

RESEARCH

Open Access



# Association of secondary metabolite gene clusters with host-specific lineages of the cereal blast fungus *Pyricularia oryzae*

Khyati Mehta<sup>1,2</sup>, Jorge C. Navarro-Muñoz<sup>2,4</sup>, Sayali Bakore<sup>3</sup>, Jérôme Collemare<sup>2\*</sup> and Rajesh Patkar<sup>1,3\*</sup>

## Abstract

Fungal plant pathogens constantly evolve and deploy novel peptide and metabolite effectors to break down plant resistance and adapt to new host plants. The blast fungal pathogen *Pyricularia oryzae* is a single species subdivided into multiple host-specific lineages that have evolved through gain and/or loss of virulence and/or effector related genes through chromosomal rearrangement. Here, we mined 68 genomes of *P. oryzae*, belonging to six host-specific lineages, to identify secondary metabolite (SM) biosynthetic gene clusters (BGCs) likely associated with potential metabolite effectors involved in host specialization. A similarity network analysis grouped a total of 4501 BGCs into 283 gene cluster families (GCFs), based on the content and architecture of the BGCs. While most of the GCFs were present in all the *P. oryzae* lineages, two (BGC-O1 and BGC-O2) were found specifically in the *Oryza* lineage and one (BGC-TLE) was found in the lineage specific to *Triticum*, *Lolium*, and *Eleusine* hosts. Further analysis of the phylogenetic relationships between core biosynthetic genes confirmed that BGC-O1, which comprises a reducing polyketide synthase gene (MGG\_08236) and four putative tailoring genes, was present only in the *Oryza* lineage. Importantly, most genes, including MGG\_08236 from the BGC-O1, were expressed specifically during pathogenesis. We propose that the *Oryza* lineage-specific BGC-O1 produces a metabolite effector likely involved in specialization of *P. oryzae* to the rice host. In addition, we identified five SM genes under positive or balancing selection only in the *Oryza* lineage, suggesting a role in the interaction with rice specifically. Our findings highlight the importance of further mining novel metabolite effectors in specialization and virulence of the blast fungus to different cereal hosts.

**Keywords** Biosynthetic diversity, Population genomics, *Magnaporthe oryzae*, Gene cluster family, Polyketide synthase, Effector, Polymorphism, Divergence

## Background

Many plant pathogenic fungi have adapted to infect specific host plants. The emergence of new pathogenic strains, as a result of host range expansion or specialization to a new host, can cause severe disease outbreaks (Giraud et al. 2010). It is crucial to understand the molecular basis of such host specialization to increase the durability of crop resistance to diseases (Giraud et al. 2010; Gladieux et al. 2014). The filamentous fungus *Pyricularia oryzae* (synonym *Magnaporthe oryzae*) is the causative agent of the economically important cereal blast disease worldwide, and is a model organism to study

\*Correspondence:

Jérôme Collemare

j.collemare@wi.knaw.nl

Rajesh Patkar

rajeshpatkar@iitb.ac.in

<sup>1</sup> Bharat Chattoo Genome Research Centre, Department of Microbiology and Biotechnology Centre, The Maharaja Sayajirao University of Baroda, Vadodra 390002, India

<sup>2</sup> Present Address: Westerdijk Fungal Biodiversity Institute, Utrecht 3584 CT, The Netherlands

<sup>3</sup> Department of Biosciences & Bioengineering, Indian Institute of Technology Bombay, Mumbai 40076, India

<sup>4</sup> Present Address: Wageningen University and Research, Wageningen 6708 PB, The Netherlands



host–pathogen interactions (Talbot 2003; Ebbole 2007). Although being capable of infecting several grasses, *P. oryzae* is classified as a single species, and is subdivided into several host-specialized lineages with limited primary host range (Gladieux et al. 2018).

Host selection pressure could be one of the factors that shape the virulence of the population of a pathogen, whereby specific virulence-related genes that encode secreted proteins and/or enzymes for secondary metabolite production – collectively known as effectors – are either gained or lost (Zhong et al. 2016; Hartmann et al. 2018). Effectors manipulate the host physiology or immunity to the pathogen's advantage, and thus are major determinants of specialization on a host (Sánchez-Vallet et al. 2018). Indeed, gain- or loss-of-function mutation in avirulence (AVR) genes of *P. oryzae* has resulted in the overcoming of host resistance or non-host resistance. For example, a widespread cultivation of wheat cultivar IAC-5 carrying the *RWT3* resistance gene led to a loss of function of the corresponding *PWT3* AVR gene in *P. oryzae* and subsequent emergence of wheat blast (Inoue et al. 2017). Similarly, although gain of the AVR gene *PWL1* led to a loss of pathogenesis on weeping lovegrass in the *Eleusine*-lineage of *P. oryzae*, it conferred virulence on the new host, finger millet (Asuke et al. 2020). Further, *AVR-Pita*, flanked by retrotransposons, has been found to be translocated to multiple genomic locations, including supernumerary chromosomes, as a result of dynamic adaptation towards its corresponding resistance gene *Pita* in certain rice varieties (Chuma et al. 2011).

Fungal secondary metabolites (SMs) have also been shown to exhibit important effector functions during plant-fungal interactions (Collemare et al. 2019). For example, ACE1, a secondary metabolite (SM) produced by *P. oryzae*, confers avirulence towards the rice cultivars carrying the corresponding resistance gene *Pi33* (Böhnert et al. 2004). Various fungal pathogens, including *P. oryzae*, are also known to produce analogues of plant hormones such as jasmonic acid, gibberellin, and cytokinins, to manipulate plant growth or subvert plant hormone-based defense signalling pathways during invasion (Patkar & Naqvi 2017; Shen et al. 2018). A derivative of jasmonic acid (12-hydroxyjasmonic acid) secreted by *P. oryzae* suppresses the host innate immunity specifically in rice but not in barley or wheat (Patkar et al. 2015). Many other fungal SMs such as host-selective toxins (HST) also act as virulence factors – for example, HC-toxin in *Cochliobolus carbonum*, T-toxin in *C. heterostrophus*, AM-toxin and ACR-toxin in *Alternaria alternata* (Lim & Hooker 1971; Walton 2006; Izumi et al. 2012; Tsuge et al. 2013). These HSTs are essential for pathogenesis on specific hosts carrying the corresponding susceptible gene. The blast fungus produces SMs such

as tenuazonic acid and pyriculol during pathogenesis (Chen & Qiang 2017; Jacob et al. 2017).

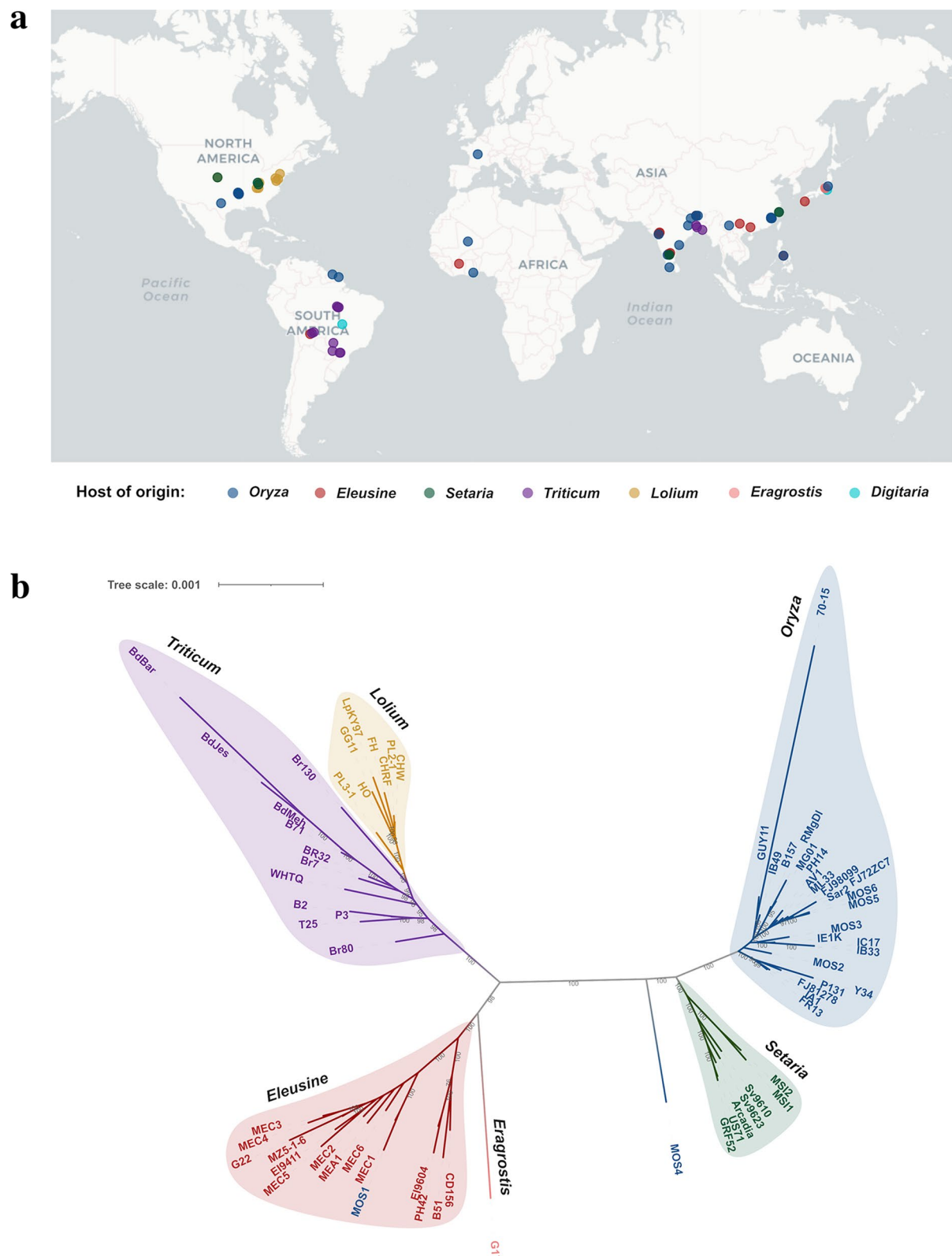
The availability of an increasing number of fungal genomes and advances in computational biology facilitates the exploration of molecular determinants of host specialization or host range expansion. For instance, population genomics of the wheat fungal pathogen *Zymoseptoria tritici* has revealed a possible role of SMs in pathogen lineage diversification and adaptation to wheat host (Hartmann et al. 2018). In *P. oryzae* isolates, specialization to rice host is likely associated with a small number of lineage-specific gene families, which could include genes involved in biosynthesis of SMs and secreted proteins of unknown functions (Chiapello et al. 2015). While most studies have focused on the effector function of certain secreted proteins in the blast fungus, the link between diverse SM biosynthetic gene clusters (BGCs) and *P. oryzae* host-specific lineages has not been explored.

Here, we embarked on the exploration of genomes from the different lineages of the blast fungus, in search for SM BGCs associated with host specialization. Using similarity network, comparative genomics, and phylogenetic analyses, we identified a candidate BGC that is likely involved in pathogenesis and specialization to rice and weeping lovegrass host plants.

## Results

### Genome sequencing of host-specific strains of *P. oryzae*

In this study, we used 68 *P. oryzae* genomes isolated from 6 different host plant species, rice (*Oryza sativa*, n=26), finger millet (*Eleusine* sp., n=14), foxtail millet (*Setaria* sp., n=7), wheat (*Triticum aestivum*, n=12), perennial ryegrass (*Lolium* sp., n=8), and weeping lovegrass (*Eragrostis curvula*, n=1) (Fig. 1a, Additional file 1: Table S1). These genomes included 15 field isolates that were collected from infected rice, finger millet, and foxtail millet plants grown in different parts of India (Fig. 1a, Additional file 1: Table S1) and sequenced their genome using Illumina technology (Additional file 1: Table S2). The remaining 53 genomes were obtained from publicly available genome sequences (Additional file 1: Table S1, Fig. 1a). Additionally, sequence data from three *Pyricularia grisea* strains previously isolated from crabgrass (*Digitaria* sp.) was used as an outgroup (Additional file 1: Table S1, Additional file 2: Figure S1). Sixteen publicly available assemblies were obtained from long-read sequencing technology and were highly-contiguous. We performed new gene predictions for all 71 assemblies. The analysis of BUSCO genes indicated high-quality gene prediction and completeness of the assemblies with minimum 91.5%, maximum 98.6%, and average 96.6% BUSCO-score (Additional file 1: Table S1).



**Fig. 1** *Pyricularia oryzae* worldwide populations are organized in host-specific lineages. **a** Geographic distribution of sequenced *P. oryzae* strains. The color-coded dots denote the host plant of origin of the strains. **b** Phylogenetic tree was constructed based on concatenation of 2655 BUSCO genes present in all 68 *P. oryzae* genomes used in this study. Colored shapes in the background depict different host-specific genetic lineages

We used 2655 BUSCO genes present in the 68 *P. oryzae* genomes to build a phylogenomic tree (Additional file 3: Dataset S1). The topology of the obtained species tree confirms that *P. oryzae* consists of multiple genetic lineages specialized on different hosts, including *Oryza*, *Setaria*, *Eleusine*, *Eragrostis*, *Triticum*, and *Lolium* (Fig. 1b). This tree is concordant with the previously reported population structure (Gladieux et al. 2018). A similar phylogenomic analysis using additional three genomes from *P. grisea* (*Digitaria* isolates) showed that *P. grisea* is a separate species from *P. oryzae*, which is consistent with a previously reported study (Additional file 2: Figure S1; Additional file 3: Dataset S2; Gladieux et al. 2018).

We performed infection assays to check the pathogenesis of 16 Indian isolates of *P. oryzae* on three hosts, rice, finger millet, and foxtail millet. We found that all isolates except one (*P. oryzae* MOS1 strain) induced clear blast lesions on their host-of-origin, while they triggered no or few symptoms on non-hosts (Additional file 2: Figure S2). In several cases, small dark necrotic spots were observed, indicating a hypersensitive response caused by recognition of the strain (Additional file 2: Figure S2). Intriguingly, MOS1 and MOS4, two strains isolated from infected rice leaf tissue, were found to be genetically similar to that of the *Eleusine* lineage and was not related to any lineage, respectively (Fig. 1b). The whole-plant infection assays using the MOS1 strain showed differential virulence pattern, where inoculated finger millet and rice plants developed highly-susceptible and moderately-resistant lesions, respectively, similar to other isolates from finger millet (Additional file 2: Figure S2). This result indicates that MOS1 is indeed an *Eleusine* lineage strain which was isolated from infected rice tissue. It is possible that the strain MOS1 could be evolving to adapt to rice host, or it randomly landed on the collected rice leaf and benefited from a co-infection by an *Oryza* lineage strain. Further studies are required to understand whether it was associated with any co-infection of the concerned rice plant tissue with two genetically distinct *P. oryzae* strains. Further, although it is genetically different from any specific lineage, the MOS4 strain showed similar virulence pattern with a few susceptible lesions and small necrotic spots as that of the *Oryza* lineage strains MOS3 and MOS5 (Additional file 2: Figure S2). These results suggest that MOS4 strain defines a separate lineage that can infect rice, yet it may have another primary host that remains to be determined.

#### Genome mining identified lineage-specific biosynthetic gene clusters in *P. oryzae*

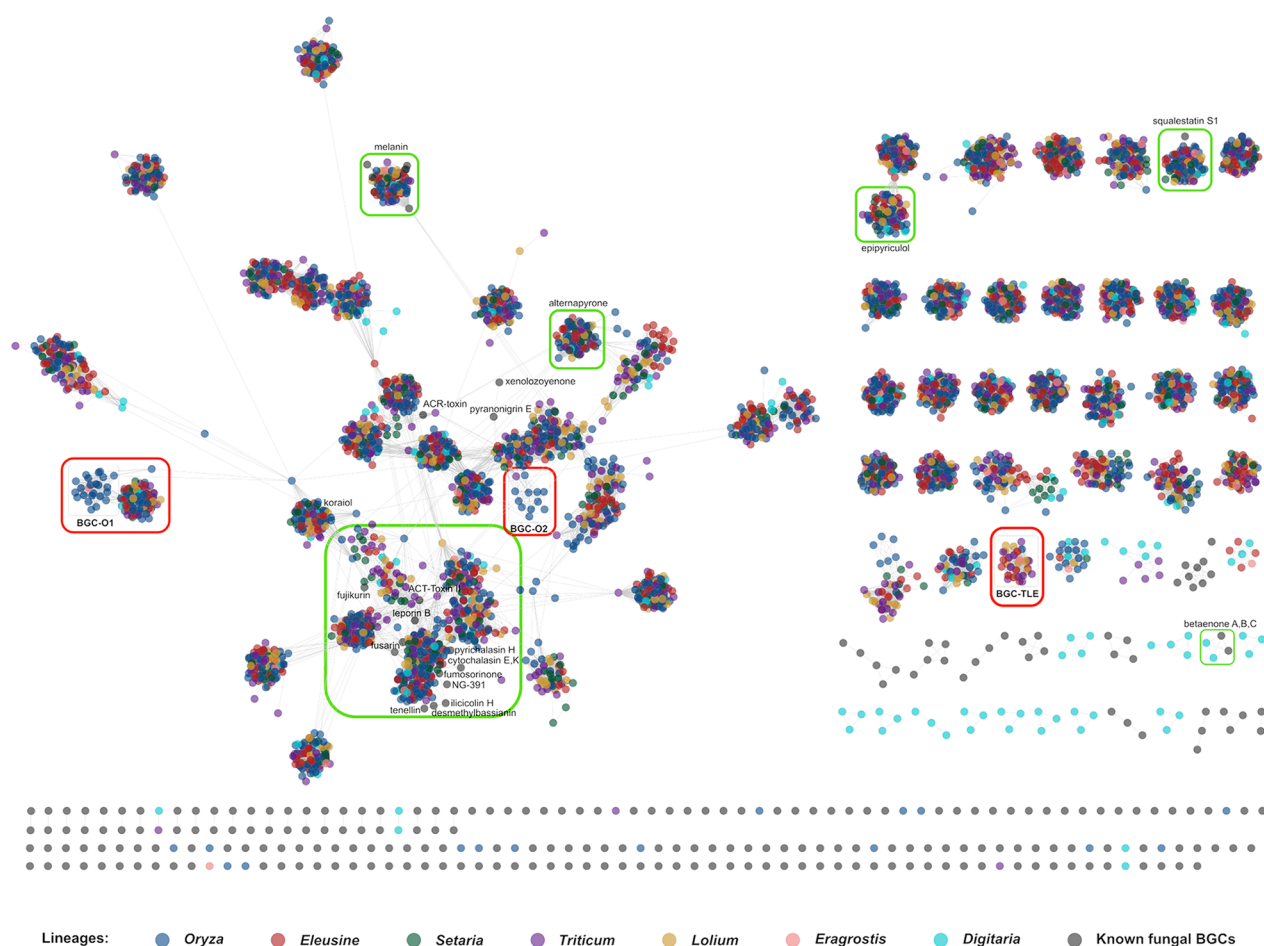
The identification of genomic regions that contain SM BGCs was performed in all 71 genomes using

antiSMASH (Blin et al. 2021). This analysis resulted in predicting a total of 4224 BGCs, with an average number of 59 BGCs per strain (Additional file 1: Table S3). The predicted BGCs were classified based on the core biosynthetic genes, such as genes encoding polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), or terpene cyclases (TCs) (Additional file 1: Table S3 and Additional file 2: Figure S3). All the host-specific lineages contained similar number of BGCs from all different classes of BGCs, with type I PKSs being the most abundant (Additional file 2: Figure S3). Overall, these analyses highlight the great potential of *P. oryzae* in producing a diverse set of SMs, some of which could be involved in virulence and/or host specialization.

To understand the role of SM genes during divergence of the *Oryza* lineages, we assessed natural selection by determining the polymorphism within populations and divergence between the *Oryza*, *Triticum*, and *Eleusine* populations in SM core biosynthetic genes shared by at least ten isolates from each host-specific lineage. We found that most SM core biosynthetic genes are either under strong or relaxed purifying selection in all three lineages (Additional file 1: Table S4). For instance, core genes in the melanin and ACE1 biosynthetic pathways are under strong purifying selection, which is consistent with their key roles in the interaction of the fungus on any host species (Chumley & Valent 1990; Fudal et al. 2007; Vy et al. 2024). Only four SM core enzyme genes (MGG\_00806, MGG\_01951, MGG\_03401, and MGG\_04775) were found to be under balancing selection in the *Oryza* lineage, and two of these genes were found to be upregulated in appressoria during host penetration (Oh et al. 2008; unpublished data). Interestingly, one gene MGG\_09239 is under positive selection in the *Oryza* lineage only (Additional file 1: Table S4). These observations suggest that only a handful of SM genes likely have a role during interaction with rice specifically. Especially, genes under balancing selection could be involved in the arms race with the rice host and represent interesting targets for future functional studies (Van Der Hoorn 2002; Oliva et al. 2015; Brunner & McDonald 2018).

To determine if any putative BGC is associated with the ability to infect specific host plants, we performed a network similarity analysis using BiG-SCAPE (Navarro-Muñoz et al. 2019). This analysis included the 4224 predicted BGCs and 277 characterized BGCs from the MIBiG database (Kautsar et al. 2019). The 4501 BGCs were clustered into 283 gene cluster families (GCFs)/subnetworks, of which 180 are singletons, including 160 from the reference characterized BGCs (Fig. 2). Twelve out of 91 GCFs present in *Pyricularia* strains comprised a characterized BGC, including largely conserved ones for DHN melanin, epipyriculol,





**Fig. 2** Similarity network analysis of biosynthetic gene clusters (BGCs) from *Pyricularia oryzae* and *Pyricularia grisea*. A BiG-SCAPE analysis with a cutoff c0.5 depicts similarity of 4224 BGCs from *P. oryzae* or *P. grisea* with 277 reference BGCs from MIBiG database. Each dot represents a BGC and is color-coded according to the lineage. Gene cluster families (GCF; subnetworks) marked with green boxes share significant homology with reference MIBiG BGCs (gray-colored circle). GCFs marked with red boxes are found to be unique to host-specific lineages. The length of the gray lines is proportional to the genetic distance between BGCs. Singletons are shown as individual dots at the bottom

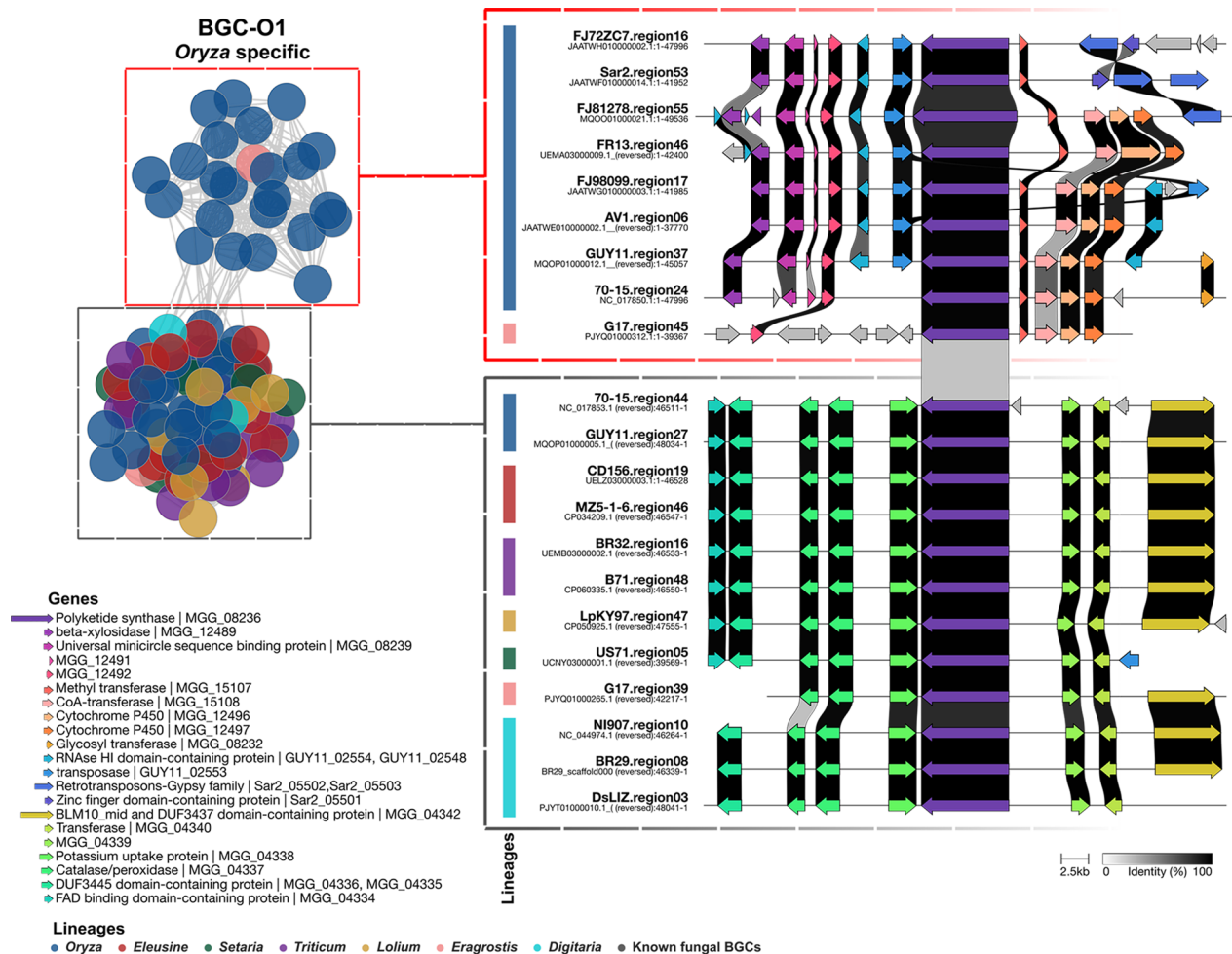
alternanapyrone, squalastatin, and cytochalasins in *P. oryzae*, and betanone in *P. grisea* (Fig. 2; Additional file 2: Figure S4–S8). Forty-four GCFs are found in several *P. oryzae* lineages as well as in *P. grisea*, most of which remain uncharacterized. Interestingly, 13 and 14 out of 91 GCFs appeared to be specific to *P. grisea* and *P. oryzae*, respectively (Fig. 2). Of the *P. oryzae*-specific GCFs, three were potentially associated with host specific lineages. While BGC-O1 and BGC-O2 were found mostly specific to the *Oryza* lineage, BGC-TLE was found specific to *Triticum*, *Lolium*, *Eleusine*, and *Eragrostis* lineages (Fig. 2), which all share a common origin and were diverged from the *Oryza* and *Setaria* lineages (Fig. 1). BGC-O1 consists of genes encoding a reducing PKS and tailoring enzymes (see below), whereas BGC-O2 is restricted to a truncated PKS gene alone that is likely not functional (Additional file 2: Figure S9). In the

case of BGC-TLE, although the standalone core biosynthetic NRPS gene therein is present in all the strains, the gene structure appeared differentially (Additional file 2: Figure S10). Manual curation of the NRPS gene showed the deletion of 557 bp in *Triticum* and *Lolium* lineages corresponding to the second exon in the *Eleusine* lineage (CD156 strain). Further, the presence of a stop codon within the adenylation domain of the NRPS gene in the *Triticum* and *Lolium* lineages suggests pseudogenization. BGC-TLE is thus likely not functional in those strains (Additional file 2: Figure S11). Yet, it could be either important for virulence on *Eleusine* host plants only, or its product might have an avirulence effector-like role in *Triticum* and *Lolium* hosts. Thus, out of three candidate BGCs, we further investigated BGC-O1 which is potentially involved in specialization on rice and weeping lovegrass hosts.

### A novel polyketide biosynthetic gene cluster was specifically found in rice/*Eragrostis*-infecting lineage of *P. oryzae*

BGC-O1 was found in 23 out of the 24 *Oryza* lineage strains used in the study, as well as in one *Eragrostis* lineage strain (Fig. 3). This GCF appeared to contain two different networks, and comparison of the loci indeed clearly differentiated two putative BGCs. One was conserved in all lineages (Fig. 3, gray box), and BGC-O1 was restricted to the *Oryza* and *Eragrostis* lineages (Fig. 3, red box). BGC-O1 contains a novel reducing PKS gene (MGG\_08236) as the core biosynthetic gene along with flanking tailoring genes, including one methyl transferase (MGG\_15107), two cytochrome P450 monooxygenases (MGG\_12496 and MGG\_12497), and one Co-A transferase (MGG\_15108).

To assess if only the BGC-O1 cluster region was specific to the *Oryza* lineage, we checked whether flanking genomic regions of BGC-O1 were conserved or variable in different lineages. We mapped representative high-quality assemblies from each lineage with the reference genome assembly of the 70–15 strain from *Oryza* lineage to assess synteny at global level. We found that BGC-O1 was located in the sub-telomeric region of chromosome 2 (NC\_017850.1), and approximately 528 kb downstream of the *ACE1* gene cluster in the reference strain 70–15. The macrosynteny analysis between 70–15 and GUY11, both strains from *Oryza* lineage, showed conservation of ~134 kb upstream and ~117 kb downstream flanking regions, along with the BGC-O1 cluster (Fig. 4a). The ~10 kb-long immediate flanking regions were found to be swapped between the GUY11 and 70–15 genomes (Fig. 4a). Similarly, comparison between 70–15 and FR13



**Fig. 3** BGC-O1 is predominantly present in *Oryza* lineage of *Pyricularia oryzae*. BiG-SCAPE network showing BGC-O1 present in 23 genomes from *Oryza* and one genome from *Eragrostis* lineages. Conservation of the BGC in selected strains is depicted using Clinker. Vertical bars besides the name of strain are colored according to the lineages. Red and gray dashed boxes separate BGC-O1 and BGC-O1-like clusters, respectively

strains revealed a partial conservation (~14 kb) of BGC-O1, with an ~9.5 kb region from BGC-O1 in 70–15 strain translocated further downstream on the same contig in the FR13 strain (Fig. 4a). Further, while ~102 kb upstream flanking region was syntenic in 70–15 and FR13, the downstream flanking region showed a few structural rearrangements (Fig. 4a). In contrast, comparison with non-*Oryza* lineages, like *Eleusine* and *Setaria*, showed a significant loss of synteny in the sub-telomeric region of chromosome 2. The upstream flanking region (~97 kb upstream of BGC-O1 in 70–15) was found to be syntenic to a corresponding region in MZ5-1–6, CD156, and US71; however, the BGC-O1 cluster and downstream sequences were absent in the genomes of these lineages (Fig. 4b). Importantly, BGC-O1 did not show any similarity with any of the reference BGCs in MIBiG, suggesting that it is a BGC likely involved in production of a novel metabolite.

#### BGC-O1 likely originates from a duplication event in the common ancestor of *P. oryzae* and *Colletotrichum eremochloae*

To understand the evolutionary history of BGC-O1, we searched for homologues of the MGG\_08236 PKS protein within *Pezizomycotina* from the MycoCosm repository (Grigoriev et al. 2014). Only two orthologues could be retrieved from the fungus *Colletotrichum eremochloae*. A phylogenetic analysis showed that the PKS Coler1|670826 (scaffold\_178) from *C. eremochloae* is an orthologue of MGG\_08236 PKS, forming an outgroup to the *P. oryzae* lineage (Fig. 5a). The other paralogue in *C. eremochloae*, Coler1|679399 (scaffold\_5), belongs to a clade that is restricted to the *Colletotrichum* genus and forms a sister clade to the MGG\_08236 PKS clade. Comparison of the genomic loci of both homologues in *C. eremochloae* and BGC-O1 in *P. oryzae* showed that these two clusters shared the PKS and two cytochrome P450 genes with BGC-O1 (Fig. 5b), while the orthologues of methyltransferase and Co-A transferase genes were present elsewhere as single copies, Coler1|671186 and Coler1|576843, on scaffold\_185 and scaffold\_10, respectively, in the *C. eremochloae* genome. The presence of the two similar BGCs in *C. eremochloae* suggests

that BGC-O1 likely originated from an ancestral duplication event. While *Pyricularia* and *Colletotrichum* were diverged ~250 MYA (calculated using TimeTree; Kumar et al. 2022), BGC-O1 is retained only in *P. oryzae* and *C. eremochloae*, suggesting a role of respective host selection pressure in its retention.

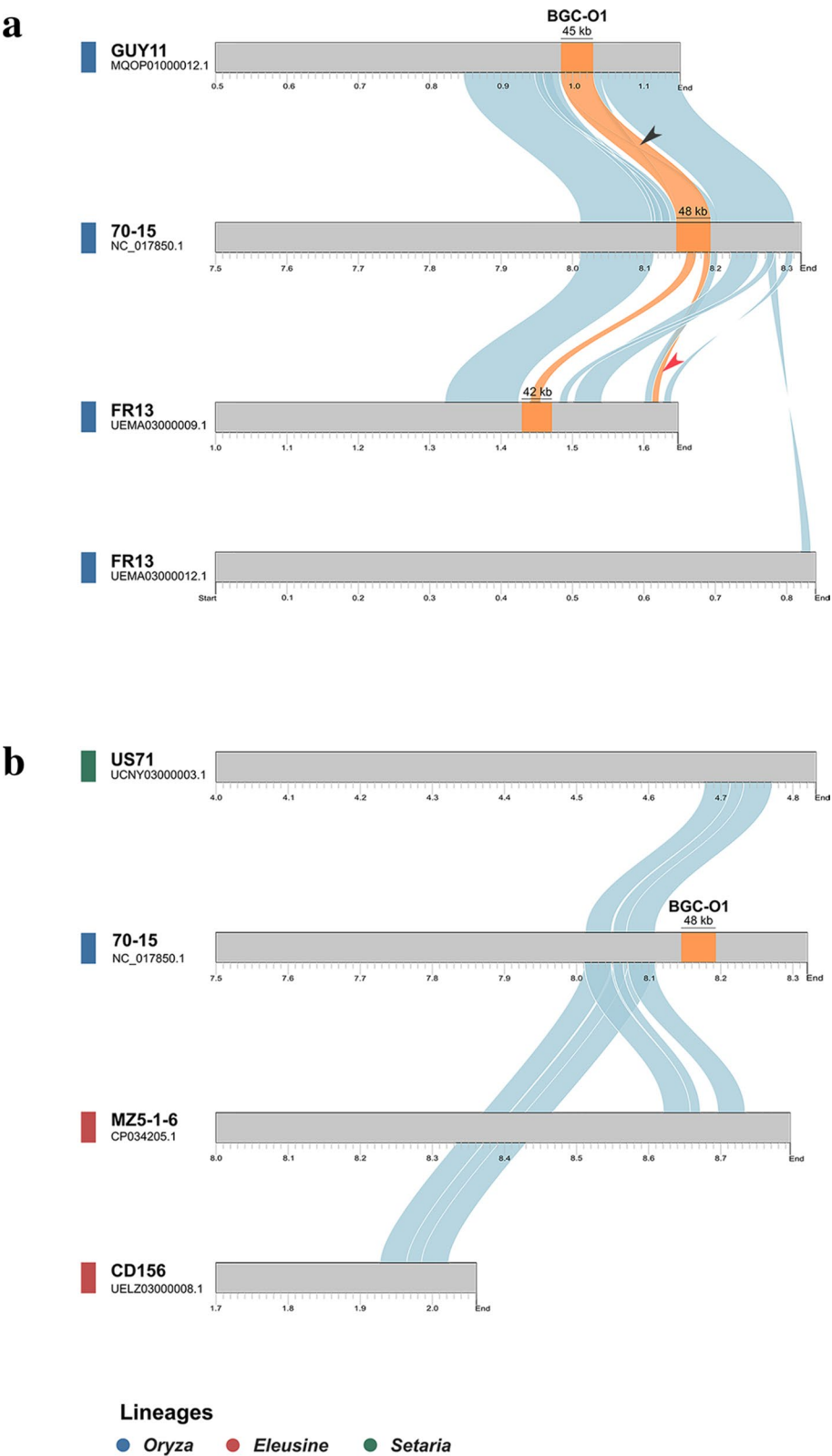
#### BGC-O1 genes are expressed specifically during host invasion

The presence of the PKS MGG\_08236 gene was checked by PCR using genomic DNA of *P. oryzae* strains from rice and finger millet host plants (Fig. 6a). The MGG\_08236 gene was present in all the *Oryza* strains studied, except for MOS3, which was identified as the only *Oryza* lineage strain lacking the BGC-O1 based on the *in-silico* analysis. The MGG\_08236 ORF was absent in all the *Eleusine* strains investigated as well as in the MOS1 and MOS4 strains, which are placed outside of *Oryza* lineage in the phylogenetic analysis (Fig. 1b and Fig. 6a).

We then studied the expression of the PKS (MGG\_08236) and three predicted tailoring genes, methyl transferase (MGG\_15107), CoA-transferase (MGG\_15108), and cytochrome P450 (MGG\_12496) of the BGC-O1 locus during different stages of infection, using RT-PCR. Total RNA was isolated from fungal vegetative mycelia grown in complete medium, the rice leaves inoculated with *P. oryzae* and incubated for different time-points, and the uninoculated rice leaves, to assess the levels of transcripts of the aforementioned genes relative to those of  $\beta$ -Tubulin as an internal control. While expression of the core PKS gene was undetected in mycelium and at 12 hpi (pre-invasive stage), its transcripts started accumulating during pathogenic development, with significantly elevated expression from 24 hpi till 96 hpi (Fig. 6b). The expression pattern of MGG\_15107 and MGG\_15108 tailoring genes was partly correlated with that of the PKS (Fig. 6b). In contrast, MGG\_12496 showed a different expression pattern, where the transcript levels were similar during both vegetative growth and host invasion (Fig. 6b), suggesting that it is not co-regulated with other genes in the predicted BGC-O1. This result shows that the MGG\_08236 PKS gene is

(See figure on next page.)

**Fig. 4** Genomic localization of BGC-O1 and synteny between the *Oryza* and non-*Oryza* lineages. **a** Synteny analysis using pairwise comparison within *Oryza* lineage. Genomes of GUY11 and FR13 aligned individually with reference genome 70–15. The syntenic BGC-O1 locus (orange) and the flanking regions (blue) on chromosome 2 of 70–15 are depicted. The differential lengths of BGC-O1 in different isolates are marked with a bar and corresponding length in kilobases. The black arrowhead depicts genomic rearrangement (swapping) in the flanking ~10 kb region. The red arrowhead marks the rearrangement of ~9 kb region of the BGC-O1 in FR13. **b** Synteny analyses of representative genome assemblies belonging to *Setaria* (US71), *Eleusine* (MZ5-1–6 and CD156), and *Oryza* (70–15) lineages. The BGC-O1 locus (orange) and flanking region (blue) sequences from US71, MZ5-1–6, and CD156 aligned with that of 70–15 are depicted



**Fig. 4** (See legend on previous page.)



specifically expressed during pathogenesis on rice and potentially has a key role to play during host colonization.

Altogether, our *in-silico* analyses identified a novel PKS gene cluster in the *Oryza*- and *Eragrostis*-specific lineages, which likely played a key role in shaping specialization of the blast fungus to rice and weeping lovegrass host plants.

## Discussion

### Population genomics identifies BGCs with potential roles in virulence and host specialization in *P. oryzae*

We identified three lineage-specific BGCs in the blast fungal pathogen *P. oryzae* using a population genomics approach. While two of them appeared to encode non-functional enzymes, the *Oryza* and *Eragrostis*-specific BGC, named BGC-O1, appears to be functional and is present in almost all the genomes from *Oryza* (23 out of 24) and *Eragrostis* (1 out of 1) lineages. Our population-genomics-based finding of the *Oryza* lineage-specific BGC (BGC-O1) is supported by gene expression analyses. Indeed, we found that the *Oryza*-lineage-specific PKS (MGG\_08236) transcript accumulated specifically during penetration and colonization of the rice host. Recent transcriptomics-based studies have shown that the expression of the PKS encoded in BGC-O1 (MGG\_08236) and neighboring tailoring genes (methyl transferase, MGG\_15107 and Co-A transferase, MGG\_15108) was induced during the initial biotrophic colonization stage of blast disease (Jeon et al. 2020; Yan et al. 2023). Importantly, the expression of MGG\_08236, along with the other pathogenesis-related genes such as the SM gene cluster *ACE1* and protein effectors SLP2, BAS2, BAS3, and AVR-Pi9, was recently found to be under the control of the bZIP transcription factor Bip1 (Lambou et al. 2024). Bip1 is found to regulate expression of a specific set of genes involved in early biotrophic stage of blast infection (Lambou et al. 2024). These findings strongly suggest that the SM produced by BGC-O1 could exhibit an effector-like function. Many plant fungal pathogens such as *Colletotrichum higginsianum*, *Fusarium graminearum*, *Zymoseptoria tritici*, and *P. oryzae* are known to produce a number of SMs during pathogenesis, especially during host penetration and colonization

(Collemare et al. 2008a, b; Patkar et al. 2015; Rudd et al. 2015; Harris et al. 2016; Dallery et al. 2017; Palma-Guerrero et al. 2017; Jeon et al. 2020; Yan et al. 2023). To study if the BGC-O1 has any role in virulence in a host-specific manner, deletion of the *PKS* gene (MGG\_08236) was attempted in an *Oryza*-specific isolate. Despite using different approaches and a number of attempts, generation of a deletion mutant has been unsuccessful so far, possibly due to its presence in the sub-telomeric region of chromosome 2 of the reference strain 70–15. A similar difficulty was faced in generating deletion mutant at the *ACE1* locus (Collemare et al. 2008a, b). Nonetheless, the MOS3 strain from *Oryza* lineage, which lacks the BGC-O1, failed to infect rice host as effectively as B157, MOS2, or MOS6 (Fig. 6a; Additional file 2: Figure S2a). This could possibly explain the key role of BGC-O1 in specialization to rice host.

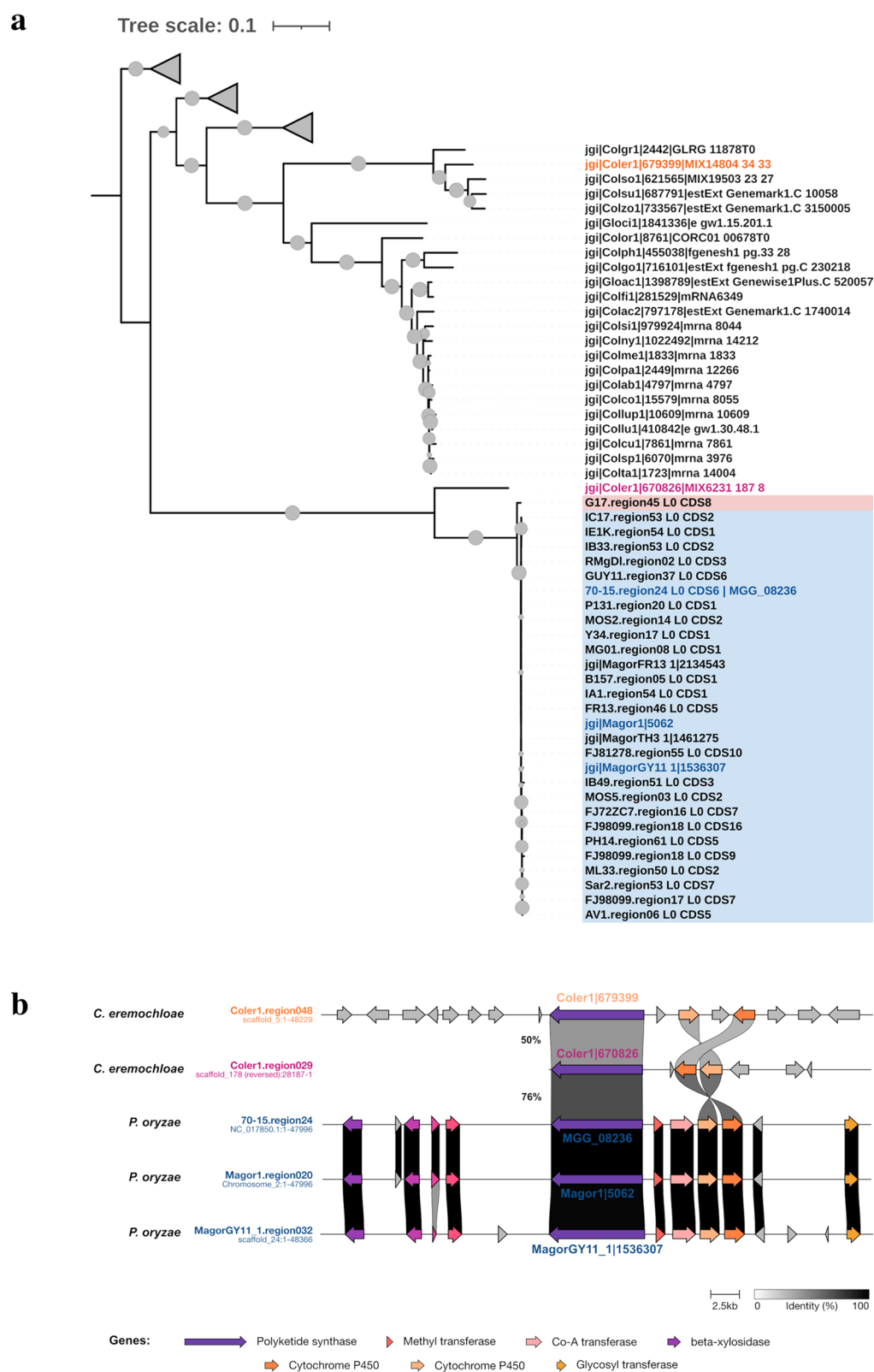
While only three BGCs were present in host-specific lineages, most of the other BGCs were found in all lineages. However, such apparently conserved BGCs show variations in their tailoring gene content. Such differences could result in the production of distinct SMs with potential effector function in a host specific manner. Alternatively, conservation of BGCs in different lineages could indicate a key role in virulence on cereals. Thus, analyses of variations among conserved SM-producing BGCs across different *P. oryzae* lineages could be useful in identifying the likely diversity in the SMs produced and their association with specific hosts.

### Conservation of BGC-O1 in an unstable genomic context

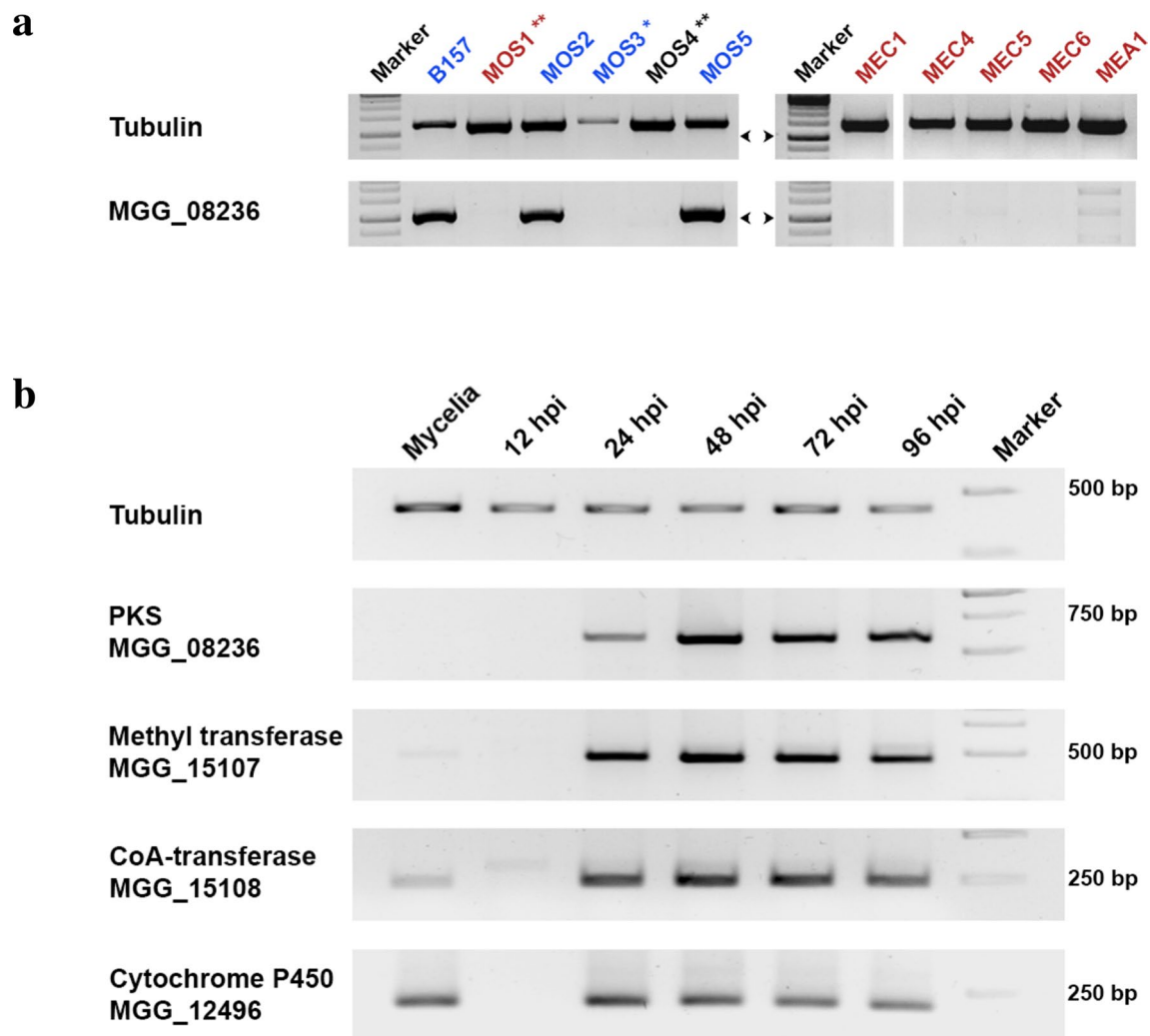
We found that the *Oryza*-specific cluster BGC-O1 is located in the sub-telomeric region on the chromosome 2 (NC\_017850.1) of the reference strain 70–15. While the genomic regions covering the BGC-O1 and its flanking sequences were highly syntenic in strains from the *Oryza* lineage, only the upstream flanking region of this locus was found conserved in all the other lineages. The chromosome ends tend to have substantially higher levels of polymorphism because they frequently undergo rearrangements due to the presence of AT-rich repeats and transposable elements (Farman & Kim 2005; Starnes et al. 2012; Rahnama et al. 2020). In absence of sexual

(See figure on next page.)

**Fig. 5** Evolutionary origin of the reducing polyketide synthase (PKS) MGG\_08236 gene. **a** Maximum-likelihood phylogenetic tree using protein sequences of reducing PKS genes from BGC-O1 and related gene cluster families, as well as homologues retrieved from MycoCosm repository. Tip labels depicting sequences from *Oryza* and *Eragrostis*-specific clade are marked with blue and pink background, respectively. The *Colletotrichum eremocloeae* ortholog (Coler1|670826), closer to MGG\_08236, is denoted with pink label; whereas the more distant paralogue (Coler1|679399) is shown in orange. Gray shaded triangles denote collapsed clades with distant sequences. Branches were supported by > 95% Bootstrap values indicated with gray circles at the nodes. The tree is rooted at midpoint. **b** Comparative analysis, using Clinker tool, of the BGC-O1 or BGC-O1-like loci from three *Oryza*-specific and *C. eremochloae* genomes



**Fig. 5** (See legend on previous page.)



**Fig. 6** BGC-O1 genes are specifically expressed during pathogenic development. **a** Gel electrophoresis of PCR products displaying presence or absence of ORF region of PKS MGG\_08236 from genomic DNA of the indicated *P. oryzae* strains. B157, MOS2, MOS3, and MOS5 belong to the *Oryza* lineage (labelled in blue). The MOS1, MEC1, MEC4, MEC5, MEC6, and MEA1 belong to the *Eleusine* lineage (labelled in red). \* indicates the only *Oryza* strain that lacked the BGC-O1 in *in-silico* analysis. \*\* indicates the *Oryza* strains placed outside *Oryza* lineage. Arrowheads correspond to the 1 kb size of band from Marker. **b** RT-PCR gel depicting expression of genes MGG\_08236 (Polyketide synthase), MGG\_15107 (Methyl transferase), MGG\_15108 (Co-A transferase), and MGG\_12496 (Cytochrome P450) in *Oryza*-specific strain B157 at different stages of rice infection (12 to 96 h post inoculation) and during vegetative growth (mycelium) in complete medium

reproduction, asexually propagating pathogens likely utilize chromosomal rearrangements as one of the mechanisms of host-specialization as shown in *Verticillium dahliae* (Jonge et al. 2013). Similarly, the regions adjacent to a telomere in *P. oryzae* are reported to be highly polymorphic and enriched with genes involved in interaction with its host plants (Rahnama et al. 2021). Gain and/or loss of effector genes, lineage-specific gene families and chromosomal rearrangements are likely the major evolutionary mechanisms involved in such host specificity and adaptation of *P. oryzae* (Chuma et al. 2011; Jonge et al. 2013; Chiapello et al. 2015; Yoshida et al. 2016; Gómez

Luciano et al. 2019). Given that BGC-O1 is also flanked by transposons in most strains of the *Oryza* lineage of *P. oryzae* (Fig. 3), we hypothesize that genomic alterations, most likely through deletions, may have resulted in the loss of the BGC in other host-specialized lineages. We further found that BGC-O1 is located on a mini-chromosome (UEMA03000009.1) in the FR13 isolate, which also carries the *ACE1* BGC. Large-scale genomic rearrangements around the loci of several virulence-related genes (*ACE1* and *AVR-Pik*) likely led to the emergence of supernumerary mini-chromosome in the rice isolate FR-13, facilitating the adaptive evolution of the blast

fungus (Langner et al. 2021). Thus, retention of BGC-O1 in *Oryza* and *Eragrostis* lineages, despite its localization in unstable sub-telomeric regions, suggests a significant role of host selection pressure in its conservation therein. The most likely selection pressure to retain BGC-O1 relates to the ability to successfully infect a host, which is also indicated by the specific expression profile during infection. The presence of homologous BGCs in *C. eremochloae*, the causal agent of anthracnose disease on centipede-grass turf in southern United States (Crouch & Tomaso-Peterson 2012), suggests a particular function for BGC-O1 in infecting certain cereals only (*Oryza*, *Eragrostis*, and *Eremochloa*). Further investigations are needed to determine if the *C. eremochloae* BGC is expressed during infection and whether it produces the same metabolite as BGC-O1 in *P. oryzae*.

## Conclusions

Our findings highlight the importance of population genomics-based studies in identifying secondary metabolite BGCs and the corresponding genomic rearrangements, likely driven by the host selection pressure, which could lead to a host-specialized lineage of the blast fungal pathogen.

## Methods

### Fungal strains and culture conditions

We isolated 15 field strains of *P. oryzae* from different geographic locations in India and from three different host plants, rice (*Oryza sativa*), finger millet (*Eleusine coracana*), and foxtail millet (*Setaria italica*). Infected tissues with blast lesions were collected and used for single spore isolation. Leaf tissues with blast disease lesions were surface sterilized using 70% ethanol and 1% sodium hypochlorite, followed by three successive washes with sterile distilled water. Surface-sterilized leaf tissues were incubated under micro-humid conditions 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, and inoculated on prune agar (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) plates and incubated at 28°C. Vegetative growth of the fungus on PA plates was continued for 10 days at 28°C, with initial 3-day incubation under dark conditions, followed by 7-day incubation under constant illumination for conidiation. Genomic DNA was extracted by grinding in liquid nitrogen. The fungal biomass was obtained from vegetative culture grown in a complete medium (CM: 6 g/L yeast extract, 6 g/L casein enzyme hydrolysate, 10 g/L sucrose) for ~3 days at 28°C, followed by

the standard protocols as described earlier (Dellaporta et al. 1983).

### Genome sequencing and assemblies

Newly isolated strains were sequenced for whole genome using a paired-end sequencing approach at >50× depth on Illumina HiSeq 2500 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 *P. oryzae* strains from NCBI (Additional file 1: Table S1). The overall quality of each genome assembly was evaluated using BUSCO v5.2.2 (Simão et al. 2015) with the Sordariomycetes dataset. 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.

### Construction of species tree

2655 BUSCO proteins conserved in 68 *P. 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 (Additional file 3: Dataset S3 and Dataset S4). A maximum likelihood tree was generated using IQ-TREE v2.1.2 (Minh et al. 2020) with partition models (Lanfear et al. 2014; Chernomor et al. 2016), model finder (Kalyaanamoorthy et al. 2017), and ultrafast bootstrap (Hoang et al. 2018; parameters -m MF -p partition. nex -bb 1000) (Additional file 3: Dataset S1 and Dataset S5). Similarly, a species tree including 3 *Pyricularia grisea* and 68 *P. oryzae* strains was generated using a total of 2557 BUSCO proteins (Additional file 3: Dataset S2, Dataset S6, and Dataset S7). Both trees were visualized and annotated using iTOL (Letunic & Bork 2021).

### 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). 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 global mode and with different cut-off values (parameters -include\_singletons -mix -no\_classify



–clans-off –cutoffs 0.3, 0.4, 0.5, 0.7). 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 such as DHN (melanin), epipyriculol, alternapyrone, squalestatin S1, and cytochalasans. The network files were visualized with Cytoscape v3.8.2 (Shannon et al. 2003). Visualization of BGC conservation was done with Clinker (Gilchrist & Chooi 2021). Gene models belonging to BGC-O1 family were manually curated (Additional file 3: Dataset S8).

### Statistical analysis for natural selection

To test for natural selection, 40 SM core biosynthetic genes shared by at least ten isolates from each host-specific lineage (*Oryza*, *Eleusine*, and *Triticum*) were used to calculate polymorphism and divergence (Additional file 1: Table S4). For each gene, the nucleotide coding region in different strains were aligned using Mafft v7.475 (parameters -reorder; Katoh et al. 2002). DnaSP v6 (Rozas et al. 2017) was used to calculate the levels of polymorphism within population ( $Pi(a)/Pi(s)$ ) and divergence between the populations ( $Ka/Ks$ ).

### Syntenic analysis

To identify the syntenic genomic regions between any genome assembly and the reference strain 70–15 assembly, global alignments were generated using nucmer (parameters -c 100 -maxmatch) program from MUM-MER 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 10,000 -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).

### Core gene phylogeny

Homologues of MGG\_08236 were identified in *Pezizomycotina* from the Joint Genome Institute MycoCosm repository (Grigoriev et al. 2014) using BLASTp. The resulting Blast hits were filtered for top 3 hits of each species. 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. 2009). A phylogenetic tree was constructed using Fasttree v2.1.11 (parameters -lg; Price et al. 2010) to identify orthologues and closest paralogues. The identified orthologous protein sequences were used to build a phylogenetic tree using IQ-TREE v2.1.2 (Minh et al. 2020) with LG substitution model and ultrafast bootstrap (Hoang et al. 2018), and the SH-like approximate likelihood-ratio test (Guindon et al. 2010) for branch support (parameters -mset LG -bb 1000 -alrt 1000) (Additional

file 3: Dataset S9–S11). The tree was visualized and annotated using iTOL (Letunic & Bork 2021).

### Whole-plant infection assay

Conidia were harvested as per the protocol described earlier (Patkar et al. 2012) from different strains collected in India. Whole-plant infection assay was carried out by spraying 10 mL spore suspension ( $\sim 10^5$  conidia/mL) containing 0.05% w/v gelatin onto ~4-week-old host plants, rice cv. CO-39 and finger millet cv. GN-4, and ~3-week-old foxtail millet cv. SIA-3088 plants, followed by incubation at 28–30°C under humid conditions, initially for 24 h in the dark and then for 4–6 days under 14 h light and 10 h dark cycles. Development of disease symptoms was monitored regularly during the entire incubation period and recorded around 5–8 days post inoculation depending on the appearance of lesions on the leaves. The infection assays were carried out with at least three biological replicates.

### 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 on ~4-week-old rice plants (blast susceptible variety PB-1509). Briefly, 20  $\mu$ L conidial suspension ( $\sim 10^5$  conidia/mL with 0.01% Tween-20) of B157 strain from *Oryza* lineage was drop-inoculated on to the surface-sterilized detached leaf blades placed on water agar plates containing 2  $\mu$ g/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-h 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  $\mu$ g 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 Additional file 1: Table S5. 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 s, 60°C (annealing step) for 30 s and 72°C (extension step); and the final extension step at 72°C for 5 min. The MGG\_08236 ORF was amplified from the genomic DNA of *P. 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.

## Abbreviations

ACE1	Avirulence Conferring Enzyme 1
AVR	Avirulence
BGC	Biosynthetic Gene Cluster
BGC-O1	Biosynthetic Gene Cluster specific to <i>Oryza</i> lineage - 1
BGC-O2	Biosynthetic Gene Cluster specific to <i>Oryza</i> lineage - 2
BGC-TLE	Biosynthetic Gene Cluster specific to <i>Triticum</i> , <i>Lolium</i> and <i>Eleusine</i> lineages
BiG-SCAPE	Biosynthetic Gene Similarity Clustering and Prospecting Engine
BUSCO	Benchmarking Universal Single-Copy Orthologs
CM	Complete Media
GCF	Gene Cluster Family
Hpi	Hours post inoculation
HST	Host Selective Toxin
MIbIG	Minimum Information about a Biosynthetic Gene cluster
NCBI	National Centre for Biotechnology Information
NRPS	NonRibosomal Peptide Synthetase
ORF	Open Reading Frame
PA	Prune Agar
PCR	Polymerase Chain Reaction
PKS	Polyketide Synthase
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SM	Secondary Metabolites
TC	Terpene Cyclases

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42483-025-00322-8>.

**Additional file 1: Table S1.** *Pyricularia oryzae* and *Pyricularia grisea* genomes used in this study. **Table S2.** *De novo* assembly statistics and gene prediction content of *Pyricularia oryzae* strains sequenced in this study. **Table S3.** Number of BGCs predicted using fungiSMASH in this study. **Table S4.** Summary of population genetic variation in 40 core biosynthetic genes in *Oryza*, *Eleusine*, and *Triticum* lineages of *P. oryzae*. **Table S5.** Oligonucleotide primers used to study the relative expression of BGC-O1 genes.

**Additional file 2: Figure S1.** *Pyricularia oryzae* and *Pyricularia grisea* are evolutionarily distinct species adapted to different hosts. **Figure S2.** Pathogenicity tests of *P. oryzae* strains isolated from different hosts and geographic locations in India. **Figure S3.** Occurrence of BGCs associated with specific classes of SM in individual genomes of *P. oryzae* and *P. grisea*. **Figure S4.** Homology of melanin-associated *P. oryzae* and *P. grisea* gene cluster family with reference BGCs in MIbIG database. **Figure S5.** Analysis of homology between ACE1 gene cluster family in *P. oryzae* and *P. grisea* and reference BGCs in MIbIG database. **Figure S6.** Analysis of homology between putative epipyriculol-associated gene cluster family in *P. oryzae* and *P. grisea* and reference BGCs in MIbIG database. **Figure S7.** Homology of putative Squalenstatin S1-associated *P. oryzae* and *P. grisea* gene cluster family with reference BGCs in MIbIG database. **Figure S8.** Analysis of homology between Tenuazonic acid associated gene cluster families in host-adapted *P. oryzae* and *P. grisea* isolates. **Figure S9.** BGC-O2 is predominantly present in *Oryza*-specific lineage of *P. oryzae*. **Figure S10.** BGC-TLE is predominantly present in *Triticum*, *Lolium*, and *Eleusine*-specific lineages of *P. oryzae*. **Figure S11.** Comparison of core biosynthetic gene region of BGC-TLE shows the pseudogenization in NRPS gene belonging to *Triticum* lineage.

**Additional file 3: Dataset S1.** Maximum-likelihood tree based on 2655 BUSCOs present in 68 strains of *P. oryzae*. **Dataset S2.** Maximum-likelihood tree based on 2557 BUSCOs present in 71 strains of *P. oryzae* and *P. grisea*. **Dataset S3.** Amino acid sequences of 2655 BUSCOs present in 68 strains of *P. oryzae*. **Dataset S4.** Amino acid sequence alignments of 2655 BUSCOs present in 68 strains of *P. oryzae*. **Dataset S5.** Best partitioning scheme used to construct Maximum-likelihood tree for *P. oryzae* strains. **Dataset S6.** Amino acid sequences of 2557 BUSCOs present in 71 strains of *P. oryzae* and *P. grisea*. **Dataset S7.** Amino acid sequence alignments of 2557 BUSCOs present in 71 strains of *P. oryzae* and *P. grisea*. **Dataset S8.** Manually curated gene models from *Oryza*-lineage-specific BGC-O1. **Dataset S9.** Amino acid sequences of the subset of rPKS (MGG\_08236) homologs in *Pezizomycotina* group. **Dataset S10.** Amino acid sequence alignments of the subset of rPKS (MGG\_08236) homologs in *Pezizomycotina* group.

**Dataset S11.** Maximum-likelihood tree based on the subset of rPKS (MGG\_08236) homologs in *Pezizomycotina* group.

## Acknowledgements

We acknowledge late Prof. Bharat B. Chattoo for his guidance and providing sophisticated laboratory facility at the Bharat Chattoo Genome Research Centre (BCGRC). We also fondly remember late Dr. Johannes Manjrekar for the useful scientific discussions during this work. We thank Dr. Malali Gowda and Dr. T. R. Sharma for providing blast-infected tissues for some of the isolates used in this study. We thank the BCGRC group at MSU for useful discussions. We thank Dr. Noppol Kobmoo (Biotec) for his useful insights on natural selection and population genetics analyses. Some sequence data were produced by the US Department of Energy (DOE) Joint Genome Institute (<https://www.jgi.doe.gov/>) in collaboration with the user community. We especially thank Dr. Rytas Vilgalys and Dr. Francis Martin for using the *Diaportheaceae* sp. PML\_573 genome generated within the framework of the 1000 Fungal Genome projects (CSP 1974). We also thank Dr. Riccardo Barancelli for letting us use the genomes of *Colletotrichum melonis* CBS 134730 (Colme1) and *Colletotrichum tamarilloi* CBS 129955 (Colta1).

## Author contributions

KM collected samples and contributed to the isolation of the Indian strains of blast fungus and phenotyping experiments. KM conducted formal analyses on diversity in BGCs, JC and JCN-M helped with the investigation of the analyses. KM, RP, and JC conducted image analyses and statistical analyses. JCN-M provided important in-house scripts for similarity network analysis. SB helped KM with the gene expression analysis. KM wrote the original draft of the manuscript. All authors contributed to the review and editing of the manuscript. KM, RP, and JC designed the study. RP provided the resources for the study. RP and JC supervised the study.

## Funding

Prof. Bharat Chattoo acquired the grant (BT/PR6949/AGIII/103/862/2012) from DBT, Gol. KM was supported by the DBT (Gol)-funded grant (BT/PR6949/AGIII/103/862/2012) and partly by the intramural University Research Scholarship from the M. S. University of Baroda. KM received EMBO—Scientific Exchange Grant (SEG no. 8866) to conduct part of this work at the Westerdijk Institute, the Netherlands. RP was awarded by SERB, Gol (EMR/2017/005303) and the Seed Grant, IIT Bombay.

## Availability of data and materials

Raw sequence data for the newly sequenced Indian strains under this study have been made accessible at SRA database under the BioProject ID PRJNA973860. All datasets generated for this study are included in the article and Additional files. All the Supplementary Datasets can be accessed through a link: [https://github.com/khyati-sb/Pyricularia\\_oryzae\\_secondary\\_metabolite\\_gene\\_clusters.git](https://github.com/khyati-sb/Pyricularia_oryzae_secondary_metabolite_gene_clusters.git)

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 25 June 2024 Accepted: 6 February 2025

Published online: 20 May 2025

## References

- Asuke S, Tanaka M, Hyon G-S, Inoue Y, Vy TTP, Niwamoto D, Nakayashiki H, Tosa Y. Evolution of an *eleusine*-specific subgroup of *Pyricularia oryzae* through a gain of an avirulence gene. *Mol Plant-Microbe Interact*. 2020;33(2):153–65. <https://doi.org/10.1094/MPMI-03-19-0083-R>.

- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, Van Wezel GP, Medema MH, Weber T. antiSMASH 60: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 2021;49(W1):W29–35. <https://doi.org/10.1093/nar/gkab335>.
- Böhnert HU, Fudal I, Dioh W, Tharreau D, Notteghem JL, Lebrun MH. A putative polyketide synthase/peptide synthetase from *magnaporthe grisea* signals pathogen attack to resistant rice. *Plant Cell.* 2004;16(9):2499–513. <https://doi.org/10.1105/tpc.104.022715>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England).* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brunner PC, McDonald BA. Evolutionary analyses of the avirulence effector AvrStb6 in global populations of *Zymoseptoria tritici* identify candidate amino acids involved in recognition. *Mol Plant Pathol.* 2018;19(8):1836–46. <https://doi.org/10.1111/mpp.12662>.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
- Chen S, Qiang S. Recent advances in tenuazonic acid as a potential herbicide. *Pestic Biochem Physiol.* 2017;143:252–7. <https://doi.org/10.1016/j.pestbp.2017.01.003>.
- Chernomor O, Von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol.* 2016;65(6):997–1008. <https://doi.org/10.1093/SYSBIO/SYW037>.
- Chiapello H, Mallet L, Guérin C, Aguilera G, Amselem J, Kroy T, Ortega-Abboud E, Lebrun M-H, Henrissat B, Gendralt A, Rodolphe F, Tharreau D, Fournier E. Deciphering genome content and evolutionary relationships of isolates from the fungus *Magnaporthe oryzae* attacking different host plants. *Genome Biol Evol.* 2015;7(10):2896–912. <https://doi.org/10.1093/gbe/evl187>.
- Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, Kusaba M, Yoshida K, Terauchi R, Fujita Y, Nakayashiki H, Valent B, Tosa Y. Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLoS Pathog.* 2011;7(7): e1002147. <https://doi.org/10.1371/journal.ppat.1002147>.
- Chumley FG, Valent B. Genetic analysis of melanin-deficient, nonpathogenic mutants of *magnaporthe grisea*. *Mol Plant Microbe Interact.* 1990;3(3):135. <https://doi.org/10.1094/MPMI-3-135>.
- Collemare J, Billard A, Böhnert HU, Lebrun M-H. Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity. *Mycol Res.* 2008a;112(2):207–15. <https://doi.org/10.1016/j.mycres.2007.08.003>.
- Collemare J, Pianfetti M, Houle A-E, Morin D, Camborde L, Gagey M-J, Barbisan C, Fudal I, Lebrun M-H, Böhnert HU. *Magnaporthe grisea* avirulence gene ACE1 belongs to an infection-specific gene cluster involved in secondary metabolism. *New Phytol.* 2008b;179(1):196–208. <https://doi.org/10.1111/j.1469-8137.2008.02459.x>.
- Collemare J, O'Connell R, Lebrun MH. Nonproteinaceous effectors: the terra incognita of plant–fungal interactions. *New Phytol.* 2019;223(2):590–6. <https://doi.org/10.1111/nph.15785>.
- Crouch JA, Tomaso-Peterson M. Anthracnose disease of centipedegrass turf caused by *Colletotrichum eremochloae*, a new fungal species closely related to *Colletotrichum sublineola*. *Mycologia.* 2012;104(5):1085–96. <https://doi.org/10.3852/11-317>.
- Dallery J-F, Lapalu N, Zampounis A, Pigné S, Luyten I, Amselem J, Wittenberg AHJ, Zhou S, de Queiroz MV, Robin GP, Auger A, Hainaut M, Henrissat B, Kim K-T, Lee Y-H, Lespinet O, Schwartz DC, Thon MR, O'Connell RJ. Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. *BMC Genomics.* 2017;18(1):667. <https://doi.org/10.1186/s12864-017-4083-x>.
- de Jonge R, Bolton M, Kombrink A, van den Berg GVD, Yadeta KA, Thomma B. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Res.* 2013. <https://doi.org/10.1101/GR.152660.112>.
- Dellaporta SL, Wood J, Hicks JB. A plant DNA miniprep: version II. *Plant Mol Biol Rep.* 1983;1(4):19–21. <https://doi.org/10.1007/BF02712670>.
- Ebbelo DJ. *Magnaporthe* as a model for understanding host-pathogen interactions. *Annu Rev Phytopathol.* 2007;45:437–56. <https://doi.org/10.1146/annurev.phyto.45.062806.094346>.
- Farman ML, Kim YS. Telomere hypervariability in *Magnaporthe oryzae*. *Mol Plant Pathol.* 2005;6(3):287–98. <https://doi.org/10.1111/J.1364-3703.2005.00285.X>.
- Fudal I, Collemare J, Böhnert HU, Melayah D, Lebrun M-H. Expression of *Magnaporthe grisea* Avirulence Gene ACE1 Is Connected to the Initiation of Appressorium-Mediated Penetration. *Eukaryot Cell.* 2007;6(3):546–54. <https://doi.org/10.1128/EC.00330-05>.
- Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics.* 2017;33(19):3088–90. <https://doi.org/10.1093/BIOINFORMATICS/BTX346>.
- Gilchrist CLM, Chooi YH. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics.* 2021;37(16):2473–5. <https://doi.org/10.1093/BIOINFORMATICS/BTAB007>.
- Giraud T, Gladieux P, Gavrillets S. Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol.* 2010;25(7):387–95. <https://doi.org/10.1016/J.TREE.2010.03.006>.
- Gladieux P, Popars J, Badouin H, Branca A, Aguilera G, De Vienne DM, Rodríguez De La Vega RC, Branco S, Giraud T. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol.* 2014;23(4):753–73. <https://doi.org/10.1111/mec.12631>.
- Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, Chen L, Terauchi R, Lebrun M, Tharreau D, Mitchell T, Pedley KF, Valent B, Talbot NJ, Farman M, Fournier E. Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *Mbio.* 2018;9(1):1–19. <https://doi.org/10.1128/mBio.01219-17>.
- Gómez Luciano LB, Tsai IJ, Chuma I, Tosa Y, Chen YH, Li JY, Li MY, Lu MYJ, Nakayashiki H, Li WH. Blast fungal genomes show frequent chromosomal changes, gene gains and losses, and effector gene turnover. *Mol Biol Evol.* 2019;36(6):1148–61. <https://doi.org/10.1093/molbev/msz045>.
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Ottillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, Smirnova T, Nordberg H, Dubchak I, Shabalov I. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 2014;42(D1):D699–704. <https://doi.org/10.1093/NAR/GKT1183>.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21. <https://doi.org/10.1093/SYSBIO/SYQ010>.
- Harris LJ, Balcerzak M, Johnston A, Schneiderman D, Ouellet T. Host-preferential Fusarium graminearum gene expression during infection of wheat, barley, and maize. *Fungal Biol.* 2016;120(1):111–23. <https://doi.org/10.1016/j.funbio.2015.10.010>.
- Hartmann FE, McDonald BA, Croll D. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Mol Ecol.* 2018;27(12):2725–41. <https://doi.org/10.1111/mec.14711>.
- Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;35(2):518–22. <https://doi.org/10.1093/MOLBEV/MSX281>.
- Inoue Y, Vy TTP, Yoshida K, Asano H, Mitsuoka C, Asuke S, Anh VL, Cumagun CJR, Chuma I, Terauchi R, Kato K, Mitchell T, Valent B, Farman M, Tosa Y. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science.* 2017;357(6346):80–3. <https://doi.org/10.1126/science.aam9654>.
- Izumi Y, Ohtani K, Miyamoto Y, Masunaka A, Fukumoto T, Gomi K, Tada Y, Ichimura K, Peever TL, Akimitsu K. A Polyketide synthase gene ACRTS2 is responsible for biosynthesis of host-selective ACR-toxin in the rough lemon pathotype of *alternaria alternata*. *Mol Plant-Microbe Interact.* 2012;25(11):1419–29. <https://doi.org/10.1094/MPMI-06-12-0155-R>.
- Jacob S, Grötsch T, Foster AJ, Schüffler A, Rieger PH, Sandjo LP, Liermann JC, Opatz T, Thines E. Unravelling the biosynthesis of pyriculariol in the rice blast fungus *magnaporthe oryzae*. *Microbiology (United Kingdom).* 2017;163(4):541–53. <https://doi.org/10.1099/mic.0.000396>.
- Jeon J, Lee GW, Kim KT, Park SY, Kim S, Kwon S, Huh A, Chung H, Lee DY, Kim CY, Lee YH. Transcriptome profiling of the rice blast fungus *magnaporthe oryzae* and its host *oryza sativa* during infection. *Mol Plant Microbe Interact.* 2020;33(2):141–4. <https://doi.org/10.1094/MPMI-07-19-0207-A>.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285>.
- Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–66. <https://doi.org/10.1093/NAR/GKF436>.
- Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hoof JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK,

- Collemare J, Linington RG, Medema MH. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 2019. <https://doi.org/10.1093/nar/gkz882>.
- Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. TimeTree 5: an expanded resource for species divergence times. *Mol Biol Evol.* 2022;39(8):174. <https://doi.org/10.1093/molbev/msac174>.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):R12. <https://doi.org/10.1186/gb-2004-5-2-r12>.
- Lambou K, Tag A, Lassagne A, Collemare J, Clergeot P-H, Barbisan C, Perret P, Tharreau D, Millazo J, Chartier E, Vries RPD, Hirsch J, Morel J-B, Beffa R, Kroj T, Thomas T, Lebrun M-H. The bZIP transcription factor BIP1 of the rice blast fungus is essential for infection and regulates a specific set of appressorium genes. *PLoS Pathog.* 2024;20(1): e1011945. <https://doi.org/10.1371/journal.ppat.1011945>.
- Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol.* 2014;14(1):1–14. <https://doi.org/10.1186/1471-2148-14-82>.
- Langner T, Harant A, Gomez-Luciano LB, Shrestha RK, Malmgren A, Latorre SM, Burbano HA, Win J, Kamoun S. Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLoS Genet.* 2021;17(2): e1009386. <https://doi.org/10.1371/journal.pgen.1009386>.
- Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021;49(W1):W293–6. <https://doi.org/10.1093/NAR/GKAB301>.
- Lim SM, Hooker AL. Southern corn leaf blight: genetic control of pathogenicity and toxin production in race T and race O of *Cochliobolus heterostrophus*. *Genetics.* 1971;69(1):115–7. <https://doi.org/10.1093/genetics/69.1.115>.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, Teeling E. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015>.
- Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappellini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. A computational framework to explore large-scale biosynthetic diversity. *Nature Chem Biol.* 2019;16(1):1–9. <https://doi.org/10.1038/s41589-019-0400-9>.
- Oh Y, Donofrio N, Pan H, Coughlan S, Brown DE, Meng S, Mitchell T, Dean RA. Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol.* 2008;9(5):R85. <https://doi.org/10.1186/gb-2008-9-5-r85>.
- Oliva RF, Cano LM, Raffaele S, Win J, Bozkurt TO, Belhaj K, Oh S-K, Thines M, Kamoun S. A recent expansion of the RXLR effector gene *Avrblb2* is maintained in global populations of *Phytophthora infestans* indicating different contributions to virulence. *Mol Plant-Microbe Interact.* 2015;28(8):901–12. <https://doi.org/10.1094/MPMI-12-14-0393-R>.
- Palma-Guerrero J, Ma X, Torriani SFF, Zala M, Francisco CS, Hartmann FE, Croll D, McDonald BA. Comparative transcriptome analyses in *zymoseptoria tritici* reveal significant differences in gene expression among strains during plant infection. *Mol Plant-Microbe Interact.* MPMI. 2017;30(3):231–44. <https://doi.org/10.1094/MPMI-07-16-0146-R>.
- Patkar RN, Naqvi NI. Fungal manipulation of hormone-regulated plant defense. *PLoS Pathog.* 2017;13(6):10–4. <https://doi.org/10.1371/journal.ppat.1006334>.
- Patkar RN, Ramos-Pamplona M, Gupta AP, Fan Y, Naqvi NI. Mitochondrial  $\beta$ -oxidation regulates organellar integrity and is necessary for conidial germination and invasive growth in *Magnaporthe oryzae*: Mitochondrial  $\beta$ -oxidation in *Magnaporthe*. *Mol Microbiol.* 2012;86(6):1345–63. <https://doi.org/10.1111/mmi.12060>.
- Patkar RN, Benke PI, Qu Z, Constance Chen YY, Yang F, Swarup S, Naqvi NI. A fungal monooxygenase-derived jasmonate attenuates host innate immunity. *Nat Chem Biol.* 2015;11(9):733–40. <https://doi.org/10.1038/nchembio.1885>.
- Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5(3): e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- Rahnama M, Novikova O, Starnes JH, Zhang S, Chen L, Farman ML. Transposon-mediated telomere destabilization: a driver of genome evolution in the blast fungus. *Nucleic Acids Res.* 2020;48(13):7197–217. <https://doi.org/10.1093/nar/gkaa287>.
- Rahnama M, Wang B, Dostart J, Novikova O, Yackzan D, Yackzan A, Bruss H, Baker M, Jacob H, Zhang X, Lamb A, Stewart A, Heist M, Hoover J, Calle P, Chen L, Liu J, Farman ML. Telomere roles in fungal genome evolution and adaptation. *Front Genet.* 2021;12: 676751. <https://doi.org/10.3389/fgene.2021.676751>.
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34(12):3299–302. <https://doi.org/10.1093/molbev/msx248>.
- Rudd JJ, Kanyuka K, Hassani-Pak K, Derbyshire M, Andongabo A, Devonshire J, Lysenko A, Saqi M, Desai NM, Powers SJ, Hooper J, Ambrosio L, Bharti A, Farmer A, Hammond-Kosack KE, Dietrich RA, Courbot M. Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. *Plant Physiol.* 2015;167(3):1158–85. <https://doi.org/10.1104/pp.114.255927>.
- Sánchez-Vallet A, Fouché S, Fudal I, Hartmann FE, Soyer JL, Tellier A, Croll D. The genome biology of effector gene evolution in filamentous plant pathogens. *Annu Rev Phytopathol.* 2018;56(1):21–40. <https://doi.org/10.1146/annurev-phyto-080516-035303>.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504. <https://doi.org/10.1101/GR.123930>.
- Shen Q, Liu Y, Naqvi NI. Fungal effectors at the crossroads of phytohormone signaling. *Curr Opin Microbiol.* 2018;46:1–6. <https://doi.org/10.1016/j.cmi.2018.01.006>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>.
- Soundararajan S, Jedd G, Li X, Ramos-pamplon M, Chua NH, Naqvi NI. Woronin body function in *Magnaporthe grisea* is essential for efficient pathogenesis and for survival during nitrogen starvation stress. *Plant Cell.* 2004;16:1564–74. <https://doi.org/10.1105/tpc.020677.2001>.
- Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics.* 2006;7(1):1–11. <https://doi.org/10.1186/1471-2105-7-62/TABLES/2>.
- Starnes JH, Thornbury DW, Novikova OS, Rehmeier CJ, Farman ML. Telomere-targeted retrotransposons in the rice blast fungus *Magnaporthe oryzae*: agents of telomere instability. *Genetics.* 2012;191(2):389–406. <https://doi.org/10.1534/genetics.111.137950>.
- Talbot NJ. On the trail of a cereal killer: exploring the biology of *magnaporthe grisea*. *Annu Rev Microbiol.* 2003;57(1):177–202. <https://doi.org/10.1146/annurev.micro.57.030502.090957>.
- Tsuge T, Harimoto Y, Akimitsu K, Ohtani K, Kodama M, Akagi Y, Egusa M, Yamamoto M, Otani H. Host-selective toxins produced by the plant pathogenic fungus *Alternaria alternata*. *FEMS Microbiol Rev.* 2013;37(1):44–66. <https://doi.org/10.1111/j.1574-6976.2012.00350.x>.
- Van Der Hoon R. Balancing selection favors guarding resistance proteins. *Trends Plant Sci.* 2002;7(2):67–71. [https://doi.org/10.1016/S1360-1385\(01\)02188-4](https://doi.org/10.1016/S1360-1385(01)02188-4).
- Vy TTP, Inoue Y, Asuke S, Chuma I, Nakayashiki H, Tosa Y. The ACE1 secondary metabolite gene cluster is a pathogenicity factor of wheat blast fungus. *Commun Biol.* 2024;7(1):812. <https://doi.org/10.1038/s42003-024-06517-7>.
- Walton JD. HC-Toxin. *Phytochemistry.* 2006;67(14):1406–13. <https://doi.org/10.1016/j.phytochem.2006.05.033>.
- Yan X, Tang B, Ryder LS, MacLean D, Were VM, Ezeola AB, Cruz-Mireles N, Ma W, Foster AJ, Osés-Ruiz M, Talbot NJ. The transcriptional landscape of plant infection by the rice blast fungus *Magnaporthe oryzae* reveals distinct families of temporally co-regulated and structurally conserved effectors. *Plant Cell.* 2023;35(5):1360–85. <https://doi.org/10.1093/plcell/koad036>.
- Yoshida K, Saunders DGO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano LM, Kamoun S, Terauchi R. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics.* 2016;17(1):370. <https://doi.org/10.1186/s12864-016-2690-6>.
- Zhong Z, Norvienyeku J, Chen M, Bao J, Lin L, Chen L, Lin Y, Wu X, Cai Z, Zhang Q, Lin X, Hong Y, Huang J, Xu L, Zhang H, Chen L, Tang W, Zheng H, Chen X, Nielsen R. Directional selection from host plants is a major force driving host specificity in *magnaporthe* species. *Sci Rep.* 2016;6:25591. <https://doi.org/10.1038/srep25591>.