# Characterization of non-coding regions in the human pathogen *Aspergillus fumigatus* and its close non-pathogenic relatives.

By Alec Brown

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

**Biological Sciences**

May 10th, 2024

Nashville, Tennessee

Approved:

Katherine Friedman, Ph.D.

Nicole Creanza, Ph.D.

Ann Tate, Ph.D.

Gustavo Goldman, Ph.D.

Antonis Rokas, Ph.D.

## Dedication

To my wife, parents, and grandparents. Thank you for everything!

**Acknowledgements**

I'd like to thank Dr. Anotnis Rokas and members of the Rokas lab for helpful discussions and feedback. Research in the Rokas lab is supported by grants from the National Institutes of Health National Institute of Allergy and Infectious Diseases (R01 AI153356), the National Science Foundation (DEB-2110404), and the Burroughs Wellcome Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. I would also like to thank Dr. Matt Mead and Dr. Jacob Steenwyk for aiding in my research efforts throughout my graduate program and to Dr. Gustavo Goldman and the members of his lab for collaboration and aiding in my professional development.

# Table of Contents

# List of Tables

# List of Figures

## Chapter 1: Introduction

### Invasive Aspergillosis, a major fungal threat to public health

Inavasive Aspergillosis (IA) is a fungal infection caused by *Aspergillus* species, most commonly *Aspergillus fumigatus* (Latge et al., 2019). These fungi are ubiquitous in nature and can be found in various indoor and outdoor environments settings (Hameed et al., 2004; Pini et al., 2008). The infections can range from mild allergic reactions to severe invasive diseases, particularly in individuals with compromised immune systems (Bongomin et al., 2017). One of the primary modes of transmission is through the inhalation of airborne *Aspergillus* conidia, which are tiny asexual spores produced by the fungi. These conidia can be present in the air at varying concentrations and can be inhaled into the lungs. In healthy individuals, the immune system usually keeps the *Aspergillus* spores in check, preventing them from causing infection. However, in people with weakened immune defenses, such as those with HIV/AIDS, organ transplant recipients, or individuals undergoing chemotherapy, these spores can lead to serious respiratory and systemic infections, commonly resulting in death (Bongomin et al., 2017) (**Figure 1**).

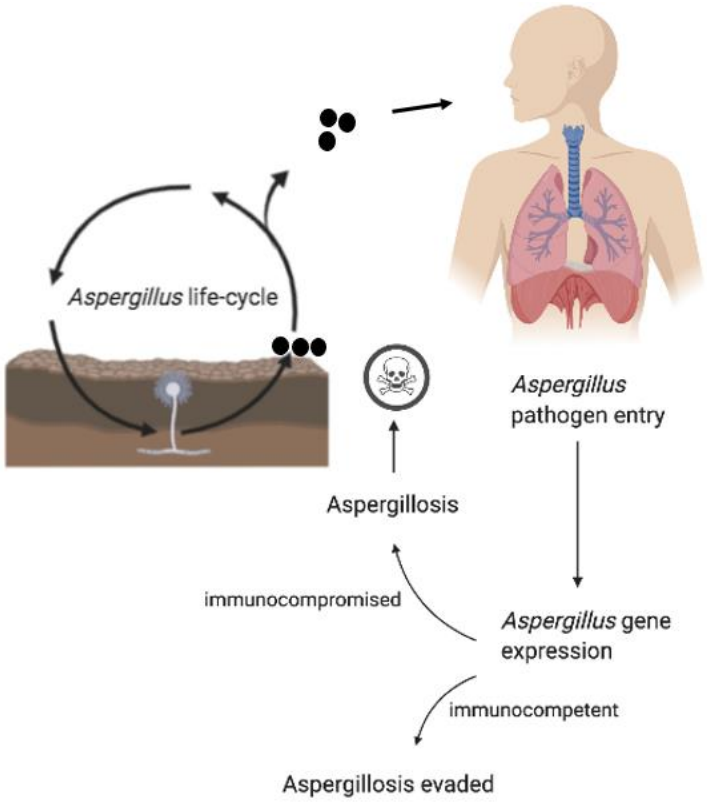



**Figure 1. Brief overview of Invasive Aspergillosis.** Aspergillus species are commonly found in soil environments, where they grow and release spores (black circles) into the air. As part of their life-cycle, spores can then land elsewhere in the environment, and the life cycle begin again. In other cases, Aspergillus spores can be inhaled by human. Once this happens, there are several dynamics and interchanges between Aspergillus gene expression and host factors. While the full story of Aspergillus pathogenicity remains unelucidated, what is known is that immunocompromised individuals are at risk of developing Aspergillosis.

## Immunocompetent patients

Different species of *Aspergillus* can result in persistent, noninvasive forms of infection with overlapping features, ranging from the development of a fungus ball (aspergilloma) to a chronic inflammatory and fibrotic process currently classified as chronic pulmonary aspergillosis (Alastruey-Izquierdo et al., 2018). In particular, colonization of a parenchymal lung cavity by *Aspergillus* is referred to as aspergilloma and consists of both dead and living mycelial elements, inflammatory cells, fibrin, mucus, and components of degenerating blood and epithelia. The mycelial mass may lie free within the cavity or be attached to the cavity wall by inflammatory/granulomatous tissue (Denning et al., 2018).

Diagnosing aspergilloma and other chronic pulmonary aspergillosis (CPA) forms leans on a set of standardized guidelines. These entail enduring symptoms lasting over three months, such as a persistent cough producing sputum and loss of weight, coupled with shortness of breath, fatigue, and potentially mild hemoptysis. Chest X-rays are the preferred diagnostic imaging technique, typically showcasing a rounded solid mass partly encircled by a clear crescent in the upper sections of the lungs. To better resolve the characteristics of an aspergilloma not visible on X-rays, a chest CT scan might be employed, facilitating a detailed understanding of the disease's pattern and reach. It is critical to distinguish CPA from a straightforward aspergilloma, as the former presents with systemic symptoms and involves more complex lung and pleural abnormalities, including the formation of cavities.

## Immunocompromised Patients

Upon the germination of conidia in the lung, hyphae of the *Aspergillus* species penetrate pulmonary arterioles and the lung parenchyma, consequently causing necrosis. This invasion can facilitate a systemic spread through the blood, leading to thrombosis and hemorrhagic infarction, as well as affecting distant organs such as the kidneys, liver, spleen, sinuses, and the central nervous system. Approximately one-third of autopsy results from individuals who suffered from IA reveal such widespread organ invasion brought about by the spread of *Aspergillus* hyphae through the arterioles (Pauw et al., 2008).

Prompt diagnosis is essential, particularly for individuals at heightened risk for IA, to enhance the prognosis of the disease. In spite of advancements over the past two decades, pinpointing a clear molecular marker for early IA detection remains a significant hurdle for both researchers and healthcare providers. Distinguishing IA from non-syndromic infections (colonization) or CPA is still a complex and imperfect process. This differentiation necessitates an amalgamation of clinical assessments, radiological evaluations, and microbiological insights (Ullmann et al., 2018). The establishment of uniform diagnostic criteria for invasive aspergillosis marks a pivotal step forward in managing IA (Pauw et al., 2008). Nevertheless, a substantial number of IA instances continue to go unnoticed by the current standards, particularly in nonneutropenic patients exhibiting uncharacteristic clinical and radiographic signs of the infection (Huang et al., 2017).

While recent years have seen a lack of substantial epidemiological research, the existing data implies a stability in clinical outcomes over the last 25 years, suggesting that the morbidity and

mortality figures cited in the 2000s remain applicable today. Since the late 1990s, IA has overtaken invasive candidiasis to become the most frequently identified fungal infection in post-mortem examinations at several facilities, with *Aspergillus* species being the cause of fungal pneumonia in about 15% to 20% of leukemia fatalities (Lehrnbecher et al., 2010). *A. fumigatus* also has implications and co-morbidity with COVID-19 (Steenwyk & Mead et al., 2021). Furthermore, IA's high prevalence and the financial burden of its treatment strategies have rendered it the most financially draining fungal illness in hospitals, with estimations of 1.2 billion dollars spent on hospitalizations caused by *Aspergillus* in the United States (Benedict et al., 2017).

## Introduction to *Aspergillus fumigatus*

*Aspergillus* fumigatus is a saprotrophic fungi that is found in soil and grows on decaying vegetation and play important roles in carbon and nitrogen recycling (Bandres et al., 2023). *A. fumigatus* is also responsible for most cases of Invasive Aspergillosis, a deadly disease which primarily affects the human lung, with a mortality rate > 50% (Bonomgin et al., 2017). *A. fumigatus* is globally distributed and can be isolated from many environments, including the soil, air and decaying vegetation. *A. fumigatus* is found in large quantities in garden and greenhouse soil and is a primary inhabitant of compost heaps and has also been found to be present in both urban and hospital settings (Hameed et al., 2004; Pini et al., 2008). *A. fumigatus* ability to survive in wide range of environments is, in large part, due to *A. fumigatus* ability to grow at a wide range of temperatures (25°C to above 37°C) and a wide range of pH (Bandres et al., 2023). *A. fumigatus* ability to survive different external pressures is intricately linked to its life cycle and its various stages.

## Life Cycle

**Conidium.** Under starvation conditions, *A. fumigatus* produces conidia (asexual spores) on specialized hyphal structures called conidiophores, which produce conidia on conidial heads. The dispersal of asexual *A. fumigatus* conidia in the environment is highly efficient compared with other *Aspergillus* species, as *A. fumigatus* conidia are much more hydrophobic than conidia produced by other species and can disperse easily (Kwon-Chung et al., 2013). This is reflected by the observation that *A. fumigatus* conidia were the dominant fungal components found after air sampling (Hameed et al., 2004), including air in hospitals (Pini et al., 2008). Humans are estimated to inhale 100–1,000 conidia each day, some of which can reach the alveoli in the lungs owing to their small diameter of 2–3μm (Latge et al., 2019) and hydrophobic outer layer (Valsecchi et al., 2018). Cilia on airway epithelial cells and resident alveolar macrophages remove inhaled conidia in immunocompetent individuals. For those who are unable to clear the inhaled conidia, conidia remain in the lungs and can germinate to cause invasive infection. The physiological human body temperature of 37°C degrees allow for the growth of *A. fumigatus*. Within 4–6 hours, conidia can germinate into short hyphae (van de Veerdonk et al., 2017).

**Ascospores**. Early literature reporting on *A. fumigatus* did not include the capacity for sexual reproduction. This changed in 2009 with the discovery of a sexual mating between *A. fumigatus* strains by O'Gorman et al., 2009. Briefly, crosses set up between two strains of opposite mating type, AFB62 (MAT11), isolated from a case of invasive aspergillosis, and AFIR928 (MAT12),

isolated from the environment, resulted in the formation of cleistothecia (completely closed fruit bodies that have no opening to the outside) that contained multiple ascospores (Sugui et al., 2011, O'Gorman et al., 2009). Further investigations have identified genes required for mating and cleistothecia formation (Yu et al., 2017). The prerequisites for sexual reproduction in *A. fumigatus* are exceedingly precise and non-physiological (O'Gorman et al., 2019). While the occurrence of ascospores in natural settings has not been documented, it may be the case that their sexual reproduction related processes enhance survival amidst adverse environmental circumstances. One such instance is the speculated role of ascospores in bolstering viability within compost heaps, where the temperature escalates significantly during the process of fermentation. Despite these suppositions, the precise function of ascospores within the life cycle of *A. fumigatus* remains unclear. However, they are unlikely to initiate aspergillosis, as ascospores are dormant and can only germinate after a 65°C thermal shock, a condition not found in humans.

**Mycelium**. Upon germination, subsequent hyphal or mycelial expansion leads to the establishment of a fungal colony. Within such colonies, the hyphae of *A. fumigatus* become embedded in an extracellular matrix (ECM), culminating in the formation of a biofilm (Morelli et al., 2021). These biofilms are engendered through either static growth conditions or in vivo scenarios during the emergence of an aspergilloma (Loussert et al., 2010). Notably, there exists a considerable divergence in the composition and configuration of the cell wall between hyphae that grow independently and those enmeshed within a biofilm (Beauvis et al., 2014). *A. fumigatus* biofilms function to increase adhesion as well as increasing antifungal resistance (Morelli et al., 2021). Mycelium cell wall architecture in filamentous fungi contrasts starkly with that of conidia and conidiophores, which are better suited for dissemination. *A. fumigatus* ranks among the fastest-growing species within the fungal realm, a trait shared with several saprotrophic counterparts, thus facilitating their colonization of diverse ecological niches. Preliminary investigations hinted at a direct correlation between the growth rates of distinct *A. fumigatus* isolates and their virulence within the host (Rhodes, 2006). However, evidence suggests this may not be the case (Puttikamonkul et al., 2010) as a more virulent strain of *A. fumigatus* virulent was shown to demonstrate a lower percent of conidia activity compared to a lesser virulent strain (Beauvais et al., 2013).

## Metabolism and stress response

*A. fumigatus* demonstrates a proficient ability to recycle both carbon and nitrogen, showcasing adaptability to a diverse spectrum of organic compounds to fuel its metabolic processes. Consequently, the fungus appears to lack stringent, specific nutritional prerequisites.
Within the human host, the tissues susceptible to *A. fumigatus* infections possess the potential to offer organic compounds and essential metals. Nonetheless, access to all indispensable nutrients is not readily available to the fungus. Host defense mechanisms encompass the concept of nutritional immunity, whereby *A. fumigatus* encounters impediments in procuring vital nutrients (Hartmann et al., 2011). For *A. fumigatus*, the lung serves as a 'sponge' necessitating degradation to liberate nutrients essential for fungal propagation. The fungus synthesizes a diverse array of proteases and enzymes that hold the capability to digest macromolecules within the environment (Perez-Cuesta et al., 2021), yielding organic components for metabolic utilization. *A. fumigatus* produces an assortment of proteases (Shemesh et al., 2017), facilitating its proliferation across a

myriad of environments. This broad biological repertoire inherent to these proteases and enzymes seemingly also contributes to the fungus's adeptness in securing essential nutrients within the human host. Yet, further investigations are required to fully understand the precise roles of specific proteases and enzymes in furnishing nutrients to *A. fumigatus* within the human lung.

## Nitrogen and carbon recycling

During the initial infection stages, obtaining nitrogen is essential. Thus, being contained inside host cells and thereby being deprived of nitrogen puts significant stress on *Aspergillus* species (Perez-Cuesta et al., 2021). Obtaining nitrogen is also required for the synthesis of amino acids, which requires regulated processes. For example, when the transcription factor CpcA is removed, it leads to increased vulnerability to amino acid deprivation. This is because CpcA plays a role in the cross-pathway control system, a significant network that primarily focuses on amino acid synthesis and helps organisms cope with starvation conditions. Its absence diminishes virulence (Krappmann et al., 2004). Strains missing the transcription factor AreA, responsible for guiding the use of nitrogen sources, show reduced virulence in mouse models of aspergillosis (Hensel et al., 1998). The transcription factor CreA is pivotal for selecting carbon sources, ensuring *A. nidulans* thrives in varied carbon and nitrogen environments (Ries et al., 2016). When CreA is absent, there is a noticeable decline in overall health and virulence in vivo (Beattie et al., 2017).

## Cation Acquisition

*A. fumigatus* requires cations like iron, zinc, and calcium for growth. Acquiring iron from host tissues is a challenge since most of it is attached to proteins, including heme and transferrin (Michels et al., 2022). To source iron, *A. fumigatus* employs a system called siderophores. This system has four key components: two extracellular siderophores (fusarinin C and triacetylfusarinin C) which are responsible for iron uptake, and two intracellular siderophores (ferricrocin and hydroxyferricrocin) that store iron in hyphae and conidia respectively (Haas 2012). Iron regulation and siderophore production are managed by transcription factors SreA and HapX. Excessive iron can harm *Aspergillus* species, so SreA acts to reduce iron acquisition during periods of ample iron (Schrettl et al., 2008), while HapX limits iron-utilizing pathways (Schrettl et al., 2010). Additionally, the “Sid” siderophore gene cluster also plays a key role in iron acquisition as well (Happacher et al., 2023).

For zinc uptake within the host, *A. fumigatus* relies on zinc transporters named ZrfA, ZrfB, and ZrfC (Amich et al., 2013). Their expression is overseen by the transcription factor ZafA, which is vital for the fungus's survival within the host (Moreno et al., 2007). Calcium is indispensable to *A. fumigatus* because it serves as a cellular signaling agent in various pathways. The balance of calcium within the cell is maintained through an array of calcium channels, pumps, and transporters (Brown & Goldman et al., 2016). The calcium/calcineurin pathway, which is in part governed by CrzA, a transcription factor which has gained interest as a required component for *A. fumigatus* virulence in mice models of infectious disease (Ries et al., 2017).

## Adaptation and survival in hypoxic conditions

*A. fumigatus* are obligate aerobes, needing oxygen to survive. However, in environments such as compost piles or human lungs, *A. fumigatus* might face reduced oxygen levels. The persistent inflammatory response during an infection can led to hypoxia by causing pulmonary tissue damage, consuming available oxygen, and blocking blood vessels (Latge et al., 2019). By lessening this inflammatory reaction, the impact of local tissue hypoxia can be diminished, improving outcomes in a mouse aspergillosis model (Gresnigt et al., 2016). Infections with *A. fumigatus* have been shown to cause local lung tissue hypoxia *in vivo* (Gresnigt et al., 2016; Grahl et al., 2011), emphasizing the organism's need to adapt to oxygen-scarce situations. The ability of the strain to grow and produce spores in low oxygen conditions is linked to its virulence (Chung et al., 2014). Moreover, growing in these hypoxic conditions triggers significant transcriptional and metabolic alterations (Barker et al., 2012; Losada et al., 2014). The sterol regulatory element-binding protein gene SrbA gets activated under conditions of low oxygen or iron scarcity. This gene is vital for effective ergosterol creation and iron balance in living organisms; when *A. fumigatus* lacks SrbA, its growth and virulence are notably diminished *in vivo* (Chung et al., 2014). The fungus's mitochondrial electron transport chain and alcohol dehydrogenase are crucial for adapting to low-oxygen environments and for maintaining virulence (Grahl et al., 2011). *Aspergillus* species can also worsen lung hypoxia by halting angiogenesis via producing the harmful compound gliotoxin, which stops the formation of new blood vessels in injured tissues (Ben-Ami et al., 2009).

## Cell wall

The infecting conidium and the growing mycelium have distinct cell wall compositions. There are variations in both the external molecules and internal components of the conidium when compared to the vegetative mycelium. The conidium's surface uniqueness and its water-repellent traits are due to its surface layer properties (Latge et al., 2019). In the conidium's cell wall, elements like α-1,3-glucan, melanin and the RodA hydrophobin create a thick shell around it (Carrion et al., 2013). The hyphal cell wall's carbohydrate distribution significantly varies from that of the conidial wall. Upon germination, the fungus discards its external melanin and rodlet layers, paving the way for hyphal growth. Yet, this shedding doesn't eliminate all melanin or hydrophobins from the hyphal wall. Depending on the surroundings, the *A. fumigatus* hyphal wall might contain melanin. Notably, the gene *pksP*, crucial for melanin synthesis, was active in the hyphae of budding conidia obtained from the lungs of immune-deficient mice (Langfelder et al., 2001), suggesting melanin a likely factor in aspergillosis virulence. As the *A. fumigatus* melanin pathways has been studied more, several additional genes, such as *arp1*, *arp2* and *arb2* have been shown to play a role in both melanin production and *A. fumigatus* virulence (Perez-Cuesta et al., 2019). However, the exact relationship between *A. fumigatus* melanin and its virulence remains a mystery.

*A. fumigatus* encounters many external antifungal agents, leading it to adapt its cell wall structure for protection. For instance, when faced with substances that disrupt the cell wall, including antifungal medications, the fungus triggers its cell wall integrity (CWI) pathway. This is primarily governed by the mitogen-activated protein kinase (MAPK) pathway, a key player in this signaling mechanism (Valiante et al., 2015). Genes involved in chitin synthase such as *csmA* and *csmB* along with genes more directly involved in the CWI pathways (such as mkk2) are also

involved in *A. fumigatus* virulence (Valiante et al., 2015). The CWI pathway plays a pivotal role in safeguarding the cell wall, guiding stress reactions, and influencing the energy allocation in growth and developmental processes.

Another essential response mechanism is the high-osmolarity glycerol (HOG) pathway (Ross et al., 2021). This pathway detects challenges like low pH, reactive oxygen species (ROS), oxygen scarcity, and antifungal medications. Calcium-mediated signaling is also involved, especially when responding to antifungal drugs targeting the cell wall (Cramer et al., 2008). This signaling engages the calcium-binding protein calmodulin and the protein phosphatase calcineurin. Furthermore, research has indicated that the target of rapamycin (TOR) signaling pathway notably amplifies the impact of the antifungal drug caspofungin on *A. fumigatus*. It's evident that the interplay between the TOR, CWI, and HOG pathways, combined with calcium signaling, modulates a dynamic cell wall based on environmental stimuli and stress (Brown & Goldman, 2016).

## Virulence / Pathogenicity

The evolutionary roots of the factors that drive the virulence of environmental fungi, like *A. fumigatus*, remain a mystery. In the environment, *A. fumigatus* confronts numerous living and non-living threats. It's theorized that thriving in such a challenging setting has served as a kind of evolutionary 'training ground' for the virulence of *A. fumigatus*, indicating that its survival skills developed for its natural habitat may be fundamental to its potency against humans. Conversely, some believe that because the natural environment varies vastly from the human body, the survival mechanisms for the former may not necessarily align with those needed for human virulence. A key point is that opportunistic pathogens from the environment, like *A. fumigatus*, might have faced various microorganisms and hosts in their ecological realm before infecting humans. The soil, where *Aspergillus* species thrive, is an intensely competitive domain, with *Aspergillus* species competing against roughly 4,000 different species for every gram of soil. In such a setting, the battle for resources is fierce, and certain adaptive mechanisms may also play a part in human virulence. The idea of 'dual use' virulence, initially proposed for *Cryptococcus neoformans* and other fungi, might be applied to *A. fumigatus* as well, illustrating how competitive traits for environmental survival might transition to human virulence. Recent research indicates similarities between *A. fumigatus*'s methods to dodge soil amoebae and its tactics to elude human immune defenses. One such method is intracellular germination, which leads to the bursting of the engulfing cell. In human immune cells, *A. fumigatus* can sprout within the phagosome, causing cell death. Moreover, *A. fumigatus* boasts an array of efflux pumps and transporters for defense against toxins, potentially offering a shield against antifungal treatments used in aspergillosis (Latge et al., 2019). Another way of framing *A. fumigatus* virulence is Arturo Casadevall's "Cards of Virulence" (Casadevall, 2006), which likens the virulome for humans and the ability of pathogenic microbes to survive within in the human host to a hand of cards. With each trait that may aid in the microbe's survival within a human as a playing card. With the correct hand of cards, the microbe can survive; and may prove to be pathogenic to the human host. Further investigations into *A. fumigatus* secondary metabolite production when compared across *Aspergillus* species has given credence to this concept (Steenwyk et al., 2020).

## Evasion of innate immunity

Conidia of *A. fumigatus* have developed various strategies to bypass the defenses of the host immune system. Elements on the surface of the conidia, like melanin and hydrophobin proteins, effectively conceal cell wall components that the immune system would typically recognize. Furthermore, melanin not only disguises the conidia but also actively hampers phagocytosis: it prevents phagosome acidification, impedes the NADPH oxidase complex (crucial for producing antifungal agents), and counters LAP, vital for the intracellular destruction of conidia by immune cells. The significant role of melanin in eluding the immune response is evident from studies showing that when melanin production is interrupted, the virulence of the fungus decreases in animal tests (Perez-Cuesta et al., 2019). Additionally, *A. fumigatus* secretes certain molecules that diminish the host's ability to phagocytose. For instance, GAG can prompt neutrophil cell death (Shevchenko et al., 2018). Toxins like fumagillin hinder neutrophil activity, while gliotoxin disrupts a signaling pathway in macrophages, weakening their function in phagocytosis and pathogen destruction (Shevchenko et al., 2018). Despite these defenses, it's intriguing to note that *A. fumigatus* typically doesn't infect healthy individuals. This is largely because the human immune system has a plethora of mechanisms to clear inhaled conidia, ensuring most people are naturally resistant to such infections. Daily, we inhale hundreds of these spores. Yet, the myriad of immune cells designed to detect and destroy these conidia in our respiratory system work efficiently to keep us safe. It's only when crucial parts of our immune defenses are compromised, as seen in certain conditions or treatments, that the fungal evasion techniques of *A. fumigatus* are potent enough to overcome human defenses and lead to an infection.

## Overview of Adaptive immune responses and adaptation of the host environment

Several T helper cells, including TH1, TH2, and cytotoxic T-cells, play roles in the body's response to *A. fumigatus*. These cells participate in both protective actions and potentially harmful immune responses (Latge et al., 2019). Which T cell group is activated hinges on factors like the specific fungal antigen, signaling pathways associated with pattern recognition receptors (PRR), and how dendritic cells present the antigen. TH2 and TH9 cells are linked with fungal allergies. In contrast, the responses of TH17 cells need precise control through IL-2 and gut microorganisms to offer effective defense against the fungus. While TH17 cells produce IL-22, which is protective against *A. fumigatus*, this production isn't limited to this T cell type; other cells like natural killer cells or innate lymphoid cells can also generate IL-22 (Dewi et al., 2017).

Furthermore, there are regulatory T cells (Treg cells) and type 1 regulatory T cells (Tr1 cells) specifically geared towards *A. fumigatus* in both humans and mice. These cells appear to have unique but interlinked roles in coordinating the immune response against this fungus. Tr1 cells control the growth of antigen-specific T cells, thereby reducing potential immune system damage, while Treg cells modulate the immune system's tolerance to fungal components associated with allergies (Seif et al., 2022). Given that humans possess specific T cells targeting *A. fumigatus*, a deeper comprehension of the adaptive immune system's reactions to this fungus can enhance patient risk assessment and treatment strategies, possibly leading to the creation of vaccines for those especially susceptible to infections. Although recent research has shed light on

the roles of T helper cells, the ideal functionality (or combination thereof) for defending against *A. fumigatus* remains uncertain.

While the ways in which *A. fumigatus* evades the innate immune system are akin to its environmental stress adaptations, the methods it uses to alter adaptive immune reactions can be both unexpected and intriguing. The cell wall molecule GAG, found in the *A. fumigatus* biofilm and released during hyphal growth, possesses significant immunomodulatory properties. In human immune cells, GAG triggers the production of the anti-inflammatory cytokine IL-1Ra, counteracting IL-1-driven immune activities (Seif et al., 2022). This includes promoting the essential TH17 cell response vital for fungal elimination. Interestingly, even though *A. fumigatus* doesn't come across cytokines in its natural environment, it has the capability to recognize and bind to the human cytokine IL-17A, subsequently adjusting its form, metabolism, and virulence (Seif et al., 2022). Furthermore, *A. fumigatus* can adjust to the host's limited nutrient supply and thrive in low oxygen environments. Its ability to stick to host cells and external matrices is vital for its ongoing presence in the host (Latge et al., 2019). Its spores can attach to diverse structures like the surface molecules of airway cells and the foundational layer, particularly in those with pre-existing lung damage, which is a risk factor for aspergillosis. Beyond influencing the host's response, GAG is essential for binding to various surfaces, inclusive of host cells, thus facilitating infection.

## Genetic Determinants of Virulence

An extensive literature and database search lead by Dr. Mead and Dr. Steenwyk identified 206 genetic determinants of virulence (Steenwyk & Mead et al., 2021). Genetic determinants of virulence (GDOV) are defined as genes that alter virulence in an animal model of disease when deleted or that are required for biosynthesis of secondary metabolites known to affect virulence. This definition resulted in a list of genes distinct from those previously published at the time, which include genes that contribute to allergy-related phenotypes and genes that are computationally predicted to contribute to virulence but have yet to be validated in an animal model of fungal disease (Steenwyk & Mead et al., 2021). The entire list of 206 GDOV be seen in **Figure 2**.

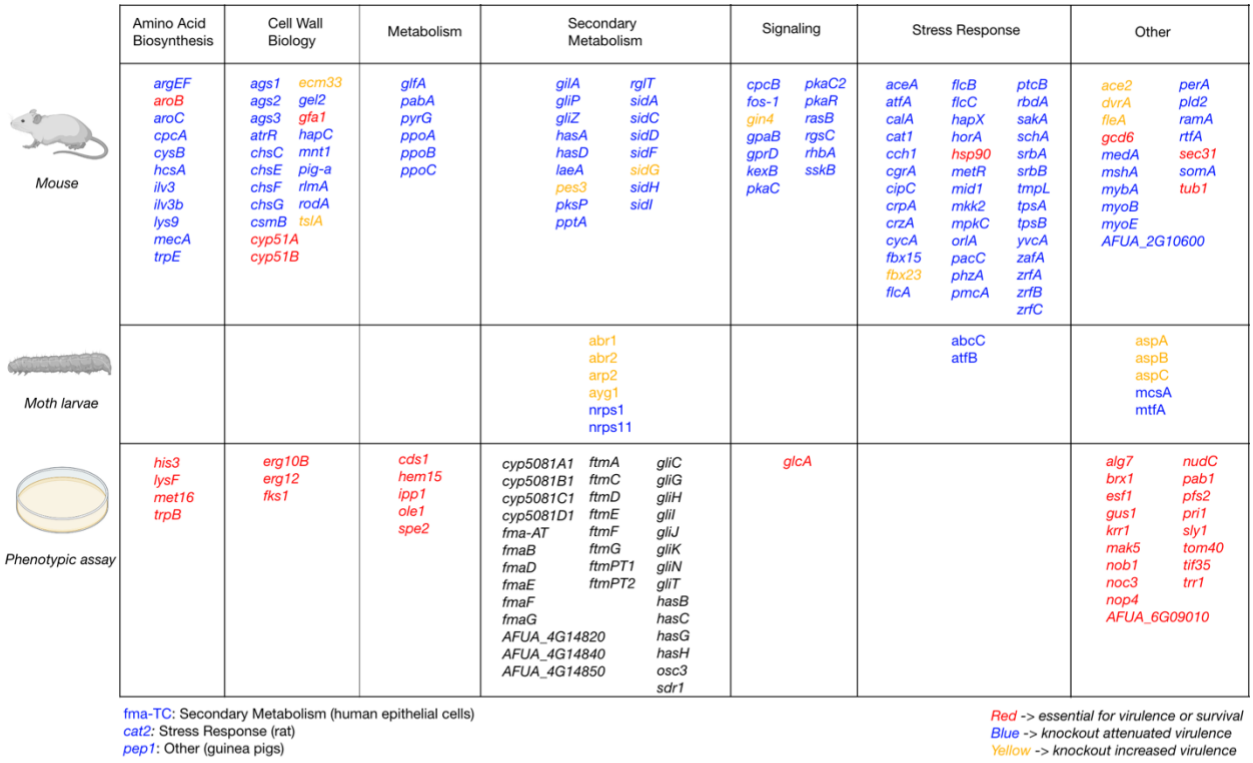



**Figure 2. The 206 Genetic Determinants of Virulence (GDOV).**

Figure 2 categorizes the 206 GDOV by gene function, the model used to determine its virulence (mouse mode, insect model or knockout phenotypic assay) and by how gene knockout impacts *A. fumigatus* virulence or survival. For those genes whose role in *A. fumigatus* virulence were determined by phenotypic assay, the gene knockouts resulted in a lack of survival, usually measured as the inability for *A. fumigatus* to grow normally on medium upon gene knockout. This is the case for GDOV involved in amino acid biosynthesis, cell wall, metabolism, signaling and other. However, this is not the case for genes involved in secondary metabolism (which are indicated in black instead of red, blue or yellow). GDOV that are involved in secondary metabolism but are not supported by evidence from an animal model of infectious disease represents 35 genes (2nd most amongst categories). The evidence for this set of genes comes from *A. fumigatus* studies showing that the production of certain metabolites plays a role in *A. fumigatus* virulence. Gliotoxin (a mycotoxin which impairs human immune cell function) has been shown to attenuate *A. fumigatus* virulence when gliZ, gliP and gliZ are knocked out (Sugui et al., 2007), as the production of gliotoxin plays a role in *A. fumigatus* virulence (Sugui et al., 2007). By extension, other *gli* genes required for gliotoxin production are also included as GDOV. Another example is the production of fumagillin (mycotoxin) which has also been demonstrated to impact *A. fumigatus* virulence (Guruceaga et al., 2021), thus the *fma* genes which regulate fumagillin production are listed as GDOV. The HAS-cluster is a secondary metabolite cluster that is silent under laboratory culture conditions (Irmer et al., 2015).

Overexpression of *hasA*, a putative transcription factor, leads to expression of all members of the cluster and consequently to synthesis of HAS (Yin et al., 2013). *HasA* and *hasD* knockout revealed attenuated virulence in mouse models, but the accessory *has* genes comprise 4 additional GDOV as requirements for HAS secretion. The final remaining large subsection of SM genes are the cyp5081set of genes, which are involved in helvolic acid biosynthesis and play an important role in oxidation-reduction processes (Mitsuguch et al., 2009). This subset of genes is considered "guilty by association" since there is no direct evidence to increment these genes, the production of the SM that these genes are involved producing / secreting have been shown to modulate *A. fumigatus* virulence, leading to their inclusion as a GDOV.

Of the 206 genetic determinants of virulence, 39 were shown to be essential to *A. fumigatus* virulence and/or survival via phenotypic assay. 31 of these genes were discerned via a classical gene knockout approach, in which knockout of the given gene led to an inability of *A. fumigatus* growth. However, for 8 of these genes, the knockout strain of *A. fumigatus* resulted in a lack of virulence in mouse models whereas all other animal model knockout assays resulted in attenuated or increase virulence, but not the removal of virulence (as measured by kill curve analysis). These 8 genes are *aroB, cyp51A, cyp51B, gfa1, hsp90, gcd6, sec31* and *tub1*.

*aroB* is a chorismate synthase which is involved in the biosynthesis of aromatic amino acids. Interestingly, of the several specific types of amino acids which are biosynthesized by *A. fumigatus*, including those whose expression play a role in *A. fumigatus* virulence and survival, it is specifically the biosynthesis of aromatics whose GDOV representative is essential for virulence in neutropenic mice (Sasse et al., 2016).

*cyp51A* and *cyp51B* are two isoforms of cyp51 as well as two of the most well studied *A. fumigatus* genes. *CYP51* is a crucial enzyme involved in the biosynthesis of ergosterol, a vital component of fungal cell membranes, including those of the pathogenic fungus "*Aspergillus* fumigatus." In the context of "*A. fumigatus*," CYP51 plays a pivotal role as it serves as a target for antifungal agents used in medical therapy (Roundtree et al., 2020). Specifically, azole antifungals like voriconazole and itraconazole target the CYP51 enzyme to inhibit the synthesis of ergosterol, which essentially impairs the growth and survival of the fungus by disrupting the integrity and function of its cell membrane (Roundtree et al., 2020). However, the extensive use of azole antifungals has led to the emergence of resistance mechanisms involving alterations in the *CYP51* gene. These mutations can lead to a reduced affinity of the drug to the *CYP51* enzyme, rendering the azole antifungals less effective (Zhang et al., 2019).

*Gfa1* encodes for glutamine-fructose-6-phosphate aminotransferase, catalyzing the first step in the chitin biosynthesis pathway. Chitin biosynthesis was shown to be essential for *A. fumigatus* virulence and is also required for *A. fumigatus* growth on certain growth media (Hu et al., 2007). The heat shock protein *hsp90* is required for virulence and acts in stress response to temperature but has also been shown to be intricately involved in the cell wall integrity pathway and cell wall stress adaptation (Rocha et al., 2021). The initiation of *A. fumigatus* infection occurs via dormant conidia deposition into the airways. Therefore, conidial germination and hyphal extension and growth occur in a sustained heat shock (HS) environment found in mammalian hosts from the *A. fumigatus* perspective (Rocha et al., 2021). Understanding the dynamics between *hsp90* and the cell wall integrity pathway remains a growing area of study (Rocha et al., 2021).

*Gcd6*, *sec31* and tub1 round out the GDOV that were shown to essential for *A. fumigatus* virulence in mice. Briefly, *gcd6* is a catalytic epsilon subunit of the translation initiation factor eIF2B; genes encoding translation factors are repressed by phagocytosis by murine macrophages (Hu et al., 2007). *Sec31* encodes for a component of the COPII coat of secretory pathway vesicles (Gautam et al., 2008) and tub1 encodes for an alpha-tubulin housekeeping gene (Mellado et al., 2007), but not much more is known about this gene.

Expanding our knowledge for any one of these 206 GDOV would provide greater insights in *A. fumigatus* virulence and survival. Recent studies have begun to focus on this list of genes (Brown et al., 2021; Mead et al., 2021). However, an even deeper understanding can be discerned by comparing the gene presence/absence and evolutionary history of *A. fumigatus* genes to that of closely related *Aspergillus* species.

**Brief overview of the diversity of *Aspergillus* species (Section *Fumigati*) and differences in their pathogenic profiles.**

Inhalation of *A. fumigatus* asexual spores along with a few other species in the genus cause a group of diseases collectively referred to as aspergillosis (see subchapter C). The most severe form of aspergillosis is invasive aspergillosis (IA), which primarily affects individuals with compromised immune systems or preexisting lung conditions (see subchapter B). Since drugs targeting IA are not always effective due to our lack of understanding of how they function inside the human host (Rosowski et al., 2019) and the evolution of drug resistance (Revie et al., 2018), infected individuals suffer high morbidity and mortality (Brown et al., 2012). Collectively, *Aspergillus* affect millions of patients and cause hundreds of thousands of life-threatening infections every year (Bongomin et al., 2017).

However, not every species within the *Aspergillus* genus contributes to infections equally. *A. fumigatus* is responsible for roughly 70% of all cases, leaving the remaining 30% to other species in the genus (Lamoth, 2016; Paulussen et al., 2016). This genus exhibits substantial diversity; while species such as *Aspergillus flavus* and *Aspergillus niger* have a distant kinship with *A. fumigatus*, each falling under different *Aspergillus* sections and displaying wide genomic divergence, others like *Aspergillus lentulus* and *Aspergillus udagawae* share a closer relationship, belonging to the same section as *A. fumigatus* - section *Fumigati* (a lineage of ~60 species that includes *A. fumigatus* and its close relatives) (Steenwyk et al., 2019; Kocsubé et al., 2016). Noteworthy is that a considerable fraction of species in this genus are either non-pathogenic or infrequently induce disease, though several can sporadically give rise to opportunistic infections in a range of hosts including mammals, birds, and occasionally other vertebrates and invertebrates (Kocsubé et al., 2016).

While substantial knowledge exists concerning the pathogenicity of *A. fumigatus* (Raffa & Keller, 2019; Rokas et al., 2020), a comprehensive understanding of the varied levels of pathogenicity across the genus remains a work-in-progress, necessitating a deeper dive into the contributing traits and genetic components. To unravel this, it is crucial to acknowledge that these species are not reliant on hosts for survival, with pathogenic effects emerging more as unintended consequences than evolved attributes. The pivotal inquiry, hence, revolves around

discerning how the traits facilitating survival in natural habitats like soil and litter environments inadvertently equipped a subset of these species with the ability to infect human hosts. By bridging this knowledge gap, we embark on a path of not only comprehending the evolutionary trajectory of pathogenicity within this genus but also fostering strategies to counter these opportunistic pathogens more effectively.

*A. fumigatus* leads to infections in over 300,000 individuals annually (Bongomin et al., 2017) and is considered the most clinically relevant amongst members of section *Fumigati*. Meanwhile, the closely related species *A. fischeri*, which showcases a considerable 95% average similarity in protein sequences to *A. fumigatus*, rarely figures in human disease cases and lacks clinical significance (Alastruey-Izquierdo et al., 2013). The evolutionary trajectory further unravels that the pathogenic trait in *A. fumigatus* emerged post its divergence from *A. fischeri* or its even nearer non-pathogenic kin A. *oerlinghausenensis*, steering to the inference that their most recent common progenitor was non-pathogenic.

Echoing a similar narrative, numerous other pathogens in this section delineate independent evolutionary paths to pathogenicity. To illustrate, although A. *udagawae* is the agent behind several thousand infections every year, its nearest relatives, including *Aspergillus aureolus*, *Aspergillus acrensis*, and *Aspergillus wyomingensis*, do not hold clinical importance (Alastruey-Izquierdo et al., 2013). This circumstance strongly advocates for the independent evolutionary acquisition of pathogenic traits by A. *udagawae*, distancing from a non-pathogenic shared ancestor with its close relatives. This evolutionary pathway spotlighting the isolated development of pathogenicity offers a rich ground for further exploration and understanding of the forces driving the emergence of pathogenic traits. Several factors, including the pervasive presence of small and easily airborne asexual spores, are believed to enhance the pathogenicity of *Aspergillus* species, particularly *A. fumigatus* (Paulussen et al., 2016). Despite the critical role these ecological traits play in human infections, they alone cannot elucidate the wide range of pathogenicity observed across species in the *Fumigati* section, as the variances in their ecologies remain apparent (Tong et al., 2017).

The ability to tolerate high temperatures is another notable factor influencing fungal pathogenicity (Robert et al., 2015). All examined species within the *Fumigati* section can survive at 37°C, depending on the medium facilitating their growth (Samson et al., 2007). This indicates that the propensity for pathogenicity is not merely a consequence of thermal tolerance. Although there is a considerable variation in their growth capabilities at this temperature, it's presumed that genetic factors predominantly influence this, hinting at a promising area for future investigations, especially focusing on growth patterns under specific stressful conditions simulating human infection environments (Samson et al., 2007).

Ascertainment bias might partly account for the dissimilarities in the pathogenicity spectrum observed, where systematic divergences from the accurate disease incidence caused by specific species result from current taxonomical identification methods of clinical isolates (Lamoth, 2016). This situation highlights a potential underestimation of the health impacts of cryptic species — those morphologically akin to major pathogens but genetically divergent. Despite this, consistent identification of known pathogens in molecular typing studies from various countries substantiates that the variations in pathogenicity are not just artifacts of misdiagnosis, suggesting

a partial genetic underpinning (Alastruey-Izquierdo et al., 2013; Negri et al., 2014). Further endorsing the genetic differences' role in pathogenicity are the distinct traits, and their related genes and pathways crucial for *A. fumigatus* pathogenicity, which showcase considerable genetic and phenotypic diversity among the *Fumigati* section species (Raffa & Keller, 2019). These traits encompass thermotolerance, response mechanisms to a variety of environmental stresses, including resistance to antifungal drugs, and the ability to produce a broad array of secondary metabolites with different structures (Frisvad & Larsen, 2016; Alastruey-Izquierdo et al., 2014). The data thus highlight the necessity for continued research to comprehensively understand the genetic bases influencing the observed spectrum of pathogenicity.

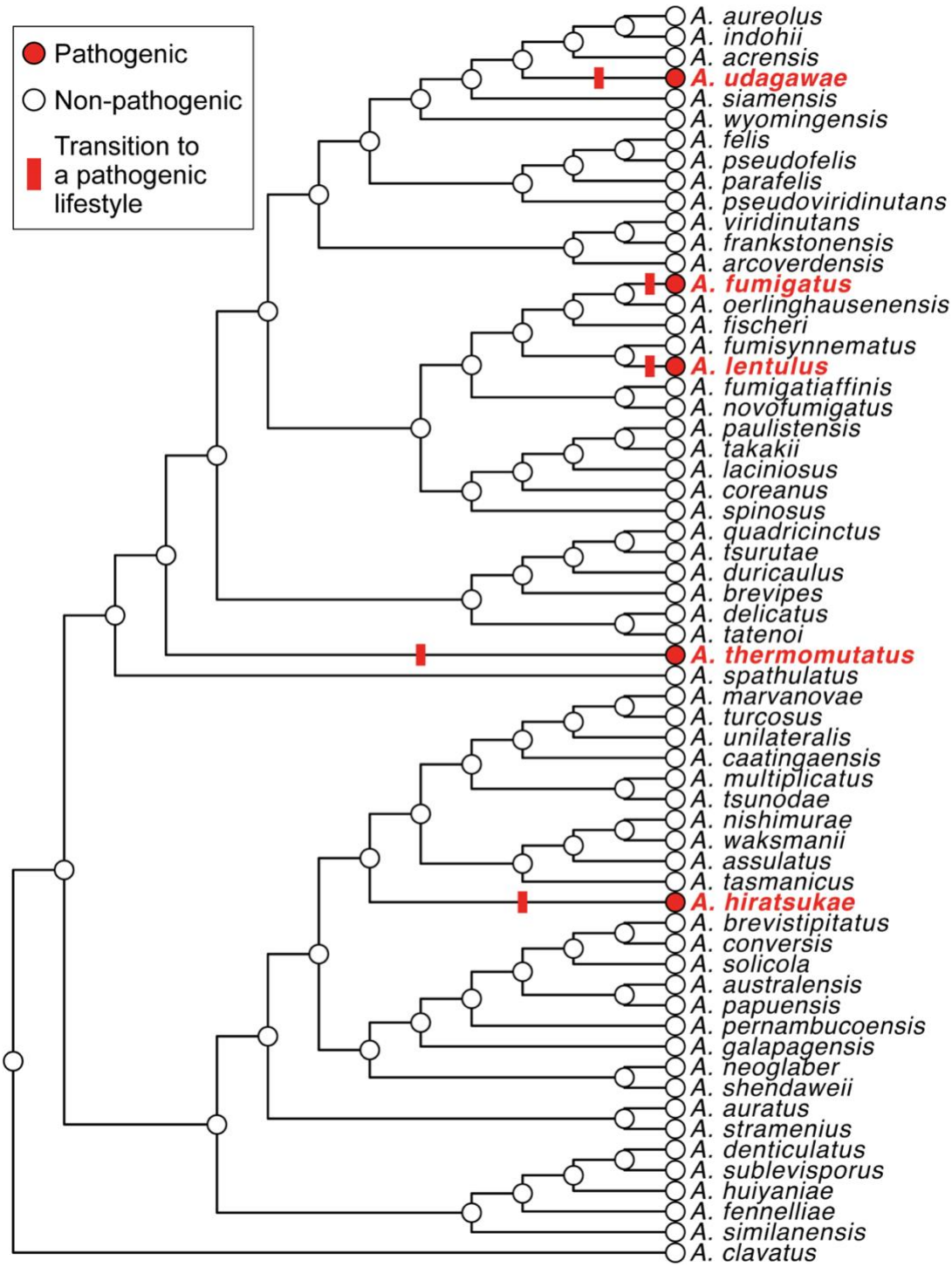



**Figure 3. *Aspergillus* phylogeny suggests that pathogenicity evolved multiple times independently in the lineage.**

Taxa marked in red represent Aspergillus species which are pathogenic to humans while those in white are considered non-pathogenic. Red bars represent a transition to a pathogenic lifestyle. From this sampling, Aspergillus pathogenicity appears to have evolved independently multiple times. Figure adapted from (Rokas et al., 2020).

**Conserved Pathogenicity and Species-Specific Models**

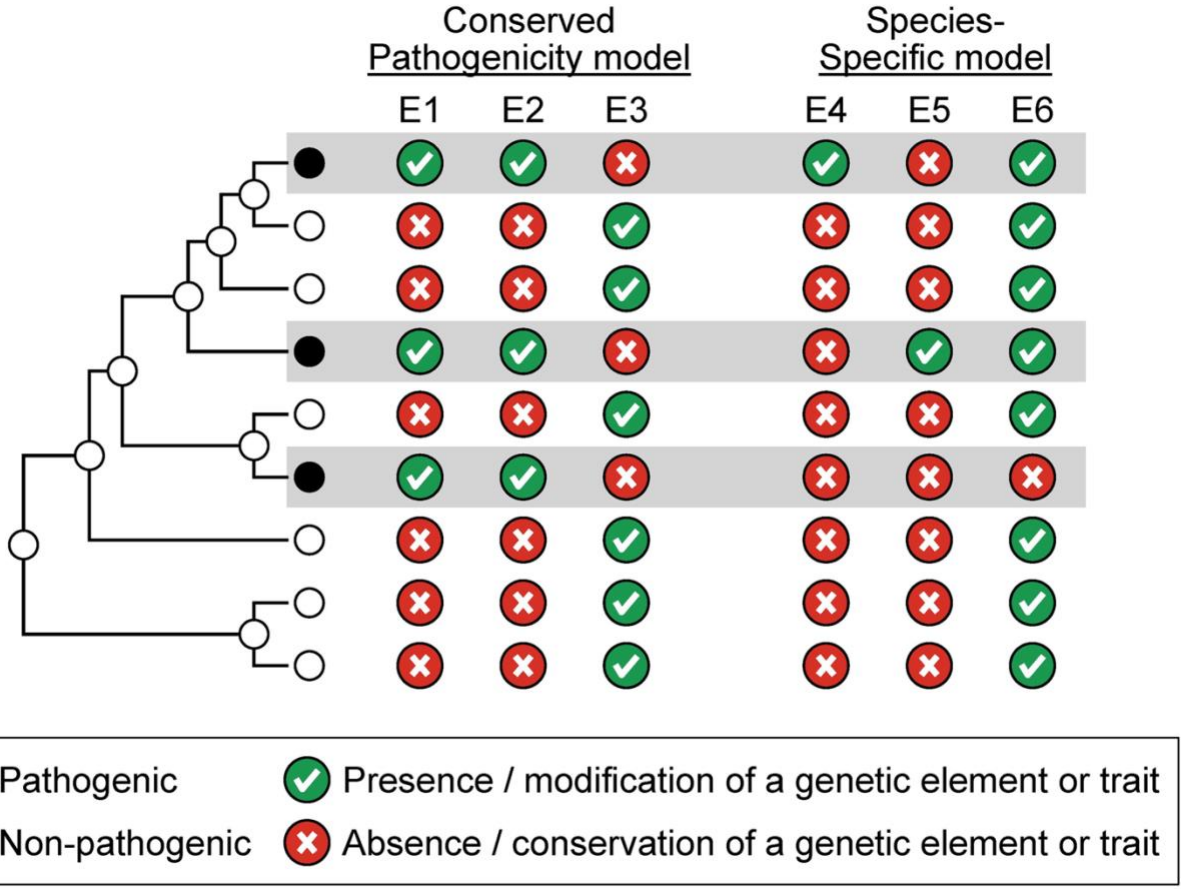



**Figure 4. Conserved pathogenicity and species-specific models. Figure adapted from (Rokas et al., 2020).**

Presented here are two potential and not necessarily exclusive theories, the "conserved pathogenicity" model and the "species-specific pathogenicity" model, to illustrate the genetic foundations of pathogenicity in *Aspergillus* section *Fumigati*. The conserved pathogenicity model suggests that *A. fumigatus* and other pathogenic species within this section harbor a set of common pathogenic traits and genetic factors (or shared differences in such elements) that are lacking in non-pathogenic species, either by possessing traits or elements such as E1 and E2 or not having elements like E3 as depicted in Figure 4.

Contrastingly, the species-specific pathogenicity model hypothesizes that individual pathogenic species have a distinctive array of traits and genetic factors (or unique variations in these elements) that set them apart from their non-pathogenic counterparts. This distinctiveness can emerge from the presence of unique traits or elements like E4 and E5 or the absence of elements such as E6 seen in Figure 4, found only in a particular pathogen and not in other pathogenic and non-pathogenic relatives.

It is essential to note that these mutual or species-specific genetic elements (or the disparities in these elements) are not just confined to variations in gene content; they encompass any genetic variation influencing pathogenic trait values. Such variations might span from minor differences, such as single or multiple nucleotide alterations in conserved protein-coding or non-coding (regulatory) regions, to more significant disparities including the presence or absence of comprehensive genetic pathways and networks. This framework accommodates a wide spectrum of genetic variations, offering a holistic approach to understanding the genetic dynamics governing pathogenicity.

Understanding which model accurately depicts the recurrent evolution of pathogenicity is crucial in crafting research approaches to delve into the fundamental molecular mechanisms present not only in the genus but also extensively in filamentous fungi. If we adhere to the conserved pathogenicity model, it anticipates that pathogenicity arises from the influence of preserved genetic components. Consequently, this implies that the acknowledged genetic factors influencing virulence in *A. fumigatus* could play a vital role in the virulence of other pathogenic entities within the *Aspergillus* species (Abad et al., 2010). This model has been the foundation for recent studies scrutinizing the extent to which genetic elements recognized as contributors to *A. fumigatus* pathogenicity find a parallel in other species (Abad et al., 2010; Rokas et al., 2020).

On the other hand, the species-specific pathogenicity model forecasts a contrasting scenario where each pathogenic entity houses a unique set of virulence determinants, thereby making the transfer of knowledge on virulence mechanisms from one species to another largely ineffective. Current genomic analyses, including those contrasting the principal pathogen *A. fumigatus* with its closely related non-pathogenic kin *A. fischeri*, and expansive evaluations of selected species throughout the genus, have started to illuminate the plausibility of both theories, presenting evidence in favor of each (Fedorova et al., 2008; de Vries et al., 2017). It's a significant stride towards a nuanced understanding of pathogenicity, laying a foundational basis for further exploratory research in this domain.

## Support for the conserved pathogenicity model

In line with the predictions set forth by the conserved pathogenicity model, studies encompassing numerous *A. fumigatus* genes traditionally linked to virulence have exhibited a high level of conservation in these virulence genetic factors amongst closely affiliated species (Mead et al., 2019; Abad et al., 2010). For instance, when comparing genomic aspects of *A. fumigatus* with its non-pathogen counterpart *A. fischeri*, 199 out of 206 GDOV including critical agents like CrzA and LaeA involved in calcium ion homeostasis and regulation of secondary metabolism respectively, are shared between the two species (Mead et al., 2019). Additionally, the known virulence factor gliotoxin (secondary metabolite) is biosynthesized in both *A. fumigatus* and *A. fischeri*, despite having vastly different pathogenic profiles (Knowles et al., 2020).

However, this research also brings to light the notion that the variable virulence witnessed across different organisms might not fundamentally originate from the distinctions in gene content; a focal point across comparative genomic studies of fungal pathogens and their non-pathogenic relatives (de Vries et al., 2017). Exploring how these preserved virulence genetic determinants

operate not just in other pathogenic but also in non-pathogenic species represents an intriguing avenue for further research, opening potential pathways to a deeper comprehension of fungal pathogenicity.

## Support for species-specific pathogenicity model

Gene content analyses across closely allied *Aspergillus* species have pinpointed numerous genes that seem to be exclusive to individual species. An expansive assessment comparing *A. fumigatus* strains Af293 and A1163 with *A. fischeri* and A. *clavatus* showed roughly 8.5% of the genes being singular to *A. fumigatus*, not found in the latter species (Fedorova et al., 2008). Interestingly, these distinctive genes of *A. fumigatus* predominantly occupy the subtelomeric regions of chromosomes, with their functionalities predominantly linked to metabolic processes including secondary metabolism, transport, and detoxification. This suggests a potential role of these genes in facilitating the survival of *A. fumigatus* within the human host (Fedorova et al., 2008). Furthermore, over two-thirds of the biosynthetic gene clusters identified in *A. fumigatus* are missing in the non-pathogenic relative *A. fischeri* (Mead et al., 2019); albeit some clusters are present in other species under the *Fumigati* section, implying a loss in *A. fischeri* rather than an emergence in *A. fumigatus*. It also seems that several clusters and their accompanying secondary metabolites have either distinctly emerged in *A. fumigatus* or have been widely lost in closely related species, rendering them unique to *A. fumigatus* now.

While recent studies have elucidated the potential of both the conserved-pathogenicity and species-specific model as viable theories, these theories have primarily focused on comparing reference *A. fumigatus* strains (typically either Af293 or A1163 (CEA10)) to close relatives. Far less is known about the comparative genomics and evolutionary histories between *A. fumigatus* strains and how these models might better explain *A. fumigatus* virulence at a population-level.

## Comparisons of *A. fumigatus* strains, a burgeoning field of study.

Comparative genomic and evolutionary analyses involving *A. fumigatus* have classically focused on comparing a reference *A. fumigatus* strain (typically Af293 or A1163 (CEA10)) to closely related species such as *A. fischeri*, A. *lentulus* and A. *clavatus* (see subchapter D). However, investigations in better understanding the differences between *A. fumigatus* strains has historically received less attention. The first comparative genomics study between *A. fumigatus* strains was a comparison between Af293 and A1163 (Fedorova et al., 2008), which revealed >98% orthologs across genes, and highly identical genomes. Despite being highly similar, A1163 was shown to be more virulent that Af293 (Colabardini et al., 2021), leading to the hypothesis that different *A. fumigatus* strains may display difference in their virulent propensities. Early studies into *A. fumigatus* strain comparisons revealed nonsynonymous mutations in several genes along with nucleotide deletions when comparing 8 *A. fumigatus* isolates from different patients with Aspergillosis, potentially implying that strains may differ from host to host (Hagiwara et al., 2014). Comparisons of secondary metabolite variation across 66 *A. fumigatus* strains revealed 5 general types of variation in SM gene clusters: nonfunctional gene polymorphisms; gene gain and loss polymorphisms; whole cluster gain and loss polymorphisms; allelic polymorphisms, in which different alleles corresponded to distinct, nonhomologous clusters; and location polymorphisms (Lind et al., 2017).

Over the past 5 years, the number of publicly available *A. fumigatus* strains with whole genome sequences and annotations has increased to > 300 strains and counting. This has opened the door to several strain comparisons with a large dataset. For example, comparisons of 300 isolates revealed that 7,563 of the 10,907 unique orthogroups (69%) are core genes and present in all isolates and the remaining 3,344 (accessory genes) show presence/absence of variation, representing 16–22% of the genome of each isolate (Barber et al., 2021). In this same research, the authors found that 43% of the virulence genes in their study (155 genes) had some degree of genetic variation expected to affect gene function and that secondary metabolism genes showed the highest variability among the virulence-associated genes, with 59 genes either being absent or showing a predicted loss of function among the 300 isolates (Barber et al., 2021).

Another question concerning *A. fumigatus* strains that had remain unclear until recently was if there were differences between *A. fumigatus* isolates from clinical patients vs those isolates collected from the environment. Larger number of *A. fumigatus* isolated has allowed deeper inquiries. Interestingly, phylogenetic inferences revealed several instances of clinical isolates being more closely grouped to environmental isolates as opposed to other clinical isolates (Barber et al., 2021). Another study involving 260 *A. fumigatus* strains found evidence for 3 primary populations of *A. fumigatus*, with recombination occurring only rarely between populations and often within them (Lofgren et al., 2022). Additionally, these 3 populations are structured by both gene variation and distinct patterns of gene presence–absence with unique suites of accessory genes present exclusively in each clade (Lofgren et al., 2022). Again, these clades included representatives of both clinical and environmental isolates. This study also found single non-synonymous variants and single nucleotide polymorphisms in both core and accessory genes, including in several GDOV (Lofgren et al., 2022). In yet another study involving 206 *A. fumigatus* strain, the researchers noted that no significant differences were observed among clinical versus environmental isolates when phylogeny was accounted for, despite observing ~40.6% of orthologs to be accessory (vary in their gene presences across all isolates (Horta et al., 2022)). Taken together, differences in clinical vs environmental and differences in geography are not drivers for differences between strains. Moreover, there exist few differences in genes across strains, and those differences that have been quantitated are either in gene presence/absence across orthologs or differences in gene nucleotide sequence, commonly observed to be SNPs. While differences in genes have been documented between *A. fumigatus* strains and between *A. fumigatus* and close relatives, what is less known is how non-coding regions may differ and what are the drivers behind such differences.

## Non-coding regions, a molecular view

It is recognized that hereditary shifts in gene expression contribute significantly to the evolutionary trajectory of various phenotypes, this is often guided by alterations in non-coding regulatory components such as promoters and enhancers (Emerson et al., 2010). This transformation hinges on two primary avenues: cis and trans effects. Cis effects refer to genetic fluctuations happening on the identical DNA molecule as the pertinent gene, influencing transcription factor binding locales in promoters or enhancers. Conversely, trans effects are driven by diffusible entities, potentially impacting any part of the genome, including transcription factors (Singor & Nuzhdin et al., 2018). The elucidation of these mechanistic underpinnings of gene expression evolution stands as a pivotal objective in evolutionary biology.

Research endeavors leveraging allele-specific RNA sequencing, encompassing the analysis of parental strains alongside F1 hybrid progenies, have significantly illuminated the role of cis and trans effects in dictating gene expression evolution. This methodology allows for the discernment of expression variations steered by cis and trans variants, unraveling their relative impacts across distinct taxa such as yeast, insects, and plants, amongst others (Emerson et al., 2010; Wittkopp et al., 2008; Goncalves et al., 2012). Investigations into non-coding regions found prevalent role of cis effects inter-species, whereas trans effects seem more pronounced intra-species, with a frequent concurrent manifestation influencing target gene expression reciprocally, a phenomenon potentially orchestrated by stabilizing selection over evolutionary spans (Mattioli et al., 2020; Gordon & Ruvinsky, 2012).

## Cis-Regulatory Elements

Cis-regulatory elements are genetic regions situated nearby (or sometimes within) the genes they control. These regions harbor DNA sequences such as promoters, enhancers, silencers, and insulators which play a pivotal role in governing the rate of gene transcription. Promoters, which are located adjacent to the gene they control, serve as binding sites for RNA polymerase and other transcription factors, facilitating the initiation of transcription. Enhancers and silencers, which are more commonly found in more complex eukaryotic organisms (Kolovos et al., 2012) can be located much farther away, and they increase or decrease the level of transcription, respectively. These cis-elements operate in a gene-specific manner, meaning their regulatory effects are limited to the nearby gene(s). They play a fundamental role in the tissue-specific and developmental stage-specific expression of genes, thereby aiding in the establishment of an organism's complex biological traits through meticulous spatial and temporal control over gene expression (Mattioli et al., 2020). The mutations in these regions can result in altered gene expression, sometimes leading to diseases and various phenotypic aberrations (Vande Zande et al., 2022).

## Trans-Regulatory Elements

In contrast, trans-regulatory elements are not confined to a specific location relative to the genes they regulate. These elements involve genes that encode for transcription factors or small RNAs that can travel through the cell and influence the expression of target genes at distant locales within the genome. The trans-regulatory factors operate through binding to cis-regulatory elements, thereby modulating their activity. Since these factors can potentially interact with multiple cis-elements across different genomic locations, they exhibit a more dispersed influence

on gene expression, contributing to the coordinated regulation of numerous genes that partake in a particular cellular pathway or process (Zhang & Emerson, 2019) This implies that changes in a single trans-regulatory factor can have ripple effects, influencing a broad spectrum of genes and thus bearing a significant potential for evolutionary changes in phenotypes. The coordinated action of trans-regulatory elements maintains the homeostasis of gene networks by orchestrating the synchronous expression of genes involved in common physiological processes (Zhang & Emerson, 2019).

The intricacy of gene regulation lies in the sophisticated interplay between cis and trans regulatory elements, orchestrating a network of gene expressions that underpin an organism's physiology. While cis-regulatory elements offer a localized control over gene expression, allowing for fine-tuned regulation, trans-regulatory elements ensure the coordinated regulation of a set of genes, bringing about a harmonized response to physiological and environmental cues (Zhang & Emerson, 2019). This dynamic and multifaceted regulatory system, comprising both cis and trans elements, fosters adaptability and complexity in biological systems, playing a cardinal role in evolutionary processes by engendering diversity in gene expression patterns, which can be the substrate for natural selection. For the research presented in this thesis document, the non-coding region exploration will focus on finding cis-regulatory elements. However, understanding both cis and trans elements is key to understanding the role of non-coding regions on gene expression.

**Role of cis-regulatory non-coding regions in gene regulation, a look at classical models.**

Much of what we know about fungal gene regulation comes from studies of the yeast *Saccharomyces cerevisiae* (Abdulrehman et al., 2011). Most yeast genes are transcribed by the enzyme RNA polymerase II (this is also true for filamentous fungi, such as *A. fumigatus*) (Kaplan, 2012). At least 50 proteins, including RNA polymerase can be involved in transcribing a gene (Kaplan, 2012). Some of these proteins bind DNA directly and can act to form a platform for polymerase. Other proteins are enzymes which can act to unwind DNA and promote polymerase initiate transcription. Other enzymes can promote or hinder transcription by chemically modifying histones. Proteins that bind to DNA and impact gene regulation (in this case, for a nearby gene(s)) are referred to as transcription factors (Rothernberg, 2022). The dynamics and intricacies between transcription factors and gene regulation are often multifaceted, complex and elegant (Rothernberg et al., 2022). A classic example of gene regulation and transcription factor binding is observed in the regulation of a GAL gene.

The GAL pathway in yeast is a well-studied regulatory system that governs the utilization of galactose as a carbon source when glucose is scarce (Harrison et al., 2022). This pathway is crucial for the yeast Saccharomyces cerevisiae to adapt to changing environmental conditions (Harrison et al., 2022).. At the heart of the GAL pathway is the Gal4 transcription factor, which plays a pivotal role in regulating the expression of genes involved in galactose metabolism (Harrison et al., 2022)..

Gal4 is a transcriptional activator that binds to specific DNA sequences in the upstream regulatory regions of GAL genes (Traven et al., 2006). These genes code for enzymes involved in the breakdown and utilization of galactose (Harrison et al., 2022).. Gal4 is inactive under

glucose-rich conditions because it is sequestered in the cytoplasm by Gal80, a repressor protein (Harrison et al., 2022).. However, in the presence of galactose, Gal3, a sensor protein, binds to galactose and interacts with Gal80, relieving its inhibition of Gal4 (Harrison et al., 2022).. This allows Gal4 to enter the nucleus and activate the transcription of GAL genes. Gal4 activates GAL gene expression by binding to upstream activating sequences (UAS) in the promoter regions of these genes. These UAS elements are specific DNA sequences recognized by Gal4, and their presence allows for the recruitment of RNA polymerase and other transcriptional machinery, leading to gene transcription (Traven et al., 2006). The binding of Gal4 to UAS is a prime example of a transcription factor recognizing and binding to specific cis-regulatory elements to modulate gene expression. Non-coding regions of DNA also play a critical role in GAL gene regulation. These regions include the promoter and enhancer sequences, as well as the UAS elements (Traven et al., 2006). Non-coding regions also provide the necessary binding sites for transcription factors like Gal4 and other regulatory proteins to interact with the DNA and control gene expression (Traven et al., 2006). Additionally, the GAL pathway involves complex regulatory loops and feedback mechanisms that ensure precise control over gene expression levels (Sellick et al., 2008). While what is written here is an extremely brief overview of the entire GAL pathway and its dynamics, the GAL pathway serves as a classic example of the interplay between transcription factors binding at specific sequences and subsequent regulation of gene expression.

### Non-coding regions and gene regulation in *Aspergillus fumigatus*

Compared to protein-coding regions, little is known about non-coding regions in *A. fumigatus* as it pertains to gene regulation. Models, such as the GAL pathway in *S. cerevisiae* provide insights on general function/pathways but much remains unknown about *A. fumigatus* specifically. What follows are some statistics, that I have found, about non-coding regions of *A. fumigatus* as of the time of this writing:

1. *A. fumigatus* genome size is 29.4 Mb, contains 9,630 protein-coding genes. For comparison *A. fischeri* genome size is 32.5Mb and contains 10,407 genes.
2. 51% of the *A. fumigatus* genome is composed of non-coding regions.
3. 80% of *A. fumigatus* genes contain at least one intron and the average number of introns per gene is ~1.8.
4. *A. fumigatus* encodes for ~400 transcription factors.
5. There are currently 6 transcription factor protein whose binding targets have been elucidated via ChIP-Seq methodology. These 6 are AtrR, CrzA, HapX, NctA, SrbA, and RglT.

When it comes to comparing non-coding regions between *A. fumigatus* and other species/strains, key questions that still remain include how (if any) changes are there across non-coding regions, how can these changes be characterized and what are the drivers of these changes? In this respect, a shift from a molecular focused approach to an evolutionary approaches are warranted to find and describe non-coding regions.

## Non-coding regions, an evolutionary view and general theory

### Characterizing sequence difference between species

In molecular population genetics or evolutionary studies, it is often assumed that DNA regions are subject to no natural selection, so that all observed mutations in each region are selectively neutral. For example, in protein-coding region synonymous mutations are often assumed to be neutral while non-coding region are usually assumed to be subject to no selective constraints. The most common test used for identifying regions (or sites) exhibiting selection between species is the dN/dS test.

The dN/dS test is a pivotal method in molecular biology and evolutionary genetics, utilized to assess the selective pressures acting on protein-coding genes. This ratio, also known as the nonsynonymous (d/N) to synonymous (d/S) substitution rate ratio, offers insights into the evolutionary forces driving the alterations in coding sequences and helps decipher whether a gene is undergoing positive selection, negative selection, or evolving neutrally. The fundamental theory underscoring the dN/dS test revolves around the different impacts of synonymous and nonsynonymous mutations. Synonymous mutations are changes in the DNA sequence that do not alter the amino acid sequence of the protein, whereas nonsynonymous mutations do result in a change in the amino acid sequence, and potentially the function of the protein.

The first step in conducting the dN/dS test is typically to align the coding sequences of the gene of interest from different species or populations and ensure that the sequences are homologous. The number of synonymous (S) and nonsynonymous (N) sites, along with the number of synonymous (s) and nonsynonymous (n) substitutions, are calculated. The rates of synonymous and nonsynonymous substitutions (d/S and d/N, respectively) are then estimated using these values. The formulae for estimating dN/dS and dN can be quite intricate, accounting for multiple substitutions at the same site and variations in the substitution rates across different codons. Several methods and models, such as the Jukes-Cantor model and the Kimura two-parameter model, can be used for these estimations. Additionally, the dN/dS can be specified for a certain branches across a given phylogeny or can be used to identify signatures of selection at specific sites. The resulting value of a dN/dS analysis can yield the following results:

1. Neutral Evolution: When dN/dS =1, it suggests that nonsynonymous and synonymous substitutions are occurring at the same rate, indicative of neutral evolution.

2. Positive Selection: When 1 dN/dS >1 this indicates that nonsynonymous substitutions are occurring more frequently than synonymous ones, suggesting that the altered protein sequences are being favored by natural selection.

3. Negative or Purifying Selection: Conversely, a dN/dS < 1 signifies that nonsynonymous substitutions are less frequent, indicating that changes in the protein sequence are selectively constrained (purifying selection).

The dN/dS test can also be adapted for non-coding regions, referred to as the $dN_{noncoding}/dS$. This test is similar to the dN/dS, except that the non-coding region (in which all sites are considered nonsynonymous, since any of these sites may play a role in gene expression) are used for dN and the dS comes from the synonymous sites of the associated protein-coding region. In this case, a dN/dS > 1 suggests that substitutions are occurring more frequently in the non-coding regions than synonymous ones, indicating the potential for positive selection in the non-coding region. Subsequently, these non-coding region then make for candidate sites for further exploration as it pertains to gene expression.

## Characterizing sequence differences within species

The tests presented in this section are for testing the neutral mutation hypothesis. This states that both the variation within populations and the differences between populations are due to neutral or nearly neutral mutations. Tests based on utilizing within and between population DNA sequence comparisons include the Hudon-Kreitman-Aguade (HKA) test and the McDonald-Kreitman (MK) test (**Fig 5**).

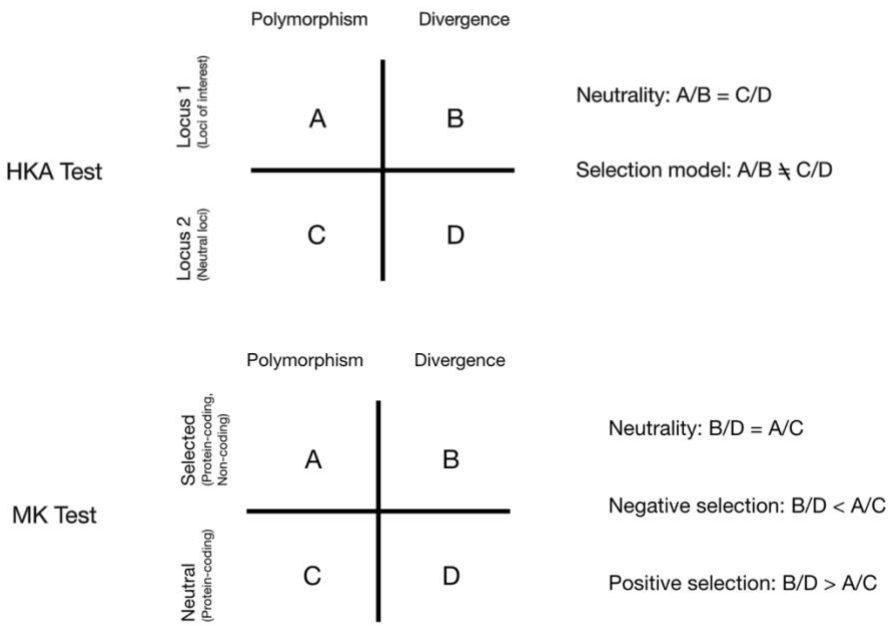



**Figure 5. Descriptions of the HKA and MK tests.**

## HKA test

The Hudson-Kreitman-Aguadé (HKA) test, conceived independently by Richard Hudson, Martin Kreitman, and Montserrat Aguadé in 1987, stands as a pivotal method in population genetics. This test has been instrumental in investigating whether the observed level of genetic polymorphism within a species aligns with the level of divergence between species, a critical assessment under the neutral theory of molecular evolution. The neutral theory assumes the following:

1. Neutral Evolution: Nearly all most mutations are selectively neutral.
2. Constant Population Size: Constant population size over time, a significant consideration given that population size fluctuations can skew results.
3. Random Mating: A panmictic population, where mating is random.
4. No Recombination: The test stipulates that there is no recombination within the regions being compared across loci.
5. No Migration: The populations compared are isolated, with no gene flow between them.

The HKA test is informative to determining whether levels of within and between population DNA variation are positively correlated, as predicted by the neutral mutation hypothesis. By comparing the observed quantity of polymorphism and divergence across different genetic loci, the HKA test determines whether the observed genetic variations are consistent with the predictions laid out by the neutral theory. This assessment allows a framework for understanding the evolutionary dynamics of populations and discerning the role of natural selection and genetic drift in shaping genetic diversity.

The basic formulation of the HKA test involves comparing observed values of polymorphism and divergence to their expected values under the neutral theory across multiple loci. A chi-square statistic is typically used to determine whether the observed data significantly deviate from the expectations under the neutral model.

Let Pi be the observed polymorphism at locus i, Di be the observed divergence at locus i, and Mi be the mutation rate at locus i. Under the neutral theory, E(Pi/Mi) should be equal to E(Di/Mi) for all loci. The expected ratio of polymorphism to divergence is thus the same across all loci, and deviations from this expectation can be tested using a chi-square statistic.

The HKA is not region-specific, meaning that the test can be used to observe polymorphism and divergence in either protein-coding or non-coding regions. One of the pioneering HKA test use-cases, levels of polymorphisms and divergence was investigated in the *Adh* (alcohol dehydrogenase) gene and 5' flanking region in 3 species of *Drosophila melanogaster* (Kreitman & Hudson, 1991). This study considered all sites in the non-coding 5' flanking region as silent sites (position at which mutations do not change polypeptide sequence encoded by the gene). In a coding region, only certain sites were considered sites (synonymous sites), these sites are represented by the 3$^{rd}$ codons, while other sites are generally considered nonsynonymous. Using the HKA tests, they found that coding regions were the cause of

departure from neutrality for the *adh* gene. Moreover, this tests provided a method to compare polymorphisms and divergence within a population for either protein-coding or non-coding regions.

## MK test

The MK test provides a method to compare the patterns of within-species polymorphisms and between-species divergence at synonymous (silent) sites and non-synonymous (replacement) sites in the coding-regions of a gene. In a singular coding-region, the synonymous sites and the nonsynonymous sites should have the same evolutionary history since the they are tightly linked. Thus, if polymorphisms and divergence are due to neutral the ratio of nonsynonymous to synonymous between species should be the same as within species. A significance difference between the two ratios is then used to reject the neutral mutation hypothesis.

McDonald and Kreitman applied this test to DNA sequences of the coding region of Adh gene from *D. melanogaster*, *D. simulans* and *D. yakuba* (McDonald & Kreitman, 1991). They found that the ratio of nonsynonymous changes to synonymous changes that are fixed between species is significantly greater than the ratio of nonsynonymous to synonymous polymorphisms, an indicator of positive selection. In contrast, examples of the ratio of nonsynonymous changes to synonymous changes that are fixed between species being significantly lower than the ratio of nonsynonymous to synonymous polymorphisms is thought to be an indicator of negative selection and represents mutations that persist in a population but are not fixed.

The MK-test has been suggested to be a conservative test since nonsynonymous mutations are subject to stronger negative selection than are synonymous mutations. Further implementations of the MK test have also considered codon usage bias.

Due to its comparisons of nonsynonymous and synonymous sites, the MK test would not be applicable to non-coding regions. While a comparison between non-coding regions would not be applicable, a comparison between a protein-coding region and an associated non-coding region (for example, an upstream promoter region) can be used. To do so requires additional assumptions. First, one must assume that the non-coding region plays a role in the expression of the gene and second, that all non-coding sites are nonsynonymous, since any of these sites may play a role in gene expression. Given these assumptions, the non-coding MK test can be applied to a given protein-coding region and an associated non-coding region to compare the ratio of nonsynonymous to synonymous changes between and within species where the nonsynonymous sites are represented by the non-coding region and synonymous sites are represented by the synonymous sites from the associated protein-coding region. In this case, if the ratio of nonsynonymous changes to synonymous changes that are fixed between species is significantly greater than the ratio of nonsynonymous to synonymous polymorphisms, then this is indicator that positive selection is due to a relatively higher number of mutations in the non-coding region.

## Thesis Aims

In the 15 years since the first *A. fumigatus* genome was published, hundreds of genes have been shown to contribute to virulence. However, efforts to identify genetic determinants of *A. fumigatus* virulence have overwhelmingly focused on protein-coding regions, whereas the impact of noncoding regions on the regulation of *A. fumigatus* genes known to contribute to virulence, and to the organism's ability to cause disease more generally, is much less understood.

**I hypothesize that there is abundant genetic variation in noncoding regions among *A. fumigatus* strains and between *Aspergillus* species and that this variation contributes to the observed differences in pathogenicity.** I will test this hypothesis by (i) comparing genome-wide sequence patterns of variation in noncoding regions between *A. fumigatus* and closely related species (Aim I), (ii) examining levels of genetic variation in noncoding regions between strains of *A. fumigatus* (Aim II)**,** and (iii) Descriptive analyses of non-coding regions comparing *A. fumigatus* Af293, *A. fumigatus* A1163 and *A. fischeri* to identify differences in putative transcription factor binding sites of known regulators involved in virulence. (Aim III).

**Specific Aim I (Chapter II): Compare genome-wide sequence patterns of variation in noncoding regions between *A. fumigatus* and closely related species.** I hypothesize that differences in noncoding regions involved in regulating the expression of orthologous virulence genes contribute to the disparity between *A. fumigatus* pathogenicity and that of closely related *Aspergillus* species. To test this, I will compare the noncoding regions upstream of the transcription start site and the protein-coding regions (as a control) of all single copy orthologous genes (orthogroups) present in *A. fumigatus* Af293 strain and 9 closely related *Aspergillus* species (5 pathogenic and 5 nonpathogenic). This set will include 206 genes in which the *A. fumigatus* gene is a known genetic determinant of virulence. For noncoding and protein-coding regions of each orthogroup, I will estimate their mutation rates and identify regions of greatest sequence divergence between species. The results of this aim will help to identify genetic differences between *A. fumigatus* and close relatives in noncoding regions upstream of genes known to be genetic determinants of *A. fumigatus* virulence.

**Specific Aim II (Chapter III): Identify and compare levels of genetic variation in noncoding regions between strains of *A. fumigatus*.** I hypothesize that differences in noncoding regions are potential contributors to differences in regulation of genetic determinants of virulence between *A. fumigatus* strains. I will test this by comparing the noncoding and protein-coding regions of all *A. fumigatus* genes, including those known to be involved in virulence, in a set of 265 strains to infer whole-genome genetic variation for single nucleotide polymorphisms, insertions-deletion polymorphisms and copy number variants. The results of this aim will aid in our understanding of noncoding region variation in *A. fumigatus* and how this variation may contribute to differences in the regulation of genes that contribute to virulence in different *A. fumigatus* strains.

**Specific Aim III (Chapter IV): : Descriptive analyses of non-coding regions comparing *A. fumigatus* Af293, *A. fumigatus* A1163 and *A. fischeri* to identify differences in putative transcription factor binding sites of known regulators involved in virulence.** I hypothesize that differences in TF binding motifs contribute to expression differences of genes

that play a role in *A. fumigatus* virulence. To test this, I will identify TF binding site presence/absence differences for a set of previously characterized TF binding sites between *A. fumigatus* and *A. fischeri* for TFs that are known genetic determinants of virulence in *A. fumigatus*. Non-coding regions that differ in the presence/absence or sequence content of transcription factor binding sites will be candidates for the observed differences in virulence; these could be experimentally tested in future experiments. These results will provide the first insights as to how non-coding region differences may contribute to differences in gene expression and virulence between *A. fumigatus* and *A. fischeri*.

It would be helpful to map these aims onto your chapters (just put the chapter numbers above).

## Chapter 2: Extensive non-coding sequence divergence between the major human pathogen *Aspergillus fumigatus* and its relatives

**(This Chapter is adapted from Brown et al., 2022)**


### Abstract

Invasive aspergillosis is a deadly fungal disease; more than 400,000 patients are infected worldwide each year and the mortality rate can be as high as 50-95%. Of the ~450 species in the genus *Aspergillus* only a few are known to be clinically relevant, with the major pathogen *Aspergillus fumigatus* being responsible for ~50% of all invasive mold infections. Genomic comparisons between *A. fumigatus* and other *Aspergillus* species have historically focused on protein-coding regions. However, most *A. fumigatus* genes, including those that modulate its virulence, are also present in other pathogenic and non-pathogenic closely related species. Our hypothesis is that differential gene regulation – mediated through the non-coding regions upstream of genes' first codon – contributes to *A. fumigatus* pathogenicity. To begin testing this, we compared non-coding regions upstream of the first codon of single-copy orthologous genes from the two *A. fumigatus* reference strains Af293 and A1163 and eight closely related *Aspergillus* section *Fumigati* species. We found that these non-coding regions showed extensive sequence variation and lack of homology across species. By examining the evolutionary rates of both protein-coding and non-coding regions in a subset of orthologous genes with highly conserved non-coding regions across the phylogeny, we identified 418 genes, including 25 genes known to modulate *A. fumigatus* virulence, whose non-coding regions exhibit a different rate of evolution in *A. fumigatus*. Examination of sequence alignments of these non-coding regions revealed numerous instances of insertions, deletions, and other types of mutations of at least a few nucleotides in *A. fumigatus* compared to its close relatives. These results show that closely related *Aspergillus* species that vary greatly in their pathogenicity exhibit extensive non-coding sequence variation and identify numerous changes in non-coding regions of *A. fumigatus* genes known to contribute to virulence.

## Introduction

Invasive aspergillosis (IA), a human disease caused by members of the fungal genus *Aspergillus*, is responsible for >400,000 cases worldwide per year with a mortality rate between 50-95% (Bongomin et al., 2017). More than 90% of IA cases are caused by *Aspergillus fumigatus*, with about a dozen other species such as *Aspergillus lentulus*, *Aspergillus thermomutatus*, and *Aspergillus udagawae* accounting for the rest (Steinbach et al., 2012; Rokas et al., 2020). Studies in both environmental (Flores et al., 2013) and hospital settings (Wirmann et al., 2018) show that asexual spores (conidia) of *A. fumigatus* and many other *Aspergillus* species are present in the air, yet *A. fumigatus* causes IA more frequently than its close relatives.

IA begins with inhalation of *Aspergillus* asexual spores and subsequent interaction between the asexual spores and the epithelium of the lung (Chotirmall et al., 2013). Several defense mechanisms including physical removal of asexual spores (Croft et al., 2016), secretion of antimicrobial peptides (Wiesner et al., 2017), and recruitment of specialized immune cells are employed by the human host to prevent spore germination (Bertuzzi et al., 2018). To cause infection, *A. fumigatus* must overcome these challenges and adapt to the host environment. The dynamics and intricacies of the interaction between *A. fumigatus* and host responses have yet to be fully elucidated. Decades of work have identified at least 206 genetic determinants of *A. fumigatus* virulence, that is genes whose deletion is known to modulate the virulence of *A. fumigatus* (for a detailed list, see: Steenwyk et al., 2021). These genetic determinants of virulence are involved in a wide range of activities including gene regulation, RNA processing, protein modification, production of secondary metabolites, amino acid biosynthesis, cell cycle regulation, morphological regulation, and others (Steenwyk et al., 2021).

The phylogeny of the genus *Aspergillus* reveals that pathogenic species are often more closely related to nonpathogenic species than to other pathogenic ones (Houbraken et al., 2014; de Vries et al., 2017; Steenwyk et al., 2019; Rokas et al., 2020; Mead & Steenwyk et al., 2021). For example, *A. fischeri* is a close relative of *A. fumigatus* (the two share >90% average nucleotide sequence similarity and >95% average amino acid sequence similarity between orthologs), yet *A. fischeri* is less virulent and is not considered clinically relevant (Mead et al., 2019; Steenwyk et al., 2020). Given the large disparity of IA cases caused by *A. fumigatus* and closely related species, early studies looked to species-specific genes in *A. fumigatus* as a potential contributor (Fedorova et al., 2008). However, a recent examination found that 206 known genetic determinants of virulence in *A. fumigatus* are shared between *A. fumigatus* and at least one other closely related species (Mead & Steenwyk et al., 2021).

Variation in non-coding regions can also contribute to phenotypic diversity (Carroll, 2008; Li & Johnson, 2012) and disease (Caron et al., 2019; Ropero et al., 2017; Jang et al., 2018). In fungi, non-coding regions found immediately upstream of genes' protein-coding regions are bound by transcription factors (TFs), impact transcriptional activity (Kim et al., 2019), and play roles in vital biological processes such as zinc homeostasis (Eide, 2020) and thermotolerance (Yamamoto et al., 2020). Differences in gene expression have become an important focus in understanding *A. fumigatus* virulence (de Castro et al., 2014, Chung et al., 2014; Furukawa et al., 2020; Ries et al., 2020; Colabardini et al., 2021; Takahashi et al., 2021). However, the role non-coding regions play in differential gene regulation between *A. fumigatus* and close relatives remains largely unknown.

Here, we perform genome-wide comparisons of intergenic, non-coding regions upstream of the first codon of single-copy orthologous genes of the reference strains *A. fumigatus* Af293 and A1163 against those of eight closely related species. We identified 5,215 single-copy orthologous genes across the 10 taxa of interest. Of the 5,215 genes, the non-coding regions of 4,483 genes either lacked homology across the ten taxa or showed extensive sequence variation, such that multiple sequence alignment was not possible. For the remaining 732 genes, each non-coding sequence was ≥500 bp long in all ten taxa and the sequence similarity of the sequence alignment between the *A. fumigatus* Af293 sequence and those of all other nine strains / species was ≥75%, enabling us to construct accurate multiple sequence alignments. Examination of the evolutionary rates of the non-coding and protein-coding regions of these 732 genes identified 418 upstream non-coding and 100 protein-coding regions whose evolutionary rate was different in *A. fumigatus* compared to close relatives. These 418 non-coding regions include 25 known genetic determinants of *A. fumigatus* virulence, such as *pkaR* (a regulatory subunit essential for protein kinase A pathway), *gliG* (glutathione S-transferase required for gliotoxin production), and *metR* (transcription factor required for sulfur assimilation).

## Methods

### Genomic data collection

All *Aspergillus* genomes are publicly available and were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). These strains include *A. fumigatus* Af293 (Nierman et al., 2005), *A. fumigatus* A1163 (Fedorova et al., 2008), *A. oerlinghausenensis* CBS139183 (Steenwyk et al., 2020), *A. fischeri* NRRL1881 (Fedorova et al., 2008), *A. lentulus* IFM54703 (Kusuya et al., 2016), *A. novofumigatus* IBT 16806 (GenBank accession: MSZS00000000.1) *A. fumigatiaffinis* 5878 (Santos et al., 2020), *A. udagawae* IFM 46973 (Kusuya et al., 2016), *A. turcosus* HMR AF 1038 (Parent-Michaud et al., 2019), and *A. thermomutatus* HMR AF 39 (Parent-Michaud et al., 2019).

### Identification of single-copy orthologous genes

To infer single-copy orthologous genes among all protein-coding sequences for all ten taxa, we used OrthoFinder, version 2.4.0 (Emms & Kelly, 2015). OrthoFinder clustered genes into orthogroups from gene-gene sequence similarity information obtained using the program DIAMOND version 2.0.9 (Buchfink et al., 2015) with the proteomes of the ten *Aspergillus* species as input. The key parameters used in DIAMOND were e-value = 1 x $10^{-3}$ with a percent identity cutoff of 30% and percent match cutoff of 70%. This approach identified 5,215 single copy orthologous genes.

### Identification of highly conserved non-coding regions

To identify highly conserved non-coding regions, we first retrieved intergenic sequences directly upstream of the first codon of all 5,215 single-copy orthologous genes for each of the ten *Aspergillus* species/strains using a custom script (https://github.com/alecbrown24/General_Bio_Scripts; this script was based on a previously available script: https://github.com/shenwei356/bio_scripts). We retrieved the first 500 bp of intergenic sequence directly upstream of each gene's first codon and used these sequences to generate FASTA files of non-coding regions, as well as FASTA files of single-copy orthologous protein-coding sequences using Python version 3.8.2 (https://www.python.org/). For some of the non-coding regions, there were <500 bp of non-coding sequence between the first codon of the gene of interest and an upstream gene; in these instances, only the intergenic region was used for subsequent analyses.

All multiple sequence alignments were constructed using MAFFT, version 7.453, with default parameter settings (Rozewicki et al., 2019). Analyses were conducted using custom Python scripts that used BioPython, version 1.78 (Cock et al., 2009), and NumPy, version 1.20.3 (Harris et al., 2020), modules. Sequence similarity in protein-coding and non-coding regions was calculated from their corresponding multiple sequence alignment files. The percent sequence similarity for each position in the alignment was calculated by determining if the nucleotide / amino acid at each position was the same as the nucleotide / amino acid for *A. fumigatus* Af293 and then dividing by 10. The percent similarity for each nucleotide / amino acid in each of the 5,215 non-coding and protein-coding regions of genes was averaged and reported. We discovered that the non-coding regions of only 732 of the 5,215 genes contained ≥500 bp of sequence directly upstream of the first codon in all ten taxa and exhibited sequence similarity ≥75%

between the *A. fumigatus* Af293 sequence and each of the other other nine strains / species, enabling us to construct accurate multiple sequence alignments. Thus, we focused our analyses on these 732 genes.

**Phylogenetic tree inference and comparisons**

To construct a phylogenomic data matrix, codon-based alignments for all 5,215 single-copy protein-coding orthologs were individually trimmed using ClipKIT, version 1.1.5 (Steenwyk et al., 2020), with the 'gappy' mode and the gaps parameter set to 0.7. The resulting trimmed codon-based alignments were then concatenated into a single matrix with 9,248,205 sites using the 'create_concat' function from PhyKIT, version 1.2.1 (Steenwyk et al., 2021). Next, the evolutionary history of the ten *Aspergillus* genomes was inferred using IQ-TREE, version 2.0.6 (Minh et al., 2020), and the "GTR+F+I+G4" model of sequence evolution, which was the best fitting one according to the Bayesian Information Criterion (Waddell and Steel, 1997; Vinet and Zhedanov, 2011). Bipartition support was assessed using ultrafast bootstrap approximations (Hoang et al., 2018). All bipartitions received full support. The inferred topology is congruent with known relationships inferred from analyses of single or a few loci as well as from genome-scale analyses (Hubka et al., 2018; Steenwyk et al., 2019; dos Santos et al., 2020).

To identify gene trees whose phylogeny was statistically different from the species phylogeny, we used the approximately unbiased test (Shimodaira, 2002). Protein-coding region and non-coding region trees were inferred using IQ-TREE, version 2.0.6 (Minh et al., 2020), with "GTR+I+G+F" as it was the best fitting substitution model (Waddell and Steel, 1997; Vinet and Zhedanov, 2011). The distributions of branch lengths for protein-coding region and non-coding region trees were determined using the "total_tree_length" function from PhyKIT version 1.2.1 (Steenwyk et al., 2021).

**Analysis of molecular evolutionary rates of protein-coding and non-coding regions between the major pathogen *A. fumigatus* and its relatives**

To determine the rate of sequence evolution in protein-coding region alignments between *A. fumigatus* and close relatives, we examined variation in the ratio of the rate of nonsynonymous (dN) to the rate of synonymous (dS) substitutions (dN/dS or $\omega$) across the phylogeny. We first obtained codon-based alignments from their corresponding protein sequence alignments using pal2nal, version 14 (Suyama et al., 2006). We next used the codon-based alignments to calculate $\omega$ values under two different hypotheses using the codeml module in paml, version 4.9 (Yang, 2007). For each gene tested, the null hypothesis ($H_0$) was that all branches of the phylogeny exhibit the same estimated $\omega$ value. We compared $H_0$ to an alternative hypothesis ($H_A$) which allows for the branch leading to *A. fumigatus* to have a distinct estimated $\omega$ value from the rest of the branches. To determine whether $H_A$ was significantly different from $H_0$ for each of the codon-based alignments, we used the likelihood ratio test with a statistical significance threshold of $\alpha = 0.01$.

To determine the rate of sequence evolution in non-coding region alignments between *A. fumigatus* and close relatives, we examined variation in the ratio of the rate of substitutions in each non-coding region (dNC) to the rate of synonymous (dS) substitutions in its corresponding protein-coding region (dNC/dS or $\zeta$) across the phylogeny. Like the analysis of the protein-coding regions, the null hypothesis ($H_0$) was that all branches of the phylogeny exhibit the same

estimated ζ value. We compared $H_0$ to an alternative hypothesis ($H_A$) which allows for the branch leading to *A. fumigatus* to have a distinct estimated ζ value from the rest of the branches. ζ values were calculated under the different hypotheses using HyPhy version 2.2.2 (Pond et al., 2004) with the "nonCodingSelection.bf" batch file as established by Oliver Fedrigo (Haygood et al., 2007; Fedrigo et al., 2011). To determine whether $H_A$ was significantly different from $H_0$ for each of the non-coding region alignments, we used the likelihood ratio test with a statistical significance threshold of $\alpha = 0.01$.

**Functional enrichment analyses of genes with signatures of different evolutionary rates**
To determine whether genes with signatures of different evolutionary rates in either their protein-coding or non-coding regions are enriched for particular functional categories, we implemented the Gene Ontology (GO) Term Finder webtool on AspGD (Cerqueira et al., 2013) using default settings. We conducted two separate analyses. The first examined those *A. fumigatus* genes that exhibited a different evolutionary rate in their non-coding regions, whereas the second examined those *A. fumigatus* genes with a different evolutionary rate in their protein-coding regions. These gene sets were compared to a general background set that includes all the features / gene names in the database with at least one GO annotation for *A. fumigatus*. Both analyses used a p-value cutoff of 0.05.

**Examination and visualization of mutational signatures**
To identify interesting examples of sequence variation between *A. fumigatus* and the other species for non-coding regions of genes of interest, we visualized and compared multiple sequence alignments using the MView function in EMBL-EBI (Madeira et al., 2019). Workflow of methods can be seen in **Figure 6**.

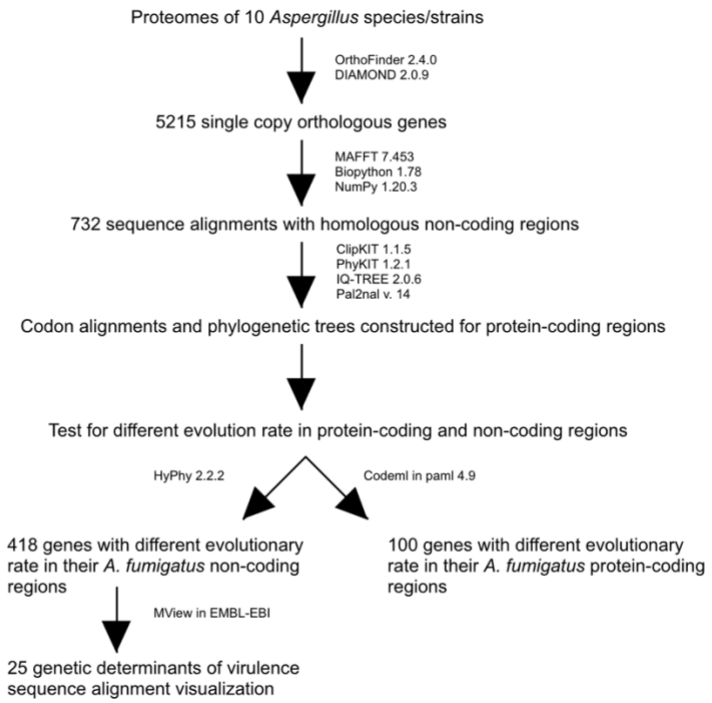



**Figure 6. Workflow of methods for Chapter 2.**

## Results

***Aspergillus* species exhibit extensive sequence variation in their non-coding regions.**
To analyze the sequence diversity of non-coding regions in section *Fumigati*, we first identified 5,215 single-copy orthologous genes amongst ten strains / species in the section (**Figure 1A**). We then computed the percent similarity between the non-coding and protein-coding regions of each *A. fumigatus* Af293 ortholog and their homologous non-coding and protein-coding regions in the other nine taxa. Those individual percent similarities were then averaged to get the final percent similarity for the non-coding and protein-coding regions of that ortholog. Averaging the non-coding region percent similarities for the 5,215 single-copy orthologous genes revealed an average similarity of ~72%; 648 alignments exhibited <50% similarity, 3,665 exhibited sequence similarity between ≥50% and <75%, and 902 exhibited ≥75% similarity. Interestingly, three genes exhibited > 90% similarity. These genes were *cnaB*, whose protein product is a calcineurin regulatory subunit and whose transcript is induced by exposure to human airway epithelial cells (Juvvadi et al., 2011; Oosthuizen et al., 2011), *AFUA_6G07800*, which is predicted to be a transcription factor with unknown function (Cerqueira et al., 2013), and *AFUA_6G04530*, which is predicted to have a role in histone acetylation (Cerqueira et al., 2013).

Average percent similarity by position in *Aspergillus* non-coding region alignments (**Figure 1B**) revealed that the percent similarity directly upstream of the first codon (-1 bp upstream) is higher than 60% and decreases as the distance from the first codon increases, approaching 40% similarity. This result suggests that potentially conserved promoter and *cis*-regulatory elements occur in these non-coding regions and is consistent with transcription factor binding location in *A. fumigatus* (de Castro et al., 2014; Chung et al., 2014). For comparison, we also calculated the average percent similarity by position in the protein-coding region alignments of all 5,215 genes (**Figure 1C**). We found that the percent similarity of the first amino acid (+1) was ~100%, indicative of the first methionine; similarity was high throughout the first 167 sites of amino acid alignment but decreased as the distance from the first amino acid increased, approaching 60% amino acid sequence similarity.

Examination of the 4,567 genes whose non-coding region alignments exhibited ≥50% similarity (i.e., the 3,665 genes whose similarity was ≥50% and <75%, and the 902 that had ≥75% similarity) revealed several instances in which one or more sequences were poorly aligned for stretches of 100 bp or more. This was especially true when sequences in these alignments exhibited large variation in their lengths. Further, we found a low level of synteny as genes immediately upstream of a non-coding region of interest generally differed between species.

We also determined the number of conserved non-coding regions with ≥75% sequence similarity and that were ≥500 bp in length between *A. fumigatus* Af293 and each of the other nine taxa, separately (**Figure 1**). We found that *A. novofumigatus* shares the fewest number of conserved non-coding regions (2,321) despite being more closely related to *A. fumigatus* than other species included in our phylogeny (Houbraken et al., 2014; Steenwyk et al., 2019; Rokas et al., 2020); this suggests that the quality of annotations may differ across the ten genomes examined and that improvements in the gene annotation of these genomes could increase the number of conserved non-coding regions shared by these taxa. *A. fumigatus* A1163 shared the greatest number of

conserved non-coding regions with *A. fumigatus* Af293 (5,020). With the exceptions of *A. novofumigatus*, *A. oerlinghausenensis*, and *A. thermomutatus*, the closer a relative is to *A. fumigatus* Af293, the greater the number of conserved non-coding regions that are shared.

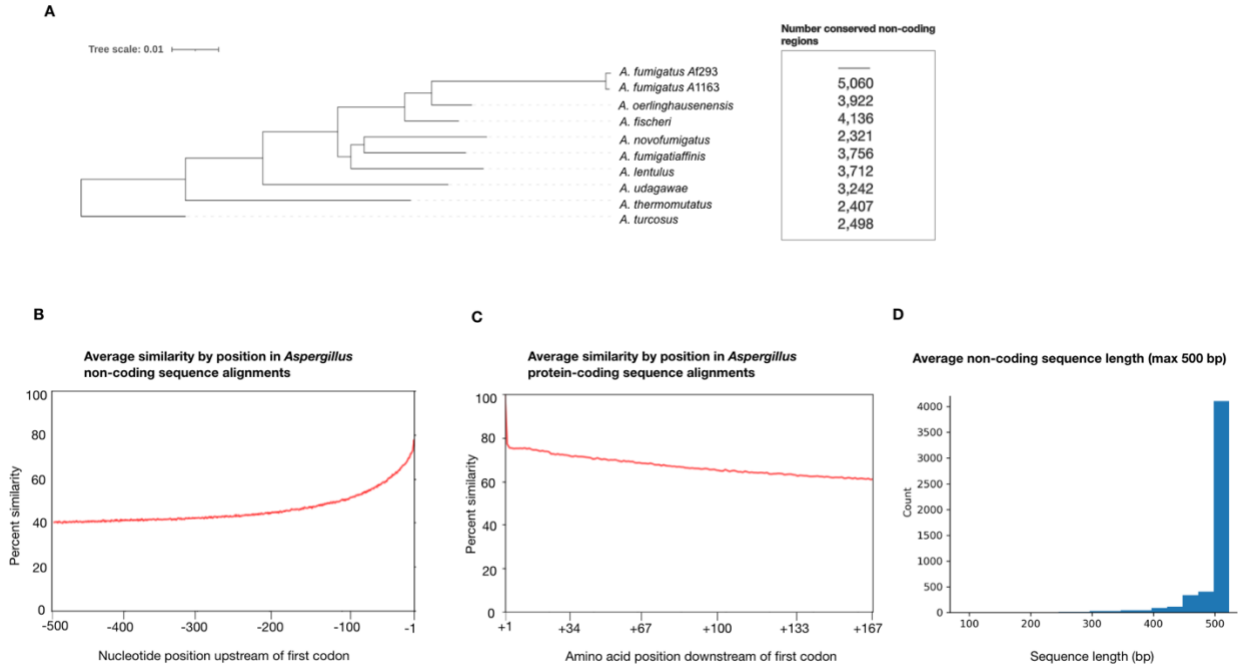



**Figure 7. *Aspergillus* section *Fumigati* species exhibit sequence variation in non-coding regions that are 500 base pairs upstream of genes' first codon.**
A. Species phylogeny of two *A. fumigatus* reference strains (Af293 and A1163) and closely related *Aspergillus* section *Fumigati* species constructed from concatenation analysis of a 5,215-gene data matrix. Branch lengths correspond to nucleotide substitutions / site. Note the long branch leading to *A. fumigatus*, indicative of a greater number of nucleotide substitutions per site in this species. The number of genes whose non-coding regions were conserved (≥ 75% sequence similarity between each species and ≥ 500 bp in length) between *A. fumigatus* Af293 and each species are shown next to the corresponding taxa. B. Average percent sequence similarity of non-coding regions of 5,215 genes by position, relative to the gene's first codon. Sequence alignments of non-coding regions were compared by position and the average percent similarities for each site are reported with -1 indicating the site directly upstream of the first codon and -500 indicating the site 500 bp upstream of the first codon. C. Average percent sequence similarity by position of the first 167 amino acid sites in the alignments of 5,215 genes, relative to the gene's first codon. The average percent similarity for each site is reported, with +1 indicating the first amino acid. D. Average sequence alignment lengths of the 5,215 non-coding regions examined in this study. 4,079 of the 5,215 non-coding regions have ≥ 500 bp in all 10 strains/species used in this study.

**Phylogenetic analyses reveal differences between non-coding region trees and protein-coding region trees.**

To help determine if differences existed between non-coding and protein-coding regions across our species, we first compared total branch lengths in phylogenetic trees constructed from both non-coding and protein-coding regions from all 5,215 genes. Comparisons of the overall distributions between these two groups revealed a statistically significant difference between the overall branch lengths of protein-coding and non-coding regions (Wilcoxon signed-ranked test; p-value = 0.004), suggesting that non-coding regions of single-copy orthologs evolve faster than protein-coding regions.

**Many non-coding but fewer protein-coding regions exhibit different rates of evolution in *A. fumigatus.***

Given the uncertainty regarding the homology of some sequences in these 5,215 non-coding region alignments and our finding that most sequence conservation was found near the first codon position, we focused our evolutionary rate analyses only on the 732 non-coding region alignments whose sequences were all ≥500 bp long and exhibited ≥75% sequence similarity between *A. fumigatus* Af293 and each other strain / species in the phylogeny. Briefly, to determine the rate of sequence evolution in protein-coding and non-coding region alignments between *A. fumigatus* and close relatives, we examined variation in the ratio of the rate of nonsynonymous (dN) to the rate of synonymous (dS) substitutions (ω value) for protein-coding regions and the variation in the ratio of the rate of non-coding (dNC) to the rate of synonymous (dS) substitutions (dNC/dS or ζ value) for non-coding regions across the phylogeny. To test whether the molecular evolutionary rates of protein-coding and non-coding regions differed between the major pathogen *A. fumigatus* and its relatives, we statistically examined whether protein-coding and non-coding *A. fumigatus* sequences evolved at a similar ($H_0$) or different ($H_A$) rate as those of other taxa (**Figure 8**).

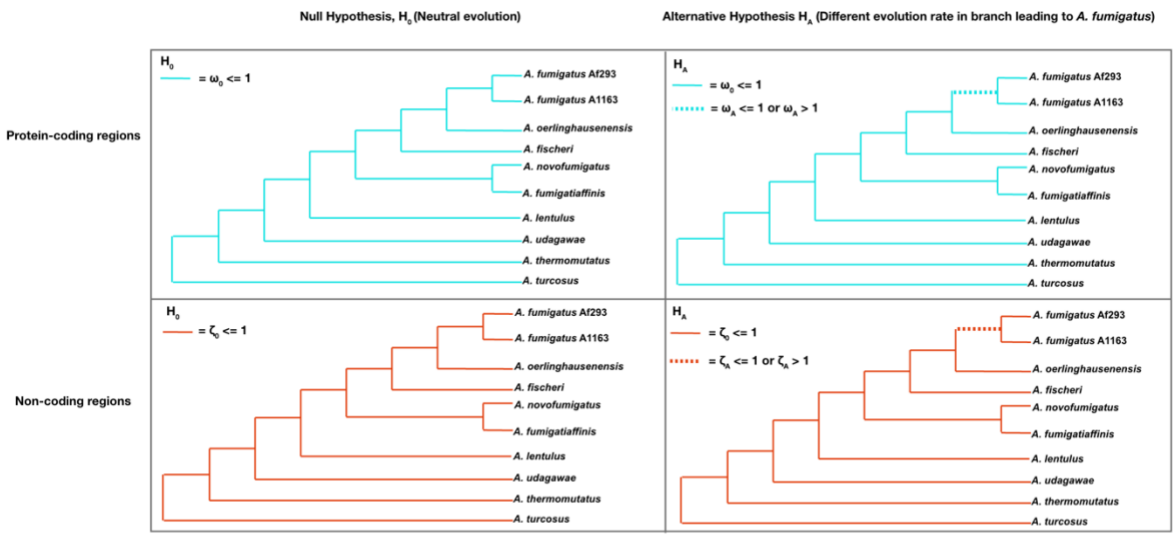



**Figure 8. Examining whether the non-coding and protein-coding regions of 732 genes have different evolutionary rates in the major pathogen *A. fumigatus*.**
The top two panels present the null and alternative hypotheses for evolutionary rate difference in the protein-coding regions of *A. fumigatus* genes relative to the other species. The null hypothesis ($H_0$, upper left) constrains all ω (dN/dS) values across all branches to be less than or equal to 1, the neutral evolutionary rate. The alternative hypothesis ($H_A$, upper right) allows the branch leading to *A. fumigatus* (dashed branch) to have an ω value lower than, equal to, or greater than 1 (indicative of evolutionary rate difference) compared to the background branches. The bottom two panels present the null and alternative hypotheses for evolutionary rate difference in the non-coding regions of *A. fumigatus* genes relative to other species. Similarly, the null hypothesis ($H_0$, bottom left) constrains all ζ (dNC/dS) values across all branches to less than or equal to 1. The alternative hypothesis ($H_A$, bottom right) allows for the branch leading to *A. fumigatus* (dashed branch) to have a ζ value lower than, equal to, or greater than 1 (evolutionary rate difference) compared to the background branches. For each protein-coding and non-coding region, a likelihood ratio test was used to determine which hypothesis best fits the data.

Examination of protein-coding regions identified 100 / 732 genes (13.7% of examined genes) that significantly rejected $H_0$ (under which all branches exhibited the same ω value) over $H_A$ (which postulates that the ω value of the branch leading to *A. fumigatus* was distinct from the background ω value of all other branches) (**Figure 9**). Examination of non-coding regions identified 418 / 732 genes (57.1% of examined genes) that significantly rejected $H_0$ (under which all branches exhibited the same ζ value) over $H_A$ (which postulates that the ζ value of the branch leading to *A. fumigatus* was distinct from the background ζ value of all other branches) (**Figure 3C**). Taken together, these results suggest a much higher amount of variation in non-coding regions than in protein-coding regions between *A. fumigatus* and relatives. The p-value distribution of protein-coding regions is uniform, while the p-value distribution of non-coding

regions is bimodal with nearly all p-values being either under 0.05 or 1.0. This result suggests that the 418 non-coding regions exhibited major differences in their relative fit for the two hypotheses, whereas protein-coding regions exhibited much smaller differences.

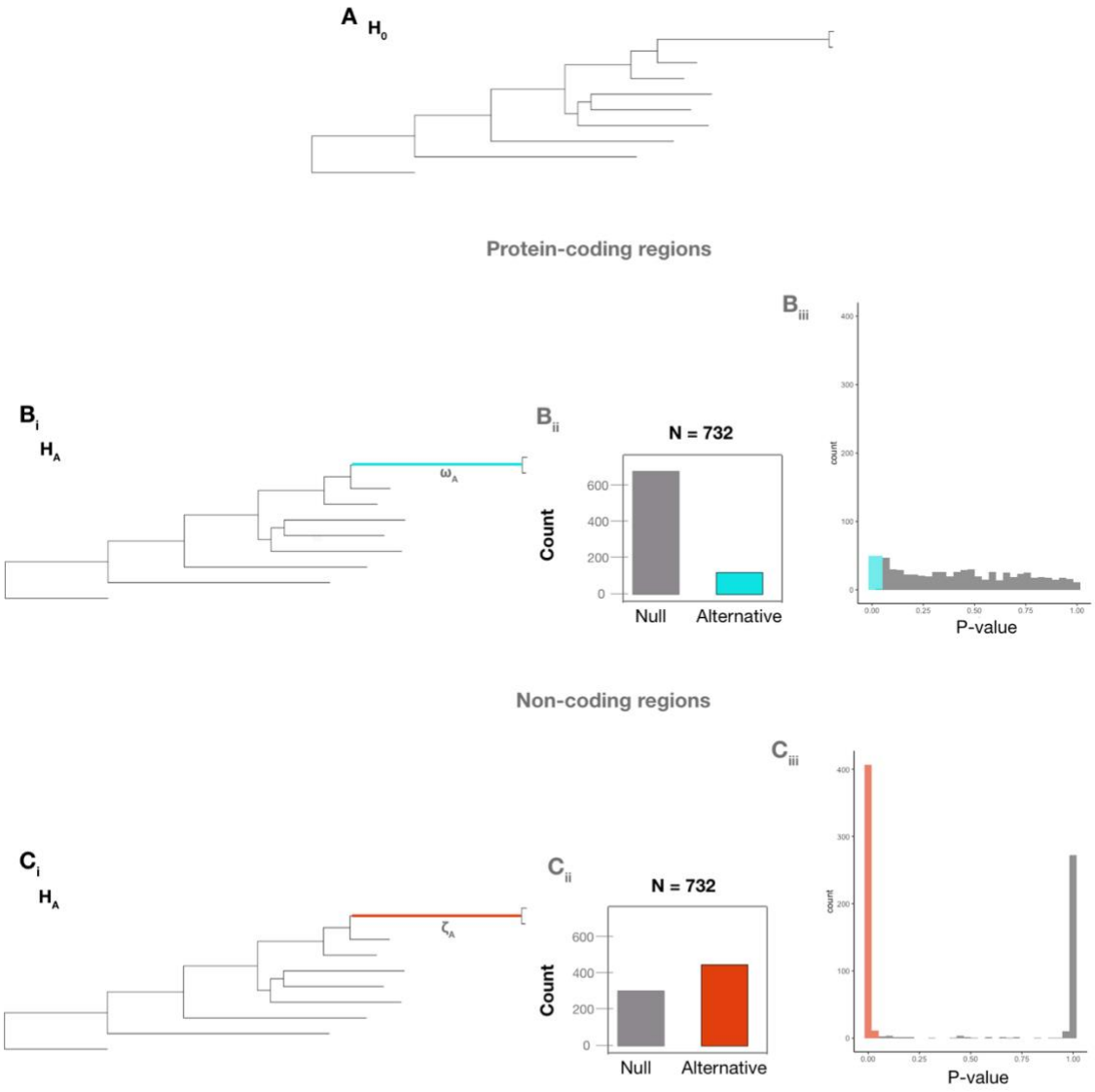



**Figure 9. Non-coding regions of *A. fumigatus* genes exhibit many more signatures of evolutionary rate difference than their corresponding protein-coding regions.**
A. The null hypothesis ($H_0$) that all branches have the same evolutionary rate. B-C. The alternative hypotheses assume that the ω value (Bi) or the ζ value (Ci) in the branch leading to *A. fumigatus* differs from the value in the rest of the branches of the phylogeny. Bii. 632 of 732 protein-coding regions (84.34%) did not reject $H_0$ (gray) and 100 of 732 (16.66%) rejected $H_0$ (blue). Biii. The distribution of p-values for protein-coding regions that did not (gray) and did (blue) reject $H_0$. Cii. 314 of 732 non-coding regions (42.90%) did not reject $H_0$ (gray) and 418 of 732 non-coding regions (57.10%) rejected $H_0$ (red), which suggests a greater amount of variation

in non-coding regions than in protein-coding regions between *A. fumigatus* and relatives. Ciii The distribution of p-values for non-coding regions that did not (gray) and did (red) reject $H_0$.

**Genes with signatures of a different evolutionary rate in non-coding regions are enriched for regulatory functions in *A. fumigatus.***

To identify functions that were over-represented in the list of genes that rejected $H_0$ for either their coding or non-coding regions, we conducted gene ontology (GO) enrichment analyses. Examination of significantly over-represented GO terms for the 418 genes with signatures of different evolutionary rates in non-coding regions revealed numerous biological processes related to regulation, metabolism, and development (e.g., "cellular component organization or biogenesis", p = 0.00016; "regulation of protein metabolic process", p = 0.00101; "hyphal growth", p = 0.00352; "regulation of cell cycle"; p = 0.0076; "developmental process", p = 0.01233; "reproduction". p = 0.00361). Of the 418 genes, 71 lacked any functional GO annotation. In comparison, for the 314 genes that did not exhibit a different evolutionary rate in their non-coding regions, the only term that was enriched was "nucleotide binding" (p = 0.06129) and was found associated with 48 genes. For the 100 genes with signatures of different evolutionary rate in their protein-coding regions, only one function was found enriched ("regulation of cellular process", p = 0.02647). Of note, 74 of the protein-coding genes lacked any functional GO annotation.

**Four genes whose non-coding regions exhibit different evolutionary rates in *A. fumigatus* also bind transcription factors that are known genetic determinants of virulence.**

To identify if any of the *A. fumigatus* genes with different evolutionary rates in their respective non-coding regions also contain known TF binding sites, we compared the list of 418 non-coding regions to binding sites of two TFs known to be genetic determinants of virulence, CrzA and SrbA (Cramer et al., 2008; Willger et al., 2008). ChIP-seq analysis of CrzA (de Castro et al., 2014) uncovered 110 genes that are directly bound by the TF in *A. fumigatus* strain Af293. Of these, 28 were reported to exhibit CrzA binding within 500 bp of the first codon, and two genes, *AFUA_8G05090* (a putative MFS transporter) and *AFUA_3G09960* (Aureobasidin resistance protein), exhibited a different evolutionary rate in the non-coding regions of *A. fumigatus* strains in our analysis. ChIP-seq analysis of SrbA (Chung et al., 2014) revealed 112 genes directly bound by the TF in *A. fumigatus* strain A1163. Of these, 57 were reported to exhibit SrbA binding within 500 bp of the first codon, and two genes, *AFUB_074100* (a gene of unknown function which appears to interact with *sldA*, a checkpoint protein kinase) and *AFUB_012300* (a gene predicted to be involved in nitrate assimilation), exhibited a different evolutionary rate in the non-coding regions of *A. fumigatus* strains in our analysis. Importantly, we found that for all four genes (*AFUA_8G05090*, *AFUA_3G09960*, *AFUB_074100*, and *AFUB_012300*) there was at least a 2 bp sequence difference between *A. fumigatus* and at least one close relative in their sequences at the TF binding site location (**Figure 10**). Together, our results suggest that intergenic non-coding regions that bind known TFs can exhibit substantial differences in their evolutionary rates between *A. fumigatus* and close relatives, which raises the hypothesis that these differences may lead to differences in gene expression.

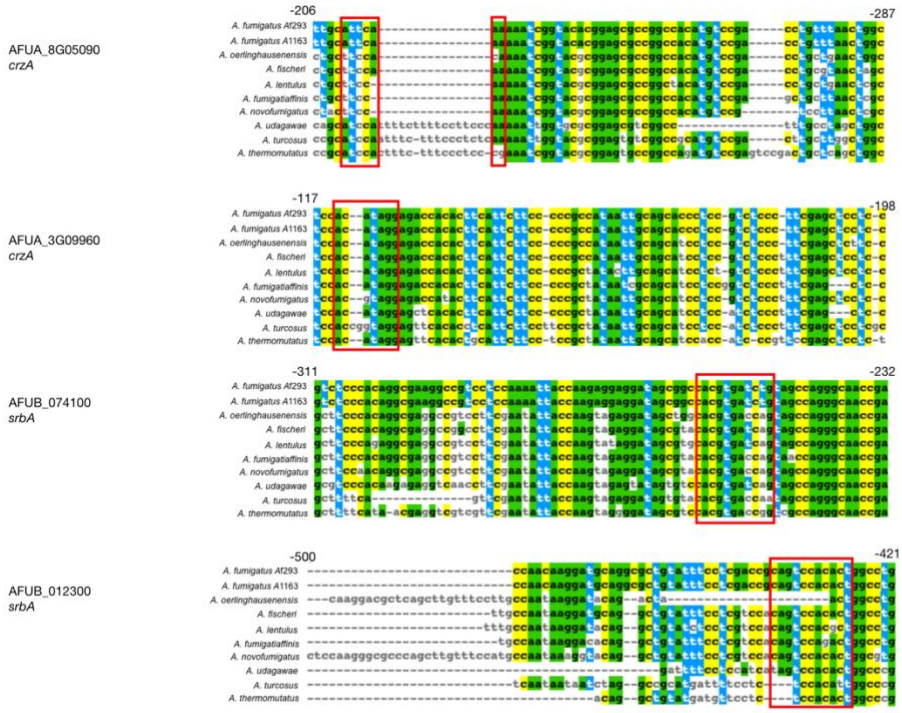



**Figure 10. CrzA and SrbA binding locations in 4 genes that exhibit a different evolutionary rate in their non-coding regions.**
Two *A. fumigatus* genes (*AFUA_8G05090* and *AFUA_3G09960*) known to bind CrzA in their non-coding regions and two *A. fumigatus* genes (*AFUB_074100* and *AFUB_012300*) known to bind SrbA in their non-coding regions have at least a 2 bp difference in the TF binding site between *A. fumigatus* and one or more relatives. Red boxes represent the binding site locations found in previous ChIP-seq experiments.

**Non-coding regions upstream of genetic determinants of virulence with different rates of evolution in *A. fumigatus*.**
We identified 25 genetic determinants of virulence whose non-coding regions exhibited a different rate of evolution in *A. fumigatus* (**Table 1**). Three genes (*metR, his3,* and *met16*) are involved in amino acid biosynthesis, eight genes (*chsF, calA, gel2, nrps1, gfa1, csmB, rlmA,* and *rodA*) are involved in cell wall biosynthesis, nine genes (*noc3, spe2, gus1, pri1, AFUA_2G10600*, *mak5, pkaR, ramA,* and *somA*) are involved in cellular metabolism, two genes (*aspB* and *tom40*) are involved in hyphal growth, and three genes (*gliG, gliI,* and *gliJ*) are involved in gliotoxin biosynthesis. Of the 25, 14 genes (*metR, chsF, calA, gel2, nrps1, gfa1, csmB, rlmA, rodA, AFUA_2G10600*, *pkaR, ramA, somA,* and *aspB*) have been shown to modulate virulence in an animal model of infectious disease**.** Three genes (*gliG, gliI,* and *gliJ*) are required for the biosynthesis of gliotoxin, a secondary metabolite involved in *A. fumigatus* virulence (Brakhage & Langfelder, 2002; Dagenais & Keller, 2009), and deletions of the eight remaining genes (*his3, met16, noc3, spe2, gus1, pri1, mak5,* and *tom40*) have been previously shown to be important for viability and therefore, likely virulence (for a detailed list, see:

Steenwyk et al., 2021). Interestingly, only two of these 25 genes (*csmB* and *rodA*) exhibit a signature of different evolutionary rate in their protein-coding region as well, a finding consistent with our result that there are more changes in non-coding regions than in protein-coding regions of *Aspergillus* genes.

**Table 1: Twenty-five genetic determinants of *A. fumigatus* virulence have a different evolutionary rate in their non-coding regions.**

| Gene ID | Gene Name | Specific Function | General Pathway | Reference |
|---|---|---|---|---|
| *AFUA_4G06530* | *metR* | putative bZip transcription factor | amino acid biosynthesis | Amich et al., 2013 |
| *AFUA_6G04700* | *his3* | Putative imidazoleglycerol-phosphate dehydratase | amino acid biosynthesis | Hu et al., 2007 |
| *AFUA_3G06540* | *met16* | phosphoadenylyl-sulfate reductase (thioredoxin) activity | amino acid biosynthesis | Hu et al., 2007 |
| *AFUA_8G05630* | *chsF* | putative chitin synthase | cell wall biology | Muszkieta et al., 2014 |
| *AFUA_5G09360* | *calA* | calcineurin a catalytic subunit | cell wall biology | Juvvadi et al., 2013 |
| *AFUA_6G11390* | *gel2* | GPI-anchored 1,3-beta-glucanosyltransferase | cell wall biology | Mouyna et al., 2005 |
| *AFUA_1G10380* | *nrps1* | non-ribosomal peptide synthase | cell wall biology | Reeves et al., 2006 |
| *AFUA_6G06340* | *gfa1* | glutamine-fructose-6-phosphate transaminase activity | cell wall biology | Hu et al., 2007 |
| *AFUA_2G13430* | *csmB* | putative chitin synthase | cell wall biology | Muszkieta et al., 2014 |
| *AFUA_3G08520* | *rlmA* | cell wall organization, cellular response to stress | cell wall biology | Rocha et al., 2016 |
| *AFUA_5G09580* | *rodA* | Asexual spores hydrophobin | cell wall biology | Shibuya et al., 1999 |
| *AFUA_2G17050* | *noc3* | rRNA processing | metabolism | Hu et al., 2007 |
| *AFUA_5G03670* | *spe2* | role in spermidine biosynthetic process | metabolism | Hu et al., 2007 |

| | | | | |
|---|---|---|---|---|
| *AFUA_5G03560* | *gus1* | Putative glutamyl-tRNA synthetase | metabolism | Hu et al., 2007 |
| *AFUA_3G09020* | *pri1* | DNA primase small subunit | metabolism | Hu et al., 2007 |
| *AFUA_2G10600* | N/A | Complex I NADH oxidoredutase | metabolism | Bromley et al., 2016 |
| *AFUA_6G08900* | *mak5* | role in maturation of 5.8S rRNA | metabolism | Hu et al., 2007 |
| *AFUA_3G10000* | *pkaR* | cAMP-dependent protein kinase regulatory subunit | metabolism | Zhao et al., 2006 |
| *AFUA_4G10330* | *ramA* | role in protein farnesylation | metabolism | Norton et al., 2017 |
| *AFUA_7G02260* | *somA* | putative role in lipid homeostasis | metabolism | Lin et al., 2015 |
| *AFUA_7G05370* | *aspB* | Putative septin | hyphal growth | Vargas-Muniz et al., 2015 |
| *AFUA_6G05110* | *tom40* | role in conidium formation, hyphal growth | hyphal growth | Hu et al., 2007 |
| *AFUA_6G09690* | *gliG* | gliotoxin production, Glutathione S-transferase | secondary metabolism | Scharf et al., 2011 |
| *AFUA_6G09640* | *gliI* | gliotoxin production, Aminotransferase | secondary metabolism | Forseth et al., 2011 |
| *AFUA_6G09650* | *gliJ* | gliotoxin production, Dipeptidase | secondary metabolism | Scharf et al., 2011 |

Examination of sequence alignments of non-coding regions of these 25 genes (**Table 1**) revealed several interesting patterns (**Figure 11**). For example, the non-coding region of *pkaR* exhibits a 10 bp stretch from -434 bp to -424 bp upstream of the first codon, which is deleted exclusively in *A. fumigatus* and present and largely conserved in all other species (**Figure 11A**). The non-coding region of g*fa1* also has a stretch of 5 bp exclusively deleted in *A. fumigatus* and present in all other species. Sequence alignment of the *gliG* non-coding region revealed an 11 bp G-rich insertion that is unique to the two *A. fumigatus* strains (**Figure 11B**). In addition to *A. fumigatus*-specific indels, we also observed that the non-coding regions of several genes known to be involved in *A. fumigatus* virulence exhibited indel variation across the other *Aspergillus* species examined as well. For example, the non-coding region of *metR* exhibits a 7 bp pyrimidine rich insertion that is found only in *A. fumigatus*, *A. oerlinghausenensis*, and *A. lentulus* (**Figure 11C**), while the non-coding regions of *calA* and *pri1* both have small sequences (12 bp and 5 bp, respectively) that are exclusively absent in *A. fumigatus* strains Af293 andA1163, and in *A. oerlinghausenensis.*

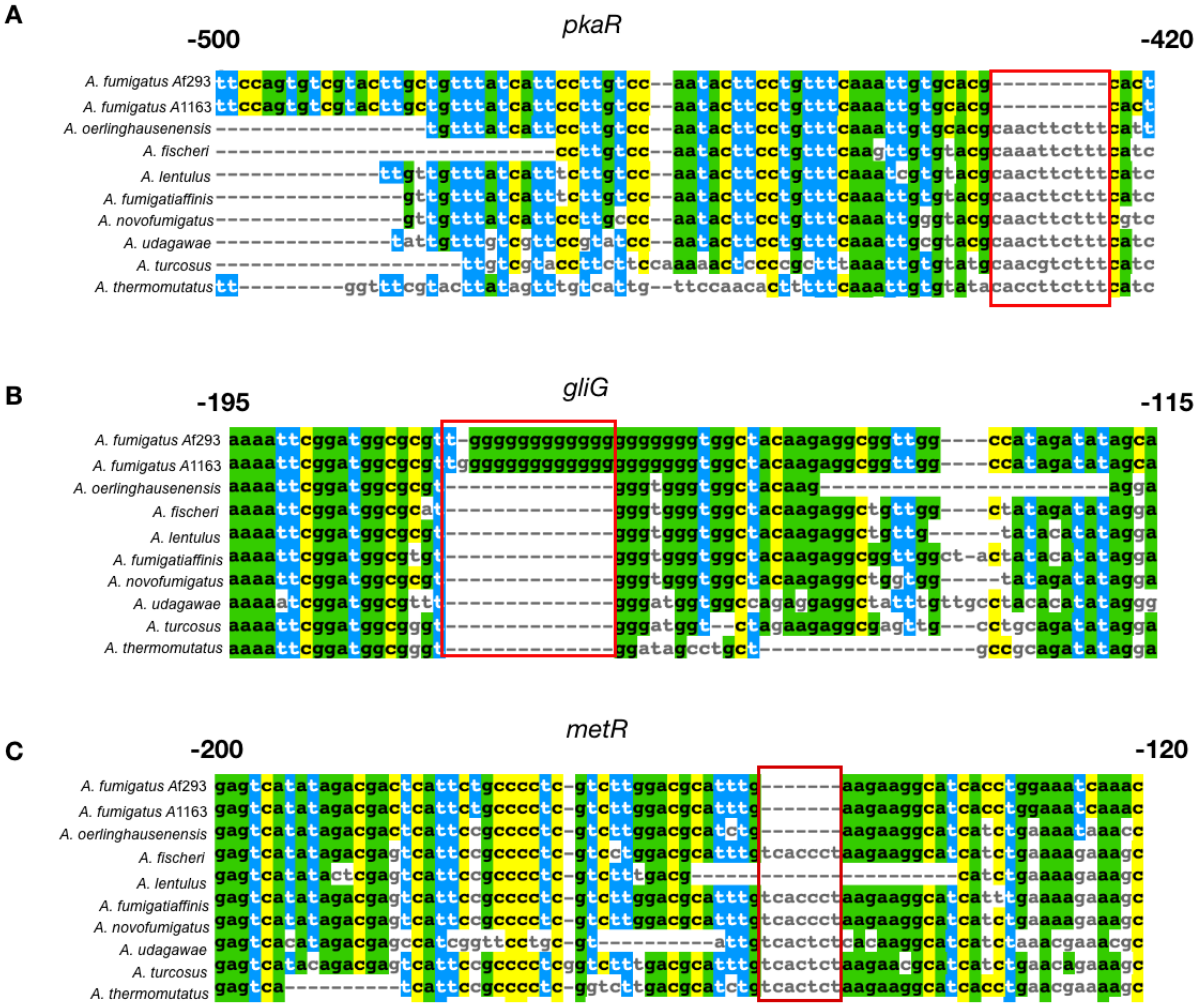



**Figure 11. Notable examples of sequence differences between *A. fumigatus* and its close relatives in non-coding regions located upstream of three known genetic determinants of *A. fumigatus* virulence that exhibited signatures of a different evolutionary rate.**
A. *pkaR* encodes a regulatory subunit involved in the regulation of the cyclic AMP – dependent protein kinase pathway; deletion of *pkaR* in *A. fumigatus* has been shown to attenuate virulence in a neutropenic mouse model (Zhao et al., 2006). The sequence alignment of the non-coding region of the *pkaR* gene exhibits a 10 bp region (red box) that is uniquely deleted in *A. fumigatus* but present in all other close relatives. B. *gliG* encodes for a glutathione S-transferase (GST) that is part of the gliotoxin biosynthetic gene cluster and is required for gliotoxin production (Scharf et al., 2011). Gliotoxin contributes to the virulence of *A. fumigatus*, inactivating host vital proteins via conjugation (Scharf et al., 2011). The sequence alignment of the non-coding region upstream of the *gliG* gene exhibits a 15 bp G-rich region that has been inserted in *A. fumigatus* and is absent from all other species. C. *metR* encodes for a TF involved in sulfur assimilation and is a known *A. fumigatus* virulence factor (Amich et al., 2013). The sequence alignment of the non-coding region of the *metR* gene contains a 7 bp region that is absent only in *A. fumigatus, A. oerlinghausenensis*, and *A. lentulus*. Colored nucleotides represent sites that are present in *A.*

*fumigatus* Af293 and shared between other species, highlighting differences between the reference strain and close relatives.

## Discussion

We identified a set of 732 genes whose non-coding regions were conserved between the genomes of reference strain *A. fumigatus* Af293, the more virulent *A. fumigatus* reference strain A1163, and eight closely related species. In these 732 genes, we also tested whether the branch leading to *A. fumigatus* exhibited a difference in the evolutionary rate in either its protein-coding or its non-coding regions compared to the other species. We found that the non-coding regions of 418 of these genes exhibit signatures of a different evolutionary rate in *A. fumigatus*. These 418 genes include 25 that are known genetic determinants of *A. fumigatus* virulence (Steenwyk et al., 2021) (**Table 1**). Given the differences in reported invasive aspergillosis cases caused by *A. fumigatus* compared to other *Aspergillus* species (Steinbach et al., 2012), genetic differences in the non-coding regions of these 418 genes, and especially of these 25 genes previously connected to virulence, may play a role in varying pathogenic potentials of *Aspergillus* section *Fumigati* species.

Gene Ontology (GO) analysis of the 418 genes which exhibit signatures of a different evolutionary rate in non-coding regions revealed an enrichment for genes involved in regulation of metabolism and development. This is consistent with previous studies of the evolution of non-coding regions in humans (Haygood et al., 2007), which experienced positive selection in non-coding regions for genes involved in metabolism regulation (particularly glucose metabolism) and regulation of development (particularly the nervous system) compared to close relatives. These results raise the possibility that non-coding regions associated with particular functions in diverse taxa are more likely to experience changes in their evolutionary rates.

We compared our list of genes with signatures of different evolutionary rates with previous ChIP-seq studies of the transcription factors CrzA and SrbA, both of which are well studied genetic determinants of virulence for *A. fumigatus* (Cramer et al., 2008; de Castro et al., 2014; Colabardini et al., 2021; Willger et al., 2008; Chung et al., 2014). We found that the non-coding regions of two genes bound by CrzA (*AFUA_8G05090* and *AFUA_3G09960*) and two genes bound by SrbA (*AFUB_074100* and *AFUB_012300*) also exhibited different evolutionary rates in *A. fumigatus*. We found several nucleotide differences in these non-coding regions that likely contributed to the observed differences in evolutionary rate. Interestingly, when we examined the sequence alignments of the non-coding regions of these genes, we found differences at the TF binding sites (TFBS) between *A. fumigatus* and relatives. We hypothesize that the different evolutionary rate we observed in these *A. fumigatus* genes are due, in part, to changes in the associated TFBS, which may influence the regulation of these genes.

Comparisons between *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* non-coding regions that were similar to the ones we report here, revealed that TFBS tend be more conserved in the proximal promoter region (within 200bp of the transcription start site) than the distal region, yet some differences in TFBS were reported between the two species in their respective proximal promoter regions (Schaefke et al., 2015). Evolutionary differences near the

transcription start site have also been reported in *Drosophila* species, with certain species (such as *Drosophila pseudoobscura*) exhibiting an increased mutation rate upstream of the transcription start site when compared to *Drosophila melanogaster* (Main et al., 2013). Combined with the results presented here, it is likely that the evolution of non-coding regions is not uniform across closely related species and that these differences may play a functional role in downstream gene expression.

We compared our list of 418 genes with signatures of a different evolutionary rate in the non-coding regions of *A. fumigatus* to a previously curated set of 206 genetic determinants of *A. fumigatus* virulence (Steenwyk et al., 2021) and found that 25 of the 206 exhibited a different evolutionary rate their non-coding regions between *A. fumigatus* and close relatives. We found that the most represented general function amongst these 25 genes was "metabolism", which raises the question of their impact on virulence, given the role that metabolism has been shown to play in *A. fumigatus* virulence (Willger et al., 2009). In particular, *pkaR* is essential for proper protein kinase A signaling (Griffioen & Thevelein, 2002) and plays a key role in the germination and growth of *A. fumigatus* asexual spores (Zhao et al., 2006). Moreover, *pkaR* has been shown to be required for *A. fumigatus* virulence (Lin et al., 2015) in an immunocompromised murine model of invasive aspergillosis (Fuller et al., 2011). We found that the protein-coding region of *pkaR* does not exhibit a different evolutionary rate in *A. fumigatus*, suggesting that it is conserved. However, analysis of the sequence alignment of the non-coding region revealed a 11 bp region (CAACTTCTTT) absent in *A. fumigatus* but present in all other species (**Figure 11**). Interestingly, this binding site is similar to the predicted TFBS for the *S. cerevisiae* TF Ste12 (Badis et al., 2008), a homolog to SteA in *A. fumigatus*. While it has yet to be elucidated if this region is involved in SteA binding, it may be that its absence changed the regulation of *pkaR* and thus somehow contributed to the evolution of *A. fumigatus* virulence.

The role of gliotoxin in *A. fumigatus*-mediated disease has been of increasing interest, due to its ability to inhibit the host immune response (Raffa & Keller, 2019). However, the gliotoxin biosynthetic gene cluster is found in both *A. fumigatus* and its non-pathogenic close relatives *A. oerlinghausenensis* and *A. fischeri*, and all three species are known to produce gliotoxin (Knowles et al., 2020; Steenwyk et al., 2020). Here, we identify that three genes in the gliotoxin biosynthetic gene cluster (*gliG, gliJ, gliI*) exhibit a different evolutionary rate in their non-coding regions in *A. fumigatus*. Gliotoxin genes have been shown to require certain TFs (GliZ and RglT for example) for gliotoxin biosynthesis and/or self-protection (Schrettl et al., 2010; Ries et al., 2020; de Castro et al., 2021). Interestingly, analysis of the sequence alignment of the non-coding region of *gliG* revealed a G-rich region unique to *A. fumigatus* (**Figure 11**). G-rich regions have been previously reported to be found in biologically active sites and to play important roles in regulating cellular processes such as gene expression (Maizels & Gray, 2013; Maity et al., 2020). This *A. fumigatus*-specific G-rich region may contribute to some unknown gliotoxin expression pattern that contributes to *A. fumigatus* virulence or the lack of disease caused by other closely related *Aspergillus* species.

*metR* encodes a bZIP DNA binding protein required for sulfur metabolism in *A. fumigatus* and whose gene expression is regulated by LaeA, a major regulator of secondary metabolism (Jain et al., 2018. Pertaining to *A. fumigatus* virulence, sulfur assimilation plays key roles in oxidative stress response and gliotoxin biosynthesis (Traynor et al., 2019). Recent efforts have identified

differences in the transcriptional profiles of *A. fumigatus* and relatives in response to exogenous gliotoxin, highlighting the pathways relating sulfur assimilation and gliotoxin production (de Castro et al., 2021). The non-coding region of *metR* contains a 7 bp region (TCACCT) in *A. fischeri* and five other species; in contrast, the two strains of *A. fumigatus*, *A. oerlinghausenensis* and *A. lentulu*s all lack this 7 bp motif (**Figure 5**). While it remains unclear if this 7 bp motif has a functional role in the expression of *metR*, this result nicely illustrates the complex patterns of sequence evolution of non-coding regions in this clade of pathogens and non-pathogens.

A major outstanding question emanating from our work is whether this extensive non-coding sequence variation of closely related *Aspergillus* species that vary in their pathogenicity functionally contributes to differences in gene expression between strains and species. Currently, there are no datasets available that report genome-wide differential expression data for *A. fumigatus* and close relatives; to our knowledge, the only published differential expression study of *A. fumigatus* and close relatives focused on expression differences only for genes involved in secondary metabolism (Takahashi et al., 2021). Designing and performing differential gene expression experiments in diverse *Aspergillus* species will be a future aim. Additional future work will include functionally test if the non-coding region differences we report here play a role in *A. fumigatus* expression and virulence. Further, testing if non-coding regions in a larger set of *A. fumigatus* strains exhibit differences in evolutionary rates would help to elucidate more recent evolutionary changes in *A. fumigatus* and the pathogenic differences observed in these strains as well.

## Chapter 3: Genome-wide patterns of non-coding sequence variation in the major fungal pathogen *Aspergillus fumigatus*

**(This Chapter is adapted from Brown et al., 2024)**

### Abstract

*A. fumigatus* is a deadly fungal pathogen, responsible for >400,000 infections per year and high mortality rates. *A. fumigatus* strains exhibit multiple phenotypic differences. For example, the A1163 strain is more virulent than the Af293 strain in multiple animal models of fungal disease. However, most *A. fumigatus* protein-coding genes, including those that modulate its virulence, are shared between *A. fumigatus* strains as well as with closely related non-pathogenic relatives. We hypothesized that *A. fumigatus* genes exhibit substantial genetic variation in the non-coding regions immediately upstream to the start codons of genes, which could reflect differences in gene regulation between strains. To begin testing this hypothesis, we identified 5,812 single-copy orthologs across the genomes of 263 *A. fumigatus* strains. *A. fumigatus* non-coding regions showed higher levels of sequence variation compared to their corresponding protein-coding regions. Specifically, we found that 1,274 non-coding regions exhibited <75% nucleotide sequence similarity (compared to 928 protein-coding regions) and 3,721 non-coding regions exhibited between 75% and 99% similarity (compared to 2,482 protein-coding regions) across strains. Only 817 non-coding regions exhibited ≥ 99% sequence similarity compared to 2,402 protein-coding regions. By examining 2,482 genes whose protein-coding sequence identity scores ranged between 75% and 99%, we identified 478 total genes with signatures of positive selection only in their non-coding regions and 65 total genes with signatures only in their protein-coding regions. 28 of the 478 non-coding regions and 5 of the 65 protein-coding regions under selection are associated with genes known to modulate *A. fumigatus* virulence. Non-coding region variation between *A. fumigatus* strains included single nucleotide polymorphisms and insertions or deletions of at least a few nucleotides. These results show that non-coding regions of *A. fumigatus* genes harbor greater amounts of sequence variation than protein-coding regions, raising the hypothesis that this variation may substantially contribute to *A. fumigatus* phenotypic strain heterogeneity.

## Introduction

Invasive Aspergillosis (IA) is one of the deadliest fungal diseases to humans, responsible for over 400,000 infection per year with mortality rate >50% (Bongomin et al., 2017). Most IA cases (>90%) are caused by *Aspergillus fumigatus* (Steinbach et al., 2012; Rokas et al., 2020), a saprophytic fungus commonly found in the soil (Flores et al., 2014) as well as urban environments, such as waste piles and hospitals (Wirmann et al., 2018). In its natural environment, *A. fumigatus* plays an important role in nitrogen and carbon recycling (Latge et al., 2019). *A. fumigatus* has adapted over time to survive environmental pressures, such as high temperatures, variation in pH, and low oxygen availability (Bhabhra & Askew, 2005; Park & Yu, 2016; Rees et al., 2017), and to compete with other microorganisms for resources (Latge et al., 2019). Recently, the World Health Organization included *A. fumigatus* in its first ever list of fungal "priority pathogens", a testament to its seriousness as a threat to public health (WHO, 2022).

*A. fumigatus* typically reproduces via asexual spores (conidia), which are released into the air for eventual germination. While some spores eventually return to the soil, others are inhaled by humans, where they interact with the epithelium of the lung (Chotirmall et al., 2013). Aided by their small diameter (2-3 μm) and hydrophobic outer layer, these spores can subsequentially reach the lung alveoli (Croft et al., 2016). Once in the lung, *A. fumigatus* must survive a hostile environment and host defense system (Bertuzzi et al., 2018). Immunocompetent individuals clear these spores, but immunocompromised ones are at risk of developing IA (Cadena et al., 2021).

Several species closely related to *A. fumigatus* are not considered pathogenic (de Vries et al., 2017; Rokas et al., 2020; Mead et al., 2021). For example, *Aspergillus fischeri* is a close relative of *A. fumigatus* (the two species share >90% average nucleotide sequence identity and >95% average amino acid sequence identity between orthologs), yet *A. fischeri* is less virulent and is not considered clinically relevant (Mead et al., 2019; Steenwyk et al., 2020). Early genomic comparisons between two strains of *A. fumigatus* (Af293 and A1163) and one strain of *A. fischeri* (NRRL 181) revealed a set of genes uniquely present in *A. fumigatus* (Fedorova et al., 2008). However, a more recent genomic examination of 18 *Aspergillus* section *Fumigati* strains representing 13 species found that 206 known genetic determinants of virulence in *A. fumigatus* are all shared between *A. fumigatus* and at least one other closely related, non-pathogenic species (Mead et al., 2021). Finally, recent examinations of genomic variation between the genomes of hundreds of *A. fumigatus* isolates (Barber et al., 2021; Lofgren et al., 2021, Horta et al., 2022) have revealed that *A. fumigatus* has an open pangenome with ~70% of its genes being highly conserved across strains (core) and that orthologs from both clinical and environmental strains exhibit a high degree of sequence conservation.

Variation in non-coding regions can also contribute to phenotypic variation, including variation in gene expression, between and within species (Caroll, 2005). We have previously demonstrated that non-coding regions between two *A. fumigatus* reference strains, Af293 and A1163 (Brown et al., 2022; Colabardini et al., 2022), as well as between *A. fumigatus* and its non-pathogenic close relatives are highly variable (Brown et al., 2022). For example, we found that 418 *A. fumigatus* genes exhibit a different rate of evolution in their non-coding regions (relative to non-pathogenic close relatives), including the non-coding regions of 25 genes that are known genetic

determinants of *A. fumigatus* virulence. Examination of these non-coding regions revealed numerous single nucleotide and insertion/deletion (indel) differences between *A. fumigatus* and closely related non-pathogenic species (Brown et al., 2022).

To increase our knowledge of non-coding region variation *within A. fumigatus* and how levels of non-coding sequence variation compare to levels of protein-coding sequence variation, we examined the genomes of 263 *A. fumigatus* strains (using 2 *A. fischeri* strains as an outgroup). Of the 5,812 single copy orthologs identified across all strains, 2,402 genes had identical or near identical sequences (≥ 99%); the same was true for 817 non-coding regions; 3,721 non-coding regions exhibited between 75% and 99% similarity (compared to 2,482 protein-coding regions); and 1,274 genes had a percent identity < 75% in their respective non-coding regions and 928 in their protein-coding regions; regions with low sequence similarity tend to yield unreliable sequence alignments, and were not included in subsequent analyses. Instead, we focused our analyses on 2,482 genes whose protein-coding regions exhibited percent sequence identities between 75% and 99%, performing two different tests of positive selection: the McDonald-Kreitman test (McDonald & Kreitman, 1991, Murga-Moreno et al., 2019) and the HKA test (Hudson et al., 1987, Ferretti et al., 2012). Examination of relative levels of sequence polymorphism to divergence of the non-coding and protein-coding regions of these 2,482 genes using the MK test identified 472 non-coding and 217 protein-coding regions with signatures of positive selection. These non-coding regions include 18 known genetic determinants of *A. fumigatus* virulence, such as *zrfB* (plasma membrane zinc transporter), *myoE* (class V myosin, involved in cellular morphogenesis), and *pld2* (putative phospholipase D protein). The HKA test identified 207 non-coding and 4 protein-coding regions whose polymorphism to divergence ratio was different compared to a neutral locus. The 207 upstream non-coding regions identified with the HKA test included 4 genetic determinants of *A. fumigatus* virulence; 1 of the 4 protein-coding regions was a genetic determinant of virulence. Molecular function terms enriched for genes associated with the non-coding regions that showed evidence of selection in both the MK and HKA tests, include ion-binding, transcriptional regulation, and stress response. Taken together, these results demonstrate that *A. fumigatus* non-coding regions are typically more variable and more often under positive selection than their protein-coding counterparts, raising the hypothesis that they too may contribute to phenotypic differences between *A. fumigatus* strains.

## Methods

### Genomic data collection

All *Aspergillus* genomes are publicly available and were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/).

### Identification of single-copy orthologous genes

To infer single-copy orthologous genes across all 265 taxa, we used OrthoFinder, version 2.4.0 (Emms & Kelly, 2015). OrthoFinder clustered genes into orthogroups from gene-gene sequence similarity information obtained using the program DIAMOND version 2.0.9 (Buchfink et al., 2015) with the proteomes of the 263 *A. fumigatus* strains and 2 *A. fischeri* strains as input. The key parameters used in DIAMOND were e-value = $1 \times 10^{-3}$ with a percent identity cutoff of 30% and percent match cutoff of 70%. We considered genes to be single-copy orthologs if they were within the cutoff thresholds and were present in all 265 taxa.

### Retrieval of non-coding regions

To identify highly conserved non-coding regions, we first retrieved the non-coding sequences directly upstream of the first codon of all single-copy orthologous genes from all genomes. Non-coding sequence retrieval was performed using custom python code which can be found at https://github.com/alecbrown24/General_Bio_Scripts (this script was adapted from a previously available script: https://github.com/shenwei356/bio_scripts). We retrieved the first 1,500 bp of non-coding sequence directly upstream of the first codon of each gene and used these sequences to generate FASTA files of non-coding regions, as well as FASTA files of single-copy orthologous protein-coding sequences using Python version 3.8.2. For some of the non-coding regions, there were <1,500 bp of non-coding sequence between the first codon of the gene of interest and an upstream gene; in these instances, only the intergenic region was used for subsequent analyses.

### Alignment and identification of conserved non-coding and protein-coding regions

Multiple sequence alignments for all non-coding and protein-coding regions were constructed using MAFFT, version 7.453, with default parameter settings (Rozewicki et al., 2019). Codon-based alignments were inferred from the corresponding protein sequence alignments using pal2nal, version 14 (Suyama et al., 2006). Sequence identity in protein-coding and non-coding regions was calculated from their corresponding multiple sequence alignment files using AliStat version 1.12 (Wong et al., 2020). The percent sequence identity for each position in the alignment was calculated by determining if the nucleotide at each position was the same as across all taxa.

To measure evolutionary conservation across individual alignment sites, we implemented the PhyloP program as part of the Phylogenetic Analysis with Space/Time Models (PHAST) suite of programs (Ramani et al., 2019). PhyloP scores reflect the evolutionary conservation of individual nucleotide sites relative to the degree of conservation expected under neutrality. A positive score is predictive of evolutionary conservation and a negative score is predictive of evolutionary acceleration relative to neutral expectations.

**Phylogenetic tree inference**

A phylogenetic tree of the 265 strains used in this study was generated by pruning from a larger *Aspergillus* species phylogeny (Steenwyk et al., 2022) using the Treehouse software in R using default parameters (Steenwyk & Rokas, 2019). Individual protein-coding region and non-coding region trees were inferred using IQ-TREE, version 2.0.6 (Minh et al., 2020), with "GTR+I+G+F" as it was the best fitting substitution model (Waddell and Steel, 1997; Vinet and Zhedanov, 2011).

**Identifying signatures of selection in *A. fumigatus* protein-coding and non-coding regions**

To examine signatures of selection in *A. fumigatus* protein-coding and non-coding region alignments, we examined variation in the ratio of non-synonymous / non-coding sites (likely under selection) to synonymous sites (likely neutral) (**Fig. 1**). We used the protein-coding and / or non-coding region alignments to calculate the fractions of polymorphic (differences between *A. fumigatus* strains) and divergent sites (differences between *A. fumigatus* and the outgroup *A. fischeri*) for non-synonymous, synonymous, and non-coding sites using the standard McDonald Kreitman test function as part of the iMKT software in R (Murga-Moreno et al., 2019). For protein-coding regions, the ratio of polymorphic non-synonymous to synonymous sites was compared to the ratio of divergent non-synonymous to synonymous sites. For non-coding regions, the ratio of polymorphic non-coding to synonymous sites was compared to the ratio of divergent non-coding to synonymous sites (**Figure 12**). For each MK test, the null hypothesis (H0) assumed that the ratio of selected vs neutral divergent sites was similar to the ratio of selected vs neutral polymorphic sites. We compared H0 to an alternative hypothesis (H1) in which there are more divergent sites that polymorphic sites across a given protein-coding or non-coding region, indicating positive selection. To determine whether H1 was significantly different from H0 for each of the codon-based alignments, we used Fischer's exact test with a statistical significance threshold of $p < 0.05$ and a Bonferroni-adjusted alpha value $< 0.01$ to adjust for multiple testing.

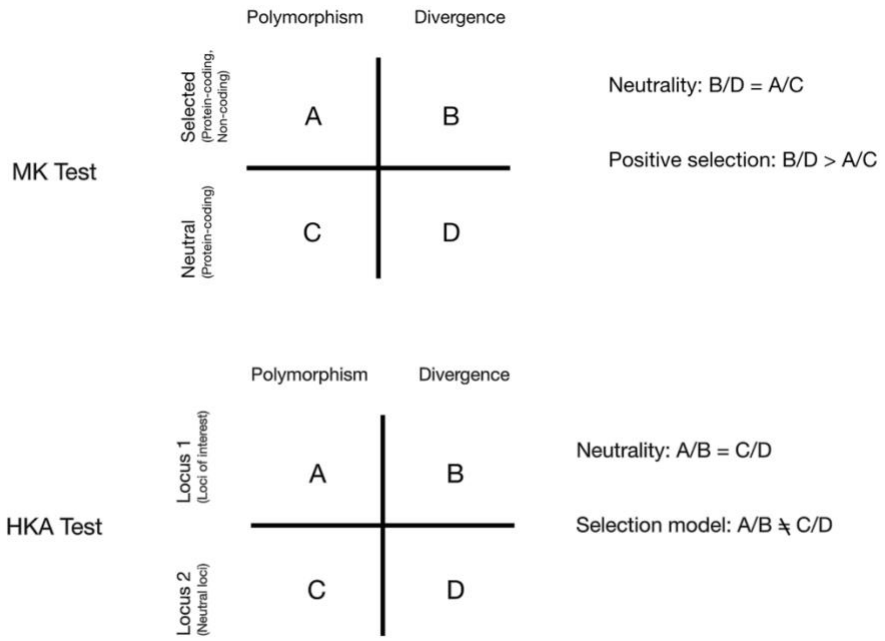



**Figure 12. Brief overview of the MK and HKA tests of selection for Chapter 3.**
The MK test (top) compares the polymorphisms (i.e., sites that vary within *A. fumigatus*) and divergence (sites that are fixed within *A. fumigatus* but differ from *A. fischeri*) between selected sites (non-synonymous or non-coding) and neutral sites (synonymous) between the protein-coding and non-coding regions of a given gene. For protein-coding regions, non-synonymous sites are compared to synonymous sites, while for non-coding regions, all sites are considered non-synonymous sites and are compared to the synonymous sites of the associated protein-coding region. Under a neutral model, the ratio of selected and neutral sites that are polymorphic is the same as the ratio of selected and neutral sites that are divergent. When the ratio of divergence is greater than the ratio of polymorphism, the MK test assumes that the selection is acting to fix advantageous non-synonymous changes, resulting in positive selection. The HKA test (bottom) compares the levels of polymorphism and divergence between two loci (the locus of interest and a reference, neutral locus). When ratio of polymorphism within species is equal to the ratio of divergence between species in the two loci, both loci are evolving neutrally. Should these ratios differ, we conclude that selection is occurring at the locus of interest.

The HKA test was also implemented, which compares the rate of polymorphism within *A. fumigatus* to divergence (between *A. fumigatus* and *A. fischeri*) at multiple loci (Hudson et al., 1987) (**Figure 12**). The HKA test assumes that if two loci are evolving neutrally, the ratio of polymorphism to divergence at these loci should be relatively constant. We compared loci of interest to neutral loci using the HKADirect program (Ferretti et al., 2012). Neutral loci were determined by comparing each of the 2,482 loci to the genomic background using Tajima D's test as part of the HKADirect program. The null hypothesis (H0) assumes that the patterns of genetic variation within a species (polymorphism) and the patterns of genetic differentiation

between species (divergence) are consistent with neutral evolution. Under these conditions, the polymorphism to divergence ratio is similar between the loci of interest and neutral loci. We compared H0 to an alternative hypothesis (H1) in which assumes that the patterns of polymorphism and divergence at the loci of interest deviate from that of neutral loci due to the action of natural selection. We used Fischer's exact test with a statistical significance threshold of $p < 0.05$ to determine significance.

Unlike the MK test, whose results can be used to directly compare the protein-coding and the non-coding regions of each gene, the HKA test instead compares each protein-coding and non-coding region to neutral loci. Thus, the MK test was used to determine differences in signatures of selection between non-coding regions and their associated protein-coding regions while the HKA test was used to detect signatures of selection in specific non-coding / protein-coding regions when compared to neutrally evolving non-coding / protein-coding regions.

**Functional enrichment analyses of genes with signatures of selection**

To determine whether genes with signatures of selection in either their protein-coding or non-coding regions were enriched for particular functional categories, we implemented the Gene Ontology (GO) tool g:PROFILER (Raudvere et al., 2019) using default settings. We performed four separate analyses for a) enrichment of genes that were significant according to the MK test of protein-coding regions, b) enrichment of genes that were significant according to the MK test of non-coding regions, c) enrichment of genes that were significant according to the HKA test of protein-coding regions, and d) enrichment of genes that were significant according to the HKA test of non-coding regions. Each of these gene sets was compared to a general background set that includes all the features / gene names in the database with at least one GO annotation for *A. fumigatus*. All functional enrichment analyses used a p-value cutoff of 0.05.

**Examination and visualization of mutational signatures**

To identify interesting examples of sequence variation between *A. fumigatus* strains for non-coding regions of genes of interest, we visualized and compared multiple sequence alignments using the MView function in EMBL-EBI (Madeira et al., 2019). Workflow of methods can be seen in **Figure 13**.

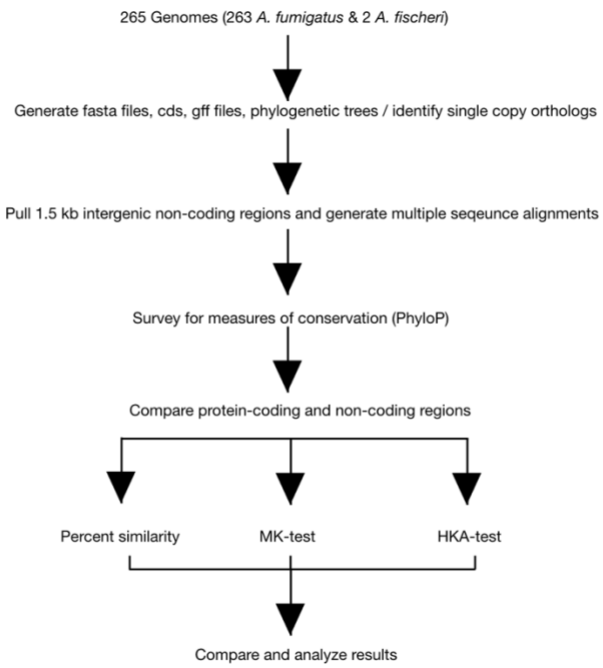



**Figure 13. Flow chart of Methods for Chapter 3.**

## Results and Discussion

### Protein-coding and non-coding regions exhibit differing levels of sequence conservation within *A. fumigatus*

To analyze the sequence diversity of non-coding regions across *A. fumigatus* strains (**Figure. 14**), we first identified 5,812 single-copy orthologous genes amongst 263 *A. fumigatus* strains and 2 *A. fischer*i strains. We then looked to measure evolutionary conservation at individual alignment sites for both protein-coding and non-coding regions across the 5,812 single copy orthologous genes of interest. Of the 5,812 single copy orthologous genes, 5,646 were found to be alignable, thus we focused our subsequent analyses around these genes.


**Figure 14. Phylogeny of the 263 A. *fumigatus* strains used in this study.**

A phylogenetic tree of the 265 strains used in this study was generated by pruning from a larger *Aspergillus* species tree inferred from analyses of 1,362 protein-coding regions (Steenwyk et al., 2022).

Examination of PhyloP scores for each protein-coding and non-coding alignment individually revealed that both protein-coding and non-coding regions exhibit varying levels of conservation across single copy orthologous genes (**Figure 15**). Examination of average PhyloP scores in protein-coding regions revealed a lower area of conservation near their start (first ~100 bp). This may be due to slight differences in gene annotation between strains or the presence of genuine variation; utilization of alternate start sites for the same gene has been demonstrated in *Aspergillus* (Kjærbølling et al., 2020). Beyond the first ~100 bp, conservation levels of protein-coding regions remain high throughout the first 1,500 bp. High conservation among protein-coding regions is also consistent with comparisons between *A. fumigatus* and closely related species (Fedorova et al., 2008).

The average PhyloP score in non-coding regions across *A. fumigatus* strains revealed the highest levels of sequence conservation (as indicated by a higher PhyloP score) were directly upstream of the start site, with conservation generally decreasing further away from the start site. We also found that conservation begins to fade around 1,500 bp upstream of the start site (as indicated by a PhyloP score of 0). This pattern of higher sequence conservation in non-coding regions right upstream of the transcription start site is consistent with a previous study of non-coding regions comparing A. *fumigatus* and closely related species (Brown et al., 2022).

To further examine the conservation of pairs of non-coding and protein-coding sequences, we calculated the percent identity for all single copy orthologs in both their protein-coding and associated non-coding regions. We found that percent nucleotide sequence identity of protein-coding regions exhibited weak but significant correlation ($r^2 = 0.143$ and p-value = $<0.0001$) with the percent nucleotide sequence identity of the associated non-coding regions (i.e., orthologs with higher percent identity in their protein-coding regions also exhibited higher percent identity in their non-coding regions). Thus, it was sometimes the case that highly similar protein-coding regions were associated with non-coding regions that display higher sequence variation. This result suggested that the functions of genes with highly conserved protein-coding regions may still differ between strains due to differences in the genes' non-coding regions. In a few cases, we also identified highly divergent protein-coding regions that were associated with highly similar non-coding regions.

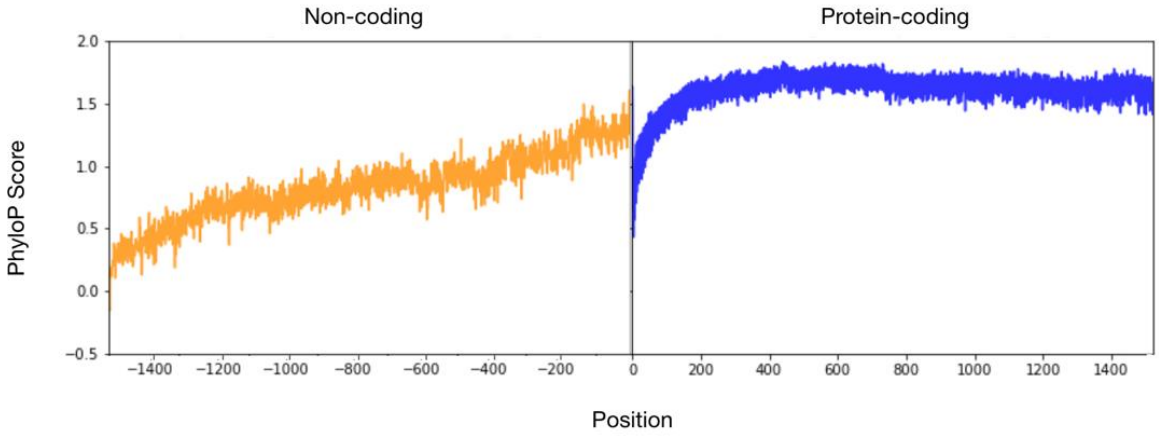



**Figure 15. Non-coding regions are less conserved than protein-coding regions in the major fungal pathogen *A. fumigatus*.**
PhyloP Score for protein-coding and non-coding regions. To determine conservation in protein-coding (blue) and non-coding (orange) regions, we calculated the PhyloP score across sites of all multiple nucleotide sequence alignments of protein-coding and upstream non-coding regions of 5,646 single copy orthologs across 263 A. *fumigatus* strains and 2 A. *fischeri* strains. Scores of conservation were measured for individual nucleotide sites and scores were then averaged across all orthologs. Conserved sites have PhyloP scores above 0 and non-conserved sites have scores below 0. We find that in non-coding regions, sites that are closer to the start of the transcription start site (TSS) exhibit a higher level of conservation and generally decrease in conservation as we move further from the TSS. In protein-coding regions, PhyloP scores are generally above 0, which are indicative of high sequence conservation; the lowest scores are observed near the start of the protein-coding regions, which is likely an artifact caused by variation in starting codon position of gene annotations across *A. fumigatus* strains.

We next computed the percent nucleotide sequence identity between the non-coding and protein-coding regions of each single-copy orthologous gene across all 265 strains (**Figure 16**) for the 5,646 genes whose protein-coding sequences were alignable. Averaging the non-coding region percent similarities for the 5,646 single-copy orthologous genes revealed an average identity of ~85%, while the average protein-coding percent identity was ~92%. Additionally, we found that 928 protein-coding alignments exhibited <75% nucleotide sequence identity (this number includes the 166 unalignable genes), 2,482 exhibited sequence identity between ≥75% and <99%, and 2,402 exhibited ≥99% identity. For non-coding region alignments, there were 1,274 non-coding alignments that exhibited a <75% identity, 3,721 that exhibited sequence identity between ≥75% and <99%, and 817 that exhibited ≥99% identity.

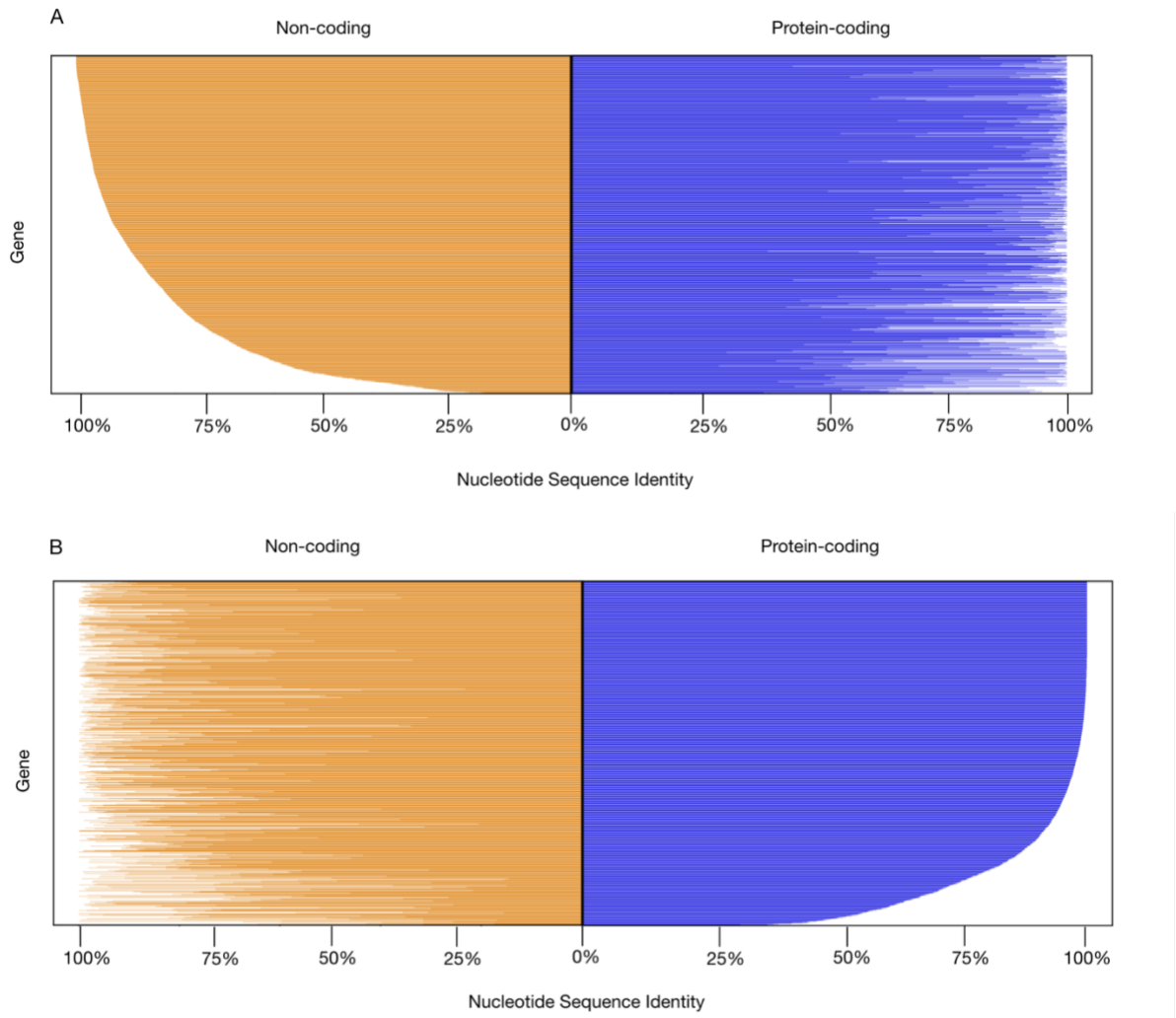



**Figure 16. *A. fumigatus* orthologs exhibit numerous instances of highly conserved protein-coding genes whose non-coding regions are poorly conserved.**
Percent identity of protein-coding and non-coding regions of 5,646 *A. fumigatus* genes. Up to 1.5kb upstream noncoding region (orange) is shown to the left were calculated and plotted by percent identity starting with 100% similar (top row) and descending. The associated protein-coding regions (blue) are shown to the right. Although the sequence conservation of protein-coding regions and the sequence conservation of their corresponding non-coding regions are correlated, there are numerous instances of genes with high protein-coding sequence identity and a lower identity in their non-coding region. A) *A. fumigatus* genes ranked by percent nucleotide sequence identity of their non-coding regions. B) *A. fumigatus* genes ranked by percent nucleotide sequence identity of their protein-coding regions.

**Many non-coding regions have signatures of positive selection**
To examine signatures of selection in *A. fumigatus* genes, we performed the MKA and HKA tests of selection in 2,482 pairs of protein-coding and non-coding region alignments. For the MK test, we found that a total of 472/2,482 (19.0%) genes exhibit signatures of selection in their non-coding regions but not in their protein-coding regions, a total of 217/2,482 (8.7%) genes experienced selection in their protein-coding regions but not in their non-coding regions, and

144/2,482 (5.80%) genes experienced selection in both their protein-coding and non-coding regions (**Figure 17**).

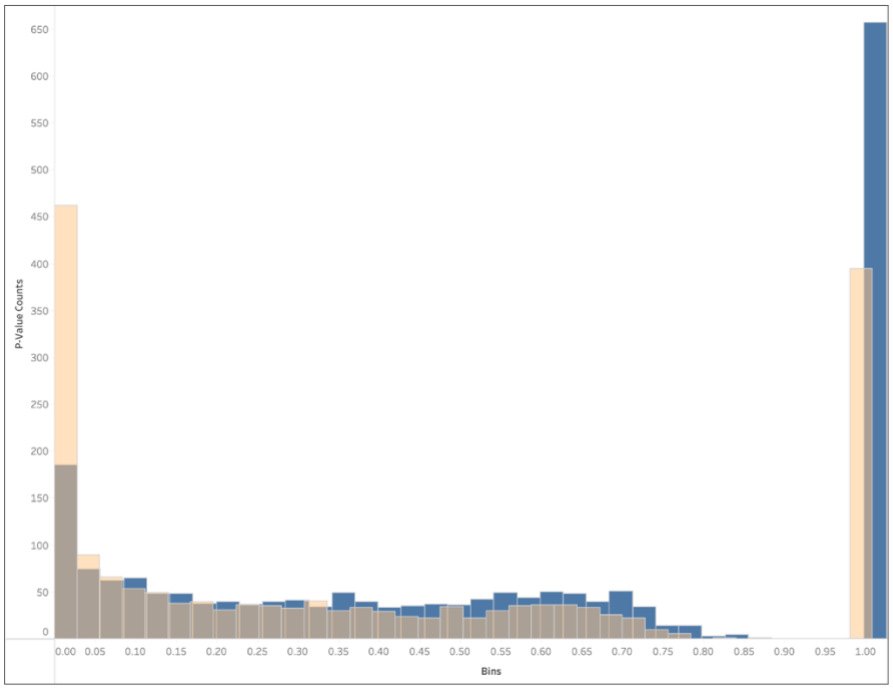



**Figure 17. A higher number of non-coding regions than protein-coding regions exhibit signatures of selection under the McDonald-Kreitman test.**
Histogram of the distribution of p-values of the MK test. The MK test was calculated for 2,482 single copy orthologs in both protein-coding (blue) and non-coding regions (orange). 217 protein-coding and 472 non-coding regions were found to be significant ($p < 0.05$).

For the HKA test, we found 4/2,482 (8.8%) and 207/2,482 genes with evidence of positive selection in their protein-coding and non-coding regions, respectively (**Figure 18**). Examination of the genes that were significant under the MK and HKA tests shows that there is relatively limited overlap for both protein-coding and non-coding regions (**Figure 18**). For example, only 36 non-coding regions exclusively exhibit evidence of selection by both tests. For protein-coding regions, the lack of overlap is largely due to the very small number of protein-coding regions that show evidence of selection in the HKA test. For the non-coding regions, the limited overlap is likely due to the differences in the neutral sites used by the two tests (the MK test uses the synonymous sites of the corresponding protein-coding region whereas the HKA test uses all the sites of a neutrally evolving non-coding region).

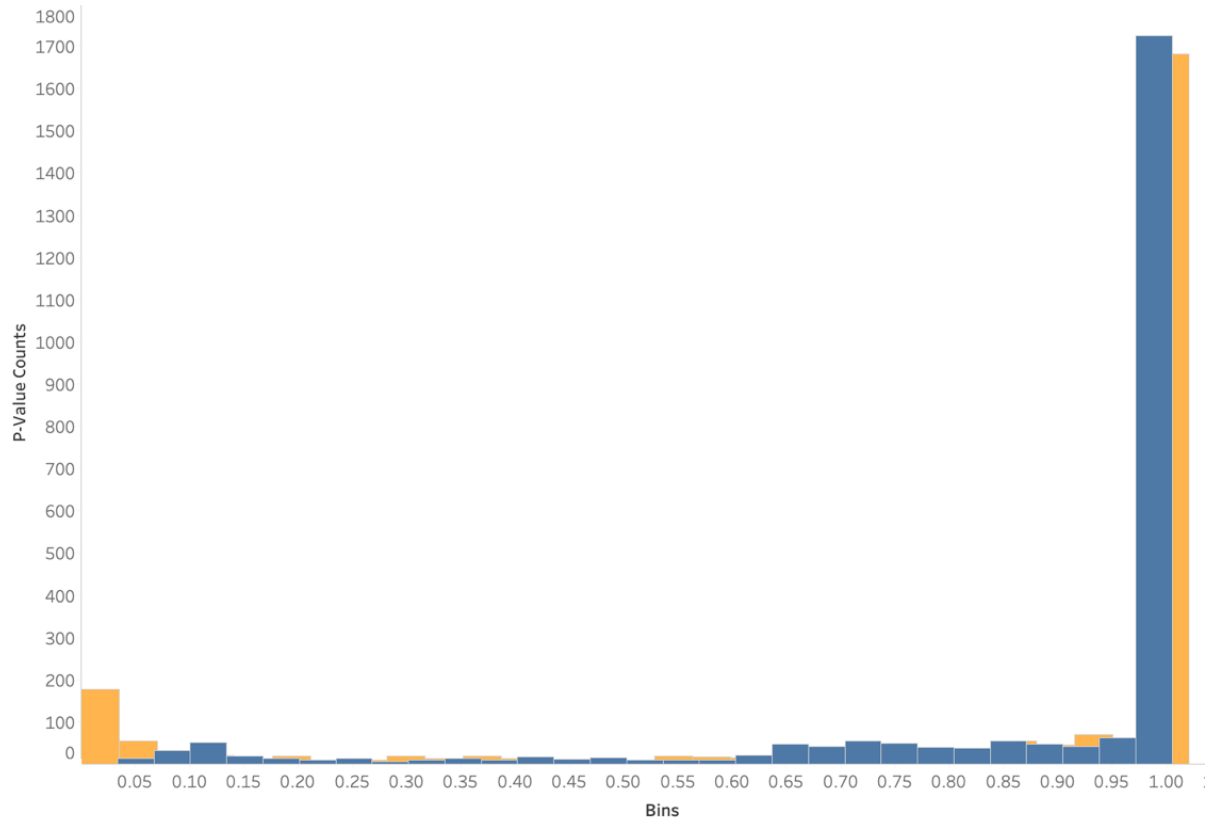



**Figure 18. HKA test identified 207 non-coding and 4 protein-coding regions that exhibit a signature of selection.**
Histogram of the distribution of p-values of the HKA test across protein-coding (blue) and non-coding (orange) regions of 2,482 genes. 207 non-coding and 4 protein-coding regions were found to be significant ($p < 0.05$).

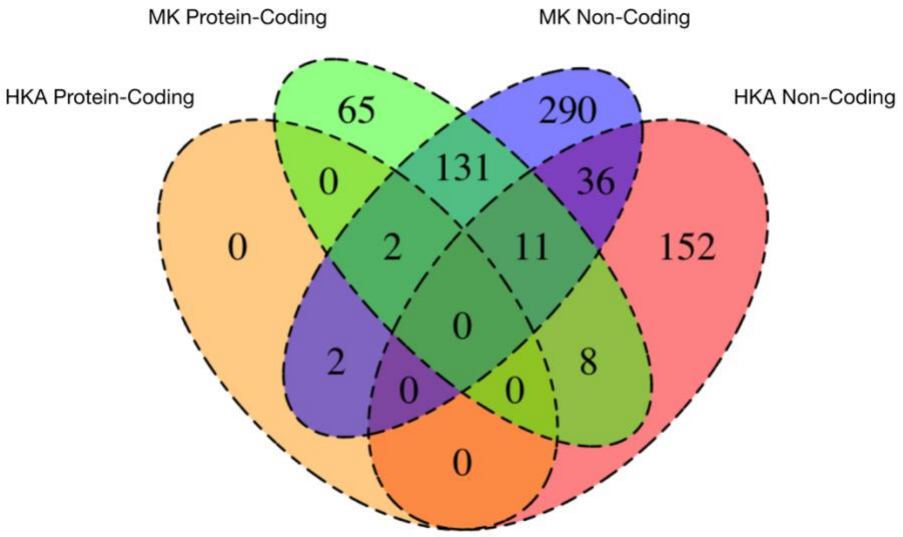



**Figure 19. Venn Diagram of significant results from HKA protein-coding, MK-protein-coding, MK non-coding and HKA protein-coding tests.**
There were 478 (290 + 36 + 152) genes with evidence of selection only in their non-coding regions compared to 65 genes with evidence of selection only in their protein-coding regions across both MK and HKA tests.

**Genes with evidence of selection in non-coding regions are enriched for binding and regulatory activity, including 21 genes involved in *A. fumigatus* virulence**
We used GO enrichment to determine if there were any functions that were overrepresented. For the 472 genes with evidence of selection under the MK test in their non-coding regions, we found 22 categories that were enriched for molecular function. "DNA Binding" was the top term identified for molecular function ($p= 9.06x10^{-8}$) and the most enriched term overall. "Cytoskeleton motor activity" ($p=6.59x10^{-6}$), "Ion Binding" ($p=7.33x10^{-5}$) and "Transcription factor binding" ($p=4.08x10^{-4}$) were also represented grouped terms for molecular function components respectively (Table S7). For the 217 genes with evidence of selection in their protein-coding regions, 7 molecular functions were overrepresented, including "ATP hydrolysis activity" ($p=1.37x10^{-3}$) and various functions involved in binding activities. For the HKA test, we find that enzyme regulator activity ($p =3.33x10^{-2}$) was the only molecular term found for the non-coding regions, and no GO terms were enriched for the HKA protein-coding results (Table S7). Additionally, the 36 genes that experienced selection in their non-coding regions under both the HKA and MK tests, GO analysis revealed enrichment for various binding processes.

We compared our list of genes under selection to a previously curated set of 206 genetic determinants of *A. fumigatus* virulence (Steenwyk et al., 2021). Given that *A. fumigatus* strains have been demonstrated to exhibit differences in virulence in mouse models of fungal disease (Keizer et al.,2021), selection in the non-coding or protein-coding regions of these genes may be

of relevance to *A. fumigatus* virulence. We found that the non-coding regions of 18 of the 206 virulence genes were under selection according to the MK test (*argEF, ags1, csmB, pabA, medA ,mtfA ,myoB ,myoE ,pld2,* g*liP, rgsC, aceA, atfA, cch1, fbx15, flcB, schA, zrfB)* and 3 according to the HKA test *(cds1*, *nop4*, *dvrA*). We found that the most represented general function amongst these 18 genes was "stress response", which raises the question of their impact on virulence, given the role that stress response has been shown to play in *A. fumigatus* virulence (Colabardini et al., 2022). We also identified 4 protein-coding regions of virulence genes (*lysF*, *myoB*, *aftA*, *fbx15*) that showed evidence of selection under the MK test. Similarly, we found 11 virulence genes (*hcsA*, *lysF*, *ags1*, *chsG*, *erg12*, *rtfA*, *tom24*, *fmaE*, *gliC*, *gliT*, *sidI*) with evidence of selection in their non-coding regions under the HKA test, and 1 (*aceA*) with evidence of selection in its protein-coding region.

**Examples of non-coding region differences between *A. fumigatus* strains**

We next sought to identify representative sequence differences in non-coding regions between *A. fumigatus* strains that exhibited signatures of selection according to the MK test or HKA test. (**Figure. 20**). One such example was the non-coding region of *AFUA_3G03310*, which was found to be under selection according to the MK test. The non-coding region exhibits a 12 bp region (AGCCACAGAACT) present in *A. fumigatus* Af293 and 91other strains but absent from the rest, including strain BA78-1137. This binding site location is an exact match to the Met31 transcription factor binding site involved in sulfur metabolism in *S. cerevisiae* (Cormier et al., 2010). The MetZ transcription factor performs a similar function in *Aspergillus nidulans* (Pilsyk et al., 2015) and may be also in *A. fumigatus* (Amich et al., 2016). Another gene with evidence of selection in its non-coding region is *AFUA_3G11330*, which encodes the putative transcription factor AftA involved in stress response and spore viability in *Aspergillus* (Lara-Rojas et al., 2011). The *AFUA_3G11330* non-coding region exhibits a 6 bp region (TACTCT) present in *A. fumigatus* Af293 and about half of the other *A. fumigatus* strains while absent in the other *A. fumigatus* strains, including the KXX37 strain. This 6 bp region is similar to the 6 bp binding site for Yap1 in *S. cerevisiae*, which is required for oxidative stress tolerance (Natkanska et al., 2017). An ortholog of Yap1 is known to be involved in voriconazole resistance in *A. flavus* (Ukai et al., 2018) and may play role in stress response in *A. fumigatus*, which is important as a mechanism for survival within the human lung. Both the MK and HKA tests found signatures of selection in *AFUA_6G03660*, an uncharacterized gene in *A. fumigatus*. Although of unknown function, this gene is of particular interest as its non-coding region differs between the two reference strains *A. fumigatus* Af293 and *A. fumigatus* A1163, which vary in their virulence in animal models of fungal disease. *A. fumigatus* A1163 exhibits a larger region of 22 bp that is absent in *A. fumigatus* A1163. This region helps to illustrate the complex non-coding sequence differences between *A. fumigatus* strains, including those that are closely related.

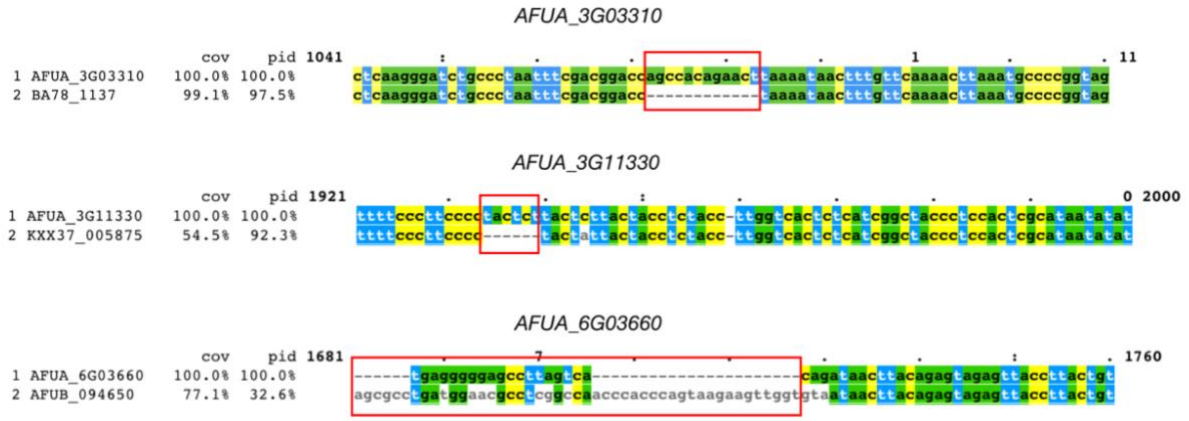



**Figure 20. Notable examples of sequence differences between A. *fumigatus* strains in non-coding regions.**
Here, we show three regions of sequence alignments of non-coding regions that differ between *A. fumigatus* Af293 and another *A. fumigatus* strain. (A) AFUA_3G03310 (RTA1 domain protein) exhibits a 12 bp region present in *A. fumigatus* Af293 but absent in several other strains, including BA78_1337. (B) AFUA_3G11330 (transcription factor AtfA) exhibits a 6 bp region absent in *A. fumigatus* Af293 but present in several other strains, including KXX37_005875. (C) AFUA_6G03660 (predicted to be involved in the production of biotin) exhibits a non-coding difference between the two reference strains of *A. fumigatus* (Af293 and A1163).

Here, we presented a comprehensive study of signatures of positive selection in both protein-coding and non-coding regions across a large number of strains from the fungal pathogen *A. fumigatus*. We identified several non-coding regions under selection in *A. fumigatus*, including several candidate transcription factor binding sites that differ between strains are await further exploration. Currently, there are no datasets available that report genome-wide differential expression data for *A. fumigatus* strains. Experiments on diverse *A. fumigatus* strains to investigate and analyze the differential expression of genes and examine the functional implications of the non-coding region disparities we have identified, specifically in relation to the expression and virulence of *A. fumigatus*, will be of great interest.

## Chapter 4: Descriptive analyses of non-coding regions comparing *A. fumigatus* Af293, *A. fumigatus* A1163 and *A. fischeri* to identify differences in putative transcription factor binding sites of known regulators involved in virulence.


### Abstract

The filamentous fungus *Aspergillus fumigatus* holds immense importance as a human pathogen, causing a range of severe infections (chiefly invasive aspergillosis), particularly in immunocompromised individuals. The understanding of its transcriptional regulatory mechanisms is crucial for developing targeted therapeutic strategies against invasive aspergillosis. Transcription factors (TFs) and upstream non-coding regions play pivotal roles in gene expression, growth, and virulence *in A. fumigatus*. *A. fumigatus* exhibits a complex regulatory network that governs the expression of virulence factors, antifungal resistance determinants, and various metabolic pathways. Transcription factors serve as regulatory switches, binding to specific DNA motifs within gene promoter regions to activate or repress gene transcription. These regulatory proteins control the expression of genes involved in stress response, metabolism, conidiation/sporulation, and host immune evasion, influencing the overall fitness and pathogenic potential of *A. fumigatus*. Moreover, the investigation of upstream non-coding regions in *A. fumigatus* has emerged as a critical area of research. These regions harbor vital cis-regulatory elements that fine-tune gene expression in response to environmental cues and host signals. Understanding the organization and function of these non-coding regions is essential for deciphering the complex gene regulatory networks that govern fungal development and pathogenesis. Additionally, the interplay between transcription factors and these non-coding regions further contributes to the intricate regulatory landscape in *A. fumigatus*.

*Aspergillus fischeri*, a close related, non-pathogens shares >95% sequence identity across coding regions and shares most of the same genes as *A. fumigatus*. Yet, they differ in their pathogenic profiles. Moreover, despite being >98% identical in their protein coding regions, *A. fumigatus* a1163 differs in its virulence propensity (more virulent) than *A. fumigatus* af293. Taken together, it may be that differences in virulence across species/strains is due to differences in gene regulation as opposed to gene presence/absence. Here, we compare the binding location for the 6 transcription factors whose binding targets have been elucidated in A. fumigatus across the reference strains *A. fumigatus* af293*, A. fumigatus* a1163 and *A. fischeri*. We identified a set of genes whose upstream binding targets in *A. fumigatus* differ from *A. fischeri* in their equivalent upstream position. The research presented here is the first study comparing known TF binding locations between *A. fumigatus* and *A. fischeri*. Deciphering the regulatory mechanisms governing gene expression may better explain pathogenic differences across these 3 *Aspergillus* species/strains.

## Introduction to the 6 transcription factors.

### CrzA

*Aspergillus fumigatus* has developed sophisticated strategies to thrive in diverse niches, including the human respiratory system and the ability of the fungus to sense and respond to fluctuating calcium levels is crucial for its survival and virulence (Latge et al., 2019). CrzA has been identified as a key mediator of calcium-mediated signaling in *A. fumigatus*. CrzA contains conserved C2H2 zinc-finger domains that enable its binding to specific DNA motifs within target gene promoters (Chang, 2008). It acts downstream of the calcium/calcineurin pathway, which includes the calcium sensor protein calmodulin and the serine/threonine phosphatase calcineurin (Shwab et al., 2019). Upon an increase in intracellular calcium levels, calcineurin dephosphorylates and activates CrzA, allowing its translocation to the nucleus, where it modulates gene expression (Shwab et al., 2019).

CrzA governs the expression of various genes involved in critical cellular processes in *A. fumigatu*s (de Castro et al., 2014, Colabardini et al., 2022). CrzA regulates the expression of cell wall-related genes, influencing cell wall integrity and morphology (de Castro et al., 2014, Colabardini et al., 2022). Additionally, CrzA plays a pivotal role in mediating responses to oxidative stress, cation homeostasis, and adaptation to antifungal drugs (Abad et al., 2010). The tight control of these cellular functions by CrzA underscores its significance in maintaining cellular homeostasis and combating adverse conditions.

CrzA's regulatory influence extends to *A. fumigatus* virulence and pathogenicity. Studies have shown that CrzA is involved in regulating the expression of genes associated with conidiation, hyphal growth, and the production of secondary metabolites, all of which are crucial for the establishment and dissemination of infection (de Castro et al., 2014, Colabardini et al., 2022). Notably, CrzA contributes to immune evasion by influencing the expression of immunomodulatory proteins, allowing *A. fumigatus* to evade host immune surveillance.

CrzA has been demonstrated to play a role in *A. fumigatus* azole resistance. More specifically, deletion of CrzA in *A. fumigatus* has been shown to impact the *A. fumigatus* transcriptome and morphology differently depending on exposure to differing concentration of caspofungin (known as the “caspofungin paradoxical effect”) (Colabardini et al., 2022). Transcriptomics revealed shared genes between *A. fumigatus* af293 and *A. fumigatus* a1163 being regulated differently given exposure to the same concentration of caspofungin (example, a higher concertation of caspofungin resulted in different genes being up/down regulated when compared across A. fumigatus strains) (Colabardini et al., 2022).

CrzA has also been shown to essential for *A. fumigatus* virulence in a mouse mode of infectious disease (de Castro et al., 2014). Several of the genes whose non-coding region is bound by CrzA in Chip-seq studies are also genetic determinants of virulence (Mead & Steenwyk et al., 2021).

### SrbA

SrbA, a member of the basic helix-loop-helix (bHLH) family, acts as a transcription factor that modulates gene expression in an oxygen-dependent manner (Chung et al., 2014). Under hypoxic conditions, SrbA accumulates and binds to specific DNA motifs within target gene promoters, activating the transcription of genes involved in the adaptation to low oxygen levels (Chung et al., 2014). Notably, SrbA controls the expression of genes related to ergosterol biosynthesis, a crucial component of fungal cell membranes, and heme biosynthesis, which is essential for various cellular processes, including respiration and iron acquisition (Gsaller et al., 2016).

SrbA's regulatory influence extends beyond cellular adaptation to oxygen availability and encompasses critical aspects of *A. fumigatus* pathogenesis (Abad et al., 2010). Studies have revealed that SrbA contributes to the regulation of genes involved in stress responses, antioxidant defense mechanisms, and virulence factor expression (Blatzer et al., 2011, Chung et al., 2014, Dhingra & Cramer, 2017). Furthermore, SrbA's role in coordinating iron homeostasis and heme biosynthesis is crucial for the acquisition of iron from the host during infection, providing *A. fumigatus* with a competitive advantage in the host environment . Given its central role in oxygen sensing and adaptation, as well as its impact on *A. fumigatus* virulence, SrbA presents a promising target for antifungal drug development (Blatzer et al., 2011). Disrupting SrbA's regulatory function could potentially impair the fungus's ability to adapt to host environments, attenuate its virulence, and enhance susceptibility to antifungal treatments. Deletion of *srbA* in *A. fumigatus* resulted in loss of virulence in mouse models (Chung et al., 2014). Thus, SrbA binding and subsequent regulation of gene targets may be essential for *A. fumigatus* virulence.

## HapX

HapX, a transcription factor belonging to the CCAAT-binding complex family, plays a central role in regulating iron homeostasis in response to changing iron levels in the environment (Schrettl et al., 2010). Understanding the regulatory functions of HapX is essential for deciphering the intricate mechanisms underlying iron acquisition and utilization in *A. fumigatus*. HapX operates as a key transcriptional repressor in iron-starved conditions, inhibiting the expression of iron-consuming proteins, including those involved in heme biosynthesis and iron-sulfur cluster assembly (Lopez-Berges et al., 2021). Under iron-replete conditions, the FepA-FepB complex acts as a sensor to detect iron availability, leading to the degradation of HapX through proteasomal pathways (Schrettl et al., 2010). This degradation enables the expression of iron-responsive genes, including siderophore biosynthetic genes, which are crucial for iron acquisition from the host and the environment.

The impact of HapX on *A. fumigatus* pathogenicity is multifaceted. The transcription factor influences the expression of genes involved in iron acquisition and utilization, essential for the establishment of infection within the host (Furukawa et al., 2020). Through its regulation of siderophore biosynthesis, HapX facilitates the acquisition of iron from the host during infection, bolstering the fungal pathogen's virulence (Furukawa et al., 2020). Additionally, HapX's repression of iron-consuming processes in iron-starved conditions helps the fungus conserve limited iron resources and maintain cellular homeostasis (Furukawa et al., 2020). Given its pivotal role in iron metabolism and its influence on *A. fumigatus* virulence, HapX represents a promising target for antifungal drug development (Abad et al., 2010). Disruption of HapX function could impair the fungus's ability to adapt to iron-limiting conditions and limit its ability to establish and sustain infection in the host. Moreover, deletion of HapX resulted in attenuated

virulence in A. fumigatus (Furakawa et al., 2020). Developing strategies to manipulate HapX-mediated iron regulation may offer novel therapeutic opportunities for managing A. fumigatus infections.

## NctA & AtrR

*Aspergillus fumigatus* azole resistance is a concerning challenge in the treatment of invasive aspergillosis, especially in immunocompromised patients (Latge et al., 2019) . One of the primary mechanisms underlying this resistance is the alteration of the Cyp51A protein, which is the target of azole antifungal drugs (Zhang et al., 2019). Cyp51A, a lanosterol 14α-demethylase, is an essential enzyme involved in the ergosterol biosynthesis pathway in fungi (Paul et al., 2017). Azole drugs bind to and inhibit Cyp51A, disrupting ergosterol production and compromising fungal cell membrane integrity (Paul et al., 2017). Mutations in the Cyp51A gene can lead to amino acid substitutions in the Cyp51A protein, reducing the drug's binding affinity and rendering azoles less effective against A. fumigatus (Bader et al., 2013). Consequently, these genetic changes confer resistance to azole antifungal agents and pose a significant therapeutic challenge (Arastehfar et al., 2021).

NctA and NctB are two CBF/NF-Y family transcription factors which have been found to contribute to azole resistance when their function is lost in *A. fumigatus* (Furakawa et al., 2020). Both NctA and NctB (Negative cofactor two A and B) function as components of the same transcriptional regulatory complex. Moreover, it was demonstrated that this NCT complex played a critical role in regulating ergosterol biosynthesis and the azole exporter CDR1B (Furukawa et al., 2020). While the interplay of NctA/NctB, Cyp51A and azole resistance has yet to be fully determined, it is likely that NctA/NctB plays a role in A. fumigatus virulence, as deletion of NctA resulted in reduced virulence (Furukawa et al., 2020).

In addition to Cyp51A being under transcriptional control of NtcA, it was also found that AtrR plays a role in its expression (Hagiwara et al., 2017). AtrR, was found to be present in A. fumigatus and to be required for iron acquisition (Yap et al., 2023). AtrR was found to be linked to regulation of both cyp51A and the ATP-binding cassette (ABC) transporter-encoding gene abcG1 (Hagiwara et al., 2017). Strains lacking *atrR* failed to drive normal transcription of either cyp51A or abcG1 and ChIP experiments demonstrated that AtrR bound to both the cyp51A and abcG1 promoter regions (Hagiwara et al., 2017). Moreover, deletion of AtrR was shown to be essential for *A. fumigatus* virulence in mice (Jackson et al., 2009).

## RglT

To defend itself against the fungicidal oxidative activities of the host immune system, *A. fumigatus* employs various defensive strategies (Latge et al., 2019). These include the secretion and production of a range of secondary metabolites (SMs), which promote fungal fitness and growth during infection (Latge et al., 2019). Among these SMs, gliotoxin is a well-characterized compound in *A. fumigatus (*Gomez-Lopez et al., 2022). The biosynthesis and secretion of gliotoxin are controlled by a cluster of 13 gli genes located on chromosome VI (Dolan et al., 2015). regulation of gli genes involves numerous proteins, including the transcription factor RglT whose deletion was shown to attenuate *A. fumigatus* virulence in mice (Ries et al., 2020).

## Methods

All *Aspergillus* genomes are publicly available and were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). These strains include *A. fumigatus* Af293 (Nierman et al., 2005), *A. fumigatus* A1163 (Fedorova et al., 2008) and *A. fischeri* NRRL1881 (Fedorova et al., 2008).

### Identification of single-copy orthologous genes

To infer single-copy orthologous genes among all protein-coding sequences for all ten taxa, we used OrthoFinder, version 2.4.0 (Emms & Kelly, 2015). OrthoFinder clustered genes into orthogroups algorithm Markov clustering models from gene similarity information using the sequence search program DIAMOND version 2.0.9 (Buchfink et al., 2015) and the proteomes of the ten *Aspergillus* species as input. The key parameters used in DIAMOND were e-value = 1 x $10^{-3}$ with a percent identity cutoff of 30% and percent match cutoff of 70%. This approach identified 7,414 single copy orthologous genes wherein all *Aspergillus* species are represented by one sequence.

### Identification of non-coding regions

We first curated intergenic sequences directly upstream of the transcription start sites of all 7,414 single-copy orthologous genes for each of the *Aspergillus* species/strains using a custom script (https://github.com/alecbrown24/General_Bio_Scripts) based on a previously available script (https://github.com/shenwei356/bio_scripts). Up to the first 1500 bp of intergenic sequence directly upstream of each gene's transcription start site were used to generate FASTA files of non-coding regions, as well as FASTA files of single-copy orthologous protein-coding sequences using Python version 3.8.2 (https://www.python.org/). For some non-coding regions, we found a coding regions existed within the 1500 bp threshold, in these instances, only the intergenic region was used for subsequent analyses. Non-coding region FASTA sequence files were used to generate multiple sequence alignments via MAFFT version 7.453 using default parameters (Katoh et al., 2019).

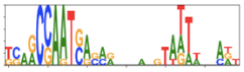
### Identification and comparison of TF binding sites.

We curated a set of 6 *A. fumigatus* TF binding motifs based on previous Chip-Seq studies, CrzA, SrbA, HapX, NctA, RgltT and AtrR along with their binding locations **(Table 2)**. To search for occurrences of the TF binding motif, we used the Find Individual Motif Occurrences (FIMO) program, version 5.3.3 (Grant et al., 2011). We specifically investigated the occurrence and conservation of the TF binding motif in the *A. fumigatus* and *A. ficheri* non-coding regions whose orthologs were shown to exhibit TF binding via ChIP-seq in the *A. fumigatus* strain (af293 or a1163) used in the original study.

### Examination and visualization of non-coding regions with differences in TF motifs

To identify examples of sequence variation between *A. fumigatus* and A. fischeri at TF binding site locations, we visualized and compared multiple sequence alignments using the Mview function in EMBL-EBI (Madeira et al., 2019).

**Table 2. Six transcription factors with known binding site locations.**

| Name | Role | Deletion phenotype | Number of genes | Binding site motif [62] | Ref. |
|---|---|---|---|---|---|
| CrzA | Involved in calcium ion homeostasis and development | Loss of virulence | 82 |  | Colabardini et al., 2022 |
| SrbA | Iron homeostasis, involved in drug resistance | Loss of virulence | 97 |  | Chung et al., 2014 |
| NctA | Involved in ergosterol biosynthesis and drug resistance | Reduced virulence | 455 |  | Furakawa et al., 2020a |
| HapX | Involved in iron homeostasis | Reduced virulence | 233 |  | Furakawa et al., 2020b |
| AtrR | ABC transporter, involved in multidrug resistance | Loss of virulence | 200 |  | Paul et al., 2019 |
| RglT | Oxidative stress resistance | Reduced virulence | 115 |  | Reis et al., 2020 |

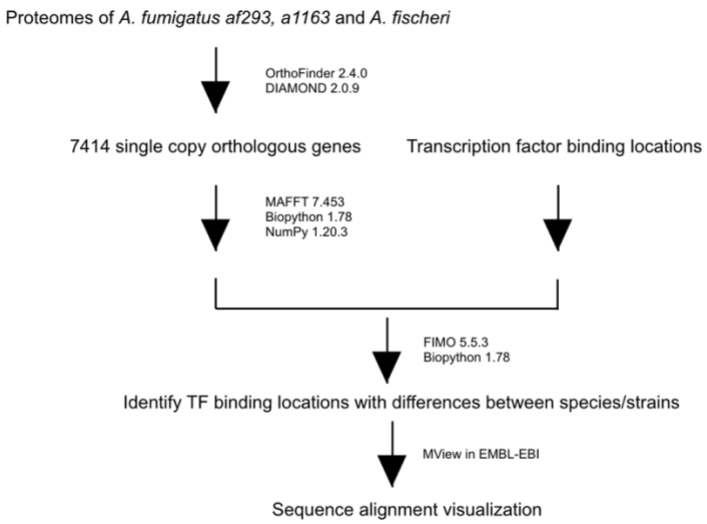



**Figure 21. Workflow of methods for Chapter 4.**

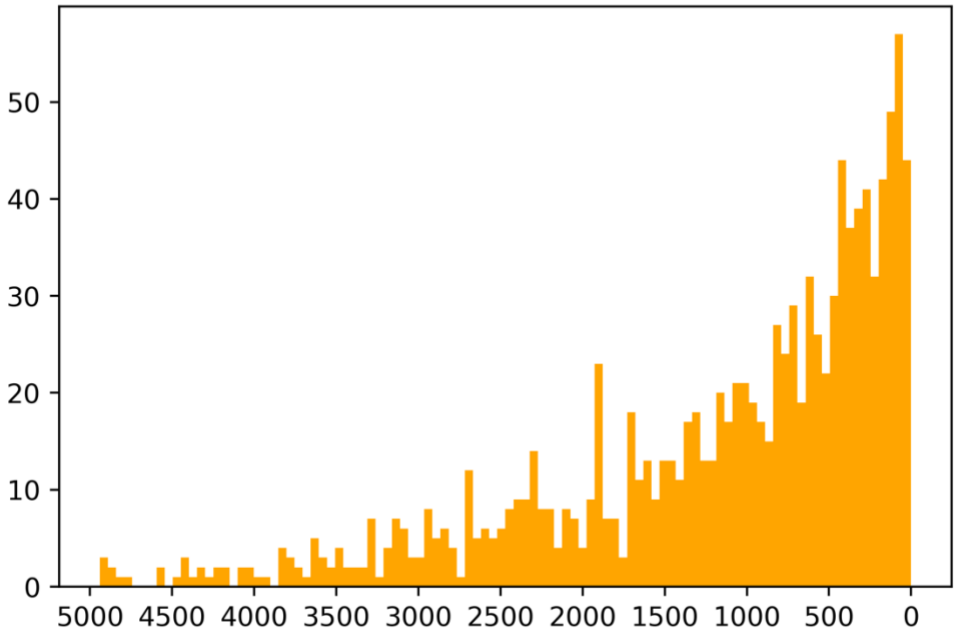


**Figure 22. Histogram of TF binding peaks across all 6 TF factors of interest.**
Position 1 represent 1 nucleotide upstream of the protein-coding region. Ranges to 5,000 bp upstream. The general trend of the histogram shows highest number of TF binding locations closer to the protein-coding regions start site and decreases further from the start site. The highest peaks are found between 50-100 bp upstream of the protein-coding region start site.

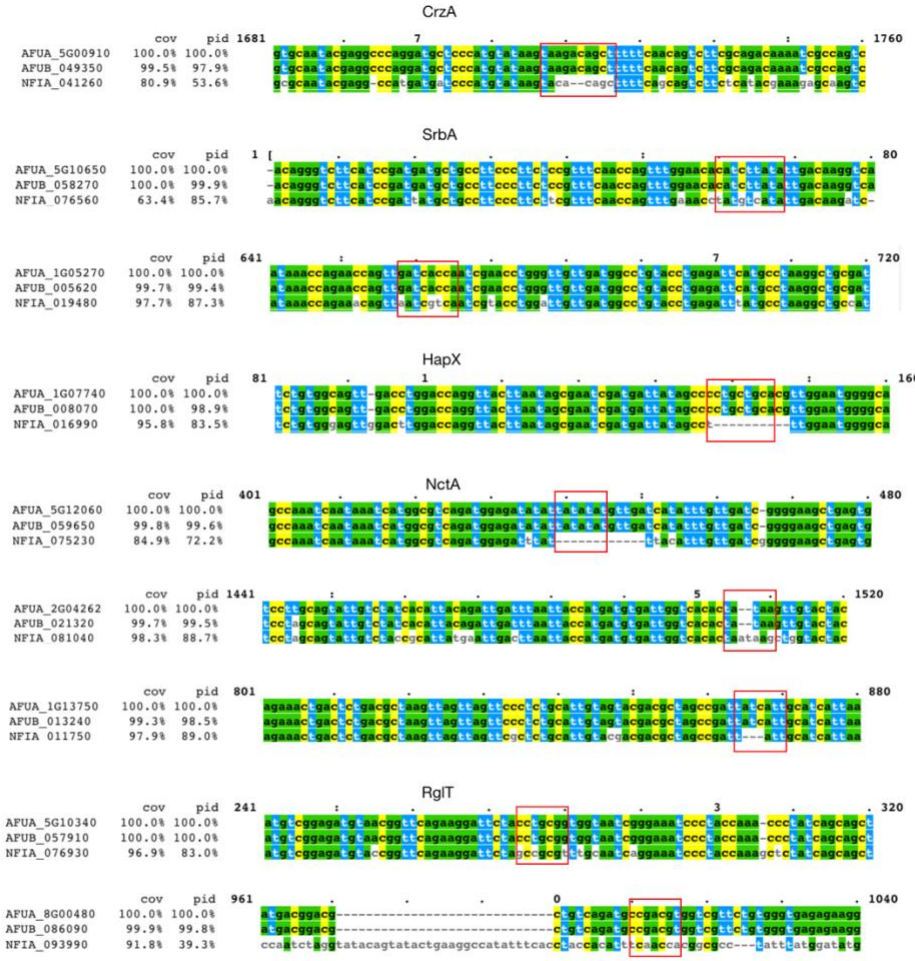



**Figure 23. Sequence alignments detailing differences at TF binding site locations.**
Red boxes represent the binding site locations found in previous ChIP-seq experiments.
We note 9 different binding location of interest. For each of these 9 binding locations there are differences in the sequence between *A. fumigatus* and *A. fischeri*. AFUA_5G00910 (CrzA), AFUA_1G07740 (SrbA), AFUA_5G12060 (NctA) and AFUA_1G13750 (NctA) exhibit a region that has been inserted in *A. fumigatus* and is absent from all other species. Sequence alignment of the non-coding region of AFUA_2G04262 (NctA) exhibits a 7 bp region (red box) which contains 3 bp that are uniquely deleted in *A. fumigatus* but present in *A. fischeri*. Sequence

alignments of the non-coding region of AFUA_5G10650 (SrbA), AFUA_1G05270 (SrbA), AFUA_5G10340 and AFUA_8G00480 all exhibit several bp differences between *A. fumigatus* and *A. fischeri.* Colored nucleotides represent sites that are present in *A. fumigatus* af293 and shared between other species, highlighting differences between the *A. fumigatus* strains and *A. fischeri*.

## Results and Discussion

Understanding the regulatory mechanisms governing gene expression is crucial for deciphering the complex biology of organisms. In this study, we investigated the binding site distribution of six key transcription factors in *Aspergillus fumigatus* using publicly available ChIP-seq data. Through a comprehensive analysis of binding site locations, we identified intriguing patterns that shed light on the regulatory landscape of this fungal pathogen.

We curated binding site locations from ChIP-seq data corresponding to six *A. fumigatus* transcription factors. Each binding site location was annotated by its distance upstream of the target gene. Constructing histograms based on these distances provided insights into the distribution of binding sites across the genome. Notably, a consistent trend emerged, wherein the number of binding locations decreased as the distance from the target gene increased. Within a proximity of 5000 bp from their respective genes, we found a total of 1182 binding locations. Of these, 416 were situated within a mere 500 bp, indicating the prevalence of proximal binding events. Additionally, 812 binding sites were identified within 1500 bp of the target genes, further emphasizing the significance of close-range regulatory interactions.

Upon closer examination of the binding site distribution, a pronounced peak emerged between 50-100 bp upstream of the genes, comprising 59 binding locations. Intriguingly, beyond the 1500 bp threshold, two distinct peaks garnered attention: one around 1700-1750 bp and another around 1900-1950 bp. Remarkably, the majority of binding locations within the 1700-1750 bp peak were attributed to the HapX transcription factor, suggesting a potential unknown regulatory role for this factor at this range. The peak centered at 1900-1950 bp exhibited a diverse representation, involving transcription factors CrzA, SrbA, HapX, and AtrR. These observed patterns of binding site localization offer introductory insights into the regulatory hierarchy of these factors and their potential roles in orchestrating gene expression.

To unravel species-specific regulatory mechanisms, we conducted a comprehensive investigation into TF binding motifs in *Aspergillus fumigatus* af293, *A. fumigatus* a1163, and *A. fischeri*. Through multiple sequence alignments of upstream regions containing TF binding locations, we employed FIMO analysis to identify binding motifs for six key transcription factors. A breakdown of genes with binding locations and the identification of candidate regions were undertaken. Our findings revealed 53 genes displaying variations in TF binding motifs between *A. fumigatus* and *A. fischeri*. Subsequent scrutiny identified nine genes with distinct binding motifs in specific upstream regions, highlighting potential species-specific regulatory events. We initiated our analysis by collating binding locations for six transcription factors across the three *Aspergillus* strains. Multiple sequence alignments were executed on upstream regions encompassing binding sites situated up to 1500 bp from the gene start. Leveraging the FIMO

tool, we computationally identified binding motifs in alignment with known TF binding motifs for each factor.

Our investigation yielded 53 genes with notable differences in TF binding motifs between *A. fumigatus* and *A. fischeri*. The distribution among the transcription factors included 3 genes for CrzA, 6 genes for SrbA, 2 genes for RglT, 17 genes for HapX, 22 genes for NctA, and 3 genes for AtrR. Although intriguing, the presence of multiple motifs within candidate regions raised the possibility of diverse binding strategies by TFs in the two species.

To focus on the most compelling results, we scrutinized cases where binding motifs differed specifically in the region unique to one species. This meticulous examination revealed nine genes exemplifying such attributes. Specifically, these genes comprised 1 for CrzA, 2 for SrbA, 2 for RglT, 1 for HapX, and 3 for NctA, while AtrR did not exhibit this phenomenon. By identifying genes with distinct motifs in specific upstream regions, we provide insights into the potential divergence of regulatory mechanisms between *A. fumigatus* and *A. fischeri*. These findings contribute to the understanding of species-specific gene regulation and offer a foundation for further investigations into the functional implications of these regulatory variations.

In our investigation of candidate regions across *A. fumigatus* and *A. fischeri*, we focused on comparing the binding site locations for the identified transcription factor binding motifs. Among the 9 candidate regions, we observed distinct patterns that could be categorized into three main groups.

Category 1: Unique Motifs in *A. fumigatus*. Within this category, we found four examples where the transcription factor binding motif was present in *A. fumigatus* but absent in *A. fischeri*. Specifically, these regions belonged to genes AFUA_5G00910 (CrzA), AFUA_1G07740 (SrbA), AFUA_5G12060 (NctA), and AFUA_1G13750 (NctA). Notably, each of these genes exhibited a binding peak at specific distances upstream of the gene start.

- AFUA_5G00910 : This gene encodes a MBL2-like secreted peptide. Its binding peak was found at 86 bp upstream of the gene start. The presence of a distinct binding motif in *A. fumigatus* suggests potential species-specific regulation of this gene.
- AFUB_004360: This gene encodes an uncharacterized protein. The binding peak was identified at 748 bp upstream. The significance of this motif in relation to the gene's function warrants further investigation.
- AFUB_008070: Encoding a LEM3/CDC50 family protein, this gene displayed a binding peak at 1093 bp upstream. The regulatory implications of this binding event require deeper exploration.
- AFUB_059650: This gene encodes a putative C2H2 transcription factor and displayed a binding peak at 550 bp upstream. The presence of a unique binding motif in *A. fumigatus* hints at specialized regulatory mechanisms.

Category 2: Variations in Binding Site Location. In the second category, we identified four examples where there were several base pair differences between the binding site locations in *A. fumigatus* and *A. fischeri*. These genes were AFUA_5G10650 (SrbA), AFUA_8G00480 (RglT), AFUB_021320 (HapX), and AFUB_013240 (HapX).

- AFUA_5G10650: This gene encodes a putative pyridoxamine phosphate oxidase, with a binding peak at 105 bp upstream. The differential binding site location between the two species suggests potential variations in regulatory networks.
- AFUA_8G00480: This gene encodes a phytanoyl-CoA dioxygenase family protein and had a binding peak at 335 bp upstream. The observed differences in binding site locations could point to divergent gene regulation mechanisms.
- AFUA_5G10340: Encoding a putative MFS transporter, this gene exhibited a binding peak at 301 bp upstream. The differences in binding sites may have implications for transporter gene regulation.
- AFUB_021320: This gene encodes a C6 transcription factor and had a binding peak at 22 bp upstream. The presence of distinct binding locations in the two species suggests potential differences in regulatory logic.
- AFUB_013240: Encoding Rpn4, a C2H2 transcription factor, this gene exhibited a binding peak at 751 bp upstream. Variations in binding site location could impact the gene's regulatory role.

Category 3: Insertion-Induced Binding Site Shift. In the final category, we identified one instance where an insertion in *A. fischeri* caused a difference in the binding site location for the gene AFUB_059650 (HapX). This gene encodes a putative C2H2 transcription factor and displayed a binding peak at 550 bp upstream. It remains unclear how an insertion may play a role in differences in HapX binding opportunity for this gene. These findings reveal intricate differences in transcription factor binding motifs and their corresponding site locations between *A. fumigatus* and *A. fischeri*. These variations suggest species-specific regulatory mechanisms, potentially influencing gene expression patterns and ultimately contributing to the distinct biology of each species. Further investigations into the functional implications of these differences are warranted to unravel the underlying regulatory networks.

In this investigation of transcription factor binding patterns across *A. fumigatus* and *A. fischeri*, we have unveiled novel insights into the regulatory landscape of these fungal species. By curating and analyzing binding site locations for six transcription factors, we uncovered distinct trends in binding motif distribution and site locations, shedding light on the intricate mechanisms governing gene expression. The comparative analysis of the candidate regions has revealed a diverse array of species-specific regulatory events that encompass both the presence and absence of transcription factor binding motifs, as well as variations in binding site distances.

Our findings underscore the dynamic nature of gene regulation within and across fungal species, offering a deeper understanding of how transcription factors orchestrate gene expression in diverse environmental contexts. The identification of genes with unique binding motifs and altered binding site locations between species implies the existence of fine-tuned regulatory strategies that contribute to the phenotypic differences observed. This research paves the way for further investigations into the functional consequences of these regulatory variations and opens new avenues for studying fungal gene regulation networks. Ultimately, our study contributes to the broader field of genomics by providing valuable insights into the molecular basis of species-specific gene expression and the intricate interplay between transcription factors and their target genes.

# Chapter 5: Conclusion

## Future Directions

This thesis outlines several non-coding regions which show a unique distinction between *A. fumigatus* and *A. fischeri*. Phenotypic validation of these candidate sites would be the first, and likely most important, next step. For an experimental design, the first step would be to examine TF binding site presence/absence between *A. fumigatus* and *A. fischeri* by examining the noncoding regions upstream of a direct gene target for each of the six TFs known to play a role in *A. fumigatus* virulence (as described in Chapter 4). I would prioritize CrzA and SrbA first since their involvement in *A. fumigatus* virulence is well described. I would start with non-coding region differences upstream of genetic determinants of virulence found in Chpater 2 such as pkaR, gliG, metR, then move on to any additional non-coding target regions. For direct target genes showing the biggest TF binding site differences between *A. fumigatus* and *A. fischeri*, RT-qPCR would be used to confirm that expression of the gene differs between the two species. Then, I would select the direct gene target gene showing the largest difference in gene expression between the two species. For this gene, I would engineer a mutant strain of *A. fumigatus* Af293 whose TF binding motif sequence for the gene of interest has been replaced with the positionally equivalent sequence from *A. fischeri* using a set of previously established procedures[57-58]. I would use this same procedure to engineer mutant strains of *A. fischeri* by replacing the native noncoding sequence with the TF binding site-containing sequence from *A. fumigatus*. *A. fumigatus* and *A. fischeri* engineered strains would be grown in triplicate. Two growth conditions would be implemented. The first would grow in both liquid and solid complete media at 37°C and pH 6.5. The second would grow in the same conditions but would also be exposed to conditions relevant to biological function of the TF of interest. For example, *crzA* is upregulated when *A. fumigatus* undergoes osmotic stress, so it would make sense to grow the strains under osmotic stress conditions (e.g., by treating strains with 1.0 M sorbitol for 60 min[59]). I would then use RT-qPCR to compare differences in gene expression *in vitro* between the engineered and wild-type (WT) strains of the two species in the two conditions using previously established methods[60].

For direct gene targets that differ in gene expression between the mutant and WT strains, I would test for differences in virulence using the insect *Galleria mellonella* model of fungal disease. *Galleria* is well established for studies of *Aspergillus* virulence[18,61], is cost effective, and many insects can be tested quickly. Briefly, spores of mutant or WT strains would be harvested in phosphate-buffered saline (PBS) and filtered through Miracloth (Calbiochem) after 2 days of growth in their respective growth conditions. Spore concentration would be estimated using a hemocytometer, and spore viability assessed through incubation on YAG medium at 37°C for 48 h. 20 *Galleria* larvae in the final instar larval stage of development would be used per condition. Each larva in the test group would be infected with a 5µl inoculum of spores from the mutant strain

(at a concentration of $1 \times 10^6$ spores/µl), whereas each larva in the control group would be inoculated with the same concentration of spores from the WT strain. The negative control would be inoculated with PBS (no spores). All injections would be performed at the hemocoel of each larva via the last left proleg. Following inoculation, all larvae would be incubated in glass petri dishes in darkness at 37°C, with larval killing scored daily. Statistical significance between survival curves of strains would be determined using Mantel-Cox log rank and Gehan-Breslow-Wilcoxon tests.

Additionally, we have yet to characterize other non-coding regions across *A. fumigatus* and other species / strains, particularly intronic sequences and downstream regions. For this, I would recommend starting with whole genome sequences and gene annotations and using the python scripts I have developed to parse and align these non-coding regions. Next, I would recommend creating multiple sequence alignments and performing $dN_{noncoding}/dS$ as described in Chapter 2. Here, it is likely we would find additional candidate sites for further exploration, including some genetic determinants of virulence.

## Outstanding Questions

While the results shown here provide a framework for better understanding non-coding regions across *A. fumigatus* and close relatives, there remain many questions that are unelucidated or not directly answered within the dissertations. Below are short lists of outstanding questions, both because they remain unanswered in the dissertation and because they are good questions, worthy of further explorations.

1. Population Structure:

With the increase in number of sequenced *A. fumigatus* strains, there has been an increased interest in how population dynamics may impact genomic differences. As of the time of this writing there are 291 *A. fumigatus* strains that are publicly available and over 300 reported. Of these, the majority of strains were isolated from Spain, Japan and the UK. Comparison of genes across a large set of *A. fumigatus* strains found several instances of strains isolated from distinct countries being more closely related to strains isolated from other countries as opposed to like countries (Barber et al., 2021, Lofgren et al., 2022). Though not shown here, we were able to verify these patterns in my own phylogenetic inferences comparing protein-coding regions across the 263 *A. fumigauts* strains presented in Chapter 3. This also holds true for non-coding regions, where we find additional strains being more closely related to strains isolated from distinct populations. The degree to which population structure plays a role in *A. fumigatus* history remains unclear. It will be interesting to see if this pattern holds true as more *A. fumigatus* strains, from more distinct countries, are isolated and sequenced.

2. Clinical vs Environmental

As an opportunistic infectious agent, there has been an interest (historically) in trying to better understand which *A. fumigatus* pathogenic features/gene expression evolved in their environment (soil, compost piles) and which are prominent with the human host.

Of the 300 *A. fumigatus* species sequenced for Barber et al., 2021, 217 are clinical isolated and 87 are environmental. The imbalance in the number of isolates between clinical and environmental makes it difficult to perform a true comparisons between the two groups. However, when generating phylogenetic trees across protein-coding and non-coding regions, we find several instances of clinical and environmental strains being more closely related between groups as opposed to closer to each other. This was also found between core genes across *A. fumigatus* strains (Barber et al., 2021, Lofgren et al., 2022).

3. Trans Effects

The regulation of genes is, in large part, due to a combination of both cis-regulatory and trans-regulatory elements. Cis regulatory elements are thought to be non-coding motifs that aid in binding transcription factors as either promoters or enhancers. In Chapter 4, we identify that the majority of transcription factor binding occurs with 1500bp of a gene's protein-coding region. While this may not necessarily be true across the >400 transcription factors present in *A. fumigatus*, this is the case for all the transcription factors whose binding target site have been elucidated. What is less known is the role that trans-regulatory elements play in *A. fumigatus* gene expression.

Trans-regulatory elements, unlike cis-regulatory elements, do not neccesarily require sequence specific motifs and can arise, in theory, from anywhere else in the genome. This makes pinpointing trans effects a difficult task, especially without a wealth of previous knowledge and established procedures to identify trans-effects, which is currently the case for *Aspergillus*. However, there have been several methodologies developed to characterize and measure trans-regulatory impacts across genes. One of these methods involves comparative genomics between two or more test groups. For example, one study compared the difference in cellular environments (tissue types) between mice and human cells. Since the genes that were focused on exhibited similar protein-coding and non-coding region sequence similarity, differences in gene expression were attributed to differences in trans regulatory elements, though what those exact elements were are not fully explained (Mattioli et al., 2020).

Other methodologies for identifying trans-regulatory elements include eQTL mapping (genome-wide) and GWAS (compare the genomes of many individuals), which involve identifying regions of the genome that contribute to variation in gene expression levels. These methods correlate genetic variation with gene expression levels across a population, helping to identify trans-regulatory elements. However, this methods, while powerful, cannot alone identify which specific trans-regulatory elements impact gene expression. RNA-seq can be used to measure gene expression levels across the genome, and by over-expressing a certain transcription factor, can use increase or decreases in gene expression across the genome as a proxy for which genes are affected by the expression of that transcription factor. But this alone does not determine whether these are due to cis or trans regulation.

One method, and potential future aim, for elucidating which proteins act as trans-regulators would be to focus on a singular *A. fumigatus* pathway. For this one pathway, we could use a combination of Chip-Seq, mass spectrometry and co-immunoprecipitation/pull down assays to

identifying binding site location for a specific protein for a well-known pathway. We can overexpress said protein and use mass spectrometry. to help identify other proteins that are expressed in the pathway, then use Immunoprecipitation to isolate these proteins along with their interaction partners.

## Conclusion

*Aspergillus fumigatus* is one of the deadliest fungal pathogens, responsible for a severe fungal disease, invasive aspergillosis, afflicting more than 400,000 patients globally each year. While there are roughly 450 species under the Aspergillus genus, only a handful have been identified as clinically pertinent, with *A. fumigatus* taking the lead, causing >90% of all IA cases. Historically, genomic comparisons between *A. fumigatus* and closely related species/strains, have often centered on gene presence/absence. However, the majority of *A. fumigatus* genes, including those that influence its virulence, can be found in both pathogenic and non-pathogenic species. This observation has led to the following hypothesis: There is abundant genetic variation in noncoding regions among *A. fumigatus* strains and between *Aspergillus* species; this variation contributes to the observed differences in pathogenicity.

The studies presented in this thesis make substantial headway in providing the first, whole genomic comparisons of non-coding regions between *A. fumigatus* and close relatives. Comparisons between non-coding regions, specifically those upstream of genes' initial codons, have unveiled significant sequence variations and an absence of homology across *Aspergillus* species. Interestingly, even among *A. fumigatus* strains, such as the highly virulent A1163 and the less virulent Af293, non-coding regions displayed greater sequence diversity than their corresponding protein-coding regions. These genetic distinctions suggest a profound role for these non-coding regions in gene regulation, potentially contributing to the phenotypic disparities observed between different *A. fumigatus* strains.

Supplementing these findings, the exploration of transcriptional regulatory mechanisms, particularly the roles of transcription factors (TFs) and non-coding regions upstream, has opened a new avenue for understanding the *A. fumigatus* pathogenesis. TFs, as regulatory switches, have been observed to bind to specific DNA motifs in gene promoter regions, dictating the activation or suppression of gene transcription. These proteins play an instrumental role in managing gene expressions critical for various facets of the fungal lifecycle, including its virulence, metabolism, and interactions with the host. Moreover, the presence of vital cis-regulatory elements within the upstream non-coding regions further underscores the intricate relationship between gene regulation and fungal pathogenicity. Comparisons between the TF binding sites across reference strains *of A. fumigatus* and its close, non-pathogenic relative, *A. fischeri*, have spotlighted a subset of genes with upstream binding targets that differ between the species. Such variations in TF binding locations may indeed be pivotal in elucidating the differential pathogenic profiles across these Aspergillus strains and species.

While the genomic composition and presence of certain genes undeniably plays a role in the pathogenesis of *A. fumigatus*, the intricate dance of gene regulation, mediated by non-coding regions and transcription factors, emerges as a potential cornerstone in understanding its virulence. Future research geared towards decoding these regulatory mechanisms promises not

only a deeper comprehension of fungal pathogenesis but also paves the way for better understanding how non-coding regions influence differences in gene regulation.

## References


Abad, A., Fernández-Molina, J. V., Bikandi, J., Ramírez, A., Margareto, J., Sendino, J., et al. (2010). What makes Aspergillus fumigatus a successful pathogen? Genes and molecules involved in invasive aspergillosis. *Rev Iberoam Micol* 27, 155–182. doi: 10.1016/j.riam.2010.10.003.

Abdulrehman, D., Monteiro, P. T., Teixeira, M. C., Mira, N. P., Lourenço, A. B., dos Santos, S. C., et al. (2011). YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in Saccharomyces cerevisiae through a web services interface. *Nucleic Acids Res* 39, D136–D140. doi: 10.1093/nar/gkq964.

Alastruey-Izquierdo, A., Alcazar-Fuoli, L., and Cuenca-Estrella, M. (2014). Antifungal Susceptibility Profile of Cryptic Species of Aspergillus. *Mycopathologia* 178, 427–433. doi: 10.1007/s11046-014-9775-z.

Alastruey-Izquierdo, A., Cadranel, J., Flick, H., Godet, C., Hennequin, C., Hoenigl, M., et al. (2018). Treatment of Chronic Pulmonary Aspergillosis: Current Standards and Future Perspectives. *Respiration* 96, 159–170. doi: 10.1159/000489474.

Alastruey-Izquierdo, A., Mellado, E., Peláez, T., Pemán, J., Zapico, S., Alvarez, M., et al. (2013). Population-Based Survey of Filamentous Fungi and Antifungal Resistance in Spain (FILPOP Study). *Antimicrobial Agents and Chemotherapy* 57, 3380–3387. doi: 10.1128/aac.00383-13.

Amich, J., Dümig, M., O'Keeffe, G., Binder, J., Doyle, S., Beilhack, A., et al. (2016). Exploration of Sulfur Assimilation of Aspergillus fumigatus Reveals Biosynthesis of Sulfur-Containing Amino Acids as a Virulence Determinant. *Infect Immun* 84, 917–929. doi: 10.1128/IAI.01124-15.

Amich, J., Schafferer, L., Haas, H., and Krappmann, S. (2013). Regulation of Sulphur Assimilation Is Essential for Virulence and Affects Iron Homeostasis of the Human-Pathogenic Mould Aspergillus fumigatus. *PLOS Pathogens* 9, e1003573. doi: 10.1371/journal.ppat.1003573.

Amich, J., Vicentefranqueira, R., Mellado, E., Ruiz-Carmuega, A., Leal, F., and Calera, J. A. (2014). The ZrfC alkaline zinc transporter is required for Aspergillus fumigatus virulence and its growth in the presence of the Zn/Mn-chelating protein calprotectin. *Cellular Microbiology* 16, 548–564. doi: 10.1111/cmi.12238.

Arastehfar, A., Carvalho, A., Houbraken, J., Lombardi, L., Garcia-Rubio, R., Jenks, J. D., et al. (2021). Aspergillus fumigatus and aspergillosis: From basics to clinics. *Stud Mycol* 100, 100115. doi: 10.1016/j.simyco.2021.100115.

Azie, N., Neofytos, D., Pfaller, M., Meier-Kriesche, H.-U., Quan, S.-P., and Horn, D. (2012). The PATH (Prospective Antifungal Therapy) Alliance® registry and invasive fungal infections: update 2012. *Diagn Microbiol Infect Dis* 73, 293–300. doi: 10.1016/j.diagmicrobio.2012.06.012.

Bader, O., Weig, M., Reichard, U., Lugert, R., Kuhns, M., Christner, M., et al. (2013). cyp51A-Based Mechanisms of Aspergillus fumigatus Azole Drug Resistance Present in Clinical Samples from Germany. *Antimicrob Agents Chemother* 57, 3513–3517. doi: 10.1128/AAC.00167-13.

Badis, G., Chan, E. T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol Cell* 32, 878–887. doi: 10.1016/j.molcel.2008.11.020.

Bandres, M. V., Modi, P., and Sharma, S. (2023). "Aspergillus Fumigatus," in *StatPearls* (Treasure Island (FL): StatPearls Publishing). Available at: http://www.ncbi.nlm.nih.gov/books/NBK482464/ [Accessed September 11, 2023].

Barber, A. E., Sae-Ong, T., Kang, K., Seelbinder, B., Li, J., Walther, G., et al. (2021). Aspergillus fumigatus pan-genome analysis identifies genetic variants associated with human infection. *Nat Microbiol* 6, 1526–1536. doi: 10.1038/s41564-021-00993-x.

Barker, B. M., Kroll, K., Vödisch, M., Mazurie, A., Kniemeyer, O., and Cramer, R. A. (2012). Transcriptomic and proteomic analyses of the Aspergillus fumigatus hypoxia response using an oxygen-controlled fermenter. *BMC Genomics* 13, 62. doi: 10.1186/1471-2164-13-62.

Beattie, S. R., Mark, K. M. K., Thammahong, A., Ries, L. N. A., Dhingra, S., Caffrey-Carr, A. K., et al. (2017). Filamentous fungal carbon catabolite repression supports metabolic plasticity and stress responses essential for disease progression. *PLoS Pathog* 13, e1006340. doi: 10.1371/journal.ppat.1006340.

Beauvais, A., Bozza, S., Kniemeyer, O., Formosa, C., Balloy, V., Henry, C., et al. (2013). Deletion of the α-(1,3)-glucan synthase genes induces a restructuring of the conidial cell wall responsible for the avirulence of Aspergillus fumigatus. *PLoS Pathog* 9, e1003716. doi: 10.1371/journal.ppat.1003716.

Beauvais, A., Fontaine, T., Aimanianda, V., and Latgé, J.-P. (2014). Aspergillus cell wall and biofilm. *Mycopathologia* 178, 371–377. doi: 10.1007/s11046-014-9766-0.

Ben-Ami, R., Lewis, R. E., Leventakos, K., and Kontoyiannis, D. P. (2009). Aspergillus fumigatus inhibits angiogenesis through the production of gliotoxin and other secondary metabolites. *Blood* 114, 5393–5399. doi: 10.1182/blood-2009-07-231209.

Benedict, K., Jackson, B. R., Chiller, T., and Beer, K. D. (2019). Estimation of direct healthcare costs of fungal diseases in the United States. *Clin Infect Dis* 68, 1791–1797. doi: 10.1093/cid/ciy776.

Bertuzzi, M., Hayes, G. E., Icheoku, U. J., van Rhijn, N., Denning, D. W., Osherov, N., et al. (2018). Anti-Aspergillus Activities of the Respiratory Epithelium in Health and Disease. *J Fungi (Basel)* 4, E8. doi: 10.3390/jof4010008.

Bhabhra, R., and Askew, D. S. (2005). Thermotolerance and virulence of Aspergillus fumigatus: role of the fungal nucleolus. *Med Mycol* 43 Suppl 1, S87-93. doi: 10.1080/13693780400029486.

Blatzer, M., Barker, B. M., Willger, S. D., Beckmann, N., Blosser, S. J., Cornish, E. J., et al. (2011). SREBP Coordinates Iron and Ergosterol Homeostasis to Mediate Triazole Drug and Hypoxia Responses in the Human Fungal Pathogen Aspergillus fumigatus. *PLoS Genet* 7, e1002374. doi: 10.1371/journal.pgen.1002374.

Bongomin, F., Gago, S., Oladele, R. O., and Denning, D. W. (2017). Global and Multi-National Prevalence of Fungal Diseases—Estimate Precision. *Journal of Fungi* 3, 57. doi: 10.3390/jof3040057.

Brakhage, A. A., and Langfelder, K. (2002). Menacing mold: the molecular biology of Aspergillus fumigatus. *Annu Rev Microbiol* 56, 433–455. doi: 10.1146/annurev.micro.56.012302.160625.

Bromley, M., Johns, A., Davies, E., Fraczek, M., Gilsenan, J. M., Kurbatova, N., et al. (2016). Mitochondrial Complex I Is a Global Regulator of Secondary Metabolism, Virulence and Azole Sensitivity in Fungi. *PLOS ONE* 11, e0158724. doi: 10.1371/journal.pone.0158724.

Brown, A., Mead, M. E., Steenwyk, J. L., Goldman, G. H., and Rokas, A. (2022). Extensive non-coding sequence divergence between the major human pathogen Aspergillus fumigatus and its relatives. *Front Fungal Biol* 3, 802494. doi: 10.3389/ffunb.2022.802494.

Brown, A., Steenwyk, J. L., and Rokas, A. (2024). Genome-wide patterns of non-coding sequence variation in the major fungal pathogen Aspergillus fumigatus. 2024.01.08.574724. doi: 10.1101/2024.01.08.574724.

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012). Hidden Killers: Human Fungal Infections. *Science Translational Medicine* 4, 165rv13-165rv13. doi: 10.1126/scitranslmed.3004404.

Brown, N. A., and Goldman, G. H. (2016). The contribution of Aspergillus fumigatus stress responses to virulence and antifungal resistance. *J Microbiol* 54, 243–253. doi: 10.1007/s12275-016-5510-4.
Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12, 59–60. doi: 10.1038/nmeth.3176.
Cadena, J., Thompson, G. R., and Patterson, T. F. (2021). Aspergillosis: Epidemiology, Diagnosis, and Treatment. *Infect Dis Clin North Am* 35, 415–434. doi: 10.1016/j.idc.2021.03.008.
Caron, B., Luo, Y., and Rausell, A. (2019). NCBoost classifies pathogenic non-coding variants in Mendelian diseases through supervised learning on purifying selection signals in humans. *Genome Biol* 20, 32. doi: 10.1186/s13059-019-1634-2.
Carrion, S. de J., Leal, S. M., Ghannoum, M. A., Aimanianda, V., Latgé, J.-P., and Pearlman, E. (2013). The RodA hydrophobin on Aspergillus fumigatus spores masks dectin-1- and dectin-2-dependent responses and enhances fungal survival in vivo. *J Immunol* 191, 2581–2588. doi: 10.4049/jimmunol.1300748.
Carroll, S. B. (2005). Evolution at Two Levels: On Genes and Form. *PLoS Biol* 3, e245. doi: 10.1371/journal.pbio.0030245.
Carroll, S. B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* 134, 25–36. doi: 10.1016/j.cell.2008.06.030.
Casadevall, A. (2006). Cards of virulence and the global virulome for humans. *ASM NEWS* 1, 359–364. doi: 10.1128/microbe.1.359.1.
Castro, P. A. de, Colabardini, A. C., Moraes, M., Horta, M. A. C., Knowles, S. L., Raja, H. A., et al. (2022). Regulation of gliotoxin biosynthesis and protection in Aspergillus species. *PLOS Genetics* 18, e1009965. doi: 10.1371/journal.pgen.1009965.
Cerqueira, G. C., Arnaud, M. B., Inglis, D. O., Skrzypek, M. S., Binkley, G., Simison, M., et al. (2014). The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic Acids Res* 42, D705-710. doi: 10.1093/nar/gkt1029.
Chakrabarti, A., Kaur, H., Savio, J., Rudramurthy, S. M., Patel, A., Shastri, P., et al. (2019). Epidemiology and clinical outcomes of invasive mould infections in Indian intensive care units (FISF study). *Journal of Critical Care* 51, 64–70. doi: 10.1016/j.jcrc.2019.02.005.
Chang, P.-K. (2008). Aspergillus parasiticus crzA, which encodes calcineurin response zinc-finger protein, is required for aflatoxin production under calcium stress. *Int J Mol Sci* 9, 2027–2043. doi: 10.3390/ijms9102027.
Chelstowska, A., and Butow, R. A. (1995). RTG genes in yeast that function in communication between mitochondria and the nucleus are also required for expression of genes encoding peroxisomal proteins. *J Biol Chem* 270, 18141–18146. doi: 10.1074/jbc.270.30.18141.
Chotirmall, S. H., Al-Alawi, M., Mirkovic, B., Lavelle, G., Logan, P. M., Greene, C. M., et al. (2013). Aspergillus-Associated Airway Disease, Inflammation, and the Innate Immune Response. *Biomed Res Int* 2013, 723129. doi: 10.1155/2013/723129.
Chung, D., Barker, B. M., Carey, C. C., Merriman, B., Werner, E. R., Lechner, B. E., et al. (2014). ChIP-seq and in vivo transcriptome analyses of the Aspergillus fumigatus SREBP SrbA reveals a new regulator of the fungal hypoxia response and virulence. *PLoS Pathog* 10, e1004487. doi: 10.1371/journal.ppat.1004487.
Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163.

Colabardini, A. C., Wang, F., Dong, Z., Pardeshi, L., Rocha, M. C., Costa, J. H., et al. (2021). Heterogeneity in the transcriptional response of the human pathogen Aspergillus fumigatus to the antifungal agent caspofungin. doi: 10.1101/2021.07.15.452449.
Cormier, L., Barbey, R., and Kuras, L. (2010). Transcriptional plasticity through differential assembly of a multiprotein activation complex. *Nucleic Acids Res* 38, 4998–5014. doi: 10.1093/nar/gkq257.
Cramer, R. A., Perfect, B. Z., Pinchai, N., Park, S., Perlin, D. S., Asfaw, Y. G., et al. (2008a). Calcineurin target CrzA regulates conidial germination, hyphal growth, and pathogenesis of Aspergillus fumigatus. *Eukaryot Cell* 7, 1085–1097. doi: 10.1128/EC.00086-08.
Cramer, R. A., Perfect, B. Z., Pinchai, N., Park, S., Perlin, D. S., Asfaw, Y. G., et al. (2008b). Calcineurin Target CrzA Regulates Conidial Germination, Hyphal Growth, and Pathogenesis of Aspergillus fumigatus. *Eukaryotic Cell* 7, 1085–1097. doi: 10.1128/ec.00086-08.
Croft, C. A., Culibrk, L., Moore, M. M., and Tebbutt, S. J. (2016). Interactions of Aspergillus fumigatus Conidia with Airway Epithelial Cells: A Critical Review. *Front Microbiol* 7, 472. doi: 10.3389/fmicb.2016.00472.
Dagenais, T. R. T., and Keller, N. P. (2009). Pathogenesis of Aspergillus fumigatus in Invasive Aspergillosis. *Clin Microbiol Rev* 22, 447–465. doi: 10.1128/CMR.00055-08.
de Castro, P. A., Chen, C., de Almeida, R. S. C., Freitas, F. Z., Bertolini, M. C., Morais, E. R., et al. (2014). ChIP-seq reveals a role for CrzA in the Aspergillus fumigatus high-osmolarity glycerol response (HOG) signalling pathway. *Mol Microbiol* 94, 655–674. doi: 10.1111/mmi.12785.
De Pauw, B., Walsh, T. J., Donnelly, J. P., Stevens, D. A., Edwards, J. E., Calandra, T., et al. (2008). Revised definitions of invasive fungal disease from the European Organization for Research and Treatment of Cancer/Invasive Fungal Infections Cooperative Group and the National Institute of Allergy and Infectious Diseases Mycoses Study Group (EORTC/MSG) Consensus Group. *Clin Infect Dis* 46, 1813–1821. doi: 10.1086/588660.
de Vries, R. P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C. A., et al. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. *Genome Biology* 18, 28. doi: 10.1186/s13059-017-1151-0.
Denning, D. W., Cadranel, J., Beigelman-Aubry, C., Ader, F., Chakrabarti, A., Blot, S., et al. (2016). Chronic pulmonary aspergillosis: rationale and clinical guidelines for diagnosis and management. *Eur Respir J* 47, 45–68. doi: 10.1183/13993003.00583-2015.
Dewi, I. M. W., van de Veerdonk, F. L., and Gresnigt, M. S. (2017). The Multifaceted Role of T-Helper Responses in Host Defense against Aspergillus fumigatus. *J Fungi (Basel)* 3, 55. doi: 10.3390/jof3040055.
Dhingra, S., and Cramer, R. A. (2017). Regulation of Sterol Biosynthesis in the Human Fungal Pathogen Aspergillus fumigatus: Opportunities for Therapeutic Development. *Frontiers in Microbiology* 8. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.00092 [Accessed August 13, 2023].
Dolan, S. K., O'Keeffe, G., Jones, G. W., and Doyle, S. (2015). Resistance is not futile: gliotoxin biosynthesis, functionality and utility. *Trends Microbiol* 23, 419–428. doi: 10.1016/j.tim.2015.02.005.
dos Santos, R. A. C., Steenwyk, J. L., Rivero-Menendez, O., Mead, M. E., Silva, L. P., Bastos, R. W., et al. (2020). Genomic and Phenotypic Heterogeneity of Clinical Isolates of the Human Pathogens Aspergillus fumigatus, Aspergillus lentulus, and Aspergillus fumigatiaffinis.

*Frontiers in Genetics* 11. Available at: https://www.frontiersin.org/article/10.3389/fgene.2020.00459 [Accessed June 20, 2022].
Eide, D. J. (2020). Transcription factors and transporters in zinc homeostasis: lessons learned from fungi. *Crit Rev Biochem Mol Biol* 55, 88–110. doi: 10.1080/10409238.2020.1742092.
Emerson, J. J., Hsieh, L.-C., Sung, H.-M., Wang, T.-Y., Huang, C.-J., Lu, H. H.-S., et al. (2010). Natural selection on cis and trans regulation in yeasts. *Genome Res* 20, 826–836. doi: 10.1101/gr.101576.109.
Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16, 157. doi: 10.1186/s13059-015-0721-2.
Fedorova, N. D., Khaldi, N., Joardar, V. S., Maiti, R., Amedeo, P., Anderson, M. J., et al. (2008). Genomic Islands in the Pathogenic Filamentous Fungus Aspergillus fumigatus. *PLOS Genetics* 4, e1000046. doi: 10.1371/journal.pgen.1000046.
Fedrigo, O., Pfefferle, A. D., Babbitt, C. C., Haygood, R., Wall, C. E., and Wray, G. A. (2011). A Potential Role for Glucose Transporters in the Evolution of Human Brain Size. *Brain Behav Evol* 78, 315–326. doi: 10.1159/000329852.
Ferretti, L., Raineri, E., and Ramos-Onsins, S. (2012). Neutrality tests for sequences with missing data. *Genetics* 191, 1397–1401. doi: 10.1534/genetics.112.139949.
Flores, M. E. B., Medina, P. G., Camacho, S. P. D., de Jesús Uribe Beltrán, M., De la Cruz Otero, M. del C., Ramírez, I. O., et al. (2014). Fungal spore concentrations in indoor and outdoor air in university libraries, and their variations in response to changes in meteorological variables. *Int J Environ Health Res* 24, 320–340. doi: 10.1080/09603123.2013.835029.
Forseth, R. R., Fox, E. M., Chung, D., Howlett, B. J., Keller, N. P., and Schroeder, F. C. (2011). Identification of cryptic products of the gliotoxin gene cluster using NMR-based comparative metabolomics and a model for gliotoxin biosynthesis. *J Am Chem Soc* 133, 9678–9681. doi: 10.1021/ja2029987.
Frisvad, J. C., and Larsen, T. O. (2016). Extrolites of Aspergillus fumigatus and Other Pathogenic Species in Aspergillus Section Fumigati. *Frontiers in Microbiology* 6. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2015.01485 [Accessed September 15, 2023].
Fuller, K. K., Richie, D. L., Feng, X., Krishnan, K., Stephens, T. J., Wikenheiser-Brokamp, K. A., et al. (2011). Divergent Protein Kinase A isoforms co-ordinately regulate conidial germination, carbohydrate metabolism and virulence in Aspergillus fumigatus. *Mol Microbiol* 79, 1045–1062. doi: 10.1111/j.1365-2958.2010.07509.x.
Furukawa, T., Scheven, M. T., Misslinger, M., Zhao, C., Hoefgen, S., Gsaller, F., et al. (2020a). The fungal CCAAT-binding complex and HapX display highly variable but evolutionary conserved synergetic promoter-specific DNA recognition. *Nucleic Acids Res* 48, 3567–3590. doi: 10.1093/nar/gkaa109.
Furukawa, T., van Rhijn, N., Fraczek, M., Gsaller, F., Davies, E., Carr, P., et al. (2020b). The negative cofactor 2 complex is a key regulator of drug resistance in Aspergillus fumigatus. *Nat Commun* 11, 427. doi: 10.1038/s41467-019-14191-1.
Gago, S., Overton, N. L. D., Ben-Ghazzi, N., Novak-Frazer, L., Read, N. D., Denning, D. W., et al. (2018). Lung colonization by Aspergillus fumigatus is controlled by ZNF77. *Nat Commun* 9, 3835. doi: 10.1038/s41467-018-06148-7.
Garcia-Rubio, R., de Oliveira, H. C., Rivera, J., and Trevijano-Contador, N. (2020). The Fungal Cell Wall: Candida, Cryptococcus, and Aspergillus Species. *Frontiers in Microbiology* 10. Available

at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02993 [Accessed September 10, 2023].
Gautam, P., Shankar, J., Madan, T., Sirdeshmukh, R., Sundaram, C. S., Gade, W. N., et al. (2008). Proteomic and transcriptomic analysis of Aspergillus fumigatus on exposure to amphotericin B. *Antimicrob Agents Chemother* 52, 4220–4227. doi: 10.1128/AAC.01431-07.
Gomez-Lopez, A., Rueda, C., Pando Pozo, R., and Sanchez Gonzalez, L. M. (2022). Dynamics of gliotoxin and bis(methylthio)gliotoxin production during the course of Aspergillus fumigatus infection. *Med Mycol* 60, myac025. doi: 10.1093/mmy/myac025.
Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., et al. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res* 22, 2376–2384. doi: 10.1101/gr.142281.112.
Gordon, K. L., and Ruvinsky, I. (2012). Tempo and Mode in Evolution of Transcriptional Regulation. *PLoS Genet* 8, e1002432. doi: 10.1371/journal.pgen.1002432.
Grahl, N., Puttikamonkul, S., Macdonald, J. M., Gamcsik, M. P., Ngo, L. Y., Hohl, T. M., et al. (2011). In vivo hypoxia and a fungal alcohol dehydrogenase influence the pathogenesis of invasive pulmonary aspergillosis. *PLoS Pathog* 7, e1002145. doi: 10.1371/journal.ppat.1002145.
Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064.
Gresnigt, M. S., Rekiki, A., Rasid, O., Savers, A., Jouvion, G., Dannaoui, E., et al. (2016). Reducing hypoxia and inflammation during invasive pulmonary aspergillosis by targeting the Interleukin-1 receptor. *Sci Rep* 6, 26490. doi: 10.1038/srep26490.
Griffioen, G., and Thevelein, J. M. (2002). Molecular mechanisms controlling the localisation of protein kinase A. *Curr Genet* 41, 199–207. doi: 10.1007/s00294-002-0308-9.
Gsaller, F., Hortschansky, P., Furukawa, T., Carr, P. D., Rash, B., Capilla, J., et al. (2016). Sterol Biosynthesis and Azole Tolerance Is Governed by the Opposing Actions of SrbA and the CCAAT Binding Complex. *PLoS Pathog* 12, e1005775. doi: 10.1371/journal.ppat.1005775.
Guruceaga, X., Perez-Cuesta, U., Pellon, A., Cendon-Sanchez, S., Pelegri-Martinez, E., Gonzalez, O., et al. (2021). Aspergillus fumigatus Fumagillin Contributes to Host Cell Damage. *J Fungi (Basel)* 7, 936. doi: 10.3390/jof7110936.
Haas, H. (2012). Iron – A Key Nexus in the Virulence of Aspergillus fumigatus. *Frontiers in Microbiology* 3. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2012.00028 [Accessed September 10, 2023].
Hagiwara, D., Miura, D., Shimizu, K., Paul, S., Ohba, A., Gonoi, T., et al. (2017). A Novel Zn2-Cys6 Transcription Factor AtrR Plays a Key Role in an Azole Resistance Mechanism of Aspergillus fumigatus by Co-regulating cyp51A and cdr1B Expressions. *PLOS Pathogens* 13, e1006096. doi: 10.1371/journal.ppat.1006096.
Hagiwara, D., Takahashi, H., Watanabe, A., Takahashi-Nakaguchi, A., Kawamoto, S., Kamei, K., et al. (2014). Whole-Genome Comparison of Aspergillus fumigatus Strains Serially Isolated from Patients with Aspergillosis. *J Clin Microbiol* 52, 4202–4209. doi: 10.1128/JCM.01105-14.
Hameed, A. A. A., Yasser, I. H., and Khoder, I. M. (2004). Indoor air quality during renovation actions: a case study. *J. Environ. Monit.* 6, 740–744. doi: 10.1039/B402995J.
Happacher, I., Aguiar, M., Yap, A., Decristoforo, C., and Haas, H. (2023). Fungal siderophore metabolism with a focus on Aspergillus fumigatus: impact on biotic interactions and potential translational applications. *Essays in Biochemistry*, EBC20220252. doi: 10.1042/EBC20220252.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2.

Harrison, M.-C., LaBella, A. L., Hittinger, C. T., and Rokas, A. (2022). The evolution of the GALactose utilization pathway in budding yeasts. *Trends Genet* 38, 97–106. doi: 10.1016/j.tig.2021.08.013.

Harrison, N., Mitterbauer, M., Tobudic, S., Kalhs, P., Rabitsch, W., Greinix, H., et al. (2015). Incidence and characteristics of invasive fungal diseases in allogeneic hematopoietic stem cell transplant recipients: a retrospective cohort study. *BMC Infect Dis* 15, 584. doi: 10.1186/s12879-015-1329-6.

Hartmann, T., Sasse, C., Schedler, A., Hasenberg, M., Gunzer, M., and Krappmann, S. (2011). Shaping the fungal adaptome--stress responses of Aspergillus fumigatus. *Int J Med Microbiol* 301, 408–416. doi: 10.1016/j.ijmm.2011.04.008.

Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K.-D., and Wray, G. A. (2007). Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* 39, 1140–1144. doi: 10.1038/ng2104.

Hensel, M., Arst, H. N., Aufauvre-Brown, A., and Holden, D. W. (1998). The role of the Aspergillus fumigatus areA gene in invasive pulmonary aspergillosis. *Mol Gen Genet* 258, 553–557. doi: 10.1007/s004380050767.

Herbrecht, R., Bories, P., Moulin, J.-C., Ledoux, M.-P., and Letscher-Bru, V. (2012). Risk stratification for invasive aspergillosis in immunocompromised patients. *Ann N Y Acad Sci* 1272, 23–30. doi: 10.1111/j.1749-6632.2012.06829.x.

Hill, M. S., Vande Zande, P., and Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet* 22, 203–215. doi: 10.1038/s41576-020-00304-w.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35, 518–522. doi: 10.1093/molbev/msx281.

Horta, M. A. C., Steenwyk, J. L., Mead, M. E., Dos Santos, L. H. B., Zhao, S., Gibbons, J. G., et al. (2022). Examination of Genome-Wide Ortholog Variation in Clinical and Environmental Isolates of the Fungal Pathogen Aspergillus fumigatus. *mBio* 13, e0151922. doi: 10.1128/mbio.01519-22.

Houbraken, J., de Vries, R. P., and Samson, R. A. (2014). "Chapter Four - Modern Taxonomy of Biotechnologically Important Aspergillus and Penicillium Species," in *Advances in Applied Microbiology*, eds. S. Sariaslani and G. M. Gadd (Academic Press), 199–249. doi: 10.1016/B978-0-12-800262-9.00004-4.

Houbraken, J., Kocsubé, S., Visagie, C. M., Yilmaz, N., Wang, X.-C., Meijer, M., et al. (2020). Classification of Aspergillus, Penicillium, Talaromyces and related genera (Eurotiales): An overview of families, genera, subgenera, sections, series and species. *Studies in Mycology* 95, 5–169. doi: 10.1016/j.simyco.2020.05.002.

Hu, W., Sillaots, S., Lemieux, S., Davison, J., Kauffman, S., Breton, A., et al. (2007). Essential Gene Identification and Drug Target Prioritization in Aspergillus fumigatus. *PLoS Pathog* 3, e24. doi: 10.1371/journal.ppat.0030024.

Huang, L., He, H., Jin, J., and Zhan, Q. (2017). Is Bulpa criteria suitable for the diagnosis of probable invasive pulmonary Aspergillosis in critically ill patients with chronic obstructive pulmonary

disease? A comparative study with EORTC/ MSG and ICU criteria. *BMC Infect Dis* 17, 209. doi: 10.1186/s12879-017-2307-y.
Hubka, V., Barrs, V., Dudová, Z., Sklenář, F., Kubátová, A., Matsuzawa, T., et al. (2018). Unravelling species boundaries in the Aspergillus viridinutans complex (section Fumigati): opportunistic human and animal pathogens capable of interspecific hybridization. *Persoonia* 41, 142–174. doi: 10.3767/persoonia.2018.41.08.
Hudson, R. R., Kreitman, M., and Aguadé, M. (1987). A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* 116, 153–159. doi: 10.1093/genetics/116.1.153.
Irmer, H., Tarazona, S., Sasse, C., Olbermann, P., Loeffler, J., Krappmann, S., et al. (2015). RNAseq analysis of Aspergillus fumigatus in blood reveals a just wait and see resting stage behavior. *BMC Genomics* 16, 640. doi: 10.1186/s12864-015-1853-1.
Jackson, J. C., Higgins, L. A., and Lin, X. (2009). Conidiation Color Mutants of Aspergillus fumigatus Are Highly Pathogenic to the Heterologous Insect Host Galleria mellonella. *PLOS ONE* 4, e4224. doi: 10.1371/journal.pone.0004224.
Jain, S., Sekonyela, R., Knox, B. P., Palmer, J. M., Huttenlocher, A., Kabbage, M., et al. (2018). Selenate sensitivity of a laeA mutant is restored by overexpression of the bZIP protein MetR in Aspergillus fumigatus. *Fungal Genet Biol* 117, 1–10. doi: 10.1016/j.fgb.2018.05.001.
Jang, Y. J., LaBella, A. L., Feeney, T. P., Braverman, N., Tuchman, M., Morizono, H., et al. (2018). Disease-causing mutations in the promoter and enhancer of the ornithine transcarbamylase gene. *Hum Mutat* 39, 527–536. doi: 10.1002/humu.23394.
Juvvadi, P. R., Fortwendel, J. R., Rogg, L. E., Burns, K. A., Randell, S. H., and Steinbach, W. J. (2011). Localization and activity of the calcineurin catalytic and regulatory subunit complex at the septum is essential for hyphal elongation and proper septation in Aspergillus fumigatus. *Mol Microbiol* 82, 1235–1259. doi: 10.1111/j.1365-2958.2011.07886.x.
Juvvadi, P. R., Gehrke, C., Fortwendel, J. R., Lamoth, F., Soderblom, E. J., Cook, E. C., et al. (2013). Phosphorylation of Calcineurin at a Novel Serine-Proline Rich Region Orchestrates Hyphal Growth and Virulence in Aspergillus fumigatus. *PLOS Pathogens* 9, e1003564. doi: 10.1371/journal.ppat.1003564.
Kaplan, C. D. (2013). Basic Mechanisms of RNA Polymerase II Activity and Alteration of Gene Expression in Saccharomyces cerevisiae. *Biochim Biophys Acta* 1829, 39–54. doi: 10.1016/j.bbagrm.2012.09.007.
Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20, 1160–1166. doi: 10.1093/bib/bbx108.
Keizer, E. M., Valdes, I. D., Forn-Cuni, G., Klijn, E., Meijer, A. H., Hillman, F., et al. (2021). Variation of virulence of five Aspergillus fumigatus isolates in four different infection models. *PLoS One* 16, e0252948. doi: 10.1371/journal.pone.0252948.
Kim, M. S., Cho, K. H., Park, K. H., Jang, J., and Hahn, J.-S. (2019). Activation of Haa1 and War1 transcription factors by differential binding of weak acid anions in Saccharomyces cerevisiae. *Nucleic Acids Res* 47, 1211–1224. doi: 10.1093/nar/gky1188.
Kjærbølling, I., Vesth, T., Frisvad, J. C., Nybo, J. L., Theobald, S., Kildgaard, S., et al. (2020). A comparative genomics study of 23 Aspergillus species from section Flavi. *Nat Commun* 11, 1106. doi: 10.1038/s41467-019-14051-y.
Knowles, S. L., Mead, M. E., Silva, L. P., Raja, H. A., Steenwyk, J. L., Goldman, G. H., et al. (2020). Gliotoxin, a Known Virulence Factor in the Major Human Pathogen Aspergillus fumigatus, Is

Also Biosynthesized by Its Nonpathogenic Relative Aspergillus fischeri. *mBio* 11, e03361-19. doi: 10.1128/mBio.03361-19.
Kocsubé, S., Perrone, G., Magistà, D., Houbraken, J., Varga, J., Szigeti, G., et al. (2016). Aspergillus is monophyletic: Evidence from multiple gene phylogenies and extrolites profiles. *Studies in Mycology* 85, 199–213. doi: 10.1016/j.simyco.2016.11.006.
Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin* 5, 1. doi: 10.1186/1756-8935-5-1.
Krappmann, S., Bignell, E. M., Reichard, U., Rogers, T., Haynes, K., and Braus, G. H. (2004). The Aspergillus fumigatus transcriptional activator CpcA contributes significantly to the virulence of this fungal pathogen. *Mol Microbiol* 52, 785–799. doi: 10.1111/j.1365-2958.2004.04015.x.
Kuster, S., Stampf, S., Gerber, B., Baettig, V., Weisser, M., Gerull, S., et al. (2018). Incidence and outcome of invasive fungal diseases after allogeneic hematopoietic stem cell transplantation: A Swiss transplant cohort study. *Transpl Infect Dis* 20, e12981. doi: 10.1111/tid.12981.
Kusuya, Y., Sakai, K., Kamei, K., Takahashi, H., and Yaguchi, T. (2016). Draft Genome Sequence of the Pathogenic Filamentous Fungus Aspergillus lentulus IFM 54703T. *Genome Announc* 4, e01568-15. doi: 10.1128/genomeA.01568-15.
Kwon-Chung, K. J., and Sugui, J. A. (2013). Aspergillus fumigatus--what makes the species a ubiquitous human fungal pathogen? *PLoS Pathog* 9, e1003743. doi: 10.1371/journal.ppat.1003743.
Lamoth, F. (2016). Aspergillus fumigatus-Related Species in Clinical Practice. *Frontiers in Microbiology* 7. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2016.00683 [Accessed September 15, 2023].
Langfelder, K., Philippe, B., Jahn, B., Latgé, J. P., and Brakhage, A. A. (2001). Differential expression of the Aspergillus fumigatus pksP gene detected in vitro and in vivo with green fluorescent protein. *Infect Immun* 69, 6411–6418. doi: 10.1128/IAI.69.10.6411-6418.2001.
Lara-Rojas, F., Sánchez, O., Kawasaki, L., and Aguirre, J. (2011). Aspergillus nidulans transcription factor AtfA interacts with the MAPK SakA to regulate general stress responses, development and spore functions. *Mol Microbiol* 80, 436–454. doi: 10.1111/j.1365-2958.2011.07581.x.
Latgé, J.-P., and Beauvais, A. (2014). Functional duality of the cell wall. *Curr Opin Microbiol* 20, 111–117. doi: 10.1016/j.mib.2014.05.009.
Latgé, J.-P., and Chamilos, G. (2019). Aspergillus fumigatus and Aspergillosis in 2019. *Clin Microbiol Rev* 33, e00140-18. doi: 10.1128/CMR.00140-18.
Lehrnbecher, T., Frank, C., Engels, K., Kriener, S., Groll, A. H., and Schwabe, D. (2010). Trends in the postmortem epidemiology of invasive fungal infections at a university hospital. *J Infect* 61, 259–265. doi: 10.1016/j.jinf.2010.06.018.
Li, H., and Johnson, A. D. (2010). Evolution of Transcription Networks — Lessons from Yeasts. *Curr Biol* 20, R746–R753. doi: 10.1016/j.cub.2010.06.056.
Li, X., Gao, M., Han, X., Tao, S., Zheng, D., Cheng, Y., et al. (2012). Disruption of the Phospholipase D Gene Attenuates the Virulence of Aspergillus fumigatus. *Infection and Immunity* 80, 429–440. doi: 10.1128/iai.05830-11.
Lin, C.-J., Sasse, C., Gerke, J., Valerius, O., Irmer, H., Frauendorf, H., et al. (2015). Transcription Factor SomA Is Required for Adhesion, Development and Virulence of the Human Pathogen Aspergillus fumigatus. *PLOS Pathogens* 11, e1005205. doi: 10.1371/journal.ppat.1005205.

Lind, A. L., Wisecaver, J. H., Lameiras, C., Wiemann, P., Palmer, J. M., Keller, N. P., et al. (2017). Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biol* 15, e2003583. doi: 10.1371/journal.pbio.2003583.

Lofgren, L. A., Ross, B. S., Cramer, R. A., and Stajich, J. E. (2021). Combined Pan-, Population-, and Phylo-Genomic Analysis of Aspergillus fumigatus Reveals Population Structure and Lineage-Specific Diversity. doi: 10.1101/2021.12.12.472145.

Lofgren, L. A., Ross, B. S., Cramer, R. A., and Stajich, J. E. (2022). The pan-genome of Aspergillus fumigatus provides a high-resolution view of its population structure revealing high levels of lineage-specific diversity driven by recombination. *PLoS Biol* 20, e3001890. doi: 10.1371/journal.pbio.3001890.

López-Berges, M. S., Scheven, M. T., Hortschansky, P., Misslinger, M., Baldin, C., Gsaller, F., et al. (2021). The bZIP Transcription Factor HapX Is Post-Translationally Regulated to Control Iron Homeostasis in Aspergillus fumigatus. *Int J Mol Sci* 22, 7739. doi: 10.3390/ijms22147739.

Losada, L., Barker, B. M., Pakala, S., Pakala, S., Joardar, V., Zafar, N., et al. (2014). Large-scale transcriptional response to hypoxia in Aspergillus fumigatus observed using RNAseq identifies a novel hypoxia regulated ncRNA. *Mycopathologia* 178, 331–339. doi: 10.1007/s11046-014-9779-8.

Loussert, C., Schmitt, C., Prevost, M.-C., Balloy, V., Fadel, E., Philippe, B., et al. (2010). In vivo biofilm composition of Aspergillus fumigatus. *Cell Microbiol* 12, 405–410. doi: 10.1111/j.1462-5822.2009.01409.x.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47, W636–W641. doi: 10.1093/nar/gkz268.

Main, B. J., Smith, A. D., Jang, H., and Nuzhdin, S. V. (2013). Transcription Start Site Evolution in Drosophila. *Molecular Biology and Evolution* 30, 1966–1974. doi: 10.1093/molbev/mst085.

Maity, A., Winnerdy, F. R., Chang, W. D., Chen, G., and Phan, A. T. (2020). Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Research* 48, 3315–3327. doi: 10.1093/nar/gkaa008.

Maizels, N., and Gray, L. T. (2013). The G4 genome. *PLoS Genet* 9, e1003468. doi: 10.1371/journal.pgen.1003468.

Mattioli, K., Oliveros, W., Gerhardinger, C., Andergassen, D., Maass, P. G., Rinn, J. L., et al. (2020). Cis and trans effects differentially contribute to the evolution of promoters and enhancers. *Genome Biol* 21, 210. doi: 10.1186/s13059-020-02110-3.

McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351, 652–654. doi: 10.1038/351652a0.

Mead, M. E., Knowles, S. L., Raja, H. A., Beattie, S. R., Kowalski, C. H., Steenwyk, J. L., et al. (2019). Characterizing the Pathogenic, Genomic, and Chemical Traits of Aspergillus fischeri, a Close Relative of the Major Human Fungal Pathogen Aspergillus fumigatus. *mSphere* 4, e00018-19. doi: 10.1128/mSphere.00018-19.

Mead, M. E., Steenwyk, J. L., Silva, L. P., de Castro, P. A., Saeed, N., Hillmann, F., et al. (2021). An evolutionary genomic approach reveals both conserved and species-specific genetic elements related to human disease in closely related Aspergillus fungi. *Genetics* 218, iyab066. doi: 10.1093/genetics/iyab066.

Mellado, E., Garcia-Effron, G., Alcázar-Fuoli, L., Melchers, W. J. G., Verweij, P. E., Cuenca-Estrella, M., et al. (2007). A New Aspergillus fumigatus Resistance Mechanism Conferring In Vitro

Cross-Resistance to Azole Antifungals Involves a Combination of cyp51A Alterations. *Antimicrob Agents Chemother* 51, 1897–1904. doi: 10.1128/AAC.01092-06.
Michels, K., Solomon, A. L., Scindia, Y., Sordo Vieira, L., Goddard, Y., Whitten, S., et al. (2022). Aspergillus Utilizes Extracellular Heme as an Iron Source During Invasive Pneumonia, Driving Infection Severity. *J Infect Dis* 225, 1811–1821. doi: 10.1093/infdis/jiac079.
Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37, 1530–1534. doi: 10.1093/molbev/msaa015.
Mitsuguchi, H., Seshime, Y., Fujii, I., Shibuya, M., Ebizuka, Y., and Kushiro, T. (2009). Biosynthesis of steroidal antibiotic fusidanes: functional analysis of oxidosqualene cyclase and subsequent tailoring enzymes from Aspergillus fumigatus. *J Am Chem Soc* 131, 6402–6411. doi: 10.1021/ja8095976.
Morelli, K. A., Kerkaert, J. D., and Cramer, R. A. (2021). Aspergillus fumigatus biofilms: Toward understanding how growth as a multicellular network increases antifungal resistance and disease progression. *PLoS Pathog* 17, e1009794. doi: 10.1371/journal.ppat.1009794.
Moreno, M. A., Ibrahim-Granet, O., Vicentefranqueira, R., Amich, J., Ave, P., Leal, F., et al. (2007). The regulation of zinc homeostasis by the ZafA transcriptional activator is essential for Aspergillus fumigatus virulence. *Mol Microbiol* 64, 1182–1197. doi: 10.1111/j.1365-2958.2007.05726.x.
Mouyna, I., Morelle, W., Vai, M., Monod, M., Léchenne, B., Fontaine, T., et al. (2005). Deletion of GEL2 encoding for a beta(1-3)glucanosyltransferase affects morphogenesis and virulence in Aspergillus fumigatus. *Mol Microbiol* 56, 1675–1688. doi: 10.1111/j.1365-2958.2005.04654.x.
Murga-Moreno, J., Coronado-Zamora, M., Hervas, S., Casillas, S., and Barbadilla, A. (2019). iMKT: the integrative McDonald and Kreitman test. *Nucleic Acids Research* 47, W283–W288. doi: 10.1093/nar/gkz372.
Muszkieta, L., Aimanianda, V., Mellado, E., Gribaldo, S., Alcàzar-Fuoli, L., Szewczyk, E., et al. (2014). Deciphering the role of the chitin synthase families 1 and 2 in the in vivo and in vitro growth of Aspergillus fumigatus by multiple gene targeting deletion. *Cell Microbiol* 16, 1784–1805. doi: 10.1111/cmi.12326.
Natkańska, U., Skoneczna, A., Sieńko, M., and Skoneczny, M. (2017). The budding yeast orthologue of Parkinson's disease-associated DJ-1 is a multi-stress response protein protecting cells against toxic glycolytic products. *Biochim Biophys Acta Mol Cell Res* 1864, 39–50. doi: 10.1016/j.bbamcr.2016.10.016.
Negri, C. E., Gonçalves, S. S., Xafranski, H., Bergamasco, M. D., Aquino, V. R., Castro, P. T. O., et al. (2020). Cryptic and Rare Aspergillus Species in Brazil: Prevalence in Clinical Samples and In Vitro Susceptibility to Triazoles. *Journal of Clinical Microbiology* 52, 3633–3640. doi: 10.1128/jcm.01582-14.
Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., et al. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus. *Nature* 438, 1151–1156. doi: 10.1038/nature04332.
Norton, T. S., Al Abdallah, Q., Hill, A. M., Lovingood, R. V., and Fortwendel, J. R. (2017). The Aspergillus fumigatus farnesyltransferase β-subunit, RamA, mediates growth, virulence, and antifungal susceptibility. *Virulence* 8, 1401–1416. doi: 10.1080/21505594.2017.1328343.
O'Gorman, C. M., Fuller, H. T., and Dyer, P. S. (2009). Discovery of a sexual cycle in the opportunistic fungal pathogen Aspergillus fumigatus. *Nature* 457, 471–474. doi: 10.1038/nature07528.

Oosthuizen, J. L., Gomez, P., Ruan, J., Hackett, T. L., Moore, M. M., Knight, D. A., et al. (2011). Dual organism transcriptomics of airway epithelial cells interacting with conidia of Aspergillus fumigatus. *PLoS One* 6, e20527. doi: 10.1371/journal.pone.0020527.

Parent-Michaud, M., Dufresne, P. J., Fournier, É., Martineau, C., Moreira, S., Perkins, V., et al. (n.d.). Draft Genome Sequences of Azole-Resistant and Azole-Susceptible Aspergillus turcosus Clinical Isolates Recovered from Bronchoalveolar Lavage Fluid Samples. *Microbiology Resource Announcements* 8, e01446-18. doi: 10.1128/MRA.01446-18.

Park, H.-S., and Yu, J.-H. (2016). Developmental regulators in Aspergillus fumigatus. *J Microbiol* 54, 223–231. doi: 10.1007/s12275-016-5619-5.

Patterson, T. F., Thompson, G. R., Denning, D. W., Fishman, J. A., Hadley, S., Herbrecht, R., et al. (2016). Practice Guidelines for the Diagnosis and Management of Aspergillosis: 2016 Update by the Infectious Diseases Society of America. *Clin Infect Dis* 63, e1–e60. doi: 10.1093/cid/ciw326.

Paul, S., Diekema, D., and Moye-Rowley, W. S. (2017). Contributions of both ATP-Binding Cassette Transporter and Cyp51A Proteins Are Essential for Azole Resistance in Aspergillus fumigatus. *Antimicrob Agents Chemother* 61, e02748-16. doi: 10.1128/AAC.02748-16.

Paul, S., Stamnes, M., Thomas, G. H., Liu, H., Hagiwara, D., Gomi, K., et al. (2019). AtrR Is an Essential Determinant of Azole Resistance in Aspergillus fumigatus. *mBio* 10, e02563-18. doi: 10.1128/mBio.02563-18.

Paulussen, C., Hallsworth, J. E., Álvarez-Pérez, S., Nierman, W. C., Hamill, P. G., Blain, D., et al. (2017). Ecology of aspergillosis: insights into the pathogenic potency of Aspergillus fumigatus and some other Aspergillus species. *Microbial Biotechnology* 10, 296–322. doi: 10.1111/1751-7915.12367.

Perez-Cuesta, U., Aparicio-Fernandez, L., Guruceaga, X., Martin-Souto, L., Abad-Diaz-de-Cerio, A., Antoran, A., et al. (2020). Melanin and pyomelanin in Aspergillus fumigatus: from its genetics to host interaction. *Int Microbiol* 23, 55–63. doi: 10.1007/s10123-019-00078-0.

Perez-Cuesta, U., Guruceaga, X., Cendon-Sanchez, S., Pelegri-Martinez, E., Hernando, F. L., Ramirez-Garcia, A., et al. (2021). Nitrogen, Iron, and Zinc Acquisition: Key Nutrients to Aspergillus fumigatus Virulence. *J Fungi (Basel)* 7, 518. doi: 10.3390/jof7070518.

Piłsyk, S., Natorff, R., Sieńko, M., Skoneczny, M., Paszewski, A., and Brzywczy, J. (2015). The Aspergillus nidulans metZ gene encodes a transcription factor involved in regulation of sulfur metabolism in this fungus and other Eurotiales. *Curr Genet* 61, 115–125. doi: 10.1007/s00294-014-0459-5.

Pini, G., Faggi, E., Donato, R., Sacco, C., and Fanci, R. (2008). Invasive pulmonary aspergillosis in neutropenic patients and the influence of hospital renovation. *Mycoses* 51, 117–122. doi: 10.1111/j.1439-0507.2007.01453.x.

Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079.

Puttikamonkul, S., Willger, S. D., Grahl, N., Perfect, J. R., Movahed, N., Bothner, B., et al. (2010). Trehalose-6-Phosphate Phosphatase is required for cell wall integrity and fungal virulence but not trehalose biosynthesis in the human fungal pathogen Aspergillus fumigatus. *Mol Microbiol* 77, 891–911. doi: 10.1111/j.1365-2958.2010.07254.x.

Raffa, N., and Keller, N. P. (2019). A call to arms: Mustering secondary metabolites for success and survival of an opportunistic pathogen. *PLOS Pathogens* 15, e1007606. doi: 10.1371/journal.ppat.1007606.

Ramani, R., Krumholz, K., Huang, Y.-F., and Siepel, A. (2019). PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. *Bioinformatics* 35, 2320–2322. doi: 10.1093/bioinformatics/bty966.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* 47, W191–W198. doi: 10.1093/nar/gkz369.

Rees, C. A., Stefanuto, P.-H., Beattie, S. R., Bultman, K. M., Cramer, R. A., and Hill, J. E. (2017). Sniffing out the hypoxia volatile metabolic signature of Aspergillus fumigatus. *J Breath Res* 11, 036003. doi: 10.1088/1752-7163/aa7b3e.

Reeves, E. P., Reiber, K., Neville, C., Scheibner, O., Kavanagh, K., and Doyle, S. (2006). A nonribosomal peptide synthetase (Pes1) confers protection against oxidative stress in Aspergillus fumigatus. *FEBS J* 273, 3038–3053. doi: 10.1111/j.1742-4658.2006.05315.x.

Revie, N. M., Iyer, K. R., Robbins, N., and Cowen, L. E. (2018). Antifungal drug resistance: evolution, mechanisms and impact. *Current Opinion in Microbiology* 45, 70–76. doi: 10.1016/j.mib.2018.02.005.

Rhodes, J. C. (2006). Aspergillus fumigatus: growth and virulence. *Med Mycol* 44 Suppl 1, S77-81. doi: 10.1080/13693780600779419.

Ries, L. N. A., Beattie, S. R., Espeso, E. A., Cramer, R. A., and Goldman, G. H. (2016). Diverse Regulation of the CreA Carbon Catabolite Repressor in Aspergillus nidulans. *Genetics* 203, 335–352. doi: 10.1534/genetics.116.187872.

Ries, L. N. A., Pardeshi, L., Dong, Z., Tan, K., Steenwyk, J. L., Colabardini, A. C., et al. (2020). The Aspergillus fumigatus transcription factor RglT is important for gliotoxin biosynthesis and self-protection, and virulence. *PLoS Pathog* 16, e1008645. doi: 10.1371/journal.ppat.1008645.

Ries, L. N. A., Rocha, M. C., de Castro, P. A., Silva-Rocha, R., Silva, R. N., Freitas, F. Z., et al. (2017). The Aspergillus fumigatus CrzA Transcription Factor Activates Chitin Synthase Gene Expression during the Caspofungin Paradoxical Effect. *mBio* 8, e00705-17. doi: 10.1128/mBio.00705-17.

Robert, V., Cardinali, G., and Casadevall, A. (2015). Distribution and impact of yeast thermal tolerance permissive for mammalian infection. *BMC Biology* 13, 18. doi: 10.1186/s12915-015-0127-3.

Rocha, M. C., Fabri, J. H. T. M., Franco de Godoy, K., Alves de Castro, P., Hori, J. I., Ferreira da Cunha, A., et al. (2016). Aspergillus fumigatus MADS-Box Transcription Factor rlmA Is Required for Regulation of the Cell Wall Integrity and Virulence. *G3 (Bethesda)* 6, 2983–3002. doi: 10.1534/g3.116.031112.

Rocha, M. C., Minari, K., Fabri, J. H. T. M., Kerkaert, J. D., Gava, L. M., da Cunha, A. F., et al. (2021). Aspergillus fumigatus Hsp90 interacts with the main components of the cell wall integrity pathway and cooperates in heat shock and cell wall stress adaptation. *Cell Microbiol* 23, e13273. doi: 10.1111/cmi.13273.

Rokas, A., Mead, M. E., Steenwyk, J. L., Oberlies, N. H., and Goldman, G. H. (2020). Evolving moldy murderers: Aspergillus section Fumigati as a model for studying the repeated evolution of fungal pathogenicity. *PLOS Pathogens* 16, e1008315. doi: 10.1371/journal.ppat.1008315.

Ropero, P., Erquiaga, S., Arrizabalaga, B., Pérez, G., de la Iglesia, S., Torrejón, M. J., et al. (2017). Phenotype of mutations in the promoter region of the β-globin gene. *J Clin Pathol* 70, 874–878. doi: 10.1136/jclinpath-2017-204378.

Rosowski, E. E., He, J., Huisken, J., Keller, N. P., and Huttenlocher, A. (2020). Efficacy of Voriconazole against Aspergillus fumigatus Infection Depends on Host Immune Function. *Antimicrobial Agents and Chemotherapy* 64, 10.1128/aac.00917-19. doi: 10.1128/aac.00917-19.

Ross, B. S., Lofgren, L. A., Ashare, A., Stajich, J. E., and Cramer, R. A. (2021). Aspergillus fumigatus In-Host HOG Pathway Mutation for Cystic Fibrosis Lung Microenvironment Persistence. *mBio* 12, e0215321. doi: 10.1128/mBio.02153-21.

Rothenberg, E. V. (2022). Transcription factors specifically control change. *Genes Dev* 36, 1097–1099. doi: 10.1101/gad.350308.122.

Roundtree, M. T., Juvvadi, P. R., Shwab, E. K., Cole, D. C., and Steinbach, W. J. (2020). Aspergillus fumigatus Cyp51A and Cyp51B Proteins Are Compensatory in Function and Localize Differentially in Response to Antifungals and Cell Wall Inhibitors. *Antimicrob Agents Chemother* 64, e00735-20. doi: 10.1128/AAC.00735-20.

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., and Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research* 47, W5–W10. doi: 10.1093/nar/gkz342.

Samson, R. A., Hong, S., Peterson, S. W., Frisvad, J. C., and Varga, J. (2007). Polyphasic taxonomy of Aspergillus section Fumigati and its teleomorph Neosartorya. *Studies in Mycology* 59, 147–203. doi: 10.3114/sim.2007.59.14.

Sasse, A., Hamer, S. N., Amich, J., Binder, J., and Krappmann, S. (2016). Mutant characterization and in vivo conditional repression identify aromatic amino acid biosynthesis to be essential for Aspergillus fumigatus virulence. *Virulence* 7, 56–62. doi: 10.1080/21505594.2015.1109766.

Schaefke, B., Wang, T.-Y., Wang, C.-Y., and Li, W.-H. (2015). Gains and Losses of Transcription Factor Binding Sites in Saccharomyces cerevisiae and Saccharomyces paradoxus. *Genome Biol Evol* 7, 2245–2257. doi: 10.1093/gbe/evv138.

Scharf, D. H., Chankhamjon, P., Scherlach, K., Heinekamp, T., Willing, K., Brakhage, A. A., et al. (2013). Epidithiodiketopiperazine biosynthesis: a four-enzyme cascade converts glutathione conjugates into transannular disulfide bridges. *Angew Chem Int Ed Engl* 52, 11092–11095. doi: 10.1002/anie.201305059.

Scharf, D. H., Remme, N., Habel, A., Chankhamjon, P., Scherlach, K., Heinekamp, T., et al. (2011). A Dedicated Glutathione *S* -Transferase Mediates Carbon–Sulfur Bond Formation in Gliotoxin Biosynthesis. *J. Am. Chem. Soc.* 133, 12322–12325. doi: 10.1021/ja201311d.

Schrettl, M., Beckmann, N., Varga, J., Heinekamp, T., Jacobsen, I. D., Jöchl, C., et al. (2010a). HapX-mediated adaption to iron starvation is crucial for virulence of Aspergillus fumigatus. *PLoS Pathog* 6, e1001124. doi: 10.1371/journal.ppat.1001124.

Schrettl, M., Carberry, S., Kavanagh, K., Haas, H., Jones, G. W., O'Brien, J., et al. (2010b). Self-Protection against Gliotoxin—A Component of the Gliotoxin Biosynthetic Cluster, GliT, Completely Protects Aspergillus fumigatus Against Exogenous Gliotoxin. *PLOS Pathogens* 6, e1000952. doi: 10.1371/journal.ppat.1000952.

Schrettl, M., Kim, H. S., Eisendle, M., Kragl, C., Nierman, W. C., Heinekamp, T., et al. (2008). SreA-mediated iron regulation in Aspergillus fumigatus. *Mol Microbiol* 70, 27–43. doi: 10.1111/j.1365-2958.2008.06376.x.

Seif, M., Kakoschke, T. K., Ebel, F., Bellet, M. M., Trinks, N., Renga, G., et al. (2022). CAR T cells targeting Aspergillus fumigatus are effective at treating invasive pulmonary aspergillosis in preclinical models. *Sci Transl Med* 14, eabh1209. doi: 10.1126/scitranslmed.abh1209.

Sellick, C. A., Campbell, R. N., and Reece, R. J. (2008). Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. *Int Rev Cell Mol Biol* 269, 111–150. doi: 10.1016/S1937-6448(08)01003-4.

Shemesh, E., Hanf, B., Hagag, S., Attias, S., Shadkchan, Y., Fichtman, B., et al. (2017). Phenotypic and Proteomic Analysis of the Aspergillus fumigatus ΔPrtT, ΔXprG and ΔXprG/ΔPrtT Protease-Deficient Mutants. *Frontiers in Microbiology* 8. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02490 [Accessed September 9, 2023].

Shevchenko, M. A., Bogorodskiy, A. O., Troyanova, N. I., Servuli, E. A., Bolkhovitina, E. L., Büldt, G., et al. (2018). Aspergillus fumigatus Infection-Induced Neutrophil Recruitment and Location in the Conducting Airway of Immunocompetent, Neutropenic, and Immunosuppressed Mice. *J Immunol Res* 2018, 5379085. doi: 10.1155/2018/5379085.

Shibuya, K., Takaoka, M., Uchida, K., Wakayama, M., Yamaguchi, H., Takahashi, K., et al. (1999). Histopathology of experimental invasive pulmonary aspergillosis in rats: pathological comparison of pulmonary lesions induced by specific virulent factor deficient mutants. *Microb Pathog* 27, 123–131. doi: 10.1006/mpat.1999.0288.

Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology* 51, 492–508. doi: 10.1080/10635150290069913.

Shwab, E. K., Juvvadi, P. R., Waitt, G., Soderblom, E. J., Barrington, B. C., Asfaw, Y. G., et al. (2019). Calcineurin-dependent dephosphorylation of the transcription factor CrzA at specific sites controls conidiation, stress tolerance, and virulence of Aspergillus fumigatus. *Mol Microbiol* 112, 62–80. doi: 10.1111/mmi.14254.

Signor, S. A., and Nuzhdin, S. V. (2018). The evolution of gene expression in cis and trans. *Trends Genet* 34, 532–544. doi: 10.1016/j.tig.2018.03.007.

Steenwyk, J. L., Balamurugan, C., Raja, H. A., Gonçalves, C., Li, N., Martin, F., et al. (2022). Phylogenomics reveals extensive misidentification of fungal strains from the genus Aspergillus. 2022.11.22.517304. doi: 10.1101/2022.11.22.517304.

Steenwyk, J. L., Buida, T. J., Labella, A. L., Li, Y., Shen, X.-X., and Rokas, A. (2021a). PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics*. doi: 10.1093/bioinformatics/btab096.

Steenwyk, J. L., Iii, T. J. B., Li, Y., Shen, X.-X., and Rokas, A. (2020a). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biology* 18, e3001007. doi: 10.1371/journal.pbio.3001007.

Steenwyk, J. L., Mead, M. E., de Castro, P. A., Valero, C., Damasio, A., dos Santos, R. A. C., et al. (2021b). Genomic and Phenotypic Analysis of COVID-19-Associated Pulmonary Aspergillosis Isolates of Aspergillus fumigatus. *Microbiol Spectr* 9. doi: 10.1128/Spectrum.00010-21.

Steenwyk, J. L., Mead, M. E., Knowles, S. L., Raja, H. A., Roberts, C. D., Bader, O., et al. (2020b). Variation Among Biosynthetic Gene Clusters, Secondary Metabolite Profiles, and Cards of Virulence Across Aspergillus Species. *Genetics* 216, 481–497. doi: 10.1534/genetics.120.303549.

Steenwyk, J. L., Mead, M. E., Knowles, S. L., Raja, H. A., Roberts, C. D., Bader, O., et al. (2020c). Variation Among Biosynthetic Gene Clusters, Secondary Metabolite Profiles, and Cards of Virulence Across Aspergillus Species. *Genetics* 216, 481–497. doi: 10.1534/genetics.120.303549.

Steenwyk, J. L., Shen, X.-X., Lind, A. L., Goldman, G. H., and Rokas, A. (n.d.). A Robust Phylogenomic Time Tree for Biotechnologically and Medically Important Fungi in the Genera Aspergillus and Penicillium. *mBio* 10, e00925-19. doi: 10.1128/mBio.00925-19.

Steinbach, W. J., Marr, K. A., Anaissie, E. J., Azie, N., Quan, S.-P., Meier-Kriesche, H.-U., et al. (2012). Clinical epidemiology of 960 patients with invasive aspergillosis from the PATH Alliance registry. *Journal of Infection* 65, 453–464. doi: 10.1016/j.jinf.2012.08.003.

Sugui, J. A., Losada, L., Wang, W., Varga, J., Ngamskulrungroj, P., Abu-Asab, M., et al. (2011). Identification and characterization of an Aspergillus fumigatus "supermater" pair. *mBio* 2, e00234-11. doi: 10.1128/mBio.00234-11.

Sugui, J. A., Pardo, J., Chang, Y. C., Zarember, K. A., Nardone, G., Galvez, E. M., et al. (2007). Gliotoxin Is a Virulence Factor of Aspergillus fumigatus: gliP Deletion Attenuates Virulence in Mice Immunosuppressed with Hydrocortisone. *Eukaryot Cell* 6, 1562–1569. doi: 10.1128/EC.00141-07.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34, W609–W612. doi: 10.1093/nar/gkl315.

Takahashi, H., Umemura, M., Ninomiya, A., Kusuya, Y., Shimizu, M., Urayama, S., et al. (2021). Interspecies Genomic Variation and Transcriptional Activeness of Secondary Metabolism-Related Genes in Aspergillus Section Fumigati. *Frontiers in Fungal Biology* 2, 14. doi: 10.3389/ffunb.2021.656751.

Tong, X., Xu, H., Zou, L., Cai, M., Xu, X., Zhao, Z., et al. (2017). High diversity of airborne fungi in the hospital environment as revealed by meta-sequencing-based microbiome analysis. *Scientific Reports* 7. doi: 10.1038/srep39606.

Traven, A., Jelicic, B., and Sopta, M. (2006). Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep* 7, 496–499. doi: 10.1038/sj.embor.7400679.

Traynor, A. M., Sheridan, K. J., Jones, G. W., Calera, J. A., and Doyle, S. (2019). Involvement of Sulfur in the Biosynthesis of Essential Metabolites in Pathogenic Fungi of Animals, Particularly Aspergillus spp.: Molecular and Therapeutic Implications. *Frontiers in Microbiology* 10, 2859. doi: 10.3389/fmicb.2019.02859.

Ukai, Y., Kuroiwa, M., Kurihara, N., Naruse, H., Homma, T., Maki, H., et al. (2018). Contributions of yap1 Mutation and Subsequent atrF Upregulation to Voriconazole Resistance in Aspergillus flavus. *Antimicrob Agents Chemother* 62, e01216-18. doi: 10.1128/AAC.01216-18.

Ullmann, A. J., Aguado, J. M., Arikan-Akdagli, S., Denning, D. W., Groll, A. H., Lagrou, K., et al. (2018). Diagnosis and management of Aspergillus diseases: executive summary of the 2017 ESCMID-ECMM-ERS guideline. *Clin Microbiol Infect* 24 Suppl 1, e1–e38. doi: 10.1016/j.cmi.2018.01.002.

Valiante, V., Macheleidt, J., Föge, M., and Brakhage, A. A. (2015). The Aspergillus fumigatus cell wall integrity signaling pathway: drug target, compensatory pathways, and virulence. *Frontiers in Microbiology* 6. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2015.00325 [Accessed September 11, 2023].

Valsecchi, I., Dupres, V., Stephen-Victor, E., Guijarro, J. I., Gibbons, J., Beau, R., et al. (2017). Role of Hydrophobins in Aspergillus fumigatus. *J Fungi (Basel)* 4, 2. doi: 10.3390/jof4010002.

van de Veerdonk, F. L., Gresnigt, M. S., Romani, L., Netea, M. G., and Latgé, J.-P. (2017). Aspergillus fumigatus morphology and dynamic host interactions. *Nat Rev Microbiol* 15, 661–674. doi: 10.1038/nrmicro.2017.90.

Vande Zande, P., Hill, M. S., and Wittkopp, P. J. (2022). Pleiotropic effects of trans-regulatory mutations on fitness and gene expression. *Science* 377, 105–109. doi: 10.1126/science.abj7185.

Vargas-Muñiz, J. M., Renshaw, H., Richards, A. D., Lamoth, F., Soderblom, E. J., Moseley, M. A., et al. (2015). The Aspergillus fumigatus septins play pleiotropic roles in septation, conidiation, and

cell wall stress, but are dispensable for virulence. *Fungal Genet Biol* 81, 41–51. doi: 10.1016/j.fgb.2015.05.014.
Vinet, L., and Zhedanov, A. (2011). A `missing' family of classical orthogonal polynomials. *J. Phys. A: Math. Theor.* 44, 085201. doi: 10.1088/1751-8113/44/8/085201.
Waddell, P. J., and Steel, M. A. (1997). General time-reversible distances with unequal rates across sites: mixing gamma and inverse Gaussian distributions with invariant sites. *Mol Phylogenet Evol* 8, 398–414. doi: 10.1006/mpev.1997.0452.
WHO (2022). WHO releases first-ever list of health-threatening fungi. Available at: https://www.who.int/news/item/25-10-2022-who-releases-first-ever-list-of-health-threatening-fungi [Accessed September 8, 2023].
Wiesner, D. L., and Klein, B. S. (2017). Lung epithelium: barrier immunity to inhaled fungi and driver of fungal-associated allergic asthma. *Curr Opin Microbiol* 40, 8–13. doi: 10.1016/j.mib.2017.10.007.
Willger, S. D., Grahl, N., and Cramer, R. A. (2009). Aspergillus fumigatus metabolism: clues to mechanisms of in vivo fungal growth and virulence. *Med Mycol* 47 Suppl 1, S72-79. doi: 10.1080/13693780802455313.
Willger, S. D., Puttikamonkul, S., Kim, K.-H., Burritt, J. B., Grahl, N., Metzler, L. J., et al. (2008). A Sterol-Regulatory Element Binding Protein Is Required for Cell Polarity, Hypoxia Adaptation, Azole Drug Resistance, and Virulence in Aspergillus fumigatus. *PLOS Pathogens* 4, e1000200. doi: 10.1371/journal.ppat.1000200.
Wirmann, L., Ross, B., Reimann, O., Steinmann, J., and Rath, P.-M. (2018). Airborne Aspergillus fumigatus spore concentration during demolition of a building on a hospital site, and patient risk determination for invasive aspergillosis including azole resistance. *J Hosp Infect* 100, e91–e97. doi: 10.1016/j.jhin.2018.07.030.
Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. *Nat Genet* 40, 346–350. doi: 10.1038/ng.77.
Wong, T. K. F., Kalyaanamoorthy, S., Meusemann, K., Yeates, D. K., Misof, B., and Jermiin, L. S. (2020). A minimum reporting standard for multiple sequence alignments. *NAR Genomics and Bioinformatics* 2, lqaa024. doi: 10.1093/nargab/lqaa024.
Yamamoto, N., Maeda, Y., Ikeda, A., and Sakurai, H. (2008). Regulation of thermotolerance by stress-induced transcription factors in Saccharomyces cerevisiae. *Eukaryot Cell* 7, 783–790. doi: 10.1128/EC.00029-08.
Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24, 1586–1591. doi: 10.1093/molbev/msm088.
Yap, A., Volz, R., Paul, S., Moye-Rowley, W. S., and Haas, H. (n.d.). Regulation of High-Affinity Iron Acquisition, Including Acquisition Mediated by the Iron Permease FtrA, Is Coordinated by AtrR, SrbA, and SreA in Aspergillus fumigatus. *mBio* 14, e00757-23. doi: 10.1128/mbio.00757-23.
Yin, W.-B., Baccile, J. A., Bok, J. W., Chen, Y., Keller, N. P., and Schroeder, F. C. (2013). A nonribosomal peptide synthetase-derived iron(III) complex from the pathogenic fungus Aspergillus fumigatus. *J Am Chem Soc* 135, 2064–2067. doi: 10.1021/ja311145n.
Yu, Y., Amich, J., Will, C., Eagle, C. E., Dyer, P. S., and Krappmann, S. (2017). The novel Aspergillus fumigatus MAT1-2-4 mating-type gene is required for mating and cleistothecia formation. *Fungal Genetics and Biology* 108, 1–12. doi: 10.1016/j.fgb.2017.09.001.
Zhang, J., Li, L., Lv, Q., Yan, L., Wang, Y., and Jiang, Y. (2019). The Fungal CYP51s: Their Functions, Structures, Related Drug Resistance, and Inhibitors. *Frontiers in Microbiology* 10.

Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.00691 [Accessed August 13, 2023].
Zhang, J., Tan, K., Wu, X., Chen, G., Sun, J., Reck-Peterson, S. L., et al. (2011). Aspergillus myosin-V supports polarized growth in the absence of microtubule-based transport. *PLoS One* 6, e28575. doi: 10.1371/journal.pone.0028575.
Zhang, X., and Emerson, J. J. (2019). Inferring Compensatory Evolution of cis- and trans-Regulatory Variation. *Trends Genet* 35, 1–3. doi: 10.1016/j.tig.2018.11.003.
Zhao, W., Panepinto, J. C., Fortwendel, J. R., Fox, L., Oliver, B. G., Askew, D. S., et al. (2006). Deletion of the regulatory subunit of protein kinase A in Aspergillus fumigatus alters morphology, sensitivity to oxidative damage, and virulence. *Infect Immun* 74, 4865–4874. doi: 10.1128/IAI.00565-06.