Chapter 2
Introduction, Objectives and Materials & Methods

2.1 Introduction

Phytopathogenic fungi pose a significant threat to global agriculture, and their ability to adapt to specific host plants is a pressing concern. The emergence of new pathogenic strains through host range expansion or adaptation to novel hosts can lead to devastating disease outbreaks, with far-reaching economic and ecological consequences. Understanding the molecular basis of host adaptation is essential to enhance crop resistance against these diseases (Giraud et al., 2010; Gladieux et al., 2014). To grasp the intricate dynamics of disease emergence and spread, it is crucial to explore the interactions between pathogen species, lineages, genetic variability, and reproductive strategies (Taylor & Fisher, 2003). Among these fungal adversaries, cereal blast disease stands out as a formidable adversary, causing annual losses ranging from 10% to a staggering 100% under favourable conditions. This destructive disease is primarily caused by the filamentous fungus *Magnaporthe oryzae*, also known as *Pyricularia oryzae*, responsible for cereal blast infections worldwide (Talbot, 2003). *M. oryzae* occupies a prominent position in the realm of fungal pathogens and serves as a critical model organism for unravelling the complexities of host-pathogen interactions (Dean et al., 2012; Ebbole, 2007). This adaptable pathogen possesses the remarkable ability to infect various aerial parts of its host plant, including leaves, necks, panicles, collars, nodes, and roots in rice (Marcel et al., 2010; Sesma & Osbourn, 2004). Within the *Magnaporthe* species complex, a diverse array of phylogenetic species (Couch et al, 2002) afflict over 50 different grass species (Ou, 1987). The epicentre of blast fungus origination lies in South-East Asia, with historical dispersals of *M. oryzae* populations from Asia to other continents (Saleh et al, 2014). The population structure of *M. oryzae* exhibits diversity in Asian regions, characterized by numerous lineages, contrasting the simpler clonal populations observed in the USA and Europe (Skamnioti & Gurr, 2009). Remarkably, *M. oryzae*, despite its multi-host nature, has diverged into numerous lineages with limited host ranges, suggesting ongoing speciation, host shifts, or host range expansions within this species (Gladieux et al, 2018). The question of how these intraspecific lineages have emerged remains a formidable and largely unresolved challenge in the realm of plant pathology.

While the precise evolutionary mechanism remain elusive, many effectors (small secreted proteins and secondary metabolites) are believed to be involved in the host-specific infection

of this pathogen. Notably, the virulence potential of *M. oryzae* populations may hinge on effector molecules, which are believed to evolve under the selective pressure exerted by host plants (Zhong et al, 2016, Hartman et al, 2018). Effectors that manipulate host physiology or immunity represent major contributors to fungal pathogen specialization (Sánchez-Vallet et al, 2018). SSPs, Small secreted proteins, secreted by pathogens often function as avirulence (AVR) effectors, initiating incompatible host-pathogen interactions upon recognition by host resistance (R) proteins. However, the blast fungus possesses the ability to circumvent resistance in an incompatible host through gain- or loss-of-function mutations in AVR genes. This phenomenon is exemplified by the emergence of wheat blast due to the loss of function of the PWT3 AVR gene in response to the widespread cultivation of wheat carrying the RWT3 resistance gene (Inoue et al., 2017). Similarly, the gain of the AVR gene PWL1 resulted in the loss of pathogenesis on weeping lovegrass within the *Eleusine*-lineage of *M. oryzae* but conferred virulence on the new host, finger millet (Asuke et al., 2020).
Furthermore, fungal secondary metabolites (SMs) have garnered attention for their production during various stages of plant colonization, where they may act as non-proteinaceous effectors (NPEs) in plant-fungal interactions (Collemare et al, 2019). These chemical effectors can function as avirulence factors or mimic plant hormones, facilitating successful host tissue invasion. For instance, ACE1, a SM produced by *M. oryzae*, triggers avirulence towards rice cultivars carrying the corresponding resistance gene Pi33 (Böhnert et al, 2004; Collemare et al, 2008). Moreover, various fungal pathogens, including *M. oryzae,* have been found to produce analogues of plant hormones, such as jasmonic acid, gibberellin, cytokinins, and ethylene, manipulating plant growth and subverting plant hormone-based defense signalling pathways during colonization (Patkar & Naqvi, 2017; Shen et al., 2018). Intriguingly, a derivative of jasmonic acid, 12-hydroxyjasmonic acid, secreted by *M. oryzae*, specifically suppresses host innate immunity in rice but not in barley or wheat, underscoring the pivotal role of fungal SMs in shaping plant-pathogen interactions (Patkar et al., 2015). Moreover, many other fungal SMs, such as host-selective toxins (HSTs), serve as pathogenicity factors, as exemplified by HC-toxin in *Cochliobous carbonum*, T-toxin in *C. heterostrophus*, AM-toxin, and ACR-toxin in *Alternaria alternata*, all of which are crucial for pathogenesis on specific hosts harbouring susceptible genes (Izumi et al., 2012; Lim & Hooker, 1971; Tsuge et al., 2013; Walton, 2006). The blast fungus itself produces toxins like tenuazonic acid, pyriculol, and pyrichalasin H during host

invasion, with the latter being essential for infecting *Digitaria* hosts (Tsurushima et al., 2005).

Major advances in comparative and population genomics have provided the tools to dissect effector biology and its evolution (Plissonneau *et al*, 2017). These advancements have bridged the divide between molecular biology, evolutionary genetics, and epidemiology, offering unprecedented insights into the complex world of host-pathogen interactions. The proliferation of fungal genome sequences, coupled with advances in computational biology, presents a unique opportunity to identify the genetic and molecular determinants underpinning host adaptation and host range expansion. For instance, the population genomics of the wheat fungal pathogen *Zymoseptoria tritici* has revealed the possible role of SM in pathogen lineage diversification and host adaptation (Hartmann et al, 2018). In *M. oryzae* isolates, adaptation to rice host is likely associated with a small number of lineage-specific gene families, which include genes involved in biosynthesis of SMs and secreted proteins of unknown functions (Chiapello et al, 2015). While some secreted protein effectors have undergone scrutiny, the full spectrum of biosynthetic gene clusters (BGCs) responsible for producing secondary metabolites (SMs) remains uncharted territory within the diverse host-specific lineages of *M. oryzae*, despite the wealth of available large-scale genomic data. As genome mining has emerged as a prominent strategy to explore BGC diversity in numerous genomes (Medema, 2018; Navarro-Muñoz et al, 2019). It is important to underscore that while small secreted proteins (SSPs) have been a subject of investigation, they alone fall short in elucidating the molecular foundations of most other plant-fungus interactions and host specialization. Novel chemical effectors (SMs) may offer insights into these mechanisms.

In the present research endeavour, our primary focus revolved around the isolation and comprehensive characterization of 80 distinct Indian field isolates of *M. oryzae*. These isolates were obtained from rice, finger millet, and foxtail millet host plants spanning fifteen diverse geographic locations across India. We firmly believe that the examination of fungal diversity originating from various agricultural regions is a pivotal stride towards gaining insights into disease dynamics at the field level. The collected *Magnaporthe* isolates were subjected to phenotypic as well as molecular characterization. Subsequently, a subset of 15 field isolates of *M. oryzae* was meticulously selected for whole genome sequencing. In our investigation, we harnessed these newly sequenced genomes in conjunction with publicly

available genomes. Our major objective was to pinpoint novel effector molecules that may play a pivotal role in host specialization within the blast fungus. To achieve this, we embarked on an extensive exploration of genomes belonging to different lineages of the blast fungus, with a particular focus on identifying SM biosynthetic gene clusters (BGCs) associated with host specialization. Utilizing tools such as comparative genomics, similarity networks and phylogenetic analyses, we successfully identified a promising candidate BGC that is likely implicated in the specific infection of rice host plants.

2.2 Objectives

1. Collection and Isolation of *Magnaporthe* from rice and non-rice hosts from various geographical regions across India
2. Diversity assessment of isolates using phenotypic and molecular markers
3. Whole genome sequencing and comparative genome analysis of the selected isolates

2.3 Materials and Methods

2.3.1 Mono-conidial isolation of field strains and culture conditions

We isolated field strains of *M. oryzae* from different geographic locations in India and from three different host plants – rice, finger millet and foxtail millet. Blast lesions infected tissues were collected and proceeded for single spore isolation at the earliest **(Fig. 2.1)**. Plant tissues with infected lesions were surface sterilized using 70% ethanol and 1% sodium hypochlorite, followed by three successive washes of sterile distilled water. Surface sterilized plant tissues were incubated in a micro-humid condition maintained in a Petri dish for 24-48 h to induce sporulation. Spores were suspended in sterile water and spread evenly on 2% water agar plate. Single germinating spores were picked up under a microscope, inoculated on oatmeal agar (OMA) plates and incubated at 28 °C (OMA: cat # M397-500G, HiMedia). All the fungal cultures were grown and maintained on prune agar (PA) plates as described earlier (PA: 1 g/L yeast extract, 2.5 g/L lactose, 2.5 g/L sucrose, 0.04% prune

juice, pH 6.5) (Soundararajan et al. 2004). Vegetative growth of the fungus on PA plates was allowed for 10 days at 28 °C, with initial 3-day incubation under dark conditions followed by 7-day incubation under constant illumination for conidiation.

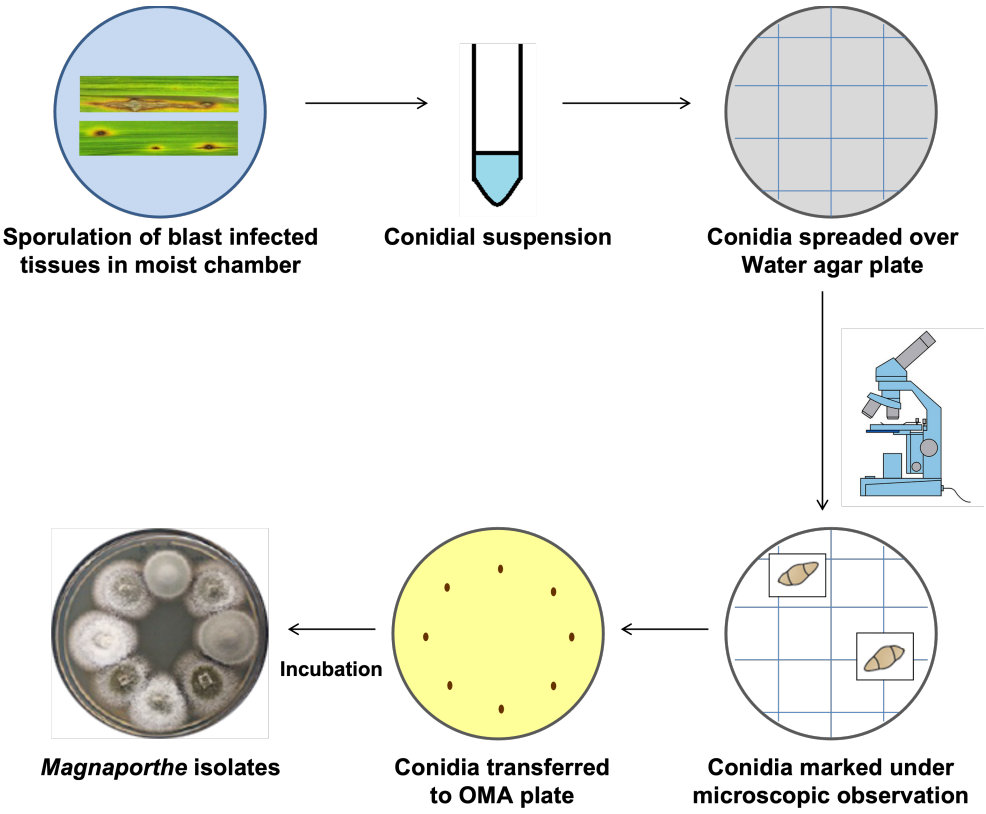

Figure 2.1: Schematic showing the steps of mono-conidial isolation from a leaf infected with blast disease

2.3.2 Filter stocks preparation for long-term storage

All the strains were grown on oat meal agar plates containing sterile Whatmann filter paper discs for ~10 days as described above. Filter papers covered with fungal biomass were transferred to sterile envelopes followed by desiccation to remove any moisture. The envelopes were placed in vacuumized poly-bags and stored at -80 °C.

2.3.3 DNA fingerprinting using transposons

Genomic DNA was extracted by grinding in liquid nitrogen the fungal biomass obtained from vegetative culture grown in an appropriate liquid medium for 2–3 days at 28 °C, followed by the standard protocols as described earlier (Dellaporta et al. 1983). In order to investigate the genetic structure of the different isolates, repetitive-PCR (George et al., 1998) was carried out targeting commonly found transposons Pot2 and Mg-SINE in *M. oryzae*. The primers were designed in outward orientation, and thus, the genomic regions between the transposons were amplified upon PCR. Pot2 and Mg-SINE are found in more than 50 copies per genome (Kachroo et al., 1994, 1995), the PCR would generate the amplicons of different sizes depending upon the distribution of these transposons **(Fig. 2.2)**. PCR was carried out using XT-5 DNA polymerase (GeNei^{TM}) on standard thermal cycler machine ABI- Veriti. The optimum thermal cycling conditions were as follows: initial denaturation step at 96^{O}C for 5 min, 4 cycles of 96 ^{O}C for 1 min, 62 ^{O}C for 10 min, 65 ^{O}C for 15 min, followed by 26 cycles of 96 ^{O}C for 30 sec, 54 ^{O}C for 1 min, 65 ^{O}C for 5 min and final extension step at 65 ^{O}C for 10 min. We used 1.2% agarose gel, pre-stained with Ethidium Bromide, and ran the gel overnight at low ~30V for better resolution of bands.

2.3.4 PCR amplifications of molecular markers (avirulence genes and mating type)

PCR targeting coding (CDS) region of six different avirulence genes (AVR-Pita, AVR1-CO39, ACE1, AVR-Pik, AVR-Pizt and PWL2) and mating types (MAT1-1 and MAT1-2) was carried out using oligo-nucleotide primers as described in **Table 2.1**. PCR was carried out using Taq DNA polymerase (GoTaq , Promega) on standard thermal cycler machines (Biorad-C1000 or ABI- Veriti). The optimum thermal cycling conditions were as follows: initial denaturation step at 95^{O}C for 5 minutes, 26 cycles of 95^{O}C for 30 seconds, annealing with variable temperature for each gene (AVR-Pita, AVR-Pik, AVR-Pizt, MAT1-1, MAT1-2 at 56 ^{O}C; AVR1-CO39, ACE1-KS at 54 ^{O}C and ACE1-AT, PWL2 at 58 ^{O}C) for 30 sec, 72^{O}C for 1 min and final extension at 72^{O}C for 5 min.

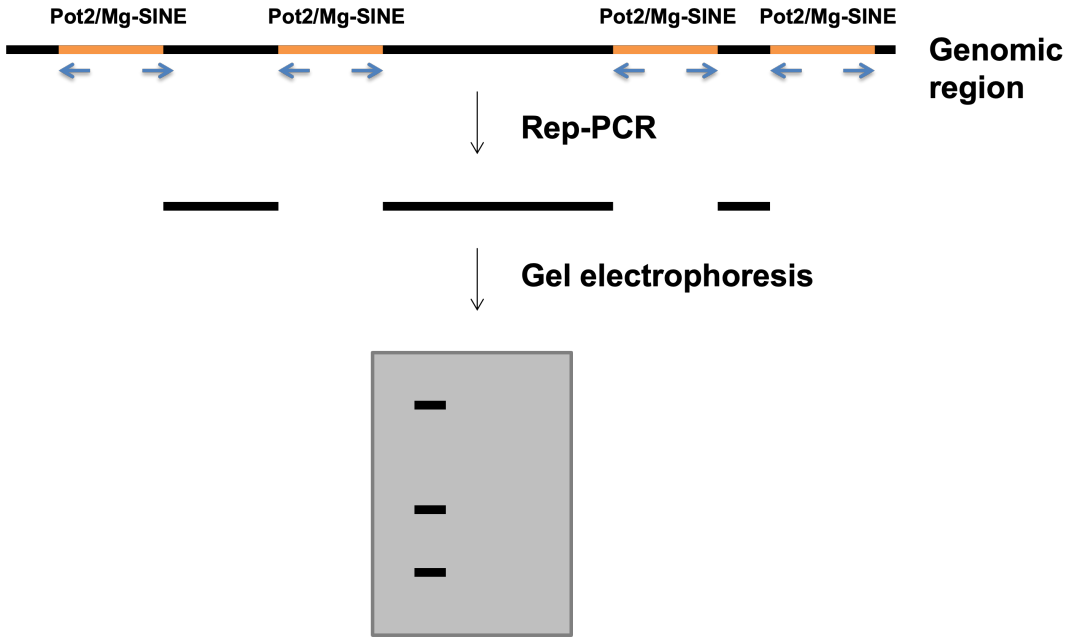

Figure 2.2: Schematic depicting the Rep-PCR methodology. Orange color bar in genomic region represents the repetitive elements Pot2/Mg-SINE. Blue arrows depict the orientation of primers annealed to the template DNA.

2.3.5 Dendrogram based on molecular markers

The presence/absence variations of AVR genes and mating types were scored into binary matrix, which was fed as an input to PAST tool. The dendrogram was generated using a distance-based method – UPGMA (Unweighted pair group method with arithmetic mean).

2.3.6 Whole-plant infection assay

To conduct the infection assay, we cultivated various plant species, including rice (cv. CO-39), finger millet (cv. GN4), foxtail millet (cv. SIA3088), and barley, as follows: First, the seeds of these plants were subjected to surface sterilization with 30% v/v ethanol for 2.5 minutes at room temperature (RT), followed by three rinses with distilled water. Subsequently, the seeds were allowed to germinate under moist conditions for 3–4 days at room temperature before being transplanted into pots.

Table 2.1: Oligo-nucleotide primers used for molecular characterization of *M. oryzae* field isolates

Primers	Sequences (5' - 3' direction)	References
Pot2-F	CGG AAG CCC TAA AGC TGT TT	Kachroo et al., 1994
Pot2-R	CCC TCA TTC GTC ACA CGT TC	
Mg-SINE-F	TTA CCG TGG CAA GGA TA	Kachroo et al., 1995
Mg-SINE-R	TGC TAA CGG TCG GCT AA	
AVR-Pita-F	GAC CCG TTT CCG CCT TTA TT	Current study
AVR-Pita-R	GAT TCC CTC CAT TCC AAC AC	
ACE1-AT-F	GAG GTG CCA GAT ATG TCG TC	Current study
ACE1-AT-R	GGA TGA GCA GAT GAG CAA CA	
ACE1-KS-F	GCA CCT TGA CGT TTG AAC AG	
ACE1-KS-R	TGA GTT TGC ATT GAG CGA GT	
AVR1-CO39-F	TGC CGC ATT TTG CTA ACC	Tosa et al., 2005
AVR1-CO39-R	GCG AAT CCA TAG ACA AGG AC	
PWL2-F	CTC CGC CAC TTT TCT CAT TC	Current study
PWL2-R	GCC CTC TTC TCG CTG TTC AC	
AVR-Pik-F	GTC AAC CAA GCG TAA ACC TC	Current study
AVR-Pik-R	CGA TTC AGA AGT TAG GCA TT	
AVRPiz-t-F	CGA CTG GTA GCA TTT GTT TC	Current study
AVRPiz-t-R	AAG TGG CTC GTT CCT AAT TG	
MAT1-1-F	TCA GCT CGC CCA AAT CAA CAA T	Wang et al., 2004
MAT1-1-R	ACT CAA GAC CCG GCA CGA ACA T	
MAT1-2-F	GAG TTG CCT GCC CGC TTC TG	
MAT1-2-R	GGC TTG GTC GTT GGG GAT TGT	

To create the soil mixture for planting, we combined the following components in the specified proportions: approximately 50% garden/potting soil, 40% coco-peat, 5% red soil, and 5% vermiculite. We used distilled water containing an insecticide (Bayer; 1 mL/2 L water) to prepare the soil mixture, which was then added to plastic pots to sow the germinated seeds. The plants were then permitted to grow for approximately 3–4 weeks in a greenhouse maintained at 28 to 30 °C under humid conditions, following a light/dark cycle of approximately 12 hours.

Whole-plant infection assay was conducted by spraying ~10^5 conidia/mL in 0.05% w/v gelatin onto ~4-week-old host plants (Rice and finger millet) and ~3-week-old foxtail millet plants. Following this, the plants were placed in an environment maintained at 28 to 30 °C with high humidity conditions. Initially, they were kept in darkness for 24 hours, and subsequently, they were exposed to alternating cycles of 14 hours of light and 10 hours of darkness for 4 to 6 days. Throughout the entire incubation period, we regularly monitored the development of disease symptoms. We recorded the disease occurrence around 5-8 days after inoculation, depending on the appearance of lesions on the leaves. These infection assays were conducted with a minimum of three biological replicates.

To assess the host invasion by the fungal strains, we examined the leaf sheaths obtained from 3–4- week-old host (rice, finger millet, wheat or barley) plants, grown as mentioned above. Leaf sheaths were prepared by excising 2-3 cm long portions of the stem region. Using fine-pointed forceps, we delicately removed the central slender tube, leaving behind a hollow cylindrical sheath suitable for the assay. Next, the leaf sheaths were inoculated with ~15 µL conidial suspension (~10^4 conidia/mL) followed by incubation at room temperature (<30 °C) under humid conditions for appropriate time points. At ~36 hpi, the midrib was removed from the sheath tissue, and observed under a microscope (Olympus BX51, Japan).

2.3.7 Mating assays

Sexual crosses were conducted on rice flour agar (RA) medium (RA: 20 g/L rice flour, 2 g/L yeast extract, 15 g/L agar). To perform sexual crosses strains of opposite mating type were used. A small mycelial plug from the test strain was placed 4 cm away from a mycelial plug of the reference strain on RA within a 90 mm petri dish, following a method described

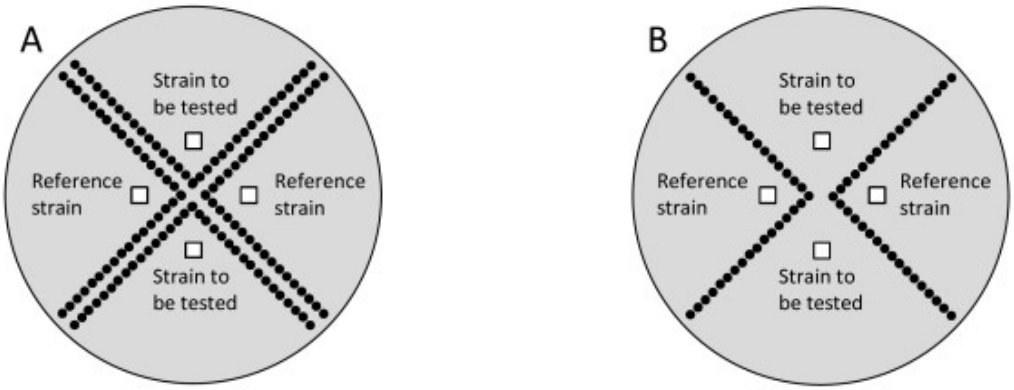

Figure 2.3: Schematic representing design of mating crosses between opposite mating type strains of *M. oryzae*. (A) The presence of two lines of perithecia between each tested strain and each reference strain indicated that both the strains are female-fertile and male fertile. (B) A single line of perithecia is produced between each tested strain and each reference strain: the tested strains are female-sterile and male-fertile. (Adapted from Saleh et al. 2012).

as previously (**Fig. 2.3**; Saleh et al., 2012). Initially, the cultures were incubated for two days at 25°C to allow the vegetative growth and were then exposed to continuous white and blue light at 20°C. After an incubation period lasting 18-21 days at 20°C, perithecia were observed.

The presence of two lines of perithecia formed between the tested strain and the reference strain indicated that both tested and reference strains were capable of male and female fertility. However, if a single line of perithecia was produced by reference strain, it indicated that the tested strain was female sterile but remained male-fertile. If perithecia were entirely absent on both sides, it indicated that the tested strain was both female-sterile and male-sterile.

2.3.8 Genome sequencing and assemblies

Newly isolated strains were sequenced for whole genome using a paired-end sequencing approach at >50X depth on Illumina HiSeq2500 platform at AgriGenome Labs Pvt. Ltd., India. Short raw reads were processed using Trimmomatic v0.38 at a threshold for the minimum read length of 80 bp for paired-end reads (parameters PE, Leading:10,

Trailing:10, Slidingwindow:4:20, Minlen:80; Bolger et al., 2014). De novo sequence assemblies were constructed using CLC Genomics Workbench v11.0 with default parameters (Word size:23, Bubble Size:50, Minimum contig length:500, Mismatch cost:2, Insertion cost:3, Deletion cost:3, Length fraction:0.5, Similarity fraction:0.8). Additionally, we used publicly available genome sequences of *M. oryzae* strains from NCBI (Supplementary table S1). The overall quality of each genome assembly was evaluated using BUSCO v5.2.2 (Simão et al., 2015) with the Sordariomycetes dataset.

2.3.9 Variant calling

Raw sequence reads after trimming were mapped to the indexed reference genome 70-15 using BWA-MEM (burrows-wheeler aligner) (H. Li & Durbin, 2009). Alignment files in BAM format were converted to SAM format using SAMtools (H. Li et al., 2009). Alignments were sorted with SAMtools, and reads with a mapping quality below 30 were removed. Variant calling was done using bcftools mpileup tool, using --ploidy 1 as *Magnaporthe* is a haploid organism. Resulted multi-samples VCF file was used for further analysis. VCF file was filtered for SNPs present in gene regions only using BEDtools (Quinlan & Hall, 2010). Fasta file was generated from multi-samples VCF file and the phylogenetic tree was plotted using MEGA X (Kumar et al., 2018). The method used was Maximum likelihood with General Time Reversible (GTR) model, as the AIC value was lowest for GTR as per the Model Test. Principal component analysis (PCA) based on SNPs was also performed using R.

2.3.10 Gene predictions and functional annotations

The repetitive regions of each genome were annotated and masked using RepeatMasker v open-4.0.6 (Smit et al., 2015). Repeat library specific to *Magnaporthe* species was used for repeat prediction. We carried out gene predictions on all the assemblies using Augustus v3.3.3 (Stanke et al., 2006) with Magnaporthe_grisea as a species model.

Protein functions were predicted for all gene models using HMMER v3.2.1 and Pfam database (Hancock & Bishop, 2004). The secretome was defined as the set of proteins with a signal peptide but no trans-membrane domain (TM) as predicted by Phobius (Käll et al.,

2004). Putative effectors were identified among the set of secreted proteins using EffectorP v 3.0 (Sperschneider & Dodds, 2022).

2.3.11 Inference of orthologous gene families

We used OrthoFinder to detect the orthologous gene families. Total proteome sequences of each species under study are analyzed for all-versus-all BLAST search to provide promising gene pairs. The BLAST bit score, for each pairwise comparison between species, gets normalized for the gene length and phylogenetic distance, called RBNHs (Reciprocal Best length-Normalised Hit). These putative cognate gene pairs are then combined in the orthogroups graphs, which are ultimately clustered into orthogroups (Emms & Kelly, 2015). Total proteome data from all 16 strains of *M. oryzae* were collected and resulted total 171940 protein sequences were used to identify the orthologous gene families using OrthoFinder v 2.5.5 with a default parameter (Emms & Kelly, 2019). A total of 11661 orthogroups were obtained.

Phylogenetic tree was constructed based on the PAV of 1953 accessory gene set using IQ-TREE v. 2.1.2 (parameters -st MORPH -bb 1000; Minh et al. 2020). Best-fit model (MK+FQ+ASC+R2) was chosen based on the Bayesian Information criterion (BIC), followed by assessment of phylogenetic tree for branch support with ultrafast bootstrap (bb). Phylogenetic tree was visualized using iTOL (Letunic and Bork 2021).

2.3.12 Distribution of feruloyl esterases (FAEs)

For phylogenetic analysis of feruloyl esterases across different host-specific strains, total protein sequences of 16 representative *M. oryzae* strains were used. Proteomes were mined for the presence of the conserved domain Tannase (Pfam ID: PF07519.13) using hmmsearch (hidden Markov model search) of HMMER suite v. 3.3.2 (parameters -cut_tc; {Citation}). The retrieved sequences were aligned using Mafft v7.475 (parameters -reorder; Katoh et al., 2002) and poorly aligned sequences were removed using TrimAl (Capella-Gutierrez et al. 2009; parameters -automated1). The Tannase domain containing protein sequences were curated manually and were aligned to construct the maximum-likelihood phylogenetic tree, using IQ-TREE v. 2.1.2 (parameters -m MFP -alrt 1000 -bb 1000 -nt AUTO; Minh et al.

2020). Best-fit model (WAG+F+I+G4) was chosen based on the Bayesian Information criterion (BIC), followed by assessment of phylogenetic tree for branch support with SH-like approximate likelihood ratio test (alrt) and ultrafast bootstrap (bb). Phylogenetic tree was visualized using iTOL (Letunic and Bork 2021).

2.3.13 Effector genes repertoire

Most effector genes are smaller in length, and our gene predictions using Augustus might have omitted their predictions. The BLASTP-based search for known effectors did not show homologs of most effectors. Thus, standalone BLASTN (parameter: evalue 1e-10, max target seq 1) was performed on all the genomes with known blast effector sequences as query (Durairaj et al., 2016). To summarize the effector content per isolate, we built a presence and absence matrix indicating presence and absence of effector genes with 1 and 0, respectively. The matrix was used to generate the heatmap displaying presence/absence polymorphisms of known blast effector genes using pheatmap R-package.

2.3.14 Construction of species tree

2655 BUSCO proteins conserved in 68 *M. oryzae* strains were retrieved from the BUSCO analysis and aligned using Mafft v7.475 (parameters -reorder; Katoh et al., 2002). TrimAL v1.4.1 (parameters -automated1; Capella-Gutiérrez et al., 2009) was used to remove poorly aligned regions (Supplementary dataset S1, S2). A maximum likelihood tree was generated using IQ-TREE v2.1.2 (Minh et al., 2020) with partition models (Chernomor et al., 2016; Lanfear et al., 2014), model finder (Kalyaanamoorthy et al., 2017) and ultrafast bootstrap (Hoang et al., 2018; parameters -m MF -p partition.nex -bb 1000) (Supplementary dataset S3, S4). Similarly, a species tree including 3 *M. grisea* and 68 *M. oryzae* strains was generated using a total of 2557 BUSCO proteins (Supplementary dataset S5-S7). Both trees were visualized and annotated using iTOL (Letunic & Bork, 2021).

2.3.15 Prediction of BGCs and similarity network analysis

BGCs were predicted using antiSMASH v6.0.1 (Blin et al., 2021) (parameters -taxon fungi -minimal -cluster_hmmer -pfam2go). fungiSMASH predicts BGCs in each dataset by identifying the presence of a core biosynthetic gene, based on conserved domains followed by empirically extending the cluster, to predict the tailoring/accessory gene(s) therein. Further, based on extensive studies on the evolutionary history of these genes, fungiSMASH would set varying rules for predicting genes coding for enzymes involved in the synthesis of different SMs such as polyketides (PK), non-ribosomal peptides (NRP), indoles, and terpenes. All the predicted BGCs and reference fungal BGCs from the MIBiG v2.1 (Kautsar et al., 2019) repository were included to perform a similarity network analysis using BiG-SCAPE v1.1.0 (Navarro-Muñoz et al., 2019) with glocal mode and with different cut-off values (parameters –include_singletons –mix –no_classify –clans-off –cutoffs 0.3 0.4 0.5 0.7). BiG-SCAPE facilitates the sequence similarity network analysis of BGCs using a combined distance metric which accounts for domain content (JI), shared adjacent domains (AI) and sequence similarity (DSS). The distance metric is calibrated based on the fact that each of the eight BiG-SCAPE product classes has different evolutionary dynamics, followed by generation of similarity networks by hierarchical clustering of BGCs to gene cluster families (GCFs). A cut-off value of 0.5 was found to be appropriate, as the relevant BGCs showed similarity with the reference BGCs from MIBiG known for metabolites like DHN (melanin), epipyriculol, alternapyrone, squalestatin and cytochalasans. The network files were visualized with Cytoscape v3.8.2 (Shannon et al., 2003). BGC conservation vizualisation of specific gene cluster families was done with Clinker (Gilchrist & Chooi, 2021).

2.3.16 Synteny analysis

To identify the syntenic genomic regions between any genome assembly and the reference assembly 70-15, global alignments were generated using nucmer (parameters –c 100 --maxmatch) program from MUMMER v3.23 package (Kurtz et al., 2004). Alignments were further filtered using delta-filter utility with length >10 kb and percent identity >80% (parameters -l 10000 -i 80) to retrieve continuous syntenic blocks. The show-coords utility was used to extract alignment coordinates in tabular format and the plots were generated in

R using KaryoploteR package (Gel & Serra, 2017). To investigate the syntenic genes between 70-15 and MZ5-1-6, bidirectional best hits were identified using BLASTP on protein sequences from both the genomes. Synteny among each chromosome belonging to these two strains was plotted using Circos tool (Krzywinski et al., 2009).

2.3.17 Core gene phylogeny

Homologs of MGG_08236 were identified in Pezizomycotina from Joint Genome Institute MycoCosm repository (Grigoriev et al., 2014) using BLASTP. The resulted amino acid sequences were aligned using MAFFT v7.475 (Katoh et al., 2002) and poorly aligned regions were removed with TrimAl v1.4.1 (Capella-Gutiérrez et al., 2009b). A phylogenetic tree was constructed using IQ-TREE v2.1.2 (Minh et al., 2020) with LG substitution model and ultrafast bootstrap (Hoang et al., 2018) as well as the SH-like approximate likelihood-ratio test (Guindon et al., 2010) for branch support (parameters -mset LG -bb 1000 -alrt 1000). The resulting tree was visualized and annotated using iTOL (Letunic & Bork, 2021).

2.3.18 Gene expression analysis by semi-quantitative RT-PCR

The expression profiles of four genes (MGG_08236, MGG_15107, MGG_15108 and MGG_12496) were studied during different stages of infection. Briefly, 20 µL conidial suspension (~10^5 conidia/mL with 0.01% Tween-20) of B157 strain from *Oryza* lineage was drop-inoculated on to the surface-sterilized 2-week-old barley detached leaf blades placed on water agar plates containing 2 µg/mL Kinetin, followed by incubation under dark (10 h) and light (14 h) cycles at 25 °C. Samples were collected by excising inoculated portions on the leaf blades, at 12-, 24-, 48-, 72- and 96-hours post inoculation (hpi). Three-day-old vegetative mycelia, grown in liquid CM, was used as a control condition. A sample from mock-inoculation (leaf blades inoculated with only 0.01% Tween-20) was used as another experimental control to check any non-specific amplification during RT-PCR. Total RNA was extracted from all samples using TRIzol® reagent (Invitrogen, USA) as per the manufacturer's instructions. Two µg of total RNA each was used for cDNA synthesis using iScript™ cDNA Synthesis Kit (Bio-Rad, USA). Oligonucleotide primers used for RT-PCR for each gene are listed in Supplementary Table S4. Optimized thermal cycling conditions

for RT-PCR were as follows: initial denaturation step at 95 °C for 5 min, followed by 30 cycling reactions each at 95 °C (denaturation step) for 30 sec, 60 °C (annealing step) for 30 sec and 72 °C (extension step); and the final extension step at 72 °C for 5 min. Expression of each gene was quantified based on the intensity of the band measured using ImageJ, and was assessed relative to that of β-tubulin (MGG_00604) as an endogenous control. The data on expression of PKS and tailoring genes are from three independent biological experiments and represented as mean ± standard deviation of mean (SDM) in the graph. The MGG_08236 ORF was amplified from the genomic DNA of *M. oryzae* strains from different lineages using the same thermal cycling conditions as above, except for the extension step, which was set at 1 min in each cycle.

Table 2.2: Primers used to study the relative gene expression of BGC-O1 genes.

Primers	Sequences (5'-3' direction)
TUB-RT-F	CATGATGGCTGCTTCTGACT
TUB-RT-R	GTACTCCTCTTCCTCCTCGT
MGG_08236-RT-F	TGGGCCAACGACAATTACGA
MGG_08236-RT-R	AAACCTTCTCCGCCGCATAGG
MGG_15107-RT-F	GTTTTGGTGCTGGCGGATTTG
MGG_15107-RT-R	CCGCTCCAACAAGCTCACATTC
MGG_15108-RT-F	TCGTTTACAGTACGAGTTGCCC
MGG_15108-RT-R	ACCGTAGTGCCCAGATTCTTGG
MGG_12496-RT-F	AGCCCACCTCACGTTTTTGACC
MGG_12496-RT-R	CCTCGTCTGAGTTGCCACGC
TUB-ORF-F	GAAATTGTTCACCTTCAGACCGGC
TUB-ORF-R	TGTTCTGGACGTTGCGCATCTGG
MGG_08236-ORF-F	ATGTCTGTTGGCGAGGCATC
MGG_08236-ORF-R	TCGTAGACCATCAGTCGGC