Chapter 6
Discussion and Conclusion

6.1 Discussion

The blast fungus infects several economically important cereal crops worldwide and has been a serious threat to the global food security (Dean et al., 2012; Talbot, 2003). Although blast fungus has ability to infect most of the crop plants, it is considered as single species diverged into multiple host-specific lineages (Gladieux, Condon, et al., 2018). Investigation on recent disease outbreak on wheat plants suggested the host jump and host range expansion as a common evolutionary mechanism to get adapted to a new plant species, often driven by directional selection pressure from the hosts (Inoue et al., 2017; Islam et al., 2016; Maciel et al., 2014). Thus, understanding the population structure dynamics of *M. oryzae* is a fundamental to underlying evolutionary processes governing host specialization including host shifts and/or host range expansion. The comparative genome analyses of pathogenic field isolates of *Magnaporthe oryzae* from different host-specific lineages is useful in identifying the molecular determinants responsible for such genetic divergence.

6.1.1 Population structure of *M. oryzae* in India

In the present study, we collected a total of 133 field strains of *M. oryzae* from three host plants across fifteen different geographic locations of India. A species tree constructed using all *M. oryzae* genomes showed a genetic divergence of *M. oryzae* species into three host-specific lineages, which is concordant with the earlier published results (Gladieux, Condon, et al., 2018). Most strains clustered into the lineages representing the host of their origin. Intriguingly, MOS1 and MOS4 strains, which were isolated from infected rice leaf tissue, were found to be outside of the *Oryza* lineage based on the phylogenetic trees constructed using genome-wide SNPs dataset **(Chapter 4, Section 4.2)** as well as 2655 single copy orthologs **(Chapter 5, Section 5.1)**. MOS1 was found to be genetically similar to that of the *Eleusine*-lineage, but not to the strains of *Oryza* lineage. Pathogenicity test data of both these strains also pointed out differential virulence pattern, where fungal inoculated rice plants developed moderately-resistant lesions **(Chapter 3, Section 3.2.2.1)**. This suggests that both the strains most likely have been isolated from infected rice tissue probably due to two reasons – i) the strains are evolving to adapt to rice host or ii) could be the case of co-infection on the concerned rice plant tissue with two genetically distinct *M. oryzae* strains,

as hypothesized earlier as one of the mechanism for emergence of new pathogenic strain (Langner et al., 2018). MOS4 strain did show any genetic similarity with any of the other host-specific lineages, one such example reported earlier for rice-infecting strain 87-120, placed further ahead to *Setaria* lineage (Gladieux, Condon, et al., 2018). These strains define a second lineage that is able to infect rice, yet it may have another primary host that remains to be determined.

Further, we identified lineage-specific gene families, found to have virulence-related gene functions, such as metalloendopeptidase activity, transporters, CWDEs and genes responsible for SM biosynthesis **(Chapter 4, Section 4.5)**. These gene functions are often found to be correlated to effectors, which has ability to modulate the host physiology and suppress the plant immune responses (Chuma et al., 2011; Durairaj et al., 2016; Huang et al., 2022; Jacob et al., 2017; Thaker et al., 2022). Fungal pathogens are known to produce a large repertoire of effectors, including proteins and secondary metabolites, which play a pivotal role in its ability to infect various host plants and thus a major determinants of host specialization (Pedroso et al., 2008; Plissonneau et al., 2017; Sánchez-Vallet et al., 2018). The roles of various protein effectors have already been studied at the host-pathogen interface. Principally, genes encoding Small Secretory Proteins (SSPs) are avirulence (*AVR*) factors, known to be involved in gene-for-gene interactions with the corresponding resistance genes of host plants. We analyzed Presence/Absence Variations (PAV) of 46 characterized Blast effectors and found out the correlation between the repertoire of effector molecules and divergence of host-specific lineages **(Chapter 4, Fig. 4.12)**. The *AVR* repertoire in a certain host-specific lineage may be the result of co-evolution between the pathogen and its host. As a result of arms-race, the allele variants are found for most of AVR genes (AVR-Pita1, AVR-Pia, AVR-Pib, AVR-Pii, AVR-Pizt, and AVR-Pik) under field conditions (Kang, 2001; Kanzaki et al., 2012; W. Li et al., 2009; Yoshida et al., 2009). Further, gain- or loss-of-function mutation in avirulence (AVR) genes of *M. oryzae* results in overcoming host resistance or non-host resistance in the case of a new host. In the early 1980s, the *M. oryzae* strains with mutations in the avirulence gene *PWT3* emerged as a result of selection pressure from widely grown wheat cultivars carrying the corresponding R gene *RWT3* (Inoue et al., 2017). The *Eleusine* (finger millet) lineage of the blast fungus has been divided in two subgroups (EC-I and EC-II); and the infectivity of these subgroups on weeping lovegrass (*Eragrostis spp.*) varied based on the presence/absence of the avirulence

gene *PWL1*. Presence of *PWL1* in EC-II caused loss of pathogenesis on weeping lovegrass, whereas the absence of *PWL1* correlated to the infection ability of EC-I subgroup on both the host plants – finger millet and weeping lovegrass (Asuke et al., 2020). Transposable elements also plays a major role in increasing diversity in effector repertoire by their deletion, pseudogenization and/or horizontal transfers (Chuma et al., 2011; Yoshida et al., 2016).

6.1.2 Identification of potential chemical effectors

We performed a large-scale population genomics-based study to identify potential BGC associated with effector-like function. We utilized 53 publicly available genomes, in addition to 15 newly sequenced genomes of Indian strains of *M. oryzae*, representing the global population of blast fungus belonging to six different host-specific lineages, Our analysis of SM-producing biosynthetic gene clusters (BGCs) in these host-specific lineages showed that *M. oryzae* genome consists of an average 59 SM-producing BGCs **(Chapter 5, Section 5.1),** including a few with known SM products such as – melanine, epipyriculol, cytochalasin, pyrichalasin, squalestatin S1, and cercosporin. Such BGCs associated with known SMs could be essential in fungal development and/or pathogenesis and would be present in almost all the strains of *M. oryzae* and *M. grisea*. However, importantly, most of the BGCs did not share similarities with any of the reference MIBiG BGCs with known products **(Chapter 5, Section 5.2)**. This suggests their potential to produce yet uncharacterized SMs, and their likely roles in diverse biological processes.

Further, our similarity network analysis based on a total 4224 BGCs belonging to six host-specific lineages of *M. oryzae*, suggested three lineage-specific BGCs **(Chapter 5, Section 5.3)**. One of the three BGCs, named BGC-O1, is present in almost all the genomes from *Oryza* (23 out of 24) and *Eragrostis* (1 out of 1) lineages studied here. Whereas another *Oryza*-specific BGC (BGC-O2), comprising of a core biosynthetic gene, but without any typical SM-tailoring genes, was found in only ten rice-infecting strains studied here. Similarly, BGC-TLE – although specifically found in *Triticum*, *Lolium*, *Eragrostis* and *Eleusine* lineages – appears to be functional only in the latter two lineages. This suggests that the SM product of BGC-TLE is most likely required for the infection on finger millet and weeping lovegrass. Importantly, our population-genomics-based finding of the *Oryza* lineage-specific BGC is well supported by the PCR-based and gene expression analyses.

We found that the *Oryza*-lineage-specific rPKS (MGG_08236) transcript accumulated specifically during pathogenesis, with a significantly elevated expression at 24 and 72 hours post inoculation **(Chapter 5, Section 5.6)**. This suggests that the rPKS has a key role to play likely during host invasion and colonisation. Importantly, transcript of *ACE1* accumulated specifically during appressorial development and maturation, with a maximum-induced expression at 17 hpi during pathogenesis (Collemare, Pianfetti, et al., 2008; Fudal et al., 2007). Recent transcriptomics-based studies have also demonstrated that the expression of PKS from BGC-O1 (MGG_08236) and neighboring tailoring genes was induced during the initial biotrophic colonization stage of blast disease (Jeon et al., 2020; Yan et al., 2023). These findings strongly suggest that the MGG_08236-based SM could have a virulence-related and/or effector-like function in *M. oryzae* infection. Many plant fungal pathogens such as *Colletotrichum higginsianum, Fusarium graminearum, Zymoseptoria tritici, and Pyricularia oryzae* are known to produce a number of SMs during pathogenesis, especially during host penetration and colonisation (Collemare, Billard, et al., 2008; Dallery et al., 2017; Harris et al., 2016; Jeon et al., 2020; Palma-Guerrero et al., 2017; Patkar et al., 2015; Rudd et al., 2015; Yan et al., 2023). While certain fungal SMs are host-selective toxins, some others are known to have virulence-related and/or effector-like function during plant-pathogen interactions (Au et al., 2000; Audenaert et al., 2013; Chen & Qiang, 2017; Collemare et al., 2019; Skellam, 2017; Wight et al., 2009). To study if the BGC-O1 has any role in virulence and/or an effector-like function 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 a deletion mutant of *ACE1* PKS gene (Collemare, Pianfetti, et al., 2008). Thus, it still remains to be studied, using a functional genomics approach, whether the *PKS* gene (MGG_08236) has a role in host specialization.

While only three BGCs were present in host-specific lineages, most of the other BGCs were found in all the lineages of *M. oryzae*. Such conservation of BGCs in different lineages could indicate a key role in their virulence on cereals. However, such apparently conserved BGCs can still be dissimilar in their tailoring gene sequences and/or content. Such differential architecture of BGCs may be involved in production of distinct SMs with

potential effector-like or virulence-related functions in a host specific manner. Genome-wide within-species study in *Aspergillus fumigatus* has identified certain variations, such as gain or loss of SM related genes/gene clusters, cluster idiomorphs and single nucleotide polymorphisms (SNPs) within conserved clusters, to be associated with the genetic diversity in the SM gene clusters (Lind et al., 2017). Thus, analyses investigating variations among conserved SM-producing BGCs across different *M. oryzae* lineages could be useful in identifying the likely diversity in the SMs produced and their association with specific hosts.

6.1.3 Evolution of *Oryza*-specific cluster BGC-O1

Additionally, 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. Interestingly, 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 **(Chapter 5, Section 5.4)**. The chromosome ends tend to have substantially higher levels of polymorphism, because they frequently undergo spontaneous rearrangements and mutations. In the absence of sexual reproduction, asexually propagating pathogens likely utilize chromosomal rearrangements as one of the mechanisms of host-specialization, by diversifying effector repertoire, as shown in *Verticillium dahliae* (Jonge et al., 2013). Similarly, the regions adjacent to a telomere in *M. oryzae* are reported to be highly polymorphic and enriched with genes involved in interaction of the blast fungus with its host plants (Rahnama et al., 2021). Indeed, 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/or adaptation of *M. oryzae* (Chiapello et al., 2015; Chuma et al., 2011; Gómez Luciano et al., 2019; Jonge et al., 2013; Yoshida et al., 2016). One of the reasons underlying such instability in the telomeric and sub-telomeric regions is the presence/activity of the transposons (Ahead, 2012; Farman & Kim, 2005; Rahnama et al., 2020). Interestingly, genes encoding protein effectors, displaying presence/absence polymorphisms among various field strains, are found to be adjacent to transposons in *M. oryzae* (Chuma et al., 2011; Xue et al., 2012; Yoshida et al., 2009). Given that BGC-O1 is also flanked by transposons in most strains of the *Oryza* lineage of *M. oryzae*, we hypothesize that genomic alterations, most likely through deletions, may have resulted in

the loss of the cluster in other host-specialized lineages. We further found that, unlike in 70-15, the BGC-O1 in FR13 strain is located on a mini chromosome (UEMA03000009.1), which also carries the *ACE1* BGC (Langner et al., 2021). Importantly, the mini chromosome in FR13 carries several genes coding for PKS, cytochrome P450 and secreted effector proteins, and thus could have a crucial role in fungal development and/or pathogenesis (Langner et al., 2021). Retention of BGC-O1 in *Oryza* and *Eragrostis* lineages, despite its localization to unstable sub-telomeric region, highlights its role in pathogenesis towards these specific hosts.

6.1.4 Origin of BGC-O1

To explore the origin of BGC-O1, we constructed phylogenetic analysis based on the orthologs of core biosynthetic gene, and we found out that the rPKS (MGG_08236) has an ortholog (jgi.p_Coler1_670826) only in *C. eremochloae* **(Chapter 5, Section 5.5)** – a taxonomically distant fungal pathogen from *Sordariomycetes* class. *C. eremochloae* is a causal agent of anthracnose disease on centipede-grass turf in southern United States (Crouch & Tomaso-Peterson, 2012).

Interestingly, the *C. eremochloae* genome carries an additional yet weak homolog of MGG_08236, which is conserved among other *Colletotrichum* species. The presence of the two similar BGCs in *C. eremochloae* suggests that BGC-O1 likely originated from an ancestral duplication event. Preliminary bidirectional blast results show that the orthologs of the two tailoring genes (MGG_15107 and MGG_15108) of BGC-O1 are present outside the corresponding BGC in *C. eremochloae*, suggesting a possible role of chromosomal rearrangements in adaptive evolution. Thus, investigation of the architectures of the BGC-O1, with differential content of tailoring genes, in *M. oryzae* and *C. eremochloae*, suggests that distinct SM products are likely synthesized by these clusters in these two fungal species. It is possible that the BGC-O1 in the *Oryza* and *Eragrostis* lineages is involved in synthesis of a specific SM required for pathogenesis towards rice and weeping lovegrass, respectively, and that the cluster is selectively maintained likely due to the selection pressure from the respective host species. While the polyketide produced by the BGC-O1 might have a key role in helping the pathogen in adapting to a particular ecological niche/host, further study is required to functionally characterize the *Oryza*-specific BGC to substantiate the in silico finding.

Altogether, our findings highlight the importance of population genomics-based studies in identifying the role of a secondary metabolite gene cluster and the corresponding genomic rearrangements, likely driven by the host selection pressure, leading to a host-specialized lineage of the blast fungal pathogen. Lastly, this study highlights the yet unexplored potential of various secondary metabolites as chemical effector molecules shaping the blast fungus–host interactions.

6.2 Concluding Remarks

Effector repertoire of a pathogen determines its ability to infect a specific host. Although the blast fungus is a single species, it has diverged into various host-specific genetic lineages. Most studies have focused on the role of protein effectors in virulence and host specialization. We believe that fungal chemical or metabolite effectors could also play a key role in virulence and/or host specialization. In the present study, we used a large-scale comparative genomics approach on various *M. oryzae* strains belonging to six different host-specific lineages, which include 15 newly sequenced Indian host-specific field isolates. **Importantly, we identified a secondary metabolite (SM) biosynthetic gene cluster (BGC) which likely has shaped the specificity of certain blast fungal strains towards rice host.**

Key highlights of our study are as follows:

- A total of three lineage-specific BGCs were identified from the various host-specific lineages of *M. oryzae*.
- BGC-O1 is specifically present in the *Oryza* lineage of *M. oryzae*.
- BGC-O1 is located in sub-telomeric regions in *Oryza* lineage, and flanked by transposons likely responsible for genomic alterations (deletions) in other *M. oryzae* lineages.
- Host-driven selection pressure likely underlies retention of BGC-O1 in rice-infecting *M. oryzae* strains.
- Gene expression analysis highlights a key role for BGC-O1 specifically during host penetration and colonization.

We believe that our findings, using a model plant pathogen, are of interest to a large scientific community studying the role of chemical effectors in virulence or host-specialization of fungal pathogens. This study shall also be of interest to the researchers interested in the evolutionary mechanisms responsible for the ecological adaptations. This study particularly highlights importance of population-genomics based approach in identifying novel SM effectors associated with virulence and/or host specificity.