



OPEN

Hybrid de novo genome-reassembly reveals new insights on pathways and pathogenicity determinants in rice blast pathogen *Magnaporthe oryzae* RMg_DI

Bhaskar Reddy¹, Aundy Kumar¹✉, Sahil Mehta², Neelam Sheoran¹, Viswanathan Chinnusamy³ & Ganesan Prakash¹

Blast disease incited by *Magnaporthe oryzae* is a major threat to sustain rice production in all rice growing nations. The pathogen is widely distributed in all rice paddies and displays rapid aerial transmissions, and seed-borne latent infection. In order to understand the genetic variability, host specificity, and molecular basis of the pathogenicity-associated traits, the whole genome of rice infecting *Magnaporthe oryzae* (Strain RMg_DI) was sequenced using the Illumina and PacBio (RSII compatible) platforms. The high-throughput hybrid assembly of short and long reads resulted in a total of 375 scaffolds with a genome size of 42.43 Mb. Furthermore, comparative genome analysis revealed 99% average nucleotide identity (ANI) with other *oryzae* genomes and 83% against *M. grisea*, and 73% against *M. poe* genomes. The gene calling identified 10,553 genes with 10,539 protein-coding sequences. Among the detected transposable elements, the LTR/Gypsy and Type LINE showed high occurrence. The InterProScan of predicted protein sequences revealed that 97% protein family (PFAM), 98% superfamily, and 95% CDD were shared among RMg_DI and reference 70-15 genome, respectively. Additionally, 550 CAZymes with high GH family content/distribution and cell wall degrading enzymes (CWDE) such endoglucanase, beta-glucosidase, and pectate lyase were also deciphered in RMg_DI. The prevalence of virulence factors determination revealed that 51 different VFs were found in the genome. The biochemical pathway such as starch and sucrose metabolism, mTOR signaling, cAMP signaling, MAPK signaling pathways related genes were identified in the genome. The 49,065 SNPs, 3267 insertions and 3611 deletions were detected, and majority of these variants were located on downstream and upstream region. Taken together, the generated information will be useful to develop a specific marker for diagnosis, pathogen surveillance and tracking, molecular taxonomy, and species delineation which ultimately leads to device improved management strategies for blast disease.

Since the traditional times, both cereal and cereal products act as pre-eminent and substantial carbohydrate food resources for much of the human population, especially in Asian countries^{1,2}. Among the pre-harvest production constraints, the global cultivation of many cereals including rice and pearl millet is mostly affected by a blast disease-causing filamentous ascomycete fungus, *Magnaporthe* (Hebert) Barr (anamorph: *Pyricularia*)³. This devastating hemibiotrophic pathogen belongs to the family *Magnaporthaceae* and is of principal concern due to the wide distribution, rapid aerial transmissions, seed-borne latent infection, and associated yield losses^{2,4,5}. Morphologically, it causes white, bluish, or greyish water-soaked lesions in all foliar parts such as grain, leaf,

¹Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India. ²Crop Improvement Group, International Centre for Genetic Engineering and Biotechnology, New Delhi 110067, India. ³Division of Plant Physiology, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India. ✉email: kumar@iari.res.in

neck, collar, nodes, and even panicles⁴. The fungus *Magnaporthe* sp. propagates via the generation of asexual spores (conidia) that disseminate the blast disease on cereal hosts^{6,7}.

Out of all reported species belonging to the genus *Magnaporthe*, both *Magnaporthe oryzae* (Anamorph: *Pyricularia oryzae*) and *Magnaporthe grisea* (Anamorph: *Pyricularia grisea*) have emerged as forage and grain production affecting pathogen that annually contributes to the substantial losses of grains^{3,4,8}. In order to control the spread and minimize the associated losses with the blast, farmers depend largely on resistant host cultivars, and application of fungicides especially tricyclazole⁹. However, these measures seem inadequate, and inefficient in blast-endemic regions because of related cost, unstable host resistance, the emergence of new pathotypes, toxic nature of the employed agrochemicals, and chemical residues on produced grains. Hence, the global sustainability of food is becoming a more challenging factor regarding the fast growth of the global population^{1,4,10}.

Due to the recent advancements in molecular techniques, genes expressed by pathogens (*M. oryzae* and *M. grisea*) for the efficient invasion, colonization, and ultimate destruction of various monocot hosts especially the rice have been characterized¹⁰. Given that both pathogenic species uniquely attack their host plants, in depth insights into underlying biological pathways and associated mechanisms are still limited with spatial concern to signaling pathways, virulence factors, and carbohydrate-active enzymes. Moreover, the detected genes or proteins can be utilized as potential targets for marker development to screening the pathogen and fungicides through in silico approach^{11,12}.

There is a wide range of pieces of literature available on fungal pathogen function and their pathogenesis. However, detailed comparative insights of tropical habitat rice blast fungus genes, family and biochemical pathways are still limited/unexplored. Therefore, to characterize such differences at the genomic and biochemical level, we performed the high-throughput whole-genome sequencing of cereal crop rice blast pathogen *M. oryzae* RMg_DI followed by hybrid de novo genome assembly and comparative functional annotation. The genome-level analysis potentially deciphers the insights of each genes, their associated biochemical pathways; network for host–pathogen interaction; growth, evolutionary relationship, and virulence genes. The detailed insights of *M. oryzae* not only about host–pathogen interaction but also in-depth knowledge of pathogenicity mechanisms can be helpful for the effective disease management strategy.

Materials and methods

Data collection and sequencing. In this study, the used data sets were collected from¹³ (Bioproject accession PRJNA330763, <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA330763>), which reported the initial draft genome assembly of rice blast disease-causing *Magnaporthe oryzae* RMg_DI. The sample processing and whole-genome analyses workflow is schematically depicted in Fig. 1. In summary, the pure fungus RMg_DI was isolated followed by genomic DNA extraction, and subsequently, the quality and quantity assessment was done using agarose gel (0.8%) and Nanodrop spectrophotometer, respectively. The pair-end library was prepared was using Illumina TruSeq Nano DNA HT chemistry, and we also prepared the single-molecule real-time (SMRT) libraries compatible for generating long reads. The quantity and quality of both libraries were checked on Agilent Bioanalyzer using high sensitivity DNACHIP. The Illumina platform quality passed libraries were sequenced on the HiSeq2500 sequencer with 125 × 2 pair-end chemistry. The SMRT-bell libraries were sequenced on PacBio-RSII using P6-C4 chemistry as described here^{13,14}.

Bioinformatic genome assembly and quality assessment. Initially, the reads were visualized in FastQC (V0.9.2) software to screen the quality of reads and identify the poor-quality reads for getting optimum read trimming and filtering parameters. Following this, the WGS reads were quality filtered to remove the adapter, poor sequences, and ambiguous bases to obtain high quality reads with filtering parameters such as reads with unknown nucleotides “N” larger than 5%, low-quality sequences (reads with more than 10% quality threshold (QV) < 20 Phred score) and reads shorter than 100 bases were trimmed using Trimmomatic v0.35¹⁵. The RMg_DI genome reads were assembled with a default setting using steps (1) PacBio read correction with LORDeC¹⁶ using Illumina reads, (2) corrected reads assembly with wtdbg2 assembler¹⁷, (3) gaps were filled with Long Read Gap closer¹⁸. Following genome assembly, the benchmark universal single-copy orthologs (BUSCO) was applied for the quantitative assessment of genome completeness against fungal lineage with gene model parameters¹⁹. Further, to determine the average nucleotide identity (ANI) between publically available genomes and generation of unweighted pair group method with arithmetic mean (UPGMA) tree, the OrthoANI calculator was utilized with default settings²⁰.

Functional genome annotation of *M. oryzae* RMg_DI. The functional annotation of assembled genome RMg_DI was performed using the GenSAS web server²¹ which provided an integrated structured pipeline for repeat masking, ab initio gene prediction, homology-based gene function determination, protein family, and superfamily. Initially, the repeat masking was performed using Repeat Masker v4.0.7²² against the fungi library with keeping GC content set at 50–52%. Further, Repeat Modeler v1.0.11 (<http://www.repeatmasker.org/>) was utilized for de novo repeat family identification and modeling. The assembled genome was submitted to AUGUSTUS v3.3.1²³ for genes and proteins prediction against the reference model *Pyricularia grisea* with both strand gene settings. The prediction of rRNA and tRNA was performed using RNAmmer v1.2²⁴ and tRNAscan-SE v2.0²⁵. The determination of simple sequence repeats (SSR) was performed with SSR Finder v1.0²¹ with parameters of 5 count repeat for di, 4 for a tri, and 3 for tetra, Penta, and hexanucleotide SSR repeats.

The InterProScan (version5.48-83.0)²⁶ of predicted protein sequences were performed to determine the occurrence of various protein families, conserved domains, superfamily, and gene ontology (GO) identifiers in the genome. Following, Blast2GO v5.2.5²⁷ was used for the mapping and annotations of GO identifiers to GO terms. The comparative carbohydrate-active enzymes (CAZymes)²⁸ were determined against the dbCAN2 database²⁹ by

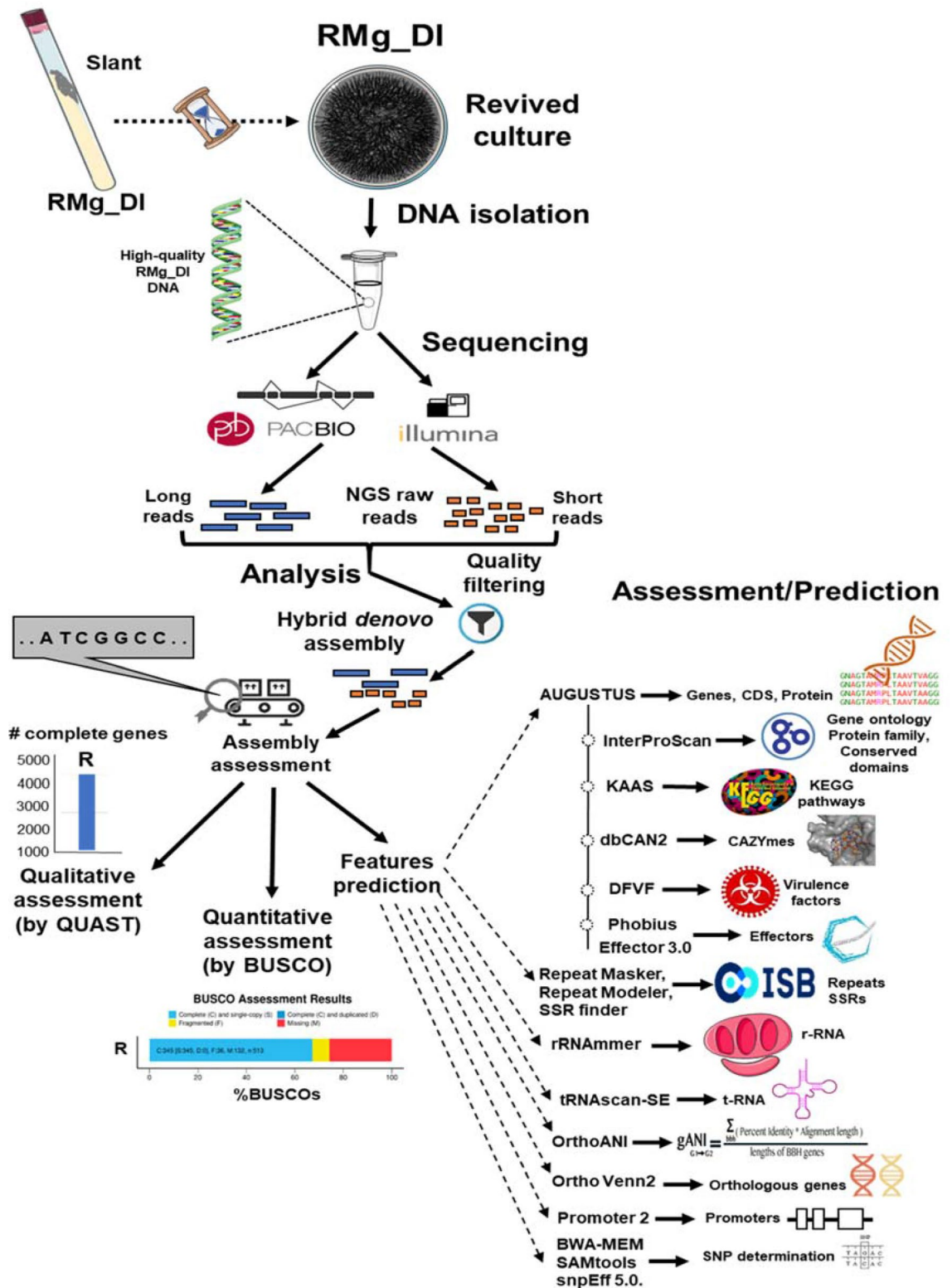


Figure 1. Schematic illustration of the WGS analysis of *M. oryzae* RMg_DI.

using the HMMER model with a setting of e-value 0.00001. The virulence factors (VFs) were identified against a database for virulence factors (DFVF) rice blast database³⁰ using DIAMOND (v0.9.26)³¹ protein aligner with parameters of max-target sequence alignment 1, ≥ 100 amino acid length, 90% identity, 60% query coverage, 60% subject coverage and e-value 0.00001. For prediction of potential effectors in the studied genomes, initially the signal peptide features were determined in the protein sequences and sequences were then submitted to EffectorP3.0 for prediction of effectors with default setting³². Further, OrthoVenn2 comparison was performed

Features	<i>M. oryzae</i> RMg_Dl
Illumina HiSeq-2500 reads (base)	65,731,812 pair-end (16.3 Gb)
PacBio-RSII reads (base)	192,820 single end (1.5 Gb)
Genome SIZE	42,437,481
Number of scaffolds	375
Scaffold N50	524,210
No of genes	10,553
Proteins	10,539
rRNA	73
tRNA	259
SSR	18,830
Virulence factors	51
CAZymes	542
Effectors	603

Table 1. Comparative statistics of the assembled whole genome of *M. oryzae* RMg_Dl.

to decipher the common and unique effector in this genome as compared to other genomes³³. In order to determine metabolic pathways, the predicted protein sequences were submitted to the KAAS webserver³⁴ using search program GHOSTX against the reference pathway model *Pyricularia grisea* 70-15³⁵. The presence of different promoters were identified using promoter2.0³⁶ with default settings, which detected 7021 promoters. For the comparative study, the used genome were GCA_000002495 (*M. oryzae* 70-15), GCA_002368485.1 (*M. oryzae* GUY11) and GCA_002924685.1 (*M. oryzae* WBKY11).

Analysis of variants and phylogenetic tree. For SNP determination, the high quality reads were mapped against *M. oryzae* 70-15 reference genome using burrows-wheeler alignment (BWA) maximal exact match (MEM) module³⁷. The mapped reads were subjected for the removal of PCR duplicate reads using Mark duplicates of Picard tool (<https://broadinstitute.github.io/picard/>). Finally, variant calling was done using SAMtools³⁸ package bcftools mpileup script with settings of phred quality score (Q) ≥ 25 , minimum variant depth (DP) ≥ 25 and mapping quality ≥ 30 . The identified SNPs were annotated to genomic region and various effects type using snpEff 5.0³⁹. For determination of genomes similarity, the comparative whole genome alignment based phylogenetic tree was generated using progressiveMauve aligner which consider gene rearrangement, gain and loss mechanism⁴⁰. For alignment we used RMg_Dl and other 13 public genomes accession were GCA_002924685.1 (WBKY11), GCA_002021675.1 (Mo-nwi-55), GCA_000969745.1 (MG01), GCA_000832285.1 (B157), GCA_002218355.1 (H08-1c), GCA_001936435.1 (MG10), GCA_003991345.1 (10,100), GCA_000734215.1 (4603.4), GCA_000292605.1 (P131), GCA_002368485.1 (GUY11), GCA_000002495.2 (70-15), GCA_000475075.1 (HN19311), GCA_000805855.1 (98-06). The revised assembled genome was deposited in the NCBI database with GenBank assembly accession number GCA_001853415.3.

Results

Genome assembly, genes identification, and assessment. Using two of the prominent high-throughput whole-genome sequencing coupled with the hybrid assembly of short and long reads resulted in a total of 375 scaffolds with a genome size of 42.42 Mb (*M. oryzae* RMg_Dl). This genome-wide gene identification showed a total of 10,555 genes with 10,539 protein sequence features (Table 1). Additionally, comparative genome analysis revealed 99% average nucleotide identity (ANI) with other *M. oryzae* genomes and 83% against *M. grisea*, and 73% against *M. poe* genomes (Fig. 2). The quantitative genome assembly finish assessment showed that complete and single-copy BUSCO was 741 (97.76%) in *M. oryzae* (RMg_Dl) strain out of total 758 BUSCO genes. Although, we found thirteen missing BUSCO in *M. oryzae* RMg_Dl (Supplementary Table S1).

Determination of transposable elements and SSRs. Determination of SSRs in RMg_Dl genomes showed that 11 different transposon families with a total of 10,250 copy numbers in the genome. The repeat types namely Type:EVERYTHING_TE occurred in the majority, whereas repeats such as DNA/TcMar-Fot1, LTR/Gypsy, LINE/Tad1, and Type:LINE is also more in copy number (Table 2). Further, SSR analysis showed that about 18,830 were present in RMg_Dl genome. Furthermore, in the case of di-nucleotides, (TA)_n was the most frequent, followed by (AG)_n and (AT)_n. Similarly, among tri-nucleotide repeats, the (CAG)_n repeats were most frequent, followed by (GCT)_n repeats. In the case of frequency for tetra-nucleotide repeats, (TACC)_n was most frequent, followed by (TAGG)_n (Supplementary Table S2).

Analysis of protein family and conserved domain in sequenced genomes. The InterProScan analysis was performed for identifying protein family (PFAM) and superfamily, which showed the total occurrence of 3774 protein families, 879 superfamilies in the RMg_Dl genome. The comparison among RMg_Dl with genomes for shared and unique PFAM, protein information resource superfamily (PIRSF), superfamily, and conserved domains in sequenced genome showed that 3724 (96.3%) PFAMs, 871 (97.4%) superfamily, 424

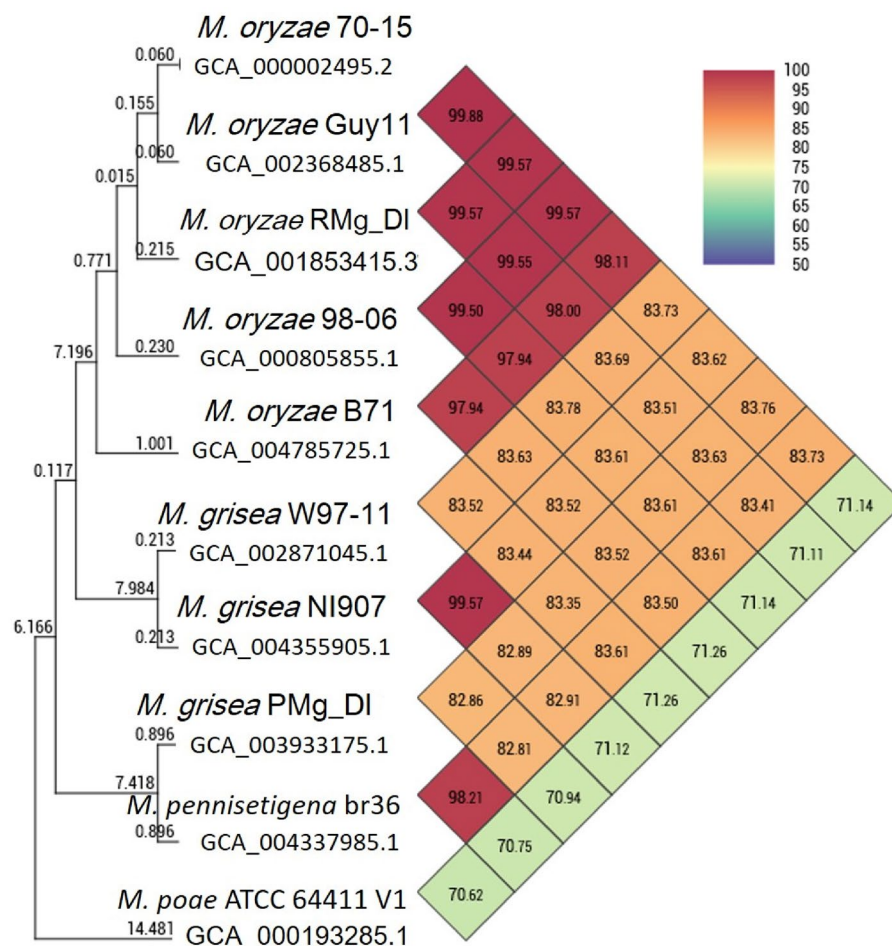


Figure 2. Average nucleotide identity (ANI) among various fungal *Magnaporthe/Pyricularia* genus

Family	<i>M. oryzae</i> RMg_DI
DNA	41
DNA/TcMar-Fot1	829
Type:DNA	870
LINE/Tad1	257
Type: LINE	257
LTR/Copia	344
LTR/Gypsy	1072
Type:LTR	1416
Type:EVERYTHING_TE	2543
Type:Simple_repeat	34
Type:Unknown	2587
Total	10,250

Table 2. The detected transposable elements in *M. oryzae* RMg_DI genome.

(84%) PIRSF and 2017 (93.1%) CDD were shared in studied genomes (Fig. 3, Supplementary Tables S3, S4 and S5). The PFAM detailed exploration showed that family/domain such as WD domain, major facilitator superfamily, cytochrome P450, fungal specific transcription factor domain, protein kinase domain, and ABC transporter was highly prevalent (Supplementary Table S3). Similarly, superfamily identification showed that P-loop containing nucleoside triphosphate hydrolase, NAD(P)-binding domain, MFS transporter, Alpha/Beta hydrolase fold, FAD/NAD(P)-binding domain, and glycoside hydrolase superfamily were prevalent in the genome (Supplementary Table S4). Further, proteins were subjected for the transmembrane, signal peptide and cytoplas-

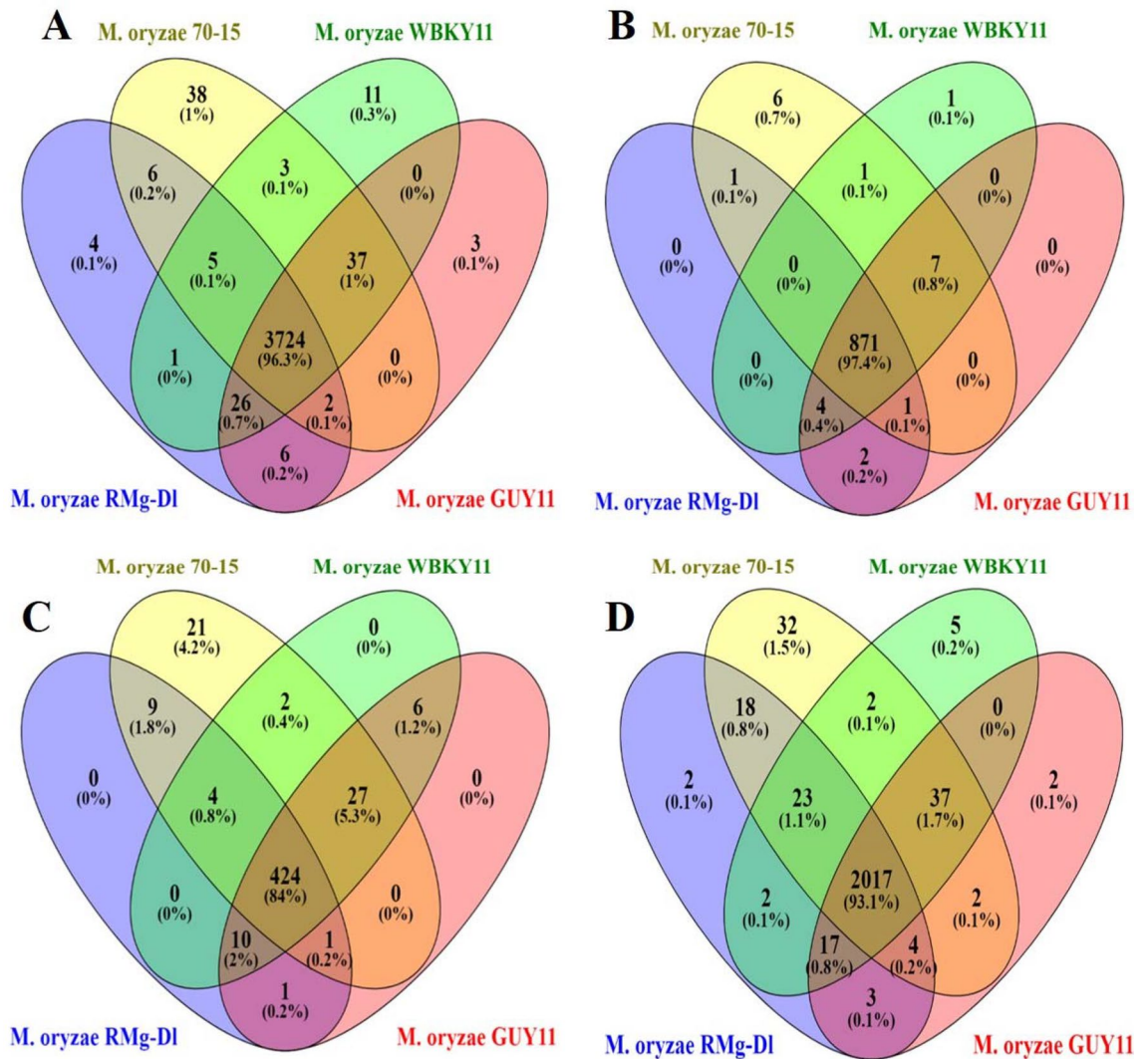


Figure 3. The predicted proteins InterProScan analysis of *M. oryzae* RMg_Dl, *M. oryzae* 70-15, *M. oryzae* WBKY11 and *M. oryzae* GUY11 genomes. The figure A, B, C and D shows the analyses against PFAM, superfamily, PIRSF and CDD databases, respectively.

mic orientation classification using Phobius tool showed that transmembrane, non-cytoplasmic domain, and cytoplasmic domain associated proteins were highly present in the genome (Supplementary Table S6).

Assembled genome functional annotation using gene ontology. The functional annotations of genes predicted were performed based on gene ontology using InterProScan and Blast2GO mapping and annotation. It provides functional signature vocabulary and hierarchical network relationships for the gene products in three classes: biological process, molecular function, and cellular component. Gene ontology (GO) mapping and annotation of sequences resulted in enrichment at level 2 category of the biological process revealed that mapped sequences ranged between 2746 and 7, molecular function ranged from 2787 to 7. The sequences assigned for cellular components ranged from 1696 to 528 (Fig. 4). Among the biological process, the majority of genes were linked with metabolic processes, cellular processes, localization, biological regulation, signaling, negative and positive regulation of biological processes (Fig. 4A). The molecular function associated GO terms were highly prevalent for catalytic function, binding, transporter activity, molecular function regulator, and structural molecule activity (Fig. 4B). Similarly, a cellular component associated terms were cellular anatomical activity, intracellular and protein-containing complex were prevalent (Fig. 4C).

Orthologous genes analysis. The orthologous features analysis for predicted proteins showed that 9216 orthologous genes were shared among RMg_Dl with 70-15, WBKY11 and GUY11 genomes, whereas 21 genes were found uniquely in RMg_Dl genome (Supplementary Fig. S1).

Identification of pathogenicity genes, virulence factors (VFs) and effectors. Further, the predicted proteins were analyzed against the pathogen-host interaction (PHI) database that revealed a total of 833

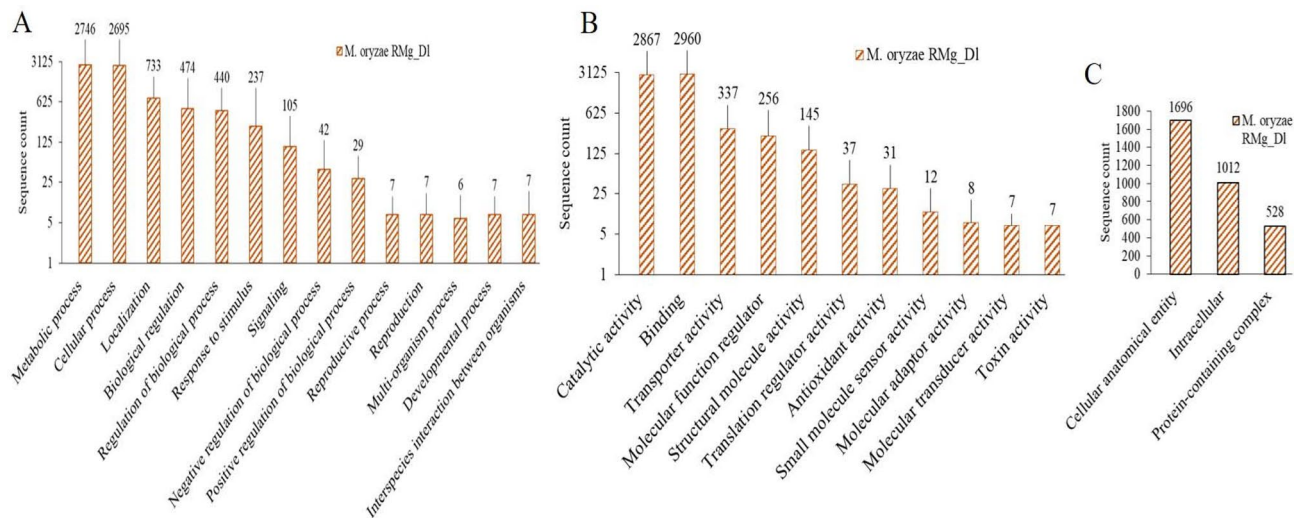


Figure 4. Functional annotation of predicted genes/proteins of *M. oryzae* RMg_DI in GO term: biological process (A) and GO term: Molecular Function (B), and cellular component (C).

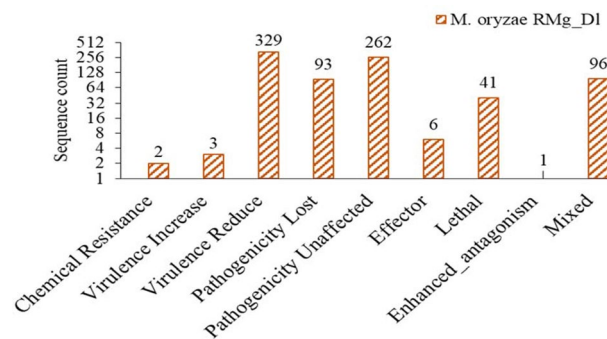


Figure 5. The assigned protein sequences to different PHI categories.

PHI genes were found in the RMg_DI genome. Among that, the majority of PHI genes were associated with reduced virulence and pathogenicity loss (Fig. 5). Also, there were a total of 51 different VFs identified, and its annotation showed that four copies of these VFs identified as gypsy retrotransposon, and function like pathogenicity, toxin activity, etc. (Supplementary Table S7). Further, these VFs consist of protein families namely Sell1 repeat, Heat-labile enterotoxin alpha chain, and WD domain, G-beta repeat, and superfamily namely ADP-ribosylation, S-adenosylmethionine synthetase, etc. (Supplementary Table S8).

The effectors were predicted through phobius mapping which showed 1986 signal peptide sequences, and then EffectorP3.0 was utilized for possible potential effector determination. This showed that a total of 603 effector genes, in which 330 were cytoplasmic and 273 were apoplasmic effectors (Supplementary Table S9). Thus, on an average, 5.72% of the total proteome content involved in effectors related functions. Upon the comparative study of these effectors with other genomes using orthologous approach, the total of 443 common effectors protein were documented, whereas 15 effectors were only shared among RMg_DI and 70-15 genomes (Supplementary Fig. S2). Further, 10 effectors genes were identified as AVR proteins, which classified to AVR-Pii in majority (4 copy), followed by Avr-Pik (2 copy) and Avr-Pita, AVR-Pita2, Avr-Pi54, AVR-Pita1 with single copy. These effectors CAZymes annotation showed that a total of 31 families were found with its total 64 copies in this genome. Among that family, AA9 was found with high occurrence, followed by CE5, CE1, AA16, and GH11 etc., in identified effectors (Supplementary Table S10). Additionally, PFAM annotation showed that a total of 142 different PFAMs were identified, which accounted for 237 copies. Among that glycosyl hydrolase family-61 was the most abundant followed by fungal cellulose binding domain, and chitin recognition protein etc. (Supplementary Table S11).

Identification of CAZymes. The CAZymes identification showed the presence of different 542 CAZyme families. These CAZymes extended analyses showed that the most abundant family was glycoside hydrolase (GH) (257 GH family). Next to GH, the auxiliary activities (AA) was the second-highest family (118 AA). The third highest abundant family was glycosyltransferase (GT) (94 GTs). Interestingly, the pectin lyase (PL) family showed the least abundance in the CAZyme family composition (Table 3). Further, the downstream analysis revealed that the AA7 and AA9 family was the most abundant, followed by GH3 and CE5 (Supplementary Table S12). Additionally, the other highly abundant families such as CE4, GH47, GH2, AA11, CE3, GH131,

Family	<i>M. oryzae</i> RMg_Dl	<i>P. oryzae</i> 70-15	<i>M. oryzae</i> WBKY11	<i>M. oryzae</i> GUY11
AA	118	130	121	115
CBM	14	14	14	14
CE	53	53	54	53
GH	257	256	263	260
GT	94	100	98	94
PL	6	7	7	7
Total	542	560	557	543

Table 3. The comparative CAZymes family profile in *M. oryzae* RMg_Dl, *P. oryzae* 70-15, *M. oryzae* WBKY11, and *M. oryzae* GUY11 genomes.

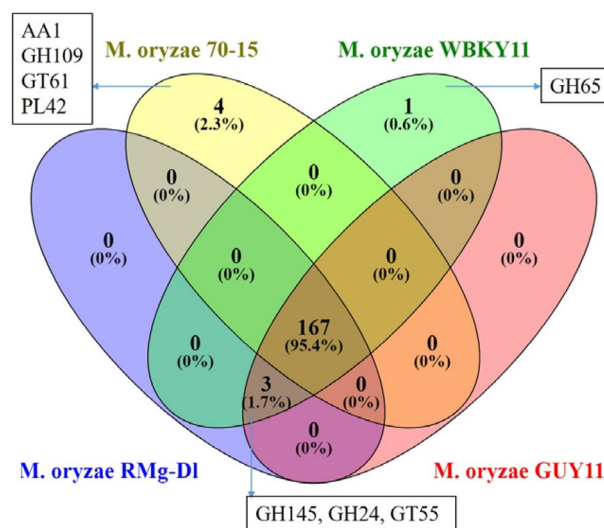


Figure 6. Identification and comparison of unique and shared CAZymes in *M. oryzae* RMg_Dl, *M. oryzae* 70-15, *M. oryzae* WBKY11 and *M. oryzae* GUY11 genomes.

GH31 were found to be equally present in studied genomes. Overall, a comparison of all CAZymes studied in genomes revealed that 167 (95.4%) families were shared between RMg_Dl, 70-15, WBKY11 and GUY11 genomes. Although, none of the CAZymes were found to be uniquely present in RMg_Dl whereas AA1, GH109, GT61, and PL42 families were uniquely detected in the 70-15 genome (Fig. 6).

Identification of metabolic pathways. The KEGG metabolic pathway analysis of the sequenced genome showed that the majority of genes were linked with metabolism and cellular process biochemical metabolic pathway categories. The detected pathways namely metabolism (KO:09100), in particular, carbohydrate metabolism (KO:09101) associated genes were 282 in the genome, and these genes were found to be distributed in fifteen different biochemical pathways (Table 4). Additionally, various genes are linked with pathways such as glycolysis/gluconeogenesis, starch, and sucrose metabolism, butanoate metabolism, propanoate metabolism, and inositol phosphate metabolism, fructose, and mannose and metabolism, pentose phosphate pathway were detected in the genome (Fig. 7, Supplementary Fig. S2). Among them, plant polysaccharide namely cellulose degradation key pathway, starch, and sucrose metabolism route-main enzyme endoglucanase (EC 3.2.1.4), cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91), and beta-glucosidase (EC 3.2.2.21) were found in the genome. Similarly, pentose and glucuronate interconversions pathways involve the pectin degradation enzyme endo-polygalacturonase (EC 3.2.1.15) and pectinesterase (EC 3.1.1.11) and pectate lyases (polygalacturonate lyase) (EC 4.2.2.2) were also found in this genome (Fig. 8, Supplementary Fig. S3). Analysis of the biochemical pathways associated with energy metabolism (KO:09102) resulted in distinguishing total of 134 genes distributed in six (6) different pathways. Among the pathways, the majority of the genes were associated with oxidative phosphorylation followed by sulfur and nitrogen metabolism (Supplementary Table S13).

The identification of environmental information processing (KO:09130) pathways revealed that 404 genes were in-particularly associated with signal transduction (KO:09132) pathways. The identified genes were involved in a total of thirty-two biological (biochemical) pathways (Table 5). These signaling pathways extended exploration showed that MAPK signaling pathway (Supplementary Fig. S4) associated genes were highest, followed by mTOR signaling pathway (KO:04150), PI3K-Akt signaling pathway (KO:04151), and AMPK signaling pathway

Carbohydrate metabolism pathways	<i>M. oryzae</i> RMg_DL (Gene count)
KO:00010 Glycolysis / Gluconeogenesis	26
KO:00020 Citrate cycle	21
KO:00030 Pentose phosphate pathway	17
KO:00040 Pentose and glucuronate inter conversions	19
KO:00051 Fructose and mannose metabolism	20
KO:00052 Galactose metabolism	15
KO:00053 Ascorbate and aldarate metabolism	5
KO:00500 Starch and sucrose metabolism	26
KO:00520 Amino sugar and nucleotide sugar metabolism	27
KO:00620 Pyruvate metabolism	29
KO:00630 Glyoxylate and dicarboxylate metabolism	23
KO:00650 Butanoate metabolism	12
KO:00640 Propanoate metabolism	18
KO:00660 C5-Branched dibasic acid metabolism	3
KO:00562 Inositol phosphate metabolism	21
Total	282

Table 4. Carbohydrate metabolism KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway for *M. oryzae* RMg_DL.

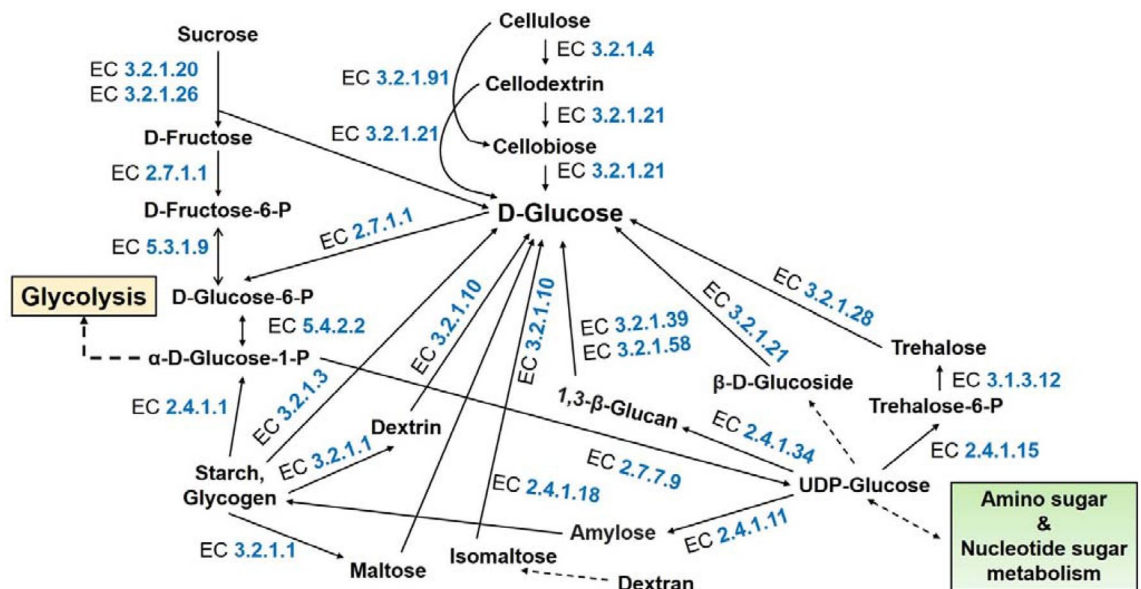


Figure 7. Starch and sucrose metabolism pathways depicting various enzymes involved including plant cell wall degrading enzymes. In here, the Sysname of enzymes related to the biosynthetic machinery are depicted in blue-colored EC numbers: The key to all of the EC numbers are as follows: EC 3.2.1.4, 4-β-D-glucan 4-glucohydrolase; EC 3.2.1.21, β-D-glucoside glucohydrolase; EC 3.2.1.91, 4-β-D-glucan cellobiohydrolase (non-reducing end); EC 3.2.1.20, α-D-glucoside glucohydrolase; EC 3.2.1.26, β-D-fructofuranoside fructohydrolase; EC 2.7.1.1, ATP:D-hexose 6-phosphotransferase; EC 5.3.1.9, D-glucose-6-phosphate aldose-ketose-isomerase; EC 2.4.1.15, UDP-α-D-glucose:D-glucose-6-phosphate 1-α-D-glucosyltransferase; EC 3.1.3.12, α, α-trehalose-6-phosphate phosphohydrolase; EC 3.2.1.28, α, α-trehalose glucohydrolase; EC 2.4.1.11, UDP-α-D-glucose: glycogen 4-α-D-glucosyltransferase; EC 2.4.1.18, (1->4)-α-D-glucan:(1->4)-α-D-glucan 6-α-D-[(1->4)-α-D-glucano]-transferase; EC 2.4.1.1, (1->4)-α-D-glucan:phosphate α-D-glucosyltransferase; EC 3.2.1.1, 4-α-D-glucan glucohydrolase; EC 3.2.1.3, 4-α-D-glucan glucohydrolase; EC 3.2.1.10, Oligosaccharide 6-α-glucohydrolase; EC 5.4.2.2, α-D-glucose 1,6-phosphomutase; EC 2.4.1.34, UDP-glucose:(1->3)-β-D-glucan 3-β-D-glucosyltransferase; EC 3.2.1.39, 3-β-D-glucan glucohydrolase; EC 3.2.1.58, 3-β-D-glucan glucohydrolase; EC 3.2.1.21, β-D-glucoside glucohydrolase and EC 2.7.7.9, UTP: α-D-glucose-1-phosphate uridylyltransferase.

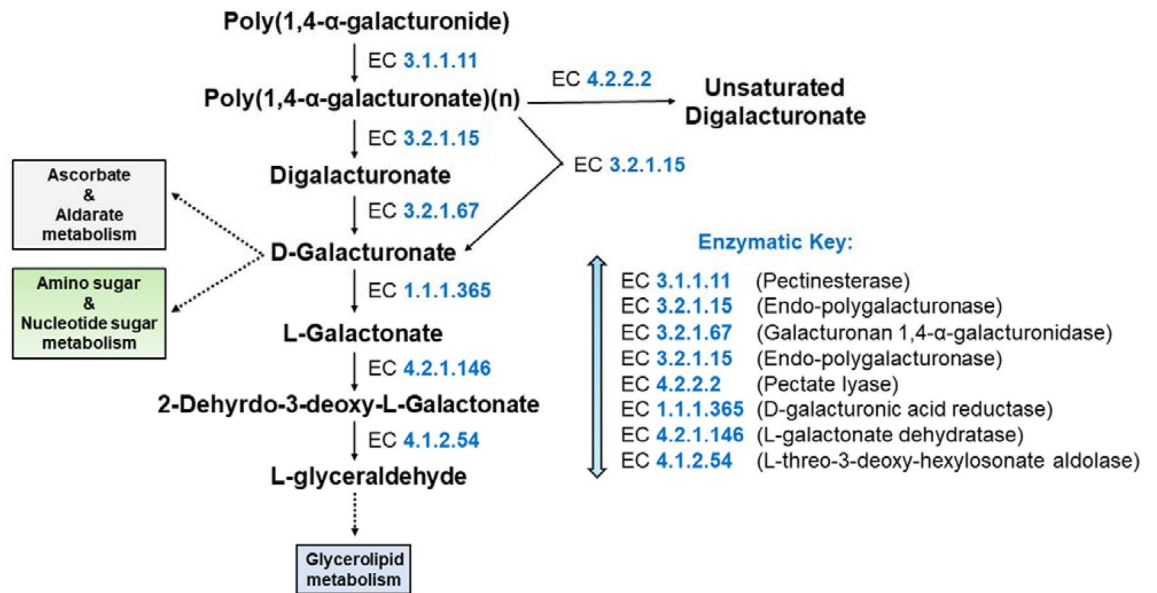


Figure 8. Pentose and glucuronate interconversion depicts the route enzymes of plant cell wall degrading enzymes.

(KO:04152). The other detected pathways were the two-component system, RAS signaling pathway, and calcium (Ca^{2+}) signaling pathway, etc.

Identification of variants. The SNPs and InDels calling showed that a total of 55,943 variants with their distribution as 49,065 SNPs, 3267 insertions and 3611 deletions in this genome. The SNPs chromosome wise classification showed that the highest SNPs were found on chromosome 1 and followed to chromosome 3 (Fig. 9A). The genome wide distribution of SNP density determination showed the occurrence of ~14 SNPs per 10 Kb in *M. oryzae* RMg_Dl genome. The identified SNPs transition and transversions were 39,326 and 9738, with their transition–transversion (Ts/TV) ratio was 4.04. The annotation of variants classified effects to region-wise, impact and functional class (Fig. 9). The majority of identified variants effect by region-wise were downstream (37%), followed to upstream (36%) and intergenic (18%) (Fig. 9B). The effect by impact showed that majority of variants depicted modifier (94.63%) followed to moderate (2.72%) and least for high (0.18%) impact (Fig. 9C). Moreover, effect by functional class, the missense type variants were 53.3% and silent were 45.78% (Fig. 9D). The conservative in-frame insertions and deletion were 0.03% and 0.03% respectively (Supplementary Table S14). Since the *M. oryzae* is a kind of rice blast disease causing pathogenic fungi, therefore we extensively performed the occurrence of variants in virulence factors and predicted potential effectors. Among the detected VFs, the total of 1340 variants were identified, in which 115 SNPs were identified as missense type in 7 different VFs. Also, majority of identified missense SNPs of VFs were found on chromosome 1, followed by chromosome 6, and VF namely *MGG_09263* was found with highest SNPs (82) (Supplementary Table S15). Similarly, for effectors, the total of 6743 variants were identified, among that 5768 were SNPs and 975 were indels. Effectors missense type SNPs determination showed that a total of 71 missense type SNPs, with highest occurrence on chromosome 1 followed by chromosome 4 and 2. The effector *MGG_17020* identified with highest missense SNPs, which was located on chromosome 4 (Supplementary Table S16). The occurrence of variants in AVR effectors were 75 in which 63 were SNPs and 12 were InDels. Among this, most of the variants were located in upstream region. The single missense SNP was found in *Avr-Pii*, and located on Chromosome2: 842166G > A (c.188C > T, p. Ala63Val). The effectors conservative and disruptive inframe InDels annotation showed that total of 29 InDels, in which conservative inframe deletion and insertions were 10 and 6, whereas disruptive inframe deletion and insertion were 6, and 7 respectively (Supplementary Table S17).

Discussion

The advent of high throughput sequencing technologies with short and long-read sequencing chemistry coupled with hybrid assembly have extensively facilitated in depth understanding of biochemical pathways, virulence factors, and conserved domains in the fungal genomes. In this study, we performed the WGS of rice blast disease-causing fungal species namely *Magnaporthe oryzae* RMg_Dl. The blast fungus is one of the wide spread pathogen reported to cause epidemics in different geographical locations and all major rice varieties^{6,41,42}. Therefore, the comparative whole genome alignment based phylogenetic tree was generated to decipher the evolutionary relationship of RMg_Dl with other 13 *Magnaporthe* genomes publically released from India, Japan, China, Thailand (Asia), USA and Guyana (Fig. S6). This analysis depicted the close relationship of Indian genomes with each other (Cluster 1), whereas moderately related to isolates representing Japan, Thailand, and shared distant relationship with Guyana (GUY11), USA (70-15) and China (HN19311, 98-06) genomes (Cluster 3).

Signal transduction pathways	<i>M. oryzae</i> RMg_DI (gene count)
KO:02020 Two-component system	19
KO:04014 Ras signaling pathway	15
KO:04015 Rap1 signaling pathway	6
KO:04010 MAPK signaling pathway	15
KO:04013 MAPK signaling pathway-fly	13
KO:04016 MAPK signaling pathway-plant	3
KO:04011 MAPK signaling pathway-yeast	56
KO:04012 ErbB signaling pathway	5
KO:04310 Wnt signaling pathway	13
KO:04330 Notch signaling pathway	4
KO:04340 Hedgehog signaling pathway	5
KO:04341 Hedgehog signaling pathway-fly	7
KO:04350 TGF-beta signaling pathway	6
KO:04390 Hippo signaling pathway	9
KO:04391 Hippo signaling pathway-fly	7
KO:04392 Hippo signaling pathway-multiple species	6
KO:04370 VEGF signaling pathway	7
KO:04371 Apelin signaling pathway	12
KO:04630 Jak-STAT signaling pathway	4
KO:04064 NF-kappa B signaling pathway	4
KO:04668 TNF signaling pathway	2
KO:04066 HIF-1 signaling pathway	15
KO:04068 FoxO signaling pathway	13
KO:04020 Calcium signaling pathway	9
KO:04070 Phosphatidylinositol signaling system	17
KO:04072 Phospholipase D signaling pathway	12
KO:04071 Sphingolipid signaling pathway	19
KO:04024 cAMP signaling pathway	9
KO:04022 cGMP-PKG signaling pathway	8
KO:04151 PI3K-Akt signaling pathway	24
KO:04152 AMPK signaling pathway	23
KO:04150 mTOR signaling pathway	37
Total	404

Table 5. Signal transduction KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway for *M. oryzae* RMg_DI.

This indicated the *M. oryzae* pathotypes representing same or proximal geographic locations (Asian continent) showed high genetic similarity (India, China, Thailand and Japan), except WBKY11 (USA) genome. In our previous study, we reported that diverse pathotypes of *Magnaporthe* were genetically homogenous indicating the trans-boundary movement across the continents⁴³. Such a pathogenic variation could be possibly associated with dynamic mechanism such as genetic mutations, recombination, geographical location and host resistance.

In the sequenced genome, we identified various genes involved in host–pathogen interaction such as virulence and pathogenicity mediators. Such occurrence of pathogenicity-related genes plays an essential role in initiating infection in the host^{42–44}. These genes mutation is believed to confer resistance against fungicides⁴⁵. Additionally, these genes have been documented for conferring fungicides resistance at the field level⁴⁶. In this study, we identified 51 different VFs containing 59 PFAM and 35 superfamilies, with four VFs identified as gypsy retrotransposon. The VFs were reported for functions like pathogenicity, toxin activity, etc. The NR database blast classified VFs namely natural trehalases acts as trehalose breakdown (component of plant cell wall) and regulated by signaling pathway and activation by phosphorylation and stimulated by Ca²⁺ and Mn²⁺⁴⁷. Similarly another VFs encode ‘ras-like protein ced-10’ involved in nutrient availability sensing and linked with cell growth and morphogenesis⁴⁸. The gypsy retrotransposon reported for DNA segment transposition, re-arrangement, genome evolution and widely distributed in ascomycota fungi and more details described here⁴⁹.

Among the total effectors, majority of effectors were cytoplasmic which acts initially upon infection enriches in biotrophic interfacial complex (BIC) prior getting transferred into plant cells, and continues its secretion after invasive hyphae development and invade adjacent plant cells⁵⁰. Whereas, apoplast effectors upon secretion, disseminated in the extracellular space of the fungal cell wall and extra-invasive-hyphal membrane (EIHM). These secreted effector are a small secreted proteins that alter host cell organization and function like manipulate plant immunity and physiology to promote infection through suppressing or activating effector-triggered

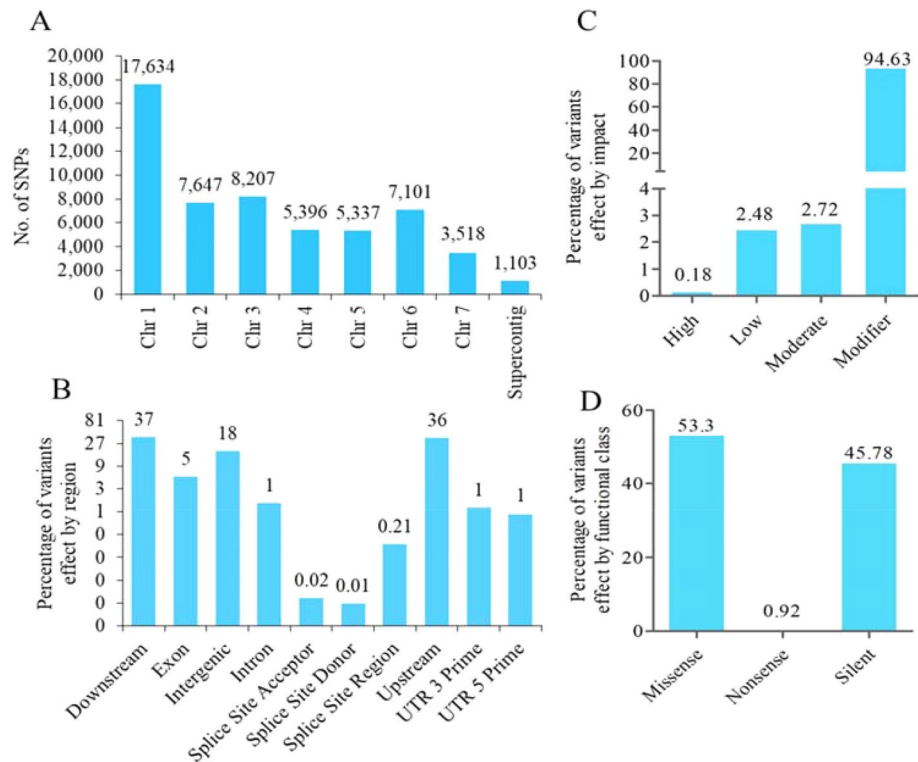


Figure 9. Identified variants and effect annotation of *M. oryzae* RMg_Dl genome. **A** = variants distribution by chromosome, **B** = effect by different region, **C** = effect by impact and **D** = effect by functional class.

immunity (ETI)⁵¹. The AVR-Pii and AVR-Pik were reported for genomic instability⁵². AVR-Pii, Avr-Pita effector reported for damaging the host innate immunity⁵³. Functional annotation of effectors using PFAM revealed that total of 237 families, but there are decreased annotation compared to total effectors, because possibly most of the effectors were classified as putative uncharacterized/hypothetical proteins. Therefore, utilizing the resources such as protein domains, families and superfamily driven classification could provide more insights of possible functional features of proteins and thus predicted effectors also²⁶. Using this approach, the documented effectors classification showed that higher occurrence of glycosyl hydrolase family 61, fungal cellulose binding domain and cutinase were observed. In particular, cutinase family associated with diverse functions such as cell surface recognition, germinal differentiation, appressorial growth, host infiltration and then virulence maturation. Moreover, role of this protein documented for cyclic AMP/protein kinase A and diacylglycerol protein kinase C signalling upon activation which eventually triggers appressorium development and infection progress by this fungus⁵⁴.

Moreover, we identified a total of 49,065 SNPs, 3267 insertions and 3611 deletions in this genome with ~14 SNPs per 10 Kb. The majority of variants were found on chromosome 1 and similar result was reported for other *M. oryzae* strains⁵⁵. Further, virulence associated SNPs were located on chromosome 1, 2, 3, 4 and 7⁵⁶. In this study, we found various SNPs on these chromosomes. The variants annotation effects classified into region wise effect showed that majority of variants were located in the gene regulatory elements namely downstream and upstream regions. The occurrence of variants in these regions potentially involve in gene expression regulation resulted in upregulation or downregulation. Interestingly, we observed various missense SNPs which causes the amino acid change in protein structure resulting to altered protein property and functionality. This could be also associated with prompting infection, severity and susceptibility with increased incidence of fungal disease⁵⁷. We documented various SNPs among genomic regions of upstream, downstream, and untranslated region (UTR) region, which influence the gene expression and regulation at the post-transcriptional level and protein synthesis⁵⁸. The 5'UTR mutation influence the binding of proteins and results to stimulation or suppression of translational regulation. Meanwhile 3'UTR mutation reported to influence the binding sites of miRNA and polyA which affect the translational deregulation⁵⁹. In this study, 54 effectors genes were documented with missense and InDels genetic variations, which accounted nearly 0.09% (54/603) of total effectors genes. Though, there were a single missense SNPs was found on Avr-Pii effector. Also we found various SNPs in AVR-Pita1 and AVR-Pita2, and previously reported for high variability and its diversification⁶⁰.

The study of genomes for families, conserved domains through InterProScan resulted in the identification of nearly 96% PFAM, 84% PIRSE, and 93% CDD were shared in the studied genomes. Further, in the present study, the majority of PFAMs namely WD domain, G-beta repeat, major facilitator superfamily, and cytochrome P450 family, P-loop containing nucleoside triphosphate hydrolase, NAD(P)-binding domain, and alpha/beta hydrolase fold superfamily were expanded the genome. These detected families and conserved domains were documented for various mechanisms viz. MAPK for growth⁶¹, virulence, and pathogenicity⁶², ABC transporter for virulence⁶³ that overall mediate multiple significant roles in pathogen-host interaction⁶⁴.

Repeats are also unique features of fungal genomes, and these identified SSRs could be utilized for fungal diversity study and disease management, species/strain differentiation, and detection⁶⁵. The rich genomic bases of GC and AT are applied for the assessment of fungal defense against transposon expansion which works through repeat-induced point mutation (RIP) and genome evolution. Such processes play a differentiating choice and facilitation of host-microbe interaction⁶⁶. In this study, we documented more than 18,000 SSRs which can be applied for the study of fungal population diversity and potential application towards making an efficient management strategy for disease prevention. Such methods are efficiently applied for controlling the citrus leaf and fruit disease caused by fungi⁶⁷.

The study on fungal genomes about the content of CAZymes showed that this organism contains distinct types of CAZyme families, which facilitates fungus to efficiently degrade host complex polysaccharides. The CAZymes also play an essential role in pathogen-host interactions (PHI) as pathogenic fungus invades host on plant cell wall via the action of cell wall degrading enzymes¹². We detected the high occurrence of GH families, and studies reported that the GH family plays a potential role in the breakdown of complex polysaccharides coupled with additional CAZymes like CBM and PL, which increase the breakdown efficacy and pave the path for other breakdown processes in the environment⁶⁹. Moreover, the genomic study of various pathogenic fungal species showed vast heterogeneity in CAZymes content because of their host specificity and nutritional requirement, which also facilitated the digestion of complex plant polysaccharides for nutritive addition and later simplify the infection process^{12,70}. The pathogenic fungus is reported to contain a distinct number of CAZymes, whereas less than saprophytic and necrophytic fungi¹². The high content of various GH family also reported in fungus invading the monocot than dicot plant because monocot enriched with polysaccharides^{12,70}.

The biochemical pathway determination detected the presence of various metabolic pathways associated with variable enzymes in the genome. Among that plant polysaccharide namely cellulose degradation key pathway, starch and sucrose metabolism route-main enzyme endoglucanase (EC 3.2.1.4), cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91), and beta-glucosidase (EC 3.2.2.21) were found in the genome. The endoglucanase (EC 3.2.1.4) initially degrades cellulose to cellobiose and then to cellobiose. Cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) degrades i) cellulose to cellobiose, ii) cellobiose to glucose. Then, beta-glucosidase (EC 3.2.2.21) generates glucose via two ways- i) cellobiose to glucose, and ii) cellobiose to glucose (Fig. 7). Further cell wall degrading enzyme endo-polygalacturonase (EC 3.2.1.15) of PL29 CAZymes, and pectate lyases (EC 4.2.2.2) of PL1 CAZymes and endo-1,4-beta-xylanase (EC:3.2.1.8) were detected which further mediate a typical role in infection and utilizes it for own growth and reproduction^{71,72}. These enzymes are recognized for cell wall degradation enzyme (CWDE) which facilitates the infection and plant cell damage^{70,71}. The enzymatically released glucose and other compounds were utilized by fungi for their growth, reproduction, and proliferation in the host⁷¹.

Functional pathway two-component system also plays an essential role in rice blast disease incidence along with the combined mechanism of MAP kinase and cAMP signaling mediated formation of infection component on rice host^{62,73}. Reports also demonstrated that cAMP signaling was associated with fungal growth and infection development⁷³⁻⁷⁵. Such kinase signaling mechanisms include the phosphorylation mediated signaling process, chemotaxis, virulence process, and secondary metabolite production in fungi⁷⁶. Additionally, the mechanism of histidine kinase-mediated environmental responses, pathogenicity, hyphal development, and then sporulation have also been documented^{8,77,78}. The MAPK pathway descriptive study revealed that various enzymes of *M. oryzae* RMg_D1 were mapped to signaling mechanisms such MAPKKK, MAPKK, and MAPK (Supplementary Fig. S4). Additionally, the protein domains are linked to each other for creating multi-domain protein structures to gain a wide range of functional property^{79,80}. This can be exemplified through flexible architectures of signaling pathways like mitogen-activated protein kinase (MAPK) cascades^{81,82}. This pathway is associated with controlling for various biological processes such as metabolism, cellular morphology, cell cycle progress, and gene expression in the influence of any extracellular signals or stimuli⁸³ and cellular signaling and pathogenesis-related structure development^{64,84}. In particular, among pathogenic fungi, metabolic pathways, ATP-binding cassette (ABC) transporters are primarily involved in defense activity against secondary metabolites or toxins secreted by the host⁸⁵.

In this sequenced genome, we observed various transport families such as the ABC transporter, phosphate transporter family, major facilitator superfamily (MFS) involved in the transport of a broad range of minerals and nutrients. The role of ABC and MFS transporters documented for the protection against natural toxic substance exists in the atmosphere or synthetic toxic compounds such as fungicides and antimycotic agents^{41,86}. Pitkin, et al.⁸⁷ reported that ABC transporters also offer defense against antimicrobial agents. Additionally, crucial role in host pathogenicity while providing protection against host defense mechanisms or releasing host-specific toxins. These transporters play as a resistance barrier against various fungicides, though prolonged uses of fungicidal agents resulted in the occurrence of resistance in ABC and MFS transporter including chemically unrelated agents and causes the decreased accumulation of these agents^{41,86}. We also detected superfamily namely cytochrome P450 (CYP) monooxygenase superfamily which is reported to be involved in a wide range of functions such as multidimensional metabolic activity and support to survival in a distinct ecological environment⁸⁸ with a contribution in infection occurrence. The fungal CYP associated with distinct kind of secondary metabolites production for its own protection and compete against attacking organism such as bacteria, plants, animals and also against other fungi^{89,90}. These compounds act as wide range of beneficial role like antibiotic, immune suppressor and mycotoxic actions^{91,92}.

Conclusion

In the present study, the whole genome sequencing of filamentous rice blast disease-causing fungus *M. oryzae* RMg_D1 was performed. The functional annotation revealed the presence of distinct enzymes linked with various metabolic pathways. The study also documented various carbohydrate metabolism-associated pathways that included starch and sucrose metabolism, pentose and glucuronate interconversion, and signaling associated

namely MAPK, cAMP pathways. We also observed various CAZymes with high content of the GH family, pathogen-host interaction-related genes, effectors and virulence factors. This information serves as a genomic architecture for the optimization of genus *Magnaporthe* mediated blast disease management and assessment of population diversity. Moreover, detected genes or proteins can be utilized as potential targets for marker development to screen the blast pathogenic organisms.

Received: 17 June 2021; Accepted: 1 November 2021

Published online: 25 November 2021

References

- Ronald, P. Plant genetics, sustainable agriculture and global food security. *Genetics* **188**, 11–20. <https://doi.org/10.1534/genetics.111.128553> (2011).
- Sharma, T. R. *et al.* Rice blast management through host-plant resistance: retrospect and prospects. *Agric. Res.* **1**, 37–52. <https://doi.org/10.1007/s40003-011-0003-5> (2012).
- Mehta, S., Singh, B., Dhakate, P., Rahman, M. & Islam, M. A. Rice, marker-assisted breeding, and disease resistance. In *Disease Resistance in Crop Plants: Molecular, Genetic and Genomic Perspectives* (ed. Wani, S. H.) 83–111 (Springer International Publishing, Cham, 2019).
- Dean, R. *et al.* The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430. <https://doi.org/10.1111/j.1364-3703.2011.00783.x> (2012).
- Miah, G. *et al.* Blast resistance in rice: a review of conventional breeding to molecular approaches. *Mol. Biol. Rep.* **40**, 2369–2388. <https://doi.org/10.1007/s11033-012-2318-0> (2013).
- Howard, R. J. & Valent, B. Breaking and entering: host penetration by the fungal rice blast pathogen *Magnaporthe grisea*. *Annu. Rev. Microbiol.* **50**, 491–512. <https://doi.org/10.1146/annurev.micro.50.1.491> (1996).
- Talbot, N. J. On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*. *Annu. Rev. Microbiol.* **57**, 177–202. <https://doi.org/10.1146/annurev.micro.57.030502.090957> (2003).
- Nemecek, J. C., Wuthrich, M. & Klein, B. S. Global control of dimorphism and virulence in fungi. *Science* **312**, 583–588. <https://doi.org/10.1126/science.1124105> (2006).
- TeBeest, D., Guerber, C. & Dittmore, M. Rice blast. In: *The Plant Health Instructor*. <https://doi.org/10.1094/PHI-I-2007-0313-07> (2007).
- Valent, B. Rice blast as a model system for plant pathology. *Phytopathology* **80**, 33–36 (1990).
- Acero, F. J. *et al.* Development of proteomics-based fungicides: new strategies for environmentally friendly control of fungal plant diseases. *Int. J. Mol. Sci.* **12**, 795–816. <https://doi.org/10.3390/ijms12010795> (2011).
- Iqbal, M. A. *et al.* Draft whole genome sequence of groundnut stem rot fungus *Helminthosporium solani* revealing genetic architect of its pathogenicity and virulence. *Sci. Rep.* **7**, 5299. <https://doi.org/10.1038/s41598-017-05478-8> (2017).
- Kumar, A. *et al.* Genome sequence of a unique *Magnaporthe oryzae* RMg_DL isolate from India that causes blast disease in diverse cereal crops, obtained using PacBio Single-Molecule and Illumina HiSeq2500 sequencing. *Genome Announc.* <https://doi.org/10.1128/genomeA.01570-16> (2017).
- Prakash, G. *et al.* First draft genome sequence of a Pearl Millet blast pathogen, *Magnaporthe grisea* strain PMg_DL, obtained using PacBio Single-Molecule Real-Time and Illumina NextSeq 500 Sequencing. *Microbiol. Resour. Announc.* <https://doi.org/10.1128/MRA.01499-18> (2019).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
- Salmela, L. & Rivals, E. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30**, 3506–3514. <https://doi.org/10.1093/bioinformatics/btu538> (2014).
- Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods.* **17**, 155–158. <https://doi.org/10.1038/s41592-019-0669-3> (2020).
- Xu, G. C. *et al.* LR_GapCloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* <https://doi.org/10.1093/gigascience/giy157> (2019).
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> (2015).
- Lee, I., Ouk Kim, Y., Park, S. C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **66**, 1100–1103. <https://doi.org/10.1099/ijsem.0.000760> (2016).
- Humann, J. L., Lee, T., Ficklin, S. & Main, D. Structural and functional annotation of eukaryotic genomes with GenSAS. *Methods Mol. Biol.* **29–51**, 2019. https://doi.org/10.1007/978-1-4939-9173-0_3 (1962).
- Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **Chapter 4**, Unit 4.10. <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
- Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–467. <https://doi.org/10.1093/nar/gki458> (2005).
- Lagesen, K. *et al.* RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108. <https://doi.org/10.1093/nar/gkm160> (2007).
- Chan, P. P. & Lowe, T. M. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1–14**, 2019. https://doi.org/10.1007/978-1-4939-9173-0_1 (1962).
- Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031> (2014).
- Conesa, A. & Gotz, S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, 619832. <https://doi.org/10.1155/2008/619832> (2008).
- Cantarel, B. L. *et al.* The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–238. <https://doi.org/10.1093/nar/gkn663> (2009).
- Huang, L. *et al.* dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* **46**, D516–D521. <https://doi.org/10.1093/nar/gkx894> (2018).
- Lu, T., Yao, B. & Zhang, C. DFVF: database of fungal virulence factors. *Database (Oxford)* **2012**, bas32. <https://doi.org/10.1093/database/bas032> (2012).
- Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60. <https://doi.org/10.1038/nmeth.3176> (2015).
- Sperschneider, J. & Dodds, P. N. EffectorP 3.0: prediction of apoplasmic and cytoplasmic effectors in fungi and oomycetes. *bioRxiv*. <https://doi.org/10.1101/2021.07.28.454080> (2021).

33. Xu, L. *et al.* OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **47**, W52–W58. <https://doi.org/10.1093/nar/gkz333> (2019).
34. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–185. <https://doi.org/10.1093/nar/gkm321> (2007).
35. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361. <https://doi.org/10.1093/nar/gkw1092> (2017).
36. Knudsen, S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* **15**, 356–361. <https://doi.org/10.1093/bioinformatics/15.5.356> (1999).
37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> (2009).
38. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
39. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92. <https://doi.org/10.4161/fly.19695> (2012).
40. Darling, A. E., Mau, B. & Perna, N. T. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLOS ONE* **5**, e11147. <https://doi.org/10.1371/journal.pone.0011147> (2010).
41. Stergiopoulos, I., Zwiers, L.-H. & De Waard, M. A. Secretion of natural and synthetic toxic compounds from filamentous fungi by membrane transporters of the atp-binding cassette and major facilitator superfamily. *Eur. J. Plant Pathol.* **108**, 719–734. <https://doi.org/10.1023/A:1020604716500> (2002).
42. Zhang, H., Zheng, X. & Zhang, Z. The *Magnaporthe grisea* species complex and plant pathogenesis. *Mol. Plant Pathol.* **17**, 796–804. <https://doi.org/10.1111/mpp.12342> (2016).
43. Sheoran, N., Ganesan, P., Mughal, N. M., Yadav, I. S. & Kumar, A. Genome assisted molecular typing and pathotyping of rice blast pathogen, *Magnaporthe oryzae*, reveals a genetically homogenous population with high virulence diversity. *Fungal Biol.* **125**, 733–747. <https://doi.org/10.1016/j.funbio.2021.04.007> (2021).
44. Gupta, L., Vermani, M., Kaur Ahluwalia, S. & Vijayaraghavan, P. Molecular virulence determinants of *Magnaporthe oryzae*: disease pathogenesis and recent interventions for disease management in rice plant. *Mycology* **12**, 174–187. <https://doi.org/10.1080/21501203.2020.1868594> (2021).
45. Cools, H. J. & Hammond-Kosack, K. E. Exploitation of genomics in fungicide research: current status and future perspectives. *Mol. Plant Pathol.* **14**, 197–210. <https://doi.org/10.1111/mpp.12001> (2013).
46. Brent, K. J. & Hollomon, D. W. Fungicide resistance: the assessment of risk. *Fungicide Resistance Action Committee 2007, FRAC Monograph No.2 second, (revised) edition* (2007).
47. Jorge, J. A., Polizeli, M. D. L. T. M., Thevelein, J. M. & Terenzi, H. F. Trehalases and trehalose hydrolysis in fungi. *FEMS Microbiol. Lett.* **154**, 165–171. <https://doi.org/10.1111/j.1574-6968.1997.tb12639.x> (1997).
48. Harispe, L., Portela, C., Scazzocchio, C., Peñalva, M. A. & Gorfinkiel, L. Ras GTPase-activating protein regulation of actin cytoskeleton and hyphal polarity in *Aspergillus nidulans*. *Eukaryot. Cell* **7**, 141–153. <https://doi.org/10.1128/EC.00346-07> (2008).
49. Muszewska, A., Hoffman-Sommer, M. & Grynberg, M. LTR retrotransposons in fungi. *PLOS ONE* **6**, e29425. <https://doi.org/10.1371/journal.pone.0029425> (2011).
50. Zhang, S. & Xu, J. R. Effectors and effector delivery in *Magnaporthe oryzae*. *PLoS Pathog.* **10**, e1003826. <https://doi.org/10.1371/journal.ppat.1003826> (2014).
51. 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. <https://doi.org/10.1094/mpmi-22-2-0115> (2009).
52. Bialas, A. *et al.* Lessons in effector and nlr biology of plant-microbe systems. *Mol. Plant Microbe Interact.* **31**, 34–45. <https://doi.org/10.1094/mpmi-08-17-0196-fi> (2018).
53. Han, J. *et al.* The fungal effector Avr-pita suppresses innate immunity by increasing COX activity in rice mitochondria. *Rice (N Y)* **14**, 12. <https://doi.org/10.1186/s12284-021-00453-4> (2021).
54. Skamnioti, P. & Gurr, S. J. *Magnaporthe grisea* cutinase2 mediates appressorium differentiation and host penetration and is required for full virulence. *Plant Cell* **19**, 2674–2689. <https://doi.org/10.1105/tpc.107.051219> (2007).
55. Cao, J. *et al.* Genome re-sequencing analysis uncovers pathogenecity-related genes undergoing positive selection in *Magnaporthe oryzae*. *Sci. China Life Sci.* **60**, 880–890. <https://doi.org/10.1007/s11427-017-9076-4> (2017).
56. Korinsak, S. *et al.* Genome-wide association mapping of virulence gene in rice blast fungus *Magnaporthe oryzae* using a genotyping by sequencing approach. *Genomics* **111**, 661–668. <https://doi.org/10.1016/j.ygeno.2018.05.011> (2019).
57. Naik, B., Ahmed, S. M. Q., Laha, S. & Das, S. P. Genetic susceptibility to fungal infections and links to human ancestry. *Front. Genet.* <https://doi.org/10.3389/fgene.2021.709315> (2021).
58. Blake, W. J. *et al.* Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* **24**, 853–865. <https://doi.org/10.1016/j.molcel.2006.11.003> (2006).
59. Chatterjee, S. & Pal, J. K. Role of 5' and 3'-untranslated regions of mRNAs in human diseases. *Biol. Cell* **101**, 251–262. <https://doi.org/10.1042/bc20080104> (2009).
60. Chuma, I. *et al.* Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast Fungus *Magnaporthe oryzae* and related species. *PLOS Pathog.* **7**, e1002147. <https://doi.org/10.1371/journal.ppat.1002147> (2011).
61. Zhao, X., Mehrabi, R. & Xu, J. R. Mitogen-activated protein kinase pathways and fungal pathogenesis. *Eukaryot. Cell* **6**, 1701–1714. <https://doi.org/10.1128/EC.00216-07> (2007).
62. Leng, Y. & Zhong, S. The role of mitogen-activated protein (MAP) kinase signaling components in the fungal development, stress response and virulence of the fungal cereal pathogen *Bipolaris sorokiniana*. *PLOS ONE* **10**, e0128291. <https://doi.org/10.1371/journal.pone.0128291> (2015).
63. Roohparvar, R., Huser, A., Zwiers, L. H. & De Waard, M. A. Control of *Mycosphaerella graminicola* on wheat seedlings by medical drugs known to modulate the activity of ATP-binding cassette transporters. *Appl. Environ. Microbiol.* **73**, 5011–5019. <https://doi.org/10.1128/AEM.00285-07> (2007).
64. Hamel, L. P., Nicole, M. C., Duplessis, S. & Ellis, B. E. Mitogen-activated protein kinase signaling in plant-interacting fungi: distinct messages from conserved messengers. *Plant Cell* **24**, 1327–1351. <https://doi.org/10.1105/tpc.112.096156> (2012).
65. Canfora, L. *et al.* Development of a method for detection and quantification of *B. brongniartii* and *B. bassiana* in soil. *Sci. Rep.* **6**, 22933. <https://doi.org/10.1038/srep22933> (2016).
66. Testa, A. C., Oliver, R. P. & Hane, J. K. OcculterCut: a comprehensive survey of AT-rich regions in fungal genomes. *Genome Biol. Evol.* **8**, 2044–2064. <https://doi.org/10.1093/gbe/evw121> (2016).
67. Moges, A. D. *et al.* Development of microsatellite markers and analysis of genetic diversity and population structure of *Colletotrichum gloeosporioides* from Ethiopia. *PLOS ONE* **11**, e0151257. <https://doi.org/10.1371/journal.pone.0151257> (2016).
68. Hervé, C. *et al.* Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc. Natl. Acad. Sci. USA* **107**, 15293–15298. <https://doi.org/10.1073/pnas.1005732107> (2010).
69. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res.* **42**, D490–D495. <https://doi.org/10.1093/nar/gkt1178> (2013).
70. Zhao, Z., Liu, H., Wang, C. & Xu, J. R. Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **15**, 6. <https://doi.org/10.1186/1471-2164-15-6> (2014).

71. Kubicek, C. P., Starr, T. L. & Glass, N. L. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu. Rev. Phytopathol.* **52**, 427–451. <https://doi.org/10.1146/annurev-phyto-102313-045831> (2014).
72. Lee, D. K. *et al.* Metabolic response induced by parasitic plant-fungus interactions hinder amino sugar and nucleotide sugar metabolism in the host. *Sci. Rep.* **6**, 37434. <https://doi.org/10.1038/srep37434> (2016).
73. Lee, Y. H. & Dean, R. A. cAMP regulates infection structure formation in the plant pathogenic fungus *Magnaporthe grisea*. *Plant Cell* **5**, 693–700. <https://doi.org/10.1105/tpc.5.6.693> (1993).
74. Xu, J. R. & Hamer, J. E. MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*. *Genes Dev.* **10**, 2696–2706. <https://doi.org/10.1101/gad.10.21.2696> (1996).
75. Adachi, K. & Hamer, J. E. Divergent cAMP signaling pathways regulate growth and pathogenesis in the rice blast fungus *Magnaporthe grisea*. *Plant Cell* **10**, 1361–1374. <https://doi.org/10.1105/tpc.10.8.1361> (1998).
76. Wolanin, P. M., Thomason, P. A. & Stock, J. B. Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biol.* **3**, REVIEWS3013. <https://doi.org/10.1186/gb-2002-3-10-reviews3013> (2002).
77. Hohmann, S. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.* **66**, 300–372. <https://doi.org/10.1128/mmb.66.2.300-372.2002> (2002).
78. Jacob, S., Foster, A. J., Yemelin, A. & Thines, E. Histidine kinases mediate differentiation, stress response, and pathogenicity in *Magnaporthe oryzae*. *Microbiol. Open* **3**, 668–687. <https://doi.org/10.1002/mbo.3.197> (2014).
79. Bornberg-Bauer, E. & Alba, M. M. Dynamics and adaptive benefits of modular protein evolution. *Curr. Opin. Struct. Biol.* **23**, 459–466. <https://doi.org/10.1016/j.sbi.2013.02.012> (2013).
80. Lees, J. G., Dawson, N. L., Sillitoe, I. & Orengo, C. A. Functional innovation from changes in protein domains and their combinations. *Curr. Opin. Struct. Biol.* **38**, 44–52. <https://doi.org/10.1016/j.sbi.2016.05.016> (2016).
81. Pawson, T. & Scott, J. D. Signaling through scaffold, anchoring, and adaptor proteins. *Science* **278**, 2075–2080. <https://doi.org/10.1126/science.278.5346.2075> (1997).
82. Rubin, G. M. The draft sequences. Comparing species. *Nature* **409**, 820–821. <https://doi.org/10.1038/35057277> (2001).
83. Chen, R. E. & Thorner, J. Function and regulation in MAPK signaling pathways: lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta.* **1773**, 1311–1340. <https://doi.org/10.1016/j.bbamcr.2007.05.003> (2007).
84. Dixon, K. P., Xu, J. R., Smirnov, N. & Talbot, N. J. Independent signaling pathways regulate cellular turgor during hyperosmotic stress and appressorium-mediated plant infection by *Magnaporthe grisea*. *Plant Cell* **11**, 2045–2058. <https://doi.org/10.1105/tpc.11.10.2045> (1999).
85. Morschhauser, J. Regulation of multidrug resistance in pathogenic fungi. *Fungal Genet. Biol.* **47**, 94–106. <https://doi.org/10.1016/j.fgb.2009.08.002> (2010).
86. de Waard, M. A. Significance of ABC transporters in fungicide sensitivity and resistance. *Pestic. Sci.* **51**, 271–275 (1997).
87. Pitkin, J. W., Panaccione, D. G. & Walton, J. D. A putative cyclic peptide efflux pump encoded by the TOXA gene of the plant-pathogenic fungus *Cochliobolus carbonum*. *Microbiology* **142**, 1557–1565. <https://doi.org/10.1099/13500872-142-6-1557> (1996).
88. Moktali, V. *et al.* Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC Genomics* **13**, 525. <https://doi.org/10.1186/1471-2164-13-525> (2012).
89. Casadevall, A. Determinants of virulence in the pathogenic fungi. *Fungal Biol. Rev.* **21**, 130–132. <https://doi.org/10.1016/j.fbr.2007.02.007> (2007).
90. Kelly, S. L. & Kelly, D. E. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us?. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20120476. <https://doi.org/10.1098/rstb.2012.0476> (2013).
91. Hoffmeister, D. & Keller, N. P. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat. Prod. Rep.* **24**, 393–416. <https://doi.org/10.1039/b603084j> (2007).
92. Shin, J., Kim, J. E., Lee, Y. W. & Son, H. Fungal cytochrome P450s and the P450 complement (CYPome) of *Fusarium graminearum*. *Toxins (Basel)* **10**, 112. <https://doi.org/10.3390/toxins10030112> (2018).

Acknowledgements

All authors are thankful to the Indian Council of Agricultural Research-Consortium Research Project On Genomics [CRP (Genomic)/VIII/2021-22/II] for funding the WGS of *Magnaporthe* and National Agricultural Higher Education Project (NAHEP)-CAAST on ‘Genomics-Assisted Crop Improvement and Management’ (NAHEP/CAAST/2018-19/07) for assisting the bioinformatic analysis. We are thankful to the Director, ICAR-IARI, New Delhi for providing all infrastructure facilities. Furthermore, the authors would also like to acknowledge the anonymous peer reviewers who gave constructive comments throughout the entire review process.

Author contributions

A.K. and V.C. conceived the grant. A.K. and B.R. designed the study. BR performed bioinformatic analysis and drafted manuscript, and N.S. isolated fungus and other wet-lab experiment. A.K., S.M., N.S., V.C. and P.G. edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-01980-2>.

Correspondence and requests for materials should be addressed to A.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021