 IA. However, the difference in values for the genes in the nucleus and the proteins involved in cellular components, biosynthetic processes and nitrogen metabolism were smaller (Supplementary Figure S4).

In the PFAM classification, the top 100 enriched PFAM domains were depicted for *R. solani* AG1 IA (Supplementary Figure S5a). The PFAM domains of fungal genes were revealed to be strikingly divergent

among the Basidiomycetes. However, the PFAM categories corresponding to Pkinase, Pkinase_Tyr, RRM_1 and RhoGEF were among the most strongly overrepresented in Basidiomycetes. Categories such as Zn_clus (83), Gmad1 (28), ABC_membrane (31), Jacalin (11), Pec_lyase_C (10), Trypan_PARP (8), and Inhibitor_I9 (14) were enriched in *R. solani* AG1 IA, and Bac_globin (10) was absent in other Basidiomycetes and *M. oryzae* (outgroup). In total, 23 PFAM domains involved in different biological functions of *R. solani* AG1 IA were detected to be the most enriched in the Basidiomycetes (Supplementary Figure S5b).

**5. Transcriptome analysis during infection**

*R. solani* AG1 IA is the casual fungus of SBD in rice plants, which penetrates into the rice plants through the inner surface of the leaf sheaths. During the infection process of *R. solani,* adhesion, penetration, colonisation and host reaction occur. *Rhizoctonia* hyphae contact the external surface of a compatible host, and the recognition results in profuse hyphal branching and the formation of infection structures. The initial steps of the infection process are characterised by the adhesion of hyphae and an altered growth pattern, which results in directed hyphal growth and the formation of penetrating hyphal structures. Infection structures including infection cushions and infection pegs that are subsequently formed, allow the fungus to penetrate intact plant tissue[77].

From the 18- to 72-hour infection stages, 5,350, 6,066, 4,369, 4,802 and 5,779 genes were up-regulated, 4,428, 3,711, 5,418, 4,939 and 4,030 genes were down-regulated, respectively (Supplementary Figure S6). The relative numbers of up-regulated genes in different GO classes were compared, the GO categories were ranked by statistical discrimination, and the top classes are shown for each of the biological processes, cellular components, and molecular functions. The enriched genes were associated with certain GO terms (Supplementary Figure S7). The genes for the following GO terms were clearly up-regulated during the host infection process: GO:0006412 translation, GO:0008152 metabolic process in biological processes, GO:0005622 intracellular, GO:0005840 ribosome, GO:0016020 membrane in cellular components, GO:0003824 catalytic activity, GO:0005506 iron ion binding, GO:0005515 protein binding, GO:0016491 oxidoreductase activity, and GO:0016787 hydrolase activity in molecular functions. Among these genes, the genes associated with catalytic activity, protein binding, and oxidoreductase activity showed a tendency for increased gene number. Some uniquely enriched genes within the GO classes showed the same expression tendency.

**6. Genes involved in pathogenicity**

Phytopathogenic fungi follow the different strategies to ensure their nutritions. Necrotrophs first kill their host cells before they can colonize them, killing is brought about either by changes in the cell metabolism after infection, or is due to the action of the toxins or extracellular enzymes produced by the infection pathogen. Necrotrophs have wide host ranges, and are non-obligate, controlled by plant jasmonate- and ethylene-dependent defence pathways[78,79]. Necrotrophic plant pathogenic fungi uses a large array of cell wall-degrading enzymes for pathogenicity[31]. Several recent studies have demonstrated a strong relationship between the repertoire of carbohydrate active enzymes in fungal genomes and their saprophytic lifestyle[18].

To study the enzymes that build and break down complex carbohydrates and glycoconjugates, we compared the distribution of CAZymes, glycoside hydrolases (GH), and polysaccharide lyases (PL) as well as the hemi-cellulous and cellulous repertoires of the other phytopathogens. Totally, the predicted CAZyme families including cell wall degrading enzymes, which are important for determining the ability of pathogens to colonise the host rice, were reduced in *R. solani* AG1 IA compared with the other pathogens. However, the especially abundant families playing specific roles in the necrotrophic life cycle were distributed such as laccase, pectinase and xylanase. In particular, laccase degrades lignin during the fungal infection progress. Four different laccases were detected in the *R. solani* AG6 strain[80]. The protein alignment and clustering of the nine laccase genes of *R. solani* AG1 IA and four reported laccase genes of *R. solani* AG6 were done. Three groups were clustered from the phylogenetic tree. Three laccase genes of *R. solani* AG1 IA were grouped with the four laccase genes of *R. solani* AG6, and the other six were clustered into the other two groups. Thus, in the *R. solani* AG1 IA genome, the laccases were more abundant and diverse than those reported for the *Rhizoctonia* genus. The absence of recognisable cutinases (CAZyme family CE5) suggests that these enzymes are not critical for penetration and infection by *R. solani* AG1 IA, which is supported by the fact that the plant cuticle generally remains intact after penetration, except in the areas of hyphal entry[69].

The expression patterns of the transcripts putatively encoding CAZymes and the cell-wall-degrading enzymes were determined during the infection process; Pearson correlation was used for the distance calculation, and the average linkage method was chosen for clustering. In the constructed clustering tree, the expressed genes during the six infection stages in the rice plant were mainly clustered into three major groups. Group A showed increased expression at 10, 18 and 24 hours after inoculation during the early infection stages, and Group B showed the highest expression, mainly at the 32-, 48- and 72-hour timepoints during the late infection stages. The last group was expressed constantly from 18 hours to 72 hours but had increased expression at 72 hours. For the CAZyme classes and family members, transcript analyses revealed large numbers of differentially regulated

genes. The GH family members involved in infection process showed the trend towards a peak at the 24-hour stage; and transcripts for the enzyme classes also showed a peak at the 24- hour stage for GH. For PL family members, the two peaks appeared at the 24- and 48-hour stages and the enzyme classes almost didn't change excluding at the 48-hour stage. In contrast, before and after 24 hours, transcripts for the total enzyme classes increased and decreased gradually. During the entire infection progress, the CBM, CE, and GT family members and classes showed a weak increasing or decreasing trend, excluding the rapid decrease for GT at 48 hours. Each GH family member was differentially expressed during the different stages: GH13 and GH5 were over-represented at 18 hours; GH13, GH5, and GH16 were over-represented at 24 and 32 hours; GH3, GH5, and GH43 were over-represented at 48 hours; and GH5 and GH13 were over-represented at 72 hours. Finally, at 72 hours, during the late infection phase, any expressed enzyme transcripts were present at low levels.

Filamentous fungi are well-known producers of secondary metabolites, which in nature fulfil various functions and most likely allow niche exploitation. Fungal pathogens generally produce an array of secondary metabolites, some of which are involved in pathogenesis[19]. Secondary metabolites are often bioactive, usually of low molecular weight, and are produced as families of related compounds at restricted parts of the life cycle, with production often correlated with a specific stage of morphological differentiation. Secondary metabolites share the enigmatic properties of cellular dispensability and restricted taxonomic distribution. 1,500 compounds from fungal metabolites were isolated and characterized thus far, which showed that more than half of these molecules had antibacterial, antifungal or antitumour activity[19]. Lots of mycotoxins from fungal pathogens have also been found and reported[19,21].

The symptoms of the rice sheath blight disease suggest possible involvement of phytotoxin in lesion symptom development. Previous research has shown that phenyl acetic acid PAA, derivatives of PAA, a phenolic compound, and a carbohydrate acted as phytotoxins of *R. solani*. PAA has been the most widely described in the literature and is known as a virulence factor for AG3 on potato[81,82]. PAA is a catabolic product of the aromatic amino acid phenylalanine, whose precursor is chorismic acid, which is the end-product of the shikimate pathway[82]. We found the 5,655 bp key AROM gene (AG1IA_04890), which participates in the synthesis of PAA. The transcript of the AROM gene increased at 32 hours during infection, which suggests an association with pathogenesis in *R. solani*.

The genes involved in a given fungal secondary metabolite biosynthetic pathway are often clustered and located at the same locus in the genome; they are also usually near telomeres and are co-expressed, thereby defining secondary metabolism gene clusters[83]. The production of secondary metabolites is catalysed by one of

five proteins, referred to here as "backbone" enzymes, which include nonribosomal peptide synthases (NRPSs), polyketide synthases (PKSs), hybrid NRPS-PKS enzymes, prenyltransferases (DMATSs), and terpene cyclases (TCs)[83]. We predicted ten secondary metabolite genes in total. The enrichment for DMATs most likely reflects the prenyltransferase-mediated addition of a dimethylallyl side chain to a variety of compounds that are associated with the production of indole alkaloids[83]. We hypothesize that *R. solani* AG1 IA may be enriched for indole alkaloid secondary metabolite classes, which have a substantial role in the parasite life cycle. These predicted encoding genes are specifically up-regulated during the infection-related development indeed (Supplementary Figure S10). Actually, secondary metabolite synthase genes are very unevenly distributed among fungi[84]. This suggests that most genes encoding these biosynthesis enzymes of fungi are relatively young in evolutionary terms and have been subject to rapid gene gain and loss. This has been attributed to diversifying selection driven by a chemical arms race between fungi and the predators, competitors, and hosts[85]. Therefore, we could come to a conclusion that the fewer key numbers and the expression profile of secondary metabolite synthase genes are consistent with the unique requirements of *R. solani* AG1 IA for adapting to diverse environments.

P450 genes are important in the metabolism of a wide variety of substrates, including endogenous chemicals, such as steroids and other small molecules[15]. Often, CYPs are involved in pathogenesis and the production of toxins and virulence factors or the utilisation of specific carbon sources[24,86]. It is reported that *Neurospora* has 41 P450s, *Aspergillus oryzae* has 152[24], and only two P450s have established roles in housekeeping functions for sterol biosynthesis (CYP51 and CYP6); the others are enzymes with specific purposes. In total, 68 putative cytochrome P450 CYPs were identified in the *R. solani* AG1 IA genome using BLASTP, with an e-value cutoff of 1e-50 in the fungal cytochrome P450 database[87]. Sixteen families were obtained, including the most enriched proteins of the CYP5037 family (13) with unknown functions. However, no members of the CYP51 and CYP6 families that are involved in sterol biosynthesis were identified, and this absence was thought to be the consequence of gene loss during evolution. A phylogenetic tree that mainly contained two large groups was constructed to include all of the families (Supplementary Figure S11a). Furthermore, the gene expression patterns of each P450 family member indicated that they could be clustered into three major clades during the host infection on the basis of their basal expression levels (Supplementary Figure S11b). In general, three main categories of expression patterns can be identified. Categories A and B contain P450 genes that were silent or had relatively low expressions during the early infection stages and showed high expression at the 32-, 48- and 72-hour stages. Category C showed constant expression at 10, 18 and 24 hours, which most likely entails

constitutive, endogenous physiological processes. This scenario suggests the involvement of these genes in important, stage-specific endogenous cellular processes. Examining the P450 biological functions, which are partly associated with the secondary metabolites during the host infection, we deduce that the highly expressed P450 genes in the different infection stages likely play a role in the metabolism and biotransformation of some toxic chemicals that attack the hosts. The specific chemical substances involved are currently under investigation.

Moreover, as another important type of pathogenicity-associated factor, transporters can export toxic molecules to the surrounding environment[88]. For toxin-producing fungi in particular, transporters are critical in the efflux of endogenously produced molecules. Sets of ABC transport proteins in the genomes of Basidiomycetes and Ascomycetes (especially members of the subphylum Pezizomycotina) share substantial commonality. However, the number of members of the specific subfamilies in the genomes of Basidiomycetes tends to be lower than that in Ascomycetes[25]. In *R. solani* AG1 IA, 187 transporters and 48 ABC transporters are predicted. The total number of predicted transporter genes was fewer than that in other selected fungi, including *C. cinereus* (268), *L. bicolor* (317), *U. maydis* (271) and *M. oryzae* (399), but it was similar to those of *A. nidulans* FGSC A4 (48) and *M. oryzae* 70-15 (48)[25]. Indeed, *R. solani* AG1 IA had the highest number of ABC genes per 1 Mb of genome among the Basidiomycete species. The ABC-A (2), ABC-B (7), ABC-C (14) and ABC-G (9) subfamily members were substantially enriched, the ABC- D (2) and ABC-E (1) subfamily members were common, and fewer ABC- F (3) subfamily members were present in the genome as the result of multiple loss events during fungal evolution. Moreover, an additional ten non-classified ABC proteins were predicted. The *M. oryzae* ABC-A transporter, Abc4 protein (MGG_00937) is required for pathogenesis and appressoria formation[25]. Therefore, we deduce that the ABC-A transporter of *R. solani* AG1 IA that is absent in most Basidiomycetes is likely linked with a unique parasitic lifestyle, which is supported by the up-regulation of the transcript during infection.

**7. The *R. solani* AG1 IA secretome**

We identified 965 potentially secreted proteins (9.17% of the proteome) in *R. solani* AG1 IA. The 518 secreted proteins annotated in PFAM can be clustered into 140 families and 444 singletons, with the largest family containing 20 members. An analysis of the secretome for pathogenicity-related proteins revealed notable families of secreted proteins, including secreted E3 ubiquitin ligases[34] (candidate effectors), protease inhibitors, cell wall-degrading enzymes, lipases, phospholipases, highly expanded proteases and cytochrome P450s, and proteins

with unknown functions. 103 small cysteine-rich proteins, which were as abundant as predicted in *F. graminearum*, were identified among the predicted secreted proteins that are less than 200 aa in length and contain at least 4% cysteine residues. These proteins are considered to be potential plant effectors that play a role in pathogenesis[30]. Indeed, small cysteine-rich proteins produced by fungi may be of ecological relevance by functioning as toxins or inhibitors of hydrolytic enzymes or by suppressing the defence responses of host plants and triggering disease resistance. Almost all avirulence factors of fungi identified to date are small, cysteine-rich proteins[89,90]. Although their biological functions are not known, we postulate that necrotrophs have an abundant inventory of secreted proteins, in particular, pathogenicity-related proteins such as proteases, glucosyl hydrolases, and exclusive potential effectors.

**8. The *R. solani* AG1 IA candidate effectors and their validation**

Effectors are emerging as the prime weapons of plant parasites and also as targets for host recognition and immunity. Many biotrophic fungal and oomycete plant pathogens deliver effector proteins directly into host cells during infection. Recent advances are revealing the extensive effector repertoires of these pathogens and their functions manipulating host cells to establish a parasitic relationship[91]. So many plant pathologists have searched formolecular and biochemical evidences of the gene-for-gene hypothesis proposed by Flor. In response to pathogen attacks, plants have evolved one of the important line of active defences, which is based on effector perception by R proteins and subsequent activation of effector-triggered immu-nity (ETI), leading to rapid and enhanced defence responses in plants, including HR[92]. However, it has been reported that necrotrophic pathogens do not comply with the gene-for-gene model. Actually, the host-specific toxins (HST) from necrotrophic pathogens, where sensitivity is restricted to specific genotypes of the host, which share all the characteristics of avirulence gene products. They have an original, primary virulence function, which involves both intracellular and extracellular targets; and they are specifically recognised by structurally diverse host resistance genes that mediate delivery to the target, undergo accelerated evolution[92,93]. These proteinaceous effectors from necrotrophic fungal pathogens including *Stagonospora nodorum*, *Pyrenophora tritici-repentis*, *Botrytis cinerea*, *Corynespora cassiicola*, *Leptosphaeria maculans*, *Mycosphaerella graminicola* and *Pyrenophora teres*, were reported[31]. The origin, evolution and biological role of proteinaceous HST effectors, have been studied, including ToxA, Tox1, Tox2, ToxB, ToxC and SnTox3 from *S. nodorum* and *Pyrenophora tritici-repentis*[31,32,90,94]. It has already become clear that the receptors of necrotrophic effectors operate as dominant disease susceptibility genes[32]. The identification of these HST-host gene interactions have indicated

the evidences for the mitochondria and chloroplast to serve as this toxin-based, inverse gene-for-gene model[31,95]. The pathogenicity of *R. solani* AG1 IA has been studied. However, no similar effector has been reported before our analysis. The genome of *R. solani* AG1 IA has large complex families of potential effector genes that encode secreted proteins that may be implicated in pathogenesis.

During the host infection stages, the expression of secreted family members was differently regulated, and some of the up-regulated genes could act as potential effectors in pathogens. To explore the transcriptional responses to rice infection, we constructed a secreted protein expression pattern based on the *R. solani* AG1 IA genome annotation. The gene expression during rice infection was monitored using samples from infected rice at 24 hours post-inoculation. In all, 145 genes were up-regulated by at least two-fold during infection relative to the initial post-infection period (Supplementary Figure S12). To identify the potential effectors, we selected 45 candidate genes that had shown two-fold or greater up-regulation from the transcriptome data. Among them, three classes of secreted effectors, AG1IA_09161 (glycosyltransferase GT family 2 domain), AG1IA_05310 (cytochrome C oxidase assembly protein CtaG/cox11 domain), and AG1IA_07795 (peptidase inhibitor I9 domain) were identified.

The glycosyltransferase GT-domain effectors were previously reported in the bacterium *Clostridium difficile* during a host-bacterium pathogen infection but have not been reported in fungi[35]. In *C. difficile*, the effectors belong to clostridial cytotoxins, which are groups in the carbohydrate-active enzyme (CAZY) family GT44. This family now has more than 30 members, including putative glycosyltransferases from other bacterial pathogens. The pathogenic mechanism triggering the host HR is clear in bacteria, but we cannot deduce the exact mechanism in fungi at present. A role of inhibitor domain effectors in plant defence has been previously reported in phytopathogens[33]. The inhibitor effectors that target host plants include the following: GIP1 (glucanase inhibitor protein-1) from *P. sojae*, which inhibits soybean *endo*-b-1,3-glucanase-A; Avr2 from *Cladosporium fulvum*, which inhibits Rcr-3 and PIP1, two closely related cysteine proteases of tomato; EPI1 and EPI10 (extracellular protease inhibitors), which target a tomato serine protease P69B; and EPIC1 and EPIC2B (extracellular cysteine protease inhibitors), which target papain-like cysteine proteases from *P. infestans*[33]. However, a peptidase inhibitor I9 domain effector that targets host plants has not been reported to date. In addition, the CtaG/cox11 domain effector is reported for the first time in fungi. Our findings reveal the possibility that there are proteins active in host cells and are delivered by the fungus to trigger a defense response, although the delivery mechanisms for all of these effectors require further investigation.

**9. G-protein signalling pathway**

In fungi, conserved signal transduction pathways control fundamental aspects of growth, development and reproduction[37]. The important classes of fungal signalling pathways are mitogen-activated protein kinase (MAPK) cascades, the calcium calcineurin pathway and cyclic AMP (cAMP) signalling. Core elements of these signalling pathways are required for virulence in a wide array of fungal pathogens of plants and mammals[96]. Understanding the molecular details of these pathways could provide opportunities for disease control. G-protein subunits and G-protein-coupled receptors are primarily responsible for transducing extracellular signals into intracellular responses that involve complex intracellular signalling networks involved in regulating morphogenesis, mating, infection and virulence in fungi[96]. The superfamily of G-protein-coupled receptors (GPCRs) is one of the largest and most diverse receptor families. These proteins have in common a seven α-helical transmembrane region (7TM) that anchors the receptor to the plasma membrane of the cell, with the N termini exposed to the extracellular space. *R. solani* AG1IA G-protein subunits were identified through assignment to the Swiss-Prot database based on an e-value of less than 6e-14 (five genes less than 1e-50). Each G protein heterotrimer is composed of α, β, and γ subunits that are associated with the plasma membrane. In contrast to yeast, which contain two Gα proteins, most characterised filamentous fungi possess three Gα proteins that are members of distinct groups[96]. Group I Gα, which was identified in filamentous fungi, was *N. crassa* GNA-1, which is 55% identical to mammalian Gαi superfamily proteins. Most of the characterised filamentous fungi possess a single Group I Gα protein[96]. The *R. solani* AG1 IA genome contains Group I Gα protein (AG1IA_03263), as well as one Group Ⅱ and one Group Ⅲ Gα protein. Notably, three novel Group IV Gpa4 (AG1IA_01029, e-value 3e-29, AG1IA_05917, e-value 6e-14 and AG1IA_01027, e-value 2e-27) subunits homologous to the one Gpa4 subunit reported in *U. maydis*[38] were predicted. These subunits are significantly less related to the other subunits, have no homolog with other fungi and independently grouped with the GPA4 from *U. maydis*. In *R. solani*, the disruption of one α subunit result in decreased vegetative growth and pathogenicity[97]. However, the other subunits functions, especially the enriched novel Group IV, require further study.

The CFEM domain is related to pathogenicity in fungi[98], and GPCRs containing the CFEM domain were analysed. CFEM- containing proteins with eight cysteine residues could function as cell-surface receptors, signal transducers or as adhesion molecules in host-pathogen interactions. In the entire genome, no GPCR protein with a specific extracellular membrane spanning the CFEM domain[15] was detected. This type is very abundant and has special pathogenic functions in *M. oryzae* and *F. graminearum*. In contrast, 60 CFEM domain GPCRs were

detected in *M. oryzae*, 25 in *N. crassa* and 70 in *A. nidulans*. This type of GPCR was actually absent in the Basidiomycete fungi, including *C. neoformans*, *U. maydis* and *P. chrysosporium*[15]. However, four proteins that contained CFEM domains were found based on Pfam data, and the four CFEM domains were compared with those of other fungi. The NCD3G domain appears to be restricted to animals (from *Caenorhabditis briggsae* to *Homo sapiens*)[99] although some other glutamate receptor family members were found in *Chytridiomycota* and *Zygomycota*. This nine-cysteine GPCR domain was first detected in fungi in the present study. The conserved cysteine residues have the potential to form disulphide bridges and could be essential for the function of NCD3G. The highly conserved glycine residues in the NCD3G domain are important for the bioactivity of human calcium-sensing receptors (CaSR)[99]. For *R. solani* AG1 IA, the novel NCD3G domain GPCR member likely originates from the glutamate receptor family in Metazoa and has a novel biological function.

Among the core elements, Ste12 can also control fungal virulence downstream of the pathogenic MAPK cascade as a master regulator of invasive growth in plant pathogenic fungi[100]. A single Ste12 ortholog was detected in all of the fungal species examined previously, except for *S. pombe* and *U. maydis*[37]. Ste12 was detected in *R. solani* AG1 IA and contains two C-terminal C2H2 zinc finger motifs, a Ste domain, and a novel protein structure found when comparing it with Ste12 in other fungi (Supplementary Figure S16). In *M. oryzae*, the Ste like region and the zinc finger region of Mst12 were both required for invasive growth and virulence[41]. The function of the novel protein structure in *R. solani* AG1 IA is not clear, although it was up-regulated during the infection phase.

**10. Mating system**

Basidiomycete fungi have evolved a unique mating system, termed tetrapolar or bifactorial incompatibility, in which mating type is determined by two unlinked loci, compatibility at both loci is required formating to occur. The ancestor of the Basidiomycetes is accepted as having a tetrapolar mating system. It is observed in all three major lineages of Basidiomycota[101,102]. However, the distribution of bipolar species is scattered throughout the Basidiomycete phylogeny. Although bipolar species appear to have multiple independent origins from tetrapolar mating systems, a model for *MAT* evolution in the Basidiomycota from an ancestral tetrapolar system to bipolarity is reported[101]. In the subphylum Agaricomycotina and Ustilaginomycotina, which encompasses the mushrooms, the majority of the species seems to be tetrapolar, the few bipolar species were derived from tetrapolar ancestors as a result of coalescence between the regions encoding the two classes of *MAT* genes, like *Cryptococcus neoformans*, *Malassezia globosa*, *Ustilago hordei*[103-105]. In the mushroom *Coprinellus*

*disseminatus* and *Pholiota nameko*, the pheromone receptor ceased to be associated with *MAT* leading to bipolarity[102,106]. It is reported that about 65% of the homobasidiomycetes possess a tetrapolar mating system, 25% of species instead have a bipolar system controlled by a single locus with multiple alleles[102]. For the third subphylum, classical mating studies indicate a predominance of bipolar systems in this subphylum for rusts, *Microbotryum violaceum*, the saprobic yeasts of the genera *Rhodosporidium* and *Sporidiobolus*[101,107]. In the well-studied bipolar species, such as the smut fungus *U. hordei* and the rust *M. violaceum*, the two mating type loci are tightly linked on the same chromosome[105,108]. Evidence has been presented demonstrating that the mating processes of AG1 IA and AG 2-2 IV can occur between homothallic and heterothallic isolates[109,110], and the mating processes of AG1 IC[111]occurs heterothallic with a bipolar mating system.

In the *R. solani* AG1 IA genome, the homologous HD1 (AG1IA_06139) and HD2 (AG1IA_08558) sequences to *Pleurotus djamor* with BLASTP (e-value less than 6e-07 and 7e-05) were predicted, and the similar HD1/HD2 specific domains were also detected according to the homologous motifs in Basidiomycetes[40]. The ste3-like pheromone receptors (AG1IA_09250, AG1IA_09224, and AG1IA_08458) were predicted. Based on the previous classic study, AG1 IA is proven to have a bipolar mating system[39], we deduced the pheromone/receptor pair genes and HD proteins are theoretically physically linked. However, in *R. solani* AG1 IA assembly, these genes were scattered on the different scaffolds (the scaffold 84, 54, 30, 1, 3 and 4). Based on the present information about our strain, we are unable to determine whether it belongs to a homothallic or a heterothallic mating system or both for the complex. Although the single spore isolates (SBIs) would be extremely difficult to obtain, the SBIs induction and single-protoplast isolates (SPIs) preparation in lab for this strain is in progress for further confirming the mating system.The Sanger sequencing of fosmids is in progress for the whole sequence between two gene locations. For the mating mechanism, the bipolar mushrooms, *C. disseminatus* and *P. nameko*, appear to utilise only homeodomain proteins rather than pheromone receptors to determine mating type[106]. The role of the *A* and *B* mating type genes in *R. solani* AG1 IA is still unclear. Therefore, in future studies, we will investigate the function of the mating proteins *in vivo* using molecular techniques.

## SUPPLEMENTARY METHODS

### Isolation of *R. solani* AG1 IA genomic DNA

The *R. solani* AG1 IA strain was grown in potato dextrose broth (PDB) medium at 28 ℃ for two days in the dark with vigorous shaking (150 rpm) and was then washed with sterile $H_2O$, frozen in liquid N2 and freeze dried, and genomic DNA was extracted using a modified CTAB method[5,112]. Mycelia (~200 mg) were ground to a fine powder under liquid nitrogen and then mixed with 10 mL of extraction buffer (100 mM Tris pH8.0, 1.4 M NaCl, 20 mM EDTA, 1%PVP, 2% β-mercaptoethanol), incubated at 65 ℃ for 50 min and added to 1 ml KAC. This solution was then incubated for 20 min and mixed with an equal volume of phenol:chloroform:isoamylol (25:24:1), followed by centrifugation at 6,000 g for 6 min. The aqueous phase was added to an equal volume of pre-cooled (-20 ℃) isoamyl alcohol and incubated for at least 40 min at -20 ℃, followed by centrifugation at 10,000 *g* for 5 min, after which the genomic DNA was transferred to 75% alcohol and dried. Then, 80 μL of TE and 3 μl of RNase were added, and the solution was incubated at 37°C in water for 40 min. The aqueous phase was added to 300 μl of chloroform:isoamylol (24:1) and centrifuged at 10,000 *g* for 1 min. Finally, the genomic DNA was washed with 70% ethanol, dried and resuspended in 400 μL of TE buffer.

### Genome sequencing and assembly

Five libraries were constructed and sequenced using a whole-genome shotgun (WGS) sequencing strategy and the Illumina Genome Analyzer (GA) sequencing technology. The insert sizes in the five libraries were 173 bp, 347 bp, 2,200 bp, 5,300 bp, and 5,600 bp, respectively (Supplementary Table S1). Low-quality data that had a qual value of less than 20 and consisted of short reads (length < 35 bp) were filtered from the raw data. Duplicate sequencing reads were then manually removed. We assembled the high-quality short reads (insert sizes of 173 bp, 347 bp) using SOAPdenovo[10]. Reads with a long insert size of 2,200 bp, 5,300 bp, and 5,600 bp were aligned to the assembly using the Burrows-Wheeler Aligner (BWA)[42], and paired-end reads with incorrect insert sizes were filtered. All of the high-quality data were then subjected to assembly. Scaffolds were constructed in a step-by-step manner using paired-end reads ranging from 2,200 bp to 5,600 bp. A connection required at least 3 consistent read pairs.

We obtained the mitochondrial genome of *R. solani* AG1 IA, which was identified by aligning the sequence to the NT database using BLAST. The 146 Kb genome sequence obtained from the SOAPdenovo[10] assembly included 29 gaps, which were filled using PCR methodology.

The heterozygosity and SNPs in the *R. solani* AG1 IA genome were analysed using the 173 bp and 347 bp library reads. Short reads were aligned to assembly sequence using BWA[42]. Then SNPs were calling using the samtools mpileup (-q 60 -Q 13) and bcftools the SAMtools[43]. The SAMtools vcfutils.pl varFilter script (-D 200 -d 10 -a 5) produces a variant call format (VCF) file.

**Assessment of the genome**

We used a total of five approaches to assess the accuracy of the *R. solani* AG1 IA genome sequence assembly. Under the mRNA-seq estimation approach, the transcriptome reads were aligned to the genome assembly with the default parameters of the Burrows-Wheeler Aligner (BWA)[43]. Because of the exon junction, some reads were not aligned to the assembly. Then, these reads were realigned to the genome using the BLAST-like Alignment Tool (BLAT)[113], and one read was considered by aligning it to the genome with a threshold of at least 97% of the bases being covered.

We next employed six fosmids, which were sequenced using Sanger technology for assessment. The fosmid library was constructed according to instructions of the CopyControl™ Fosmid Library Production Kit (EPICENTRE American). We randomly selected the 38-48 kb DNA fragments from the *R. solani* AG1 IA genome to construct the fosmid library, for which the titre was $1.8\times10^5$ cfu/lib. The base-calling process for the fosmids employed Phred[114], and Phrap (www.phrap.org) and Consed[115] were used for assembly. The sequences of six fosmids that were randomly selected from the library were compared to the genome assembly. Under this approach, the qual value of the fosmids was the Phred qual, and the coverage of the genome sequence was calculated using reads with an insert size of 173 bp. The short reads were selected to calculate the coverage with the following criteria: unique alignment to the genome sequence and no gap openings, gap extensions or mismatches. The average coverage of the scaffold sequences that were aligned to the fosmid sequences was greater than 61 fold.

We then employed 87 proteins found in the UniProt database for *R. solani* for homologous analysis. We found seven non-aligned proteins related to the *R. solani* viral genome because our strain does not naturally contain viral plasmids. The other 80 proteins were aligned to the genome based on BLASTP results. Finally, we designed 28 primer pairs for regions obtained from the assembly. Twenty-eight genomic regions could be amplified and were Sanger sequenced from both ends. We used BLASTP searching homologous genes against 246 single-copy orthologous gene families of fungi genomes[66] with thresholds of identity $\geq 0.3$, e-value cutoff of 1e-5.

**Analysis of repeats**

Repetitive elements of *R. solani* AG1 IA were identified de novo by identifying sequence elements, PILER[44] and RepeatScout[45] were used for detection. Seven families were identified by PILER, and 77 families were identified by RepeatScout. PALS[44] was applied for local alignments, and MUSCLE was used to perform multiple alignments during the PILER analysis process. In the 77 families generated by RepeatScout, 75 repeats occurred at least three times, and three were duplicated in the PILER families. Thus, a de novo repeat library that included 81 families (74 RepeatScout families) was constructed and assigned to the RepeatMasker (http://repeatmasker.org) repeat protein library with BLASTX. Twenty families were identified as DNA transposons. Additionally, 39 LTR families and 20 LINE families were identified, and two families remained unknown. The results of the de novo method for repeat annotation was analysed using RepeatMasker v3-2-9 (http://repeatmasker.org). The original repeat analysis indicated that transposable elements (TEs) constitute 2,214,756 bp of the genome (6.39%). Using the ProcessRepeats program in RepeatMasker, 1,822,825 bp of TEs and 55,336 bp of tandem repeats were checked and confirmed. The homology analysis was performed using RepeatMasker with Repbase 16.09[116]. A total of 4,858 bp of transposable elements and 56,504 bp of tandem repeats were identified with the Rebase library.

We also accounted for the annotation of predicted *R. solani* AG1 IA genes in the TE analysis when searching the Pfam database (http://pfam.sanger.ac.uk/), the Swiss-Prot database (http://www.uniprot.org/) and the NCBI RefSeq fungus proteins (ftp://ftp.ncbi.nih.gov/refseq/release/fungi/).

**Gene prediction and annotation**

To obtain a more accurate gene set, we performed integrated prediction. Homologous proteins were identified through the Swiss-Prot database[117] using BLAST v2.2.24. Novel candidate genes were obtained from running Augustus[47] and GeneMark-ES[48], which were used as ab initio programs. To apply EuGene v3.6[11] for integrated gene prediction, we obtained predicted translation start sites using NetStart[49]. Exon junctions were identified from RNA-seq using TopHat v1.1.4[50]. All of these predictions and the transcriptome assembly were combined to perform prediction using the EuGene platform, and 10,489 genes were obtained. Some genes were manually checked against transcriptome contigs and high-similarity proteins in Swiss-Prot.

We predicted the open reading frames (ORFs) of the *R. solani* AG1 IA genome using a combination of manual and automated annotation[11]. Using BLAST and HMMER v3.0 http://hmmer.janelia.org/, ORFs were annotated in the UniProt database[117], and ORF domains were found in the Pfam database[118]. ORFs were also assigned to

the EuKaryotic Orthologous Groups (KOG) database (E-value cut-off of 1e-5), and the NCBI RefSeq fungi protein set using BLASTP (E-value cut-off of 1e-6), respectively. In total, genes were annotated with integrated annotation from these databases. Meanwhile, these genes were assigned to the Pathogen-Host Interaction (PHI) database[65] using BLASTP with an e-value cut-off of 1e-50.

For the mitochondrial genome, we obtained ORFs that were highly similar to mitochondrial genes of fungi or other eukaryotes. During the prediction process, BLASTX was used to identify candidate homologous genes from the Swiss-Prot database[117], and ORFs were obtained using the exonerate[119] aligning sequence against Swiss-Prot proteins. We also identified tRNAs, rRNA and ribozymes by running tRNAscan-SE[120] and searching against the Rfam database[121,122].

**Evolution**

To perform ortholog analysis, the reciprocal best match approach based on BLASTP was used, orthoMCL http://v2.orthomcl.org was used to identify ortholog, co-ortholog and inparalog pairs as well as reciprocal better hits for each pair. The pairs and their weights were used to construct the orthoMCL graph for clustering with the MCL algorithm[51]. During the analysis, the match cut-off was 50%, and the e-value was 1e-5. More than 72.61% of the pairs were assigned to the same KOG group. KOG (ftp://ftp.ncbi.nih.gov/pub/COG/KOG/) alignment analysis of the nine fungi was performed with an e-value 1e-5. To detect gene evolution rates, we calculated the tree branch lengths for 608 clusters of 1:1 orthologs of the ten fungi (*R. solani* AG1 IA, *M. oryzae*, *C. cinereus*, *C. neoformans*, *P. graminis*, *U. maydis*, *P. placenta*, *P. chrysosporium*, *Melampsora laricis-populina*, and *L. bicolor*). The tree branch lengths indicated different evolutionary rates for these ortholog groups. The distribution of the branch length values differed from a normal distribution, as determined by the Shapiro-Wilk test ($p < 0.01$). The tree branch lengths of $\geq 10$ and $< 11$ were obtained from PAML analysis. Global alignment of the proteins of 608 ortholog groups with only one protein of one fungus was performed using CLUSTALW v2.1[52]. Along with the phylogenetic tree of the ten fungi, we present the tree branch length of each ortholog group calculated using codeml with the JTT amino acid substitution model from the PAML v4.4e package[53]. Regarding the distribution of the tree lengths, the Shapiro-Wilk normality test was used to determine whether the data were normally distributed.

**Comparative genomics**

The proteins of the *R. solani* AG1 IA genome were compared with the protein sequences of nine fungi, namely

*Magnaporthe oryzae* (MG)[15], *Coprinus cinereus* (CC)[70], *Cryptococcus neoformans* (CN)[123], *Puccinia graminis* (PG)[124], *Ustilago maydis* (UM)[16], *Postia placenta* (PP)[75], *Phanerochaete chrysosporium* (PC)[125], and *Laccaria bicolor* (LB)[126], *Melampsora laricis-populina* (ML)[124] using BLASP. We also performed a gene ontology analysis of *R. solani* AG1 IA proteins with Blast2GO[127]. The analysis was performed for the important cereal pathogens *U. maydis*, *M. oryzae* and *R. solani* AG1 IA[15,16]. The download sites, references, data versions, annotation proteins and download site hosts of the genomes used for comparison were described on our official website: http://rice.sicau.edu.cn/sequence/fungi.zip.

The annotation data were downloaded from http://bioinfo.cau.edu.cn/agriGO/download.php and http://genome.jgi-psf.org/Ustma1/Ustma1.download.ftp.html.

Chi-square test analysis resulted in a total of 808 GO terms being found ($P<0.01$). Generic GO slims (ftp://ftp.geneontology.org/pub/go/GO_slims/goslim_generic.obo) without three root classes were used for GO value analysis. To perform PFAM classification, we collected pfam database annotated results for CC, CN, LB, MG, PC, PG, PP, ML and UM, stated each pfam domain number for each species (if it did not exist, it was set to 0) and the sum for all of the species, and calculated the p-value of each pfam domain annotated result between *R. solani* AG1 IA and the other strains using the phyper function of R version 2.10.1. A Bonferroni correction was applied to the calculated p-values, and corrected $p \geq 0.05$ was used as a threshold to obtain *R. solani* AG1IA enriched pfam domains.

**Transcriptome assembly and analysis**

The transcriptome was sequenced from the *R. solani* AG1 IA library with an insert size of 180 bp using Illumina GA II technology, and 12,416,622 reads were produced. The transcriptome assembly was performed with ABySS[128] v1.2.5, which is based on the deBruijn graphing algorithm. During this process, kmers ranging from 25 to 50 bp were used for assembly. The trans-ABySS[129] v1.0.1 pipeline was used, which can address variable transcript expression and multiple isoform expression. A total of 47,327 contigs (length >200 bp) from the assembly were obtained, with an N50 size of 664 bp. These cDNA contigs were used as ESTs to provide hints for gene prediction.

To detect the interactions between rice 9311 and *R. solani* AG1 IA, we collected the strain from six infection stages and prepared mRNA[130]. The six investigated stages were as follows: mycelium on the plant surface after incubation for 10 h; light infection cushion on the rice surface after incubation for 18 h; when the mycelium had invaded the host and a heavy infection cushion had appeared after incubation for 24 h; light necrosis

post-infection after incubation for 32 h; intermediate necrosis post-infection after incubation for 48 h; and at heavy necrosis of the infected rice tissue after incubation for 72 h. We then extracted mRNA from the samples according to the instructions of the E.N.Z.A. total RNA isolation kit (Omega Bio-tek, Atlanta, GA). Fungal cDNA libraries were constructed from disease lesions at 10 hours, 18 hours, 24 hours, 32 hours, 48 hours and 72 hours after inoculation (HAI), and each library product was sequenced using a library with a read insert size of 180 bp with Illumina GA II technology. We employed mRNA-Seq for expression analysis. The RNA expression analysis was based on the predicted genes of *R. solani* AG1 IA. First, Tophat[50]was used to map mRNA reads to the genome, and Cufflinks[54] was then used to calculate the expected fragments per kilobase of transcript per million fragments sequenced (FPKM).

Hierarchical clustering methods are widely applied to analyse gene expression data, and genes with similar functions group together[36,55,56]. Hierarchical clustering analysis of *R. solani* AG1 IA proteins performed using transcriptome expression data indicated several groups sharing highly similar expression patterns (Correlation > 0.98). Cluster v3.0[57] was used to analyse the genome-wide expression data. An average linkage hierarchical clustering method was used, and Pearson correlation was employed to measure the similarity of the expressed genes, and we used this method to identify three groups of candidate effector genes.

**Genes involved in pathogenicity**

We scanned the carbohydrate active enzyme (CAZyme) genes of *S. nodorum*, *F. graminearum*, *M. oryzae*, *C. cinereus*, *P. graminis*, *U. maydis*, *M. laricis-populina*, *L. Bicolor* and *R. solani* AG1 IA using BLASTP with an e-value of less than 1e-50. The codes indicating the enzyme classes were those defined by the CAZyme database (http://www.cazy.org/). Protein function analysis using Rost[131] indicated a high accuracy (86%) of the pairwise BLAST hits with an e-value of less than 1e-50. The CAZymes database, which contains 128,641 proteins, was downloaded from the CAZymes Analysis Toolkit (CAT)[58]. The protein alignment of the nine laccase genes of *R. solani* AG1 IA and four reported laccase genes of *R. solani* AG6 was performed using MUSCLE[132], and the phylogenetic relationships were analysed using TREEBEST with the neighbour-joining method and the JTT model[60](http://treesoft.sourceforge.net/index.shtml).

We aligned the *R. solani* AG1 IA proteins to those in the Cytochrome P450 Homepage database (http://drnelson.uthsc.edu/cytochromeP450.html). The result of the alignments was filtered with a threshold e-value of less than 1e-50. The proteins were aligned using ClustalW v2.[52]. The protein distance was calculated using the PHYLIP http://evolution.genetics.washington.edu/phylip/ Protdist program with the

Jones-Taylor-Thornton model. A phylogenetic tree was constructed with the PHYLIP UPGMA program.

In addition, we predicted the transporters present in the genome using the TCDB (Transporter Classification Database http://www.tcdb.org/)[133] with BLASTP at an e-value ≤ 1e-50 among the selected fungi. ATP-binding cassette (ABC) proteins constitute one of the largest protein families. Of the protein families that have been characterised so far, most are involved in the ATP-dependent transport of a remarkably broad range of substrates across biological membranes. In toxin-producing fungi, transporters are critical for the efflux of endogenously produced molecules[88]. ABC proteins exhibit certain conservative motifs, such as the Walker A box, Walker B box, and ABC signature motif. Similarity analysis can be employed for identification of ABC proteins using known proteins and domains[25]. The *R. solani* AG1 IA proteins were assigned to the ABC transporters in the UniProt database (http://www.uniprot.org/) using BLASTP with an e-value ≤ 8e-14 (41 proteins with an e-value ≤ 1e-50) and the TCDB with an e-value ≤ 1e-50. The hits were subjected to subsequent domain analysis by searching against the Pfam database. Proteins with domains such as ABC transporters, ABC-2 family transporter proteins, an ABC transporter transmembrane region, ABC transporter transmembrane region 2, ABC-2 type transporters, and CDR ABC transporters were considered to be ABC transporters. If a protein exhibited a conserved sequence with multiple known ABC transporters in the same subfamily, we considered it to be in the same subfamily. ABC transporters in seven subfamilies (ABC-A, ABC-B, ABC-C, ABC-D, ABC-E, ABC-F, and ABC-G) were identified.

**G-protein signalling pathway**

We found homologous *R. solani* AG1 IA genes that were very similar to genes in the MAPK signalling pathway, the Ca signalling pathway and the cAMP pathway of *Saccharomyces cerevisiae* using BLASTP with a threshold e-value of less than 1e-50 (http://www.genome.jp/kegg/pathway/sce/sce04011.html).

*R. solani* AG1IA G-protein subunits were identified through assignment to the Swiss-Prot database based on an e-value of less than 6e-14 (five genes less than 1e-50). GPCRs were predicted based on a previous report[59]. GPCRs characterised in *N. crassa, M. oryza, A. nidulans*, and *F. graminearum* were used in BLASTP queries against the deduced proteomes of *R. solani* AG1 IA using an e-value of less than 1e-06. Through Pfam database annotation, other GPCRs were predicted with domain descriptions. To ensure that the putative GPCR proteins contain the typical 7-TM domain topology, with an extracellular N-terminus and an intracellular C-terminal domain, the BLASTP results were analysed using membrane-spanning prediction software.

All of these GPCR-like proteins were evaluated to verify the presence of 7 transmembrane (7-TM) helices

with TMPRED, Phobius (http://phobius.sbc.su.se/), TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), and (http://www.ch.embnet.org/software/TMPRED_form.html).

**Selection of candidate effectors in secretomes for further experiments**

The obtained transcriptome data revealed that many secreted proteins were up-regulated during the progression of infection. Among the 965 potentially secreted proteins we identified, 234 secreted proteins showed a two-fold or greater difference in expression during early infection progress in the rice-fungus interaction. Because these secreted proteins may function as effectors that manipulate the innate immunity of plants and that enable parasitic colonisation by interacting with plant receptors, we selected 145 genes that were up-regulated at least twofold with a low molecular weight during infection relative to the initial post-infection period to identify potential effectors. A total of 45 potential effector proteins were selected for verification of their function[134,135]. These proteins, expressed by *E. coli* Transetta (DE3), were applied to rice, maize and soybean leaves, and we observed whether cell death phenotypes appeared in the leaves at 48 hours post-infiltration. Three candidate effectors, AG1IA_09161, AG1IA_07795, and AG1IA_05310, were verified, which were then purified and used to inoculate rice, maize and soybean leaves; cell death phenotypes were visible 48 hours after inoculation with the purified proteins. The domains of these proteins were predicted based on annotation in the Pfam database (http://pfam.sanger.ac.uk/). Analysis of domains homologous to those of the three verified effectors was performed to predict other proteins having similar functions. Some of these proteins showed up-regulation during infection, which suggests that they could be potential effectors. Hierarchical clustering based on the three effectors was also used to predict candidate effectors[55,56] (Correlation > 0.98).

**cDNA manipulation**

RNA was prepared from the fungal mycelia using the Ambion RNAqueous kit (Ambion, USA), and CDNA was synthesized using Two-Step RT-PCR SuperMix (TransGen Biotech, China). All DNA manipulations and other procedures, including agarose gel electrophoresis, were performed according to standard protocols[112]. All of the restriction enzymes and T4 DNA ligase were used following the manufacturer's instructions (Takara, Japan).

**Plant growth conditions**

The rice 9311 cultivar, for which the genome sequence is available (http://www.rice.genomics.org.cn/), was used in this study. Rice 9311 plants were grown in the greenhouse under white light for 16 h at 27℃ and in the dark

for 8 h at 22 ℃; the relative humidity was 80% during growth to V11 phase[136], at which point the leaves were collected for detached inoculation. In addition, maize, soybean and plants of the rice R60 and S75 cultivars were grown in a natural field, and detached seedling leaves were infiltrated under conditions of white light for 16 h at 28 ℃ and darkness for 8 h at 28 ℃, with 80% relative humidity.

## SUPPLEMENTARY REFERENCES


61. Cubeta, M. A. & Vilgalys, R. Population Biology of the *Rhizoctonia solani* Complex. *Phytopathology* **87**, 480-484 (1997).
62. Yong, X. *et al*. Genetic Diversity and Pathogenicity Variation in *Rhizoctonia solani* Isolates from Rice in Sichuan Province, China. *Rice Science* **15**, 137-144 (2008).
63. Parmete J.R. *Rhizoctonia solani*, Biology and Pathology. *University of California Press*, 50-54 (1970).
64. Jaap, K. *et al*. Heterogeneity in electrophoretic karyotype within and between anastomosis groups of *Rhizoctonia solani*. *Mycol Res* **100**, 789-797 (1996).
65. Winnenburg, R. *et al*. PHI-base update: additions to the pathogen host interaction database. *Nucleic Acids Res* **36**, 572-576 (2008).
66. Aguileta, G. *et al.* Assessing the performance of single-copy genes for recovering robust phylogenies. *Syst Biol* **57**, 613-627 (2008).
67. Gabriela A. *et al*. Genome Evolution in Plant Pathogenic and Symbiotic Fungi. *Advances in Botanical Research* **49,** 151-193 (2009).
68. Hane, J. K. *et al.* Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell* **19**, 3347-3368 (2007).
69. Levesque, C. A. *et al*. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol* **11**, R73 (2010).
70. Stajich, J. E. *et al.* Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc Natl Acad Sci U S A* **107**, 11889-11894 (2010)
71. Gogarten, J. P. & Olendzenski, L. Orthologs, paralogs and genome comparisons. *Curr Opin Genet Dev* **9**, 630-636 (1999).
72. Enright, A. J., Kunin, V. & Ouzounis, C. A. Protein families and TRIBES in genome sequence space. *Nucleic Acids Res* **31**, 4632-4638 (2003).
73. MacPherson, S., Larochelle, M. & Turcotte, B. A fungal family of transcriptional regulators: the zinc cluster proteins. *Microbiology and molecular biology reviews* **70**, 583-604 (2006).
74. Coleman, J. J. & Mylonakis, E. Efflux in fungi: la piece de resistance. *PLoS Pathog* **5**, 1-7 (2009).
75. Martinez, D. *et al.* Genome, transcriptome, and secretome analysis of wood decay fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion. *Proc Natl Acad Sci U S A* **106**, 1954-1959 (2009).
76. Shelest, E. Transcription factors in fungi. *FEMS Microbiol Lett* **286**, 145-151 (2008).
77. Gonzâlez G. V., *et al*. Review. Biology and Systematics of the form genus *Rhizoctonia*. *Spanish Journal of Agricultural Research* **4,** 55-79 (2006).
78. Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol* **43**, 205-227 (2005).
79. Oliver, R. P. & Ipcho, S. V. Arabidopsis pathology breathes new life into the necrotrophs vs. biotrophs

classification of fungal pathogens. *Mol Plant Pathol* **5**, 347-352 (2004).
80. Wahleithner, J. A. *et al.* The identification and characterization of four laccases from the plant pathogenic fungus *Rhizoctonia solani*. *Curr Genet* **29**, 395-403 (1996).
81. Akoi, H., Sassa, T., and Tamura, T. Phytotoxic metabolites of *Rhizoctonia solani*. *Nature* **200,** 575 (1963).
82. Iacobellis, N. S., and DeVay, J. E. Studies on pathogenesis of *Rhizoctonia solani* in beans: An evaluation of the possible role of phenylacetic acid and its hydroxy derivatives as phytotoxins. *Physiol. Plant Pathol.* **30**, 421-432 (1987).
83. Jason C. S. & Antonis R. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc Natl Acad Sci U S A* **107**, 10136–10141 (2010).
84. Frisvad, J. C., Andersen, B. & Thrane, U. The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. *Mycol Res* **112**, 231-240 (2008).
85. Magan, N. Mycotoxin contamination of food in Europe: early detection and prevention strategies. *Mycopathologia* **162**, 245-253 (2006).
86. Cresnar, B. & Petric, S. Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta* **1814**, 29-35 (2010).
87. Park, J. *et al.* Fungal cytochrome P450 database. *BMC Genomics* **9**, 402 (2008).
88. Amnuaykanjanasin, A. & Daub, M. E. The ABC transporter ATR1 is necessary for efflux of the toxin cercosporin in the fungus *Cercospora nicotianae*. *Fungal Genet Biol* **46**, 146-158 (2009).
89. Catanzariti, A. M., Dodds, P. N. & Ellis, J. G. Avirulence proteins from haustoria-forming pathogens. *FEMS Microbiol Lett* **269**, 181-188 (2007).
90. Liu, Z. *et al.* The cysteine rich necrotrophic effector SnTox1 produced by *Stagonospora nodorum* triggers susceptibility of wheat lines harboring Snn1. *PLoS Pathog* **8**, e1002467 (2012).
91. Panstruga, R. & Dodds, P. N. Terrific protein traffic: the mystery of effector protein delivery by filamentous plant pathogens. *Science* **324**, 748-750 (2009).
92. De Wit, P. J., Mehrabi, R., Van den Burg, H. A. & Stergiopoulos, I. Fungal effector proteins: past, present and future. *Mol Plant Pathol* **10**, 735-747 (2009).
93. Hogenhout, S. A., Van der Hoorn, R. A., Terauchi, R. & Kamoun, S. Emerging concepts in effector biology of plant-associated organisms. *Mol Plant Microbe Interact* **22**, 115-122 (2009).
94. Manning, V. A. & Ciuffetti, L. M. Localization of Ptr ToxA produced by *Pyrenophora tritici-repentis* reveals protein import into wheat mesophyll cells. *Plant Cell* **17**, 3203-3212 (2005).
95. Friesen, T. L., Meinhardt, S. W. & Faris, J. D. The *Stagonospora nodorum*-wheat pathosystem involves multiple proteinaceous host-selective toxins and corresponding host sensitivity genes that interact in an inverse gene for gene manner. *Plant J* **51**, 681-692 (2007).
96. Li, L., Wright, S. J., Krystofova, S., Park, G. & Borkovich, K. A. Heterotrimeric G protein signaling in filamentous fungi. *Annu Rev Microbiol* **61**, 423-452 (2007).
97. Kasahara, S. & Nuss, D. L. Targeted disruption of a fungal G-protein beta subunit gene results in increased vegetative growth but reduced virulence. *Mol Plant Microbe Interact* **10**, 984-993 (1997).
98. Kulkarni, R. D., Kelkar, H. S. & Dean, R. A. An eight cysteine containing CFEM domain unique to a group of fungal membrane proteins. *Trends in Biochemical Sciences* **28**, 118-121 (2003).
99. Liu, X. *et al.* NCD3G: a novel nine-cysteine domain in family 3 GPCRs. *Trends Biochem Sci* **29**, 458-461 (2004).
100. Madhani, H. D. & Fink, G. R. Combinatorial control required for the specificity of yeast MAPK signaling. *Science* **275**, 1314-1317 (1997).
101. Coelho, M. A., Sampaio, J. P. & Goncalves, P. A deviation from the bipolar-tetrapolar mating paradigm in

an early diverged basidiomycete. *PLoS Genet* **6**, e1001052 (2010).
102. James, T. Y., Srivilai, P., Kues, U. & Vilgalys, R. Evolution of the bipolar mating system of the mushroom *Coprinellus disseminatus* from its tetrapolar ancestors involves loss of mating type specific pheromone receptor function. *Genetics* **172**, 1877-1891 (2006).
103. Lengeler, K. B. *et al.* Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. *Eukaryot Cell* **1**, 704-718 (2002).
104. Xu, J. *et al.* Dandruff-associated Malassezia genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci U S A* **104**, 18730-18735 (2007).
105. Lee, N., Bakkeren, G., Wong, K., Sherwood, J. E. & Kronstad, J. W. The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. *Proc Natl Acad Sci U S A* **96**, 15026-15031 (1999).
106. Yi, R. *et al.* Genomic structure of the A mating type locus in a bipolar basidiomycete, *Pholiota nameko*. *Mycol Res* **113**, 240-248 (2009).
107. Giraud, T., Yockteng, R., Lopez-Villavicencio, M., Refregier, G. & Hood, M. E. Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryot Cell* **7**, 765-775 (2008).
108. Hood, M. E. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. *Genetics* **160**, 457-461 (2002).
109. Toda, T. & Hyakumachi, M. Heterokaryon formation in *Thanatephorus cucumeris* anastomosis group 2-2 IV. *Mycologia* **98**, 726-736 (2006).
110. Qu, P. *et al.* Use of single-protoplast isolates in the study of the mating phenomena of *Rhizoctonia solani* (*Thanatephorus cucumeris*)AG-1 IC and IA. *Mycoscience* **49**, 132-137 (2008).
111. Qu, P. *et al.* Heterokaryon formation in *Thanatephorus cucumeris* (*Rhizoctonia solani*) AG-1 IC. *Mycol Res* **112**, 1088-1100 (2008).
112. Sambrook, J., Fritsch, E. F., & Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd ed. in Cold Spring Laboratory, Cold Spring Harbor, NY.* (1989).
113. Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664 (2002).
114. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194 (1998).
115. Gordon, D. Viewing and editing assembled sequences using Consed. *Curr Protoc Bioinformatics* Chapter **11**, Unit11 12 (2003).
116. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462-467 (2005).
117. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* **38**, 142-148 (2010).
118. Finn, R. D. *et al*. The Pfam protein families database. *Nucleic Acids Res* **38**, 211-222 (2010).
119. Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
120. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964 (1997).
121. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**, 121-124 (2005).
122. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335-1337 (2009).
123. Loftus, B. J. *et al.* The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321-1324 (2005).

124. Duplessis, S. *et al.* Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A* **108**, 9166-9171 (2011).
125. Martinez, D. *et al.* Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol* **22**, 695-700 (2004).
126. Martin, F. *et al.* The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* **452**, 88-92 (2008).
127. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676 (2005).
128. Simpson, J. T. *et al*. ABySS: a parallel assembler for short read sequence data. *Genome Res* **19**, 1117-1123 (2009).
129. Robertson, G. *et al*. De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**, 909-912 (2010).
130. Zhao, C. J. *et al*. Identification of defense-related genes in rice responding to challenge by *Rhizoctonia solani*. *Theor Appl Genet* **116**, 501-516 (2008).
131. Rost, B. Enzyme function less conserved than anticipated. *J Mol Biol* **318**, 595-608 (2002).
132. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).
133. Saier, M. H., Jr., Yen, M. R., Noto, K., Tamang, D. G. & Elkan, C. The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**, 274-278 (2009).
134. Win, J. *et al.* Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* **19**, 2349-2369 (2007).
135. Tuori, R. P., Wolpert, T. J. & Ciuffetti, L. M. Heterologous expression of functional Ptr ToxA. *Mol Plant Microbe Interact* **13**, 456-464 (2000).
136. Venu, R. C. *et al.* RL-SAGE and microarray analysis of the rice transcriptome after *Rhizoctonia solani* infection. *Mol Genet Genomics* **278**, 421-431 (2007).