UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie
"Lazzaro Spallanzani"

Laurea Magistralis in Molecular Biology and Genetics

Advancement of Metagenomic Sweep Sequencing for Strain-Level Surveillance of Candida spp. Colonisation in Hospitalised Patients During first wave of COVID-19 Pandemic

Sviluppo del sequenziamento metagenomico di tipo "sweep" per il monitoraggio a livello di ceppo della colonizzazione da Candida spp. nei pazienti ricoverati durante la prima ondata della pandemia di COVID-19

Supervisor:
Davide Sassera

Co-supervisor:
Michela Vumbaca

Experimental thesis by
Kumar Utkarsh

Academic Year 2025/2026

INDEX

RIASSUNTO

Le specie di Candida sono funghi commensali comunemente presenti nel microbiota umano che possono causare infezioni opportunistiche in condizioni di immunosoppressione o di alterazione dell'omeostasi ospite-microbioma, portando alla candidosi (Lass-Flörl et al., 2024). La caratterizzazione ad alta risoluzione delle popolazioni di Candida nei campioni clinici rimane difficile a causa della presenza di comunità miste e di varianti a bassa frequenza. Gli approcci di metagenomica sweep combinati con il sequenziamento ad alta profondità dopo arricchimento in coltura consentono la caratterizzazione della diversità microbica all'interno di campioni clinici e permettono una risoluzione a livello di ceppo (Maklin et al., 2021).

In questo studio sono stati analizzati 601 campioni rettali, nasali e respiratori prelevati da pazienti ricoverati presso l'Ospedale San Matteo (Pavia) tra aprile e maggio 2020. È stato utilizzato un approccio gerarchico basato su riferimenti per caratterizzare la diversità delle specie di *Candida* e determinare la struttura della popolazione a livello di ceppo. *Candida albicans*, *Candida parapsilosis* e *Candida glabrata* sono state le specie rilevate più frequentemente in tutti i tipi di campioni. *C. albicans* ha mostrato un'ampia distribuzione in più siti senza un clone dominante, mentre *C. glabrata* era in gran parte presente nei campioni rettali, risultato che riflette la sua nicchia gastrointestinale. Al contrario, *C. parapsilosis* ha mostrato una struttura di popolazione peculiare dominata da un unico cluster (C3), che rappresentava circa il 90–95% dei rilevamenti in tutti i tipi di campioni, sia in terapia intensiva che nei reparti ordinari. L'analisi filogenetica che ha confrontato gli isolati colonizzanti con isolati causanti candidemia precedentemente descritti provenienti dallo stesso ospedale (2018-2020) ha indicato che essi appartengono tutti allo stesso lignaggio genomico, confermando la persistenza di un clone resistente agli azoli sia in contesti di colonizzazione che di infezione.

Nel complesso, questo studio offre una panoramica genomica delle popolazioni di *Candida* presenti in ambito ospedaliero durante la pandemia di COVID-19, evidenziando la persistenza di un ceppo dominante di *C. parapsilosis* associato sia a casi di colonizzazione che a casi di candidemia. Infine, ma non meno importante, i risultati ottenuti mettono in luce la capacità dell'approccio metagnomico "sweep" di cogliere la diversità microbica nei campioni clinici anche per quanto riguarda i funghi, mantenendo al contempo una risoluzione sufficiente per dedurre la struttura delle popolazioni a livello di ceppo e clonale.

ABSTRACT

Candida spp. are commensal fungi commonly present in the human microbiota that can cause opportunistic infections under conditions of immune suppression or disruption of host–microbiome homeostasis, leading to candidiasis (Lass-Flörl et al., 2024). High-resolution characterization of *Candida* populations in clinical samples remains challenging due to the presence of mixed communities and low-frequency variants. Metagenomic sweep approaches combined with deep sequencing after culture enrichment allow the recovery of microbial diversity within samples and enable strain-level resolution (Maklin et al., 2021).

In this study, 601 rectal, nasal, and respiratory samples collected from hospitalized patients at San Matteo Hospital (Pavia) between April and May 2020 were analysed. A reference-based hierarchical approach was used to characterize *Candida* species diversity and resolve population structure at strain level. *Candida albicans*, *Candida parapsilosis*, and *Candida glabrata* were the most frequently detected species across all sample types. *C. albicans* showed broad, multi-site distribution with no dominant clone, and *C. glabrata* was largely confined to rectal samples, reflecting its gastrointestinal niche. Contrary, *C. parapsilosis*, showed a highly skewed structure dominated by a single cluster (C3), accounting for approximately 90–95% of detections across all sample types and both ICU and ordinary wards. Phylogenetic analysis comparing colonizing isolates with previously described candidemia-causing isolates from the same hospital (2018-2020) indicated that they belong to the same genomic lineage, confirming the persistence of an azole-resistant clone causing both colonization and infection.

Overall, this work provides a genomic overview of *Candida* populations circulating in hospital settings during the COVID-19 pandemic, highlighting the persistence of a dominant *C. parapsilosis* lineage associated with both colonization and candidemia cases. Furthermore, the results obtained highlight the ability of the metagenomic sweep approach to capture the fungal diversity in the clinical samples while maintaining sufficient resolution to infer strain-level and clonal population structure.

INTRODUCTION

The Fungi kingdom consists of organisms with important roles in ecology and a high level of biodiversity. While most fungi are non-harmful or useful, a particular group especially opportunistic pathogen has developed pathogenicity, which affects human and animal life in serious ways in terms of global health and biodiversity (Corbu et al., 2023). A disturbing trend in recent years involves emerging pathogenic fungi being dispersed across the non-native geographical ranges of origin, resulting in serious occurrences of new diseases in human populations. Such non-native pathogenic fungal strains, which spread largely in response to human factors such as global trade and travel in a climate-change reality, pose an important One Health risk with respect to emerging infectious diseases in a manner described in (Fisher et al., 2020). A taxonomic system to describe these dangers is necessary, in which a critical role is played by a definition of a fungal family.

The analysis of fungi as a whole has been revolutionized by modern taxonomic practices, which have undergone a shift with the advent of molecular phylogenetics (Borman et al., 2023). This phylogenetic framework has not only re-classified entire branches of the fungal kingdom but has also provided critical insights into their ecology, evolution, and diverse biological roles from saprobic decomposers to mutualistic symbionts and, in some cases, pathogens (Stajich, 2017).

One of the most common causes of fungal infections in humans are family Ascomycota belonging to the genus *Candida* (Turner & Butler, 2014). Although *Candida* species are widely recognised as opportunistic pathogens in immunocompromised patients, some species have increasingly emerged as a serious clinical threat due to their ability to cause invasive infections with greater severity and resistance to antifungal treatment (Wiederhold, 2022). For diagnostic perspective, it is possible to identify commonalities in virulence, antifungal resistance, and ecological pressures among related fungi. For *Candida*, understanding the biological basis of these traits is therefore important for surveillance and to develop novel and more effective therapies to enhance diagnostic techniques (Mayer et al., 2013)

1.1 Genus: *Candida*

The *Candida* genus belongs to the Saccharomycetes family, phylum Ascomycota and comprises over 200 species (Howell et al., 2015). *Candida* species are usually commensal organisms of the human microbiota, primarily colonizing mucosal surfaces, such as the oral cavity, the gastrointestinal or the genitourinary tracts (Caetano et al., 2023). However, their ecological niche extends far beyond the host, persisting in a wide range of environmental habitats (Caetano et al., 2023). Their capacity to persist both in the environment and in human hosts has contributed to their emergence as a leading cause of severe hospital-acquired fungal infections worldwide (De-la-Pinta et al., 2026). Indeed, under conditions of host susceptibility they can shift to opportunistic pathogens and lead to various forms of candidiasis.

Candidiasis refers to infections caused by *Candida* species (Pérez-García., 2017), which include superficial mucosal manifestations such as vaginal candidiasis, one of the most frequent fungal infections, but can also overgrow in the mouth, throat, or esophagus. In hospitalized patients, invasive candidiasis (IC) develops when *Candida* spreads to internal organs such the kidneys, brain, or circulation (Pappas et al., 2018). Bloodstream infections caused by *Candida* spp. are known as candidemia, and have become increasingly common and one of the most common causes of nosocomial infections, frequently occurring in hospitalized or critically ill patients and associated with high mortality rates (Pfaller and Diekema, 2007).

Nosocomial infections, also known as health-care associated infections, are infections that are acquired during hospitalization after at least 48 hours of stay and not present at the time of hospital admission (Kollef et al., 2021). Hospitalized individuals are more likely to contract candidiasis, with increased probability for inpatients in critical care units (ICUs). Gastrointestinal surgery, central venous catheters, immunosuppressive medications, and broad-spectrum antibiotic use are a few of these risk factors (McCarty & Pappas, 2016).

The prevalence of *Candida* infections and the following IC has increased with *Candida* spp., despite improvements in health systems and the development of novel antifungals. With an average yearly incidence of 626.000 cases and a mortality rate of 35% with therapy and 90% without it, candidiasis continues to be one of the most prevalent human fungal pathogens (Soriano et al., 2023). *Candida* infections continue to be a significant health concern due in

large part to challenges with early identification, the absence of antifungal vaccinations, and the rise in antifungal drug resistance (Gabaldón & Carrete, 2016; Polke et al., 2015).

While more than 15 species of *Candida* can cause disease in humans, clinical infections are dominated by six of *Candida* species. *Candida albicans* remains the most prevalent and significant opportunistic yeast, followed by the non-*albicans* species (NCAC) *Candida glabrata, Candida tropicalis, Candida parapsilosis, Candida krusei, and Candida auris* (McCarty & Pappas, 2016).

1.1.1 Morphology

Candida species exhibit notable morphological variability, changing depending on environmental conditions and stage of life. *Candida* species can grow as yeast oval form, pseudohyphae or hyphae (**Figure 1**). This phenotypic switching represents a key biological trait that affects both laboratory identification and their pathogenic potential. In *Candida*, morphogenesis is regulated by complex interactions between genetic programs and environmental cues, playing a crucial role in the fungus's ability to survive and proliferate within the host (Alves et al., 2020). Most *Candida* species grow as oval, unicellular yeast forms that divide by budding or fission. This morphology supports its persistence at commensal sites and facilitates dissemination through the bloodstream during systemic infection. Alternatively they can grow as filamentous elongated cells, which include the pseudohyphae (chains of cells with constriction between them) or hyphae, chains of cells separated by true septa. The fungus's ability to shift between these forms confers an important virulence trait and is especially evident in *C. albicans* and *C. tropicalis* (Chow et al., 2021).

Morphogenetic switching allows these species to adjust to different host microenvironments as temperature, pH, nutrient levels, and immune pressure change. Hyphae form through apical growth and have continuous, parallel cell walls, which is a defining trait of this species. *C. albicans* is capable of growing in all morphological forms, and this has been shown to be an important factor in promoting the switch from commensal to pathogenic form (Lopes et al., 2022). This feature, for pathogenic form, can offer functional advantages, including deeper tissue penetration, reduced recognition by host immune defenses, and strong biofilm growth on both living and nonliving surfaces. Hyphae apply mechanical force that helps *Candida* to pass through epithelial and endothelial barriers, and to express hypha-related virulence factors,

such as adhesins and secreted hydrolytic enzymes, together these processes lead to host cell injury and inflammation (Lopes et al., 2022).

However, not all *Candida* species are able to grow in the three different morphologies, for example *C. parapsilosi*s or *C. glabrata* do not form filaments, relying on different approaches to cause disease. These species mainly depend on yeast-phase–specific processes, such as strong adhesion, survival inside host phagocytes, and rapid development of antifungal resistance, to establish infection (Pais et al., 2019). This difference suggests that, within the *Candida* genus, changes in morphological capacity are not directly linked to higher pathogenic potential. Assessing morphogenetic behavior is helpful for species-level identification and for estimating traits linked to pathogenicity, including tissue invasion, biofilm formation, and the likely level of clinical risk to the host (Kadosh & Mundodi, 2020).

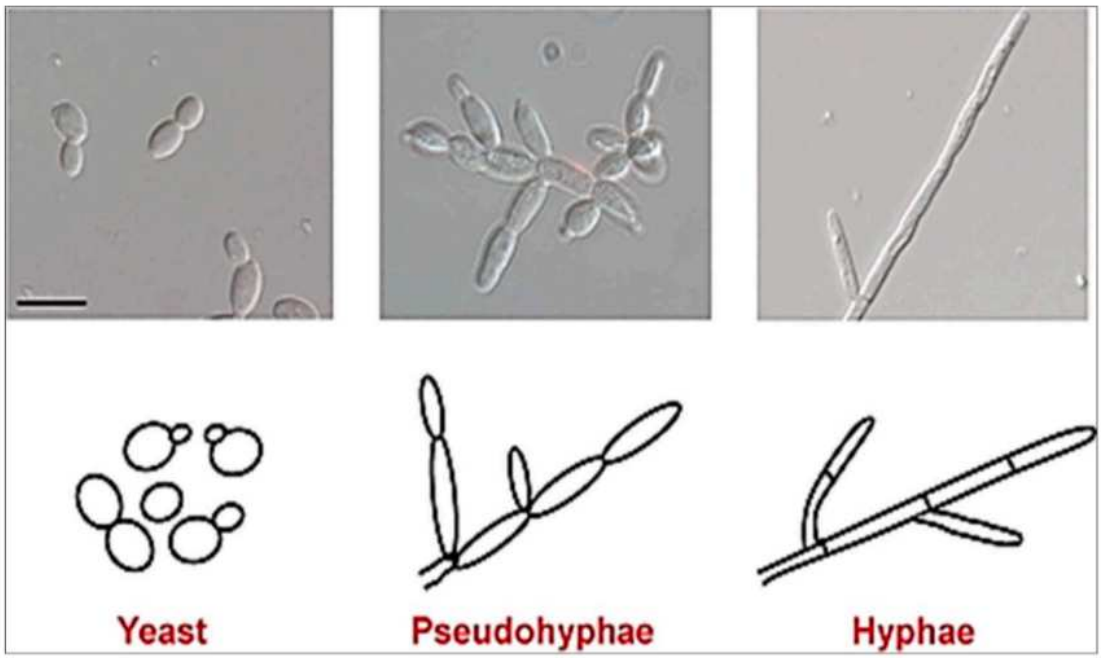

Figure 1: Morphological forms of *Candida* species (Thompson et al., 2011).

1.1.2 Genome and life cycle

The genome size of *Candida* species ranges between 11 and 14 Megabases, varying in chromosome numbers and ploidy levels (Ciurea et al., 2020). And variation in chromosome number and ploidy levels has been shown to influence several cellular characteristics such as cellular volume, genetic stability, and transcriptional level (Bennett et al., 2014). Ploidy

alterations are regulated by sexual reproduction between cells of various mating types, which is followed by meiosis (Usher, 2019). Sexual reproduction increases genetic diversity, facilitating the spread of advantageous mutations and the eradication of harmful mutations, improving an organism's capacity to adapt to various environments (Yadav et al., 2023). Under certain conditions, asexual reproduction may confer an advantage, since sexual reproduction requires the presence of two distinct mating types and can disrupt well-adapted genetic combinations (Ene & Bennett, 2014). Indeed, not every *Candida* species can reproduce clonally and finish the normal sexual life cycle (Moorhouse et al., 2021). Budding is the process by which *Candida* spp. replicate clonally. In this process a small bud develops on the mother cell's surface and grows until it reaches an appropriate size. A genetically identical daughter cell is created when the bud separates from the mother cell (Motaung et al., 2015; Wang et al., 2017).

In order to proliferate and survive inside the host, *Candida* species develop an adaptive mechanism and many of the adaptive mechanisms are underpinned by genetic variation (Alves et al., 2020). Although asexual reproduction occurs in the absence of recombination, genetic variation can still arise through the accumulation of mutations during successive replication cycles. These alterations include single nucleotide polymorphisms (SNPs), insertions, deletions, copy number variations (CNVs), chromosome loss, and duplications (Xu, 2021).

1.2 *Candida albicans*

Candida albicans is a diploid yeast, and the most common commensal species in the human microbiome, under normal conditions colonize the gastrointestinal tract, oral cavity and in the genitourinary tract (d'Enfert et al., 2021) However, it is also an opportunistic pathogen capable of causing superficial and systemic infections. Its pathogenicity is associated with several virulence factors, including the ability to transition from yeast to hyphal form, which has a role for the shift from the commensal to the pathogenic state, as well as the ability to form biofilm on both biotic and abiotic surface (Talapko et al., 2021). The complex three-dimensional biofilm structure limits antifungal penetration and protects fungal cells (Gong et al., 2020). Furthermore the genome of *C. albicans* is characterized by genomic plasticity driven by mutations and chromosomal rearrangements, contributing to phenotypic diversity and the emergence of antifungal resistance (Todd et al., 2019).

1.3 *Candida glabrata*

C. *glabrata* is an haploid asexual yeast, with 13 chromosomes, able to colonize as a commensal organism the gastrointestinal tract and epithelial surface. Until now it was considered as non-pathogenic commensal organisms of human mucosal tissues (Beardsley et al., 2024), but recently was found to cause superficial or systemic infections often in the hospital setting (Beardsley et al., 2024) and identified especially in the regions of North America and Europe, as one of the major causes of candidiasis (Angoulvant et al., 2016). Given this commensal lifestyle, most of *C. glabrata* deep infections occur through a pre-existing colonization in the patient and its ability to invade tissue and colonize bloodstream (Romo et al., 2020). This characteristic contributes to its emergence as an opportunistic pathogen, particularly in elderly individuals and immunocompromised patients (Hassan et al., 2021).

1.4 *Candida parapsilosis*

Candida parapsilosis is among the clinically significant NCA species, particularly distributed in nosocomial settings. Is a diploid organism with the genome organized in eight chromosome pairs and a total genome size of approximately 13Mb (Branco et al., 2023). It is a typical hospital-acquired pathogen transmitted through the hands of healthcare workers or by the use of contaminated medical instruments and antiseptic solutions (Franconi et al., 2023). The most clinically significant characteristic of this organism is its intrinsic and acquired resistance to azole antifungals, particularly fluconazole, which severely limits first-line treatment options and contributes directly to persistent and difficult-to-treat bloodstream infections (Amann et al., 2025). There has been worldwide monitoring of fluconazole-resistant strains, its emergence and their clonal spread. And, as a significant concern the identification of echinocandin-resistant isolates with FKS1 mutations has led to the decline of this pathogen's susceptibility to antifungal agents (Szekely et al., 2023). This resistance evolution is making the condition of this pathogen being merely a treatable infection.

1.5 *Candida tropicalis*

C. tropicalis is consistently reported as the leading or second most common NAC species in epidemiological studies across Asia and Latin America (Dos Santos et al., 2023). This species particularly affects neutropenic hosts, causing severe, widespread, and often life-threatening infections. Virulence-wise, it has the ability to form biofilm formation and also a very strong

hematogenous dissemination agent which as a result has a very high mortality rate in different groups (Dos Santos et al., 2023). However, *C. tropicalis* is mostly susceptible to azole therapy although resistance has been increasingly reported, and hence the rise of multidrug-resistant isolates especially in the oncology and transplant units becomes a very serious problem (Li et al., 2025).

1.6 *Candida auris*

The field of invasive candidiasis has been transformed by the detection and rapid international spread of *C. auris* since its first identification in 2009 (Hsu et al., 2025), leading major health organisations to classify it as a serious global health threat. Genomically, *C. auris* (genome size ~12.4 Mb) has high geographic diversity, with sequencing identifying at least five distinct clades that emerged independently across continents (Chow et al., 2021). Its genome encodes a repository of genes that are associated with efflux pump activity, biofilm formation, and cell wall remodelling mechanisms that directly points its resistance to major antifungal classes (Ruiz & Lorenz, 2021). Most isolates are resistant to fluconazole, approximately a quarter to amphotericin B, and an increasing number to echinocandins, leaving some strains pan-resistant with extremely limited treatment options (Ostrowsky et al., 2020). It survives for days on hospital surfaces while resisting standard disinfectants, spreads silently through wards via asymptomatic colonised individuals, and transmits efficiently through contaminated surfaces and medical equipment (Dire et al., 2023; Soliman S. S. M, 2023). Mortality rates associated with outbreaks range from 30% to over 60%, highlighting the urgent need for genomic surveillance and robust infection control to contain this pathogen before it becomes permanently established in healthcare systems worldwide (Soliman S. S. M, 2023).

1.7 Other *Candida* Species

1.7.1 *Candida guilliermondii* (formally known as *Meyerozyma guilliermondii*)

Candida guilliermondii is a haploid yeast with a genome size of 10.6 Mb and 8 chromosomes (Butler et al., 2009). It is commonly found in the environment, on human skin, and as a commensal organism on mucosal surfaces, where it rarely causes disease in healthy individuals (Ghasemi et al., 2022). However, in immunocompromised patients, particularly those with haematological malignancies or undergoing chemotherapy, can cause invasive infections

associated with high mortality (Tang et al., 2025), heightened by its reduced susceptibility to fluconazole and amphotericin B.

1.7.2 *Candida rugosa* (formally known as *Diutina rugosa*)

C. rugosa is increasingly recognised as an emerging agent of human infections, with cases reported across multiple regions worldwide and a particularly notable prevalence in Latin America (Pfaller et al., 2010). Human infections are predominantly reported in immunocompromised patients and are strongly associated with invasive medical procedures, including the use of central venous catheters and prior surgical intervention, which disrupt natural host barriers and create entry points for opportunistic fungal pathogens (Leite-Jr et al., 2023).

1.7.3 *Candida norvegensis* (formally known as *Pichia norvegensis*)

C. norvegensis is an emerging species associated with reduced susceptibility to fluconazole and has been reported more frequently in Europe than in other continents (Pinho et al., 2024). However, cases of invasive candidiasis caused by this species have also been described globally, predominantly in immunocompromised patients (Pfaller et al., 2010). Despite its rarity, these infections carry a serious risk of mortality if not promptly and adequately treated, echinocandins currently represent the most effective therapeutic option given the species' intrinsic reduced susceptibility to fluconazole (Kim et al., 2024).

1.7.4 *Candida krusei* (formally known as *Pichia kudriavzevii*)

Candida krusei is an emerging NAC species capable of causing both superficial and invasive infections in humans (Nguyen et al., 2024). It is widely distributed in natural environments including soil, food, vegetables, and fruits, and is inherently resistant to fluconazole (Nagarathnamma et al., 2017). It predominantly affects immunocompromised patients and is associated with a high crude mortality rate of approximately 49%, which is particularly concerning given its increasing clinical incidence (Nguyen et al., 2024). Beyond its intrinsic fluconazole resistance, *C. krusei* can develop acquired resistance to echinocandins and polyenes, making it a potentially multidrug-resistant pathogen with very limited treatment options (Karakoyun et al., 2024).

1.8 Species diversity in human microbiome

Candida species within the human microbiome exhibit considerable ecological and genetic diversity. They typically exist as commensal organisms but retain the capacity to transition into opportunistic pathogens when host or environmental conditions become favorable (Romo et al., 2020). This diversity reflects differences between body sites, variation across individuals, and the ways microbial species co-occur within the same community. Within the human mycobiota, *C. albicans* is the most frequently detected fungal species and often dominated fungal communities across multiple body sites, particularly prevalent in the oral cavity and in the gastrointestinal tract. Nevertheless, the composition of the human mycobiota is highly variable among individuals, and other *Candida* species may also contribute to fungal colonization.

Other *Candida* species, such as *Candida glabrata*, *Candida parapsilosis*, *Candida tropicalis* and *Candida krusei* are frequently detected alongside other commensal yeasts, although they typically colonize only a subset of individuals or specific body sites (Guinea, 2014) (**Figure 2**). These differences between individuals reflect several host-related factors, including age, diet, immune function, prior antibiotic use, and underlying health conditions (Ciurea et al., 2020). Furthermore, the relative abundance and composition of *Candida* species can change over time in response to alterations in the host environment, such as antimicrobial treatment or immune suppression (Ciurea et al., 2020).

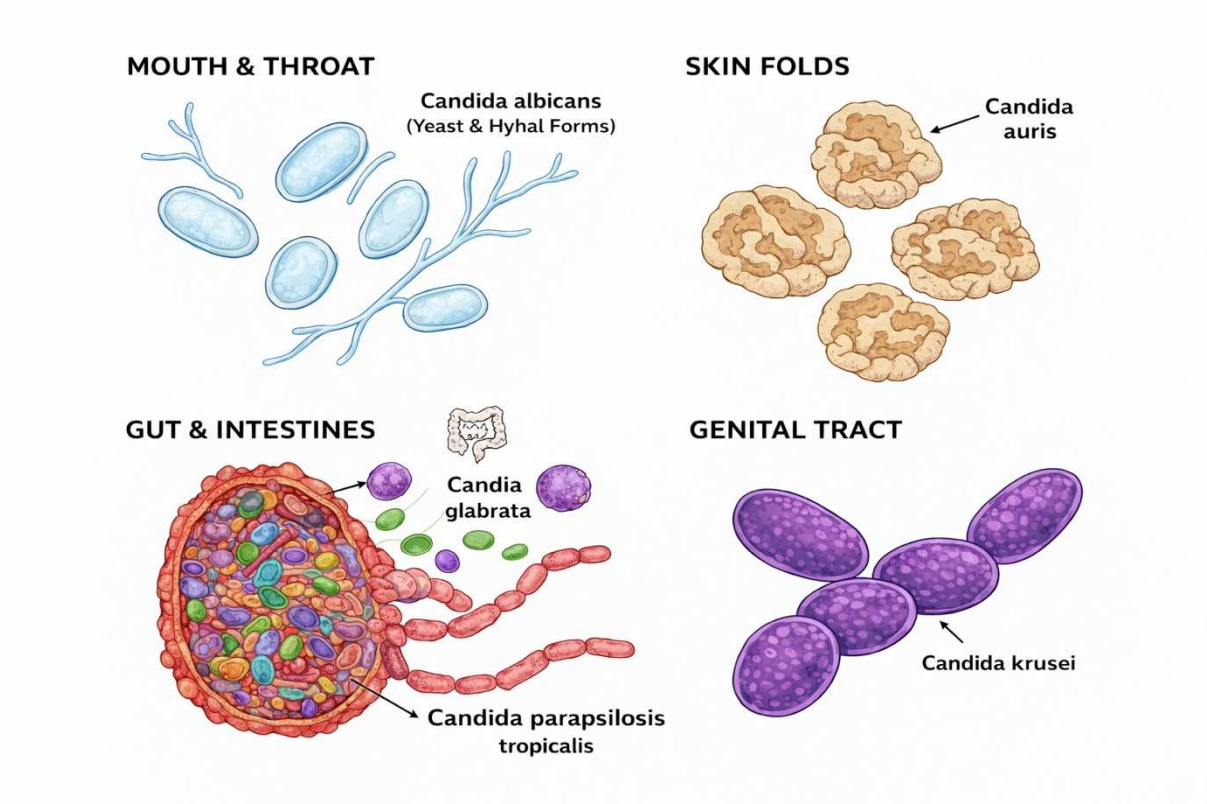

Figure 2: Distribution of clinically relevant *Candida* species across different human body sites. *C. albicans* (yeast and hyphal forms) colonizes the mouth and throat; *C. auris* is commonly associated with skin folds; *C. glabrata* and *C. parapsilosis/tropicalis* are found in the gut and intestines; and *C. krusei* predominantly affects the genital tract (Koundal et al., 2020).

1.9 *Candida* Epidemiology

1.9.1 *Candida* distribution

The distribution of *Candida* species varies across regions and patient groups, shaped by geography, host factors, healthcare practices, and antifungal selection. *C. albicans* remains most commonly isolated in mucosal and invasive candidiasis worldwide, yet studies over the past decades report a sustained, clinically relevant increase in infections caused by non-albicans *Candida* species (Pinho et al., 2024). This shift is most apparent in hospital and intensive care settings, where changing patient populations, dependence on invasive devices, longer hospital stays, and widespread use of antifungal prophylaxis have altered species patterns and the way infections develop and spread.

Among non-albicans *Candida* species, *C. glabrata*, *C. parapsilosis,* and *C. tropicalis* are leading causes of candidiasis worldwide as demonstrated (**Figure 3)**, with rates that differ across regions and patient groups. *C. glabrata* is reported more often in North America and parts of Europe, particularly in older adults and in patients with weakened immune systems (Gómez-Gaviria et al., 2022). It is also commonly linked to prior azole antifungal exposure, which is consistent with its lower baseline susceptibility and its tendency to develop resistance quickly. *C. parapsilosis* is reported more often in neonatal and pediatric patients and is frequently linked to catheter-related bloodstream infections and other indwelling devices, likely because it adheres well to synthetic materials and can form biofilms (Amann et al., 2025). *C. tropicalis* is more common in tropical and subtropical areas and is often linked to invasive infections in patients with hematologic cancers, especially when neutropenia is present (Keighley et al., 2024).

The recent emergence of *C. auris* marks a clear change in the epidemiology of candidiasis and indicates that the distribution of *Candida* species can shift over time in response to changing clinical and environmental pressures (Abdel-Hamid et al., 2023). *C. auris* is notable for resistance to several antifungal drugs, its spread in healthcare settings, and its ability to remain on hospital surfaces over time. Its rapid spread across countries creates major problems for diagnosis, treatment, and infection control.

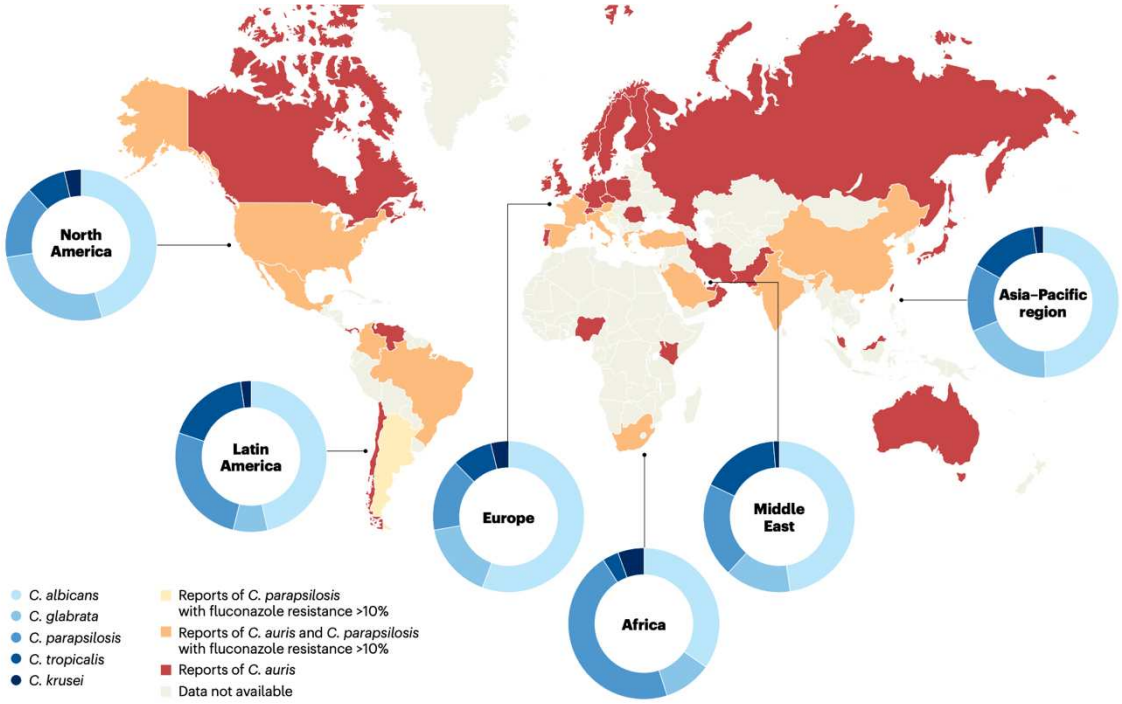

Figure 3:- Geographical variation in the distribution of *Candida* species and countries reporting the major species of concern. Species distribution as well as resistance rates are influenced by numerous factors, including the health-care system, in combination with the frequency of vulnerable patient groups, antimicrobial treatment and prophylaxis policies, demographics, and environmental factors such as climate and agricultural practices. The global spread of *C. auris* and fluconazole-resistant *C. parapsilosis* strains are worrisome developments, with both species causing large hospital outbreaks of resistant strains (Lass-Flörl et al., 2024).

In healthcare settings, the mix of *Candida* species also depends on local ecological niches within the institution and on routine clinical practices. High-risk hospital areas, including intensive care units, oncology wards, transplant centers, and surgical units, often show different patterns in the fungal species (Vazquez et al., 2025) (**Figure 4)**. These differences are shaped by patients' underlying conditions, immune function, how often invasive procedures are used and also the use of antifungal drugs. Mapping the distribution of *Candida* species is important for selecting novel antifungal treatment, infection control plans, and improving the understanding of how the epidemiology and clinical outcomes of candidiasis changes over time (Saeed et al., 2024).

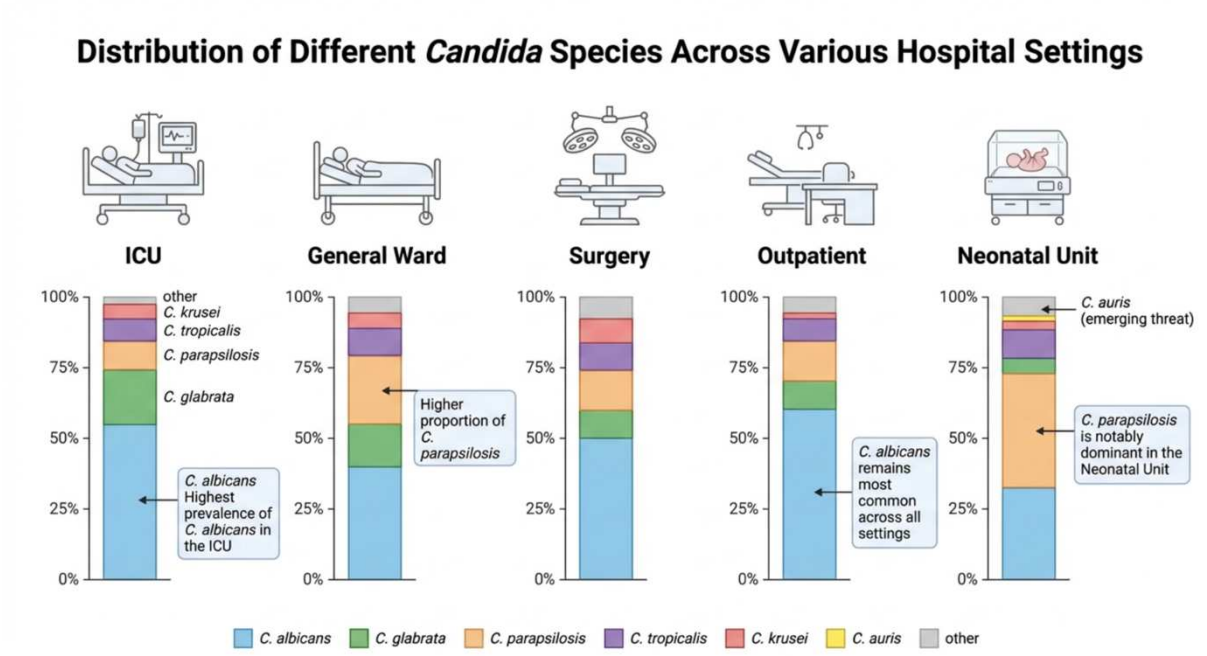

Figure 4:- The bar chart illustrates the relative prevalence of major *Candida* species in intensive care units (ICU), general wards, surgical units, outpatient departments, and neonatal units. *C. albicans* remains the predominant species across all settings, with particularly high

prevalence in the ICU and surgical wards. Non-*albicans Candida* species show setting-specific patterns, including an increased proportion of *C. parapsilosis* in general wards and marked dominance in neonatal units, reflecting its association with neonates and indwelling medical devices. *C. glabrata* and *C. tropicalis* contribute variably across adult care settings, while *C. krusei* and other species remain less frequent. The emergence of *C. auris* in neonatal and high-risk settings highlights its growing clinical significance and the need for enhanced surveillance and infection control measures.

1.9.2 *Candida* outbreak & *Candida* genomic epidemiology

In recent decades, there has been an increase in outbreaks caused by *Candida* species, notably NCAC species with antifungal resistance (Corzo-Leon et al., 2021). An epidemic happens when the number of illness cases in a certain area surpasses expected levels within a given time frame (Houlihan & Whitworth, 2019).

Fungal outbreaks, especially *Candida*, pose a considerable risk to hospitalized patients (Douglas et al., 2023). Using epidemiological tools to combat the growing incidence of nosocomial outbreaks caused by *Candida* is crucial, as these infections often result in high mortality and morbidity (Douglas et al., 2023). Fungal strain typing is a crucial stage in epidemiological investigations as it allows for genetic links between isolates, tracing transmission routes, and identifying infection origins. This information can be used to manage and prevent outbreaks (Hadrich & Ranque, 2015).

Multilocus sequence typing is one of the most used molecular approaches used for fungal strain identification. This method provides important insights into genetic diversity and strain relationships by analyzing the DNA sequences of housekeeping genes to evaluate population structure (Tönnies et al., 2021). However, whole genome sequencing (WGS) is now the recommended approach because of the decreased costs of genomic sequencing and the development of computer capacity. WGS provides a more complete picture of the isolates by sequencing the full genome, revealing more details regarding genetic variants and patterns of transmission (Rhodes & Fisher, 2019).

Genome-wide SNP analysis is the most efficient method for evaluating phylogenetic diversity since the population structure of *Candida* species is primarily clonal, resulting in a more

accurate distinction of closely related organisms (Moorhouse et al., 2021). We can therefore identify whether there is a clonal outbreak, whether individuals have been affected with distinct strains, or whether transmission is taking place between patients based on phylogenetic data. The integration of WGS into standard workflows is complicated by issues such as fungal genome size, fluctuating ploidy levels, recombination events, limited comparative genome databases, and the lack of characteristics that identify a genome reference as reliable (Douglas et al., 2023). For example, although it is generally acknowledged that a threshold of two to thirty SNPs can be used to determine clonality in *C. auris* (Douglas et al., 2023), no equivalent standard has been developed for *C. parapsilosis* because there aren't many genomic studies for this species (Misas et al., 2024).

Developing established data analysis pipelines and integrating WGS into everyday workflow are crucial to reducing and preventing the growing number of outbreaks. By improving epidemiology and disease control initiatives, these tools can save time and money (Sherry et al., 2022).

1.9.3 Methods in genomic epidemiology

Candida genus is composed of species with different virulence and antifungal resistance patterns. For this reason, early diagnosis and correct identification are crucial to define the appropriate treatment to improve the clinical outcomes (Pappas et al., 2018).

In the surveillance and to prevent and control the spread of pathogens, WGS is widely applied nowadays. Since the introduction of next generation sequencing (NGS) technologies, significant changes to the field of pathogen genomics research have been done. One of the major innovations introduced by NGS is the ability to generate huge amounts of sequencing data in shorter time than traditional sequencing and to run parallel analyses in multiple samples at the same time, reducing time and cost (Satam et al., 2023). This method is highly effective in clinical practice, as rapidly obtaining accurate data enables quick classification and detection of isolates during outbreaks, thereby enhancing monitor efforts and facilitating the implementation of preventive methods (Michel et al., 2023).

Sequencing reads are stored in FASTQ files format for subsequent analysis. The FASTQ file contains the information regarding the reads and their quality. Each sequencing read is

described in four lines, the first is the identifier for the sequence, followed by the nucleotide sequence, the third line is a symbol separator after which the quality scores is reported, which represent the confidence level for each base call (Piñeiro & Pichel, 2022). After obtaining sequencing reads from FASTQ files, mapping to a reference genome can be performed, to determine in what position of a previously sequenced reference genome our reads can be located, without the need of assembly (Schbath et al., 2012). The previously generated FASTQ reads are aligned to a reference genome generating a Sequence Alignment Map format (SAM) format (Langmead & Salzberg, 2012), a consensus format for alignments of reads to a reference genome. The header of the SAM file is characterized by the first lines starting with "@", including information about the alignment, such as the version of the format or the reference sequence name. The first lines are followed by the alignment section consisting of 11 mandatory columns, which provide details about the mapping, such as the quality, template length or chromosomal position (Liu et al., 2023). SAM files can be extremely large, especially when dealing with high-throughput sequencing data, therefore are compressed into BAM files, their binary equivalent, to significantly reduce their storage space (Long et al., 2017). The alignment of reads to a reference genome allows detecting different bases in a single position, allowing to find single nucleotide polymorphism (SNP). The process to obtain the variants is known as variant calling, which produces variant calling files (VCF) as output (Chen et al., 2023; Kidd et al., 2019). The main section of the VCF file instead contains one row for each variant, with columns that describe various features, such as the types of variants and where it is located (Bayat et al., 2017).

1.10 Virulence

Candida species, as opportunistic pathogens, can infect even healthy hosts although they cause disease in hosts with increased susceptibility to infections (Ciurea et al., 2020). This highlights the critical role of the host environment in *Candida* infections. Virulence is not an inherent trait of the pathogen, but is shaped by the interaction between the fungus and the host (Farhan et al., 2024).

The pathogenicity of *Candida* species is mediated by various virulence factors that directly interact with host cells, cause damage, and facilitate both colonization and spread of the infection (Talapko et al., 2021). The main virulence factors of *Candida* species are:

- Morphology changes (Farhan et al., 2024)

● Cell adhesion and invasion (Deorukhkar, 2017)
● Biofilm formation (Malinovská et al., 2023)

1.10.1 Morphology changes

An important virulence factor in *Candida* species is their ability to undergo phenotypic changes, altering the cell's surface behavior to adapt to host conditions, like pH variations, oxygen level and nutrition availability (Farhan et al., 2024).

Dimorphism is a phenotypic change considered one of the primary virulence factor in *C. albicans*, as it marks the transition from a commensal to a pathogenic state, with the hyphal form being invasive and able to penetrate host tissue (Talapko et al., 2021). However, this characteristic is not consistent across all *Candida* species, raising questions about its classification as a universal virulence factor for the entire genus (Kadosh & Mundodi, 2020).
The phenotypic switching enables microorganisms of a single strain to adopt different colony phenotypes, without including yeast-hyphae transition (Soll, 2014). This switching can affect different processes, like adherence to the host cells, expression of cell surface hydrophobicity and secretion of proteinases, enhancing the adaptability to stressful environments (Deorukhkar, 2017). This phenomenon has been observed primarily in *C. albicans,* with the white-to-opaque transition. Similar transitions have also been noted in other NCAC species, such as in *C. tropicalis* and *C. parapsilosis*, where different phenotypes have been associated with varying abilities to form biofilm (Malinovská et al., 2023).

1.10.2 Cell adhesion and invasion

The cell wall is the primary interface with the host and contains crucial proteins that contribute to the progression of the infection (Farhan et al., 2024). Among these proteins, adhesins play a crucial role in mediating cell adhesion, which is essential for colonization both in the host cells and on the abiotic surfaces (Talapko et al., 2021).

Adhesion is another important virulence factor in *Candida* organisms, helping to prevent the clearance of *Candida* from the host, enabling its persistence, and is essential for initiating host tissue invasion, thereby establishing the infection (Deorukhkar, 2017). Invasion can occur through two primary mechanisms: induced-endocytosis and active penetration (Arafa et al., 2023). Induced-endocytosis is a passive, host-mediated process in which *Candida* proteins, known as invasins, bind to host ligands, triggering the recruitment of acting-recognizing

proteins and facilitating the entry of the fungal cells (Richardson et al., 2018). Active penetration, on the other hand, is driven by hydrolytic enzymes, enzymes secreted in the local environment that disrupt cellular membranes. Once *Candida* has invaded the tissue, the infection can spread through dissemination, where the fungus enters the nearby blood vessels and subsequently reaches the organ targets (Strickland & Shi, 2021).

1.10.3 Biofilm formation

Following the initial attachment, and depending on environmental conditions and on the type of surface, fungal cells can divide, proliferate, and progress to biofilm formation, considered a crucial virulence factor in *Candida* species (Malinovská et al., 2023; Wall et al., 2019).

Biofilm is the result of a community of microorganisms that adhere on multiple surfaces, frequently on mucosal surfaces and non-living surfaces (Gulati & Nobile, 2016). From these biofilms, cells can disperse throughout the patient's body, potentially causing invasive candidiasis (Malinovská et al., 2023). This microbial community is embedded in an extracellular matrix, which, due to its complex structure, serves as a physical barrier against environmental challenges (Cavalheiro & Teixeira, 2018). This barrier provides resistance to antifungal drugs and protection from the host's immune response, leading to persistent infections. As a result, biofilm-forming *Candida* species are associated with higher infection persistence and mortality rates compared to non-biofilm forming strains (Gulati & Nobile, 2016; Kernien et al., 2017).

1.11 Antifungal resistance

Antifungal resistance among *Candida* species is an emerging challenge in candidiasis infections, and this is an actual concern to patients and health care systems across the world. Resistance levels among *Candida* species vary greatly and vary based on antifungal agents, and they can be classified based on whether they develop resistance intrinsically or after infection (Bhattacharya et al., 2020). Intrinsic resistance among *Candida* species is characterized by species that inherently exhibit resistance to some antifungals, and this is normally evident among those that inherently develop resistance to antifungals. Acquired resistance, in contrast, arises when *Candida* species develop reduced susceptibility following infection, typically as a result of antifungal exposure (Czajka et al., 2023). This phenomenon is most commonly observed in *C. albicans* and *C. tropicalis*, and it has become an increasingly

important concern in *C. auris*, which frequently exhibits resistance to multiple antifungal agents (Czajka et al., 2023).

The mechanisms of resistance in *Candida* are mainly linked to antifungal classes in a manner that counteracts the efficacy of antifungal agents on a cellular level. In the case of azoles (fluconazole and voriconazole), mechanisms of resistance are mainly linked to modifications in antifungal targets, namely mutations or overexpression of the ERG11 gene that encodes lanosterol 14α-demethylase (Li et al., 2025). Such antifungal targets are often coupled with increased expression of efflux pumps in the cell membrane, which are stimulated by CDR1, CDR2, or MDR1 genes to efflux antifungal agents out of cells (Li et al., 2025). The main mechanism of echinocandins (caspofungin or micafungin) involves mutations in hot-spot regions of FKS1 or FKS2 genes, which code subunits of β-1,3-D-glucan synthase (Hori et al., 2018). Such mutations lead to a significant reduction in antifungal binding sites. In polyene antifungal agents, mainly amphotericin B, mechanisms of resistance are less common but have been recorded. Such mechanisms are mainly linked to alterations in sterol composition in cell membranes (Carolus et al., 2020).

From an epidemiological perspective, antifungal resistance develops mainly in two ways. First, resistant strains can emerge within an individual patient, particularly after prolonged or repeated antifungal treatment, as drug exposure favors organisms that are able to survive. Second, and more concerning, resistant strains can spread within healthcare settings through clonal transmission between patients (Vitiello et al., 2023). This type of spread has been frequently observed with *C. parapsilosis*, which can persist on surfaces, contaminate hospital environments, and transmit easily between patients, leading to nosocomial outbreaks (Vitiello et al., 2023). These patterns highlight the importance of continuous molecular surveillance, effective infection control measures, and antifungal programs to monitor resistance, detect outbreaks early, and support appropriate treatment of *Candida* infections (Neto Junior et al., 2024).

1.12 The Hospital:Home for fungal species

While resistant *Candida* exists in nature, hospitals provide the perfect conditions because of widespread antifungal use and immunocompromised patients and for these resistant strains to dominate and spread. The major reservoir sources in a hospital setting are:

- Abiotic Surfaces
- Candidemia and Other Device-Related Candidiasis
- Healthcare Workers and Patients

1.12.1 Abiotic Surfaces

Candida species, especially *C. auris* and *C. parapsilosis*, have the ability to thrive on dried abiotic surfaces, including patient room furnishings and infusion pumps (Silva et al., 2024). On these abiotic surfaces, *Candida* species can grow into a stubborn biofilm, which protects these microorganisms against common disinfectants (Silva et al., 2024). Due to its ability to form biofilms, *C. parapsilosis* can persist for long periods on hospital surfaces and devices. This persistence can lead to clonal outbreaks, particularly in neonatal ICUs which later on is difficult to control (Schelenz et al., 2016).

1.12.2 Candidemia and Other Device-Related Candidiasis

Candidemia, a bloodstream infection, is often linked to device-related candidiasis. *Candida* species like *C. albicans* and *C. glabrata* readily form resistant biofilms on medical devices (e.g., central lines, urinary catheters) (Mallick et al., 2018). These biofilms act as a persistent source of infection, which can infect the bloodstream and cause candidemia, presenting a major treatment challenge in hospitals (Mallick et al., 2018).

1.12.3 Healthcare Workers and Patients

The primary risk is not colonization itself, but the shift of *Candida* from a harmless resident to an invasive pathogen. In critically ill patients, factors like broad-spectrum antibiotics which disrupt the protective bacterial microbiome and immune suppression allow *Candida* to overgrow and enter the bloodstream (Cortegiani et al., 2018). Healthcare workers can inadvertently transmit strains between patients that can lead to outbreaks (Cortegiani et al., 2018).

1.13 Nosocomial Infection (From Colonisation to pathogenic Infection)

Nosocomial, or hospital-acquired infections are a major challenge for healthcare systems worldwide, as they are responsible for increased patient morbidity, mortality, and healthcare costs (Liu et al., 2020). Among these infections, invasive fungal infections especially those caused by *Candida* species known as nosocomial candidiasis are recognised as the most severe

and complex ones (Kollef et al., 2021). Nosocomial candidiasis reflects the opportunistic pathogenicity of *Candida* species and primarily results from disruption of the normally balanced commensal relationship between these fungi and the human host within the hospital environment (Kollef et al., 2021). In healthy individuals, *Candida* species often reside on mucosal and skin surfaces, including the gastrointestinal tract, the oral cavity and pharynx, and the skin, without causing illness (Gerós-Mesquita et al., 2020). This commensal state is maintained by intact epithelial barriers, colonization resistance from a diverse bacterial microbiome, and well-functioning innate and adaptive immune responses.

In hospitals, the usual balance between the host and microbes often shifts in ways that can favor *Candida* growth and spread. In hospitalized patients, broad-spectrum antibiotics are used often and can disrupt the normal bacterial microbiota, which reduces bacterial competition and allows fungi to proliferate which leads to higher colonization density. This phenomenon is called colonization pressure, and is a well-known risk factor for invasive candidiasis (d'Enfert et al., 2021). Colonization is common in critically ill patients, especially those in intensive care units. And this is because of severe underlying disease, damage to epithelial barriers due to surgery or invasive devices, immunosuppression, extended hospital stays, and exposure to *Candida* strains that are adapted to hospital settings (León et al., 2018; d'Enfert et al., 2021).

The shift from *Candida* colonization to invasive disease, such as candidemia or deep-seated candidiasis, does not occur in all patients, but it becomes more likely in the presence of hospital-related risk factors. This shift results from the combined effects of host susceptibility, microbial capacity to survive and spread, and epidemiological pressures linked to healthcare settings (Soriano et al., 2023). Invasive medical procedures such as central venous catheters, urinary catheters, and many more can create direct entry routes into the body and bypass the usual anatomical barriers. At the same time, immunosuppressive treatments and severe illness weaken host defenses, and exposure to antifungal treatment applies selective pressure that favors species with lower susceptibility (Thomas-Rüddel et al., 2021). Over time, the epidemiology of hospital-acquired candidiasis has shifted from being mainly caused by *C. albicans* to a growing case due to non-*albicans Candida* species, including *C. glabrata* and *C. Parapsilosis. C. Parapsilosis* and multidrug-resistant *Candida* species are important to address in clinical and research settings because they can limit standard treatment options (Pappas et al., 2018). Nosocomial candidiasis reflects unintended effects of modern medical care, in which life-saving treatments can disturb the normal balance between the host and resident microbes and allow commensal fungi to become important causes of infection.

1.14 Candida and SARS-CoV-2

COVID-2019 is caused by SARS-CoV-2 infection and results in severe acute respiratory distress syndrome (Li et al., 2021). Bacteria and fungi co-infected with SARS-CoV-2 cases became common in the absence of specialized protocols (Pemán et al., 2020), particularly during the initial waves of the outbreak (Swaney et al., 2023). Co-infections in SARS-CoV-2 patients can worsen symptoms and increase fatality rates (Crotty et al., 2020). As a result, it is clinically vital to investigate how these co-infections occur and how to avoid them.

SARS-CoV-2-associated Candidiasis (CAC) has been reported frequently, indicating a link between the two infections (Tsai et al., 2023). In fact, patients with severe SARS-CoV-2 infections are frequently admitted to the ICU, where they are more exposed to risk factors such as immunosuppressive medications and surgical operations, increasing their chances of developing candidemia (Arastehfar et al., 2020). Early diagnosis, proper antifungal treatment can improve the clinical outcome of CAC patients, as the mortality rate associated with *Candida* infections is higher in SARS-CoV-2 patients compared to non-SARS-CoV-2 patients (Tsai et al., 2023).

1.15 Sampling methodologies

1.15.1 Single colonies

Single colonies is a key step when studying and identifying *Candida* species in clinical and research settings. It separates mixed microbial samples into discrete growth so later tests and species-level characterization can be done with better accuracy (Figueroa-Bossi et al., 2022). A colony is a visible cluster of fungal cells that is assumed to have grown from one original cell. In *Candida* research and diagnostics, isolating a single colony is necessary to ensure accurate species identification, consistent phenotypic assessment, and reliable antifungal susceptibility results (Kucukates et al., 2016). Isolation is usually done with the streak plate method, where a clinical sample is spread over a solid agar surface with a sterile loop so that the number of cells decreases across the plate and separate colonies can form. Other methods, including spread plate and pour plate techniques, can be used depending on the sample type and the laboratory's requirements (Sanders E. R., 2012).

Candida colony requires a careful review of morphology, including colony size, shape, color, elevation, margin, and surface texture. These traits provide initial guidance for identifying the species and help confirm that one uncontaminated colony is selected for later analysis. In clinical microbiology laboratories, when a blood culture is positive and the broth contains *Candida* species, the sample is first subcultured to obtain isolated colonies. These colonies are then used for final identification by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry and for antifungal susceptibility testing with standardized broth microdilution methods. Assessing mixed microbial cultures may produce unclear or misleading findings, which supports the need to isolate and study a single colony (Chen et al., 2021).

From a genomic and molecular standpoint, isolating a single colony is needed to obtain high-quality genomic DNA that is free from contamination. Working with pure cultures limits interference from other microorganisms and supports reliable downstream work, such as genome sequencing and comparative analyses. Even so, culture-based methods are limited to organisms that can grow in laboratory conditions and they miss viable but non-culturable populations. Phenotypic variation among genetically similar *Candida* cells can sometimes lead to mixed colony shapes or appearances, which may make results harder to interpret (Andrzejuk et al., 2024). Despite these limitations, isolating single *Candida* colonies remains the gold standard for definitive identification and characterization of the organism causing candidiasis (Lewis et al., 2020).

1.15.2 Sweep approach

Plate sweep metagenomics provides a practical approach between single-colony sequencing and culture-independent metagenomics, allowing analysis of mixed growth from plated samples without reducing the sample to one isolate. In this method, microbial colonies are collected together from an enrichment culture plate, so the sample retains the community diversity that developed under the selected growth conditions (Maklin et al., 2021). This approach provides a practical way to generate high-depth sequencing data for selected organisms while limiting bias when samples are compared across timepoints or experimental conditions. In this each constituent population can be quantified, depending on specialized computational tools and can be drawn on large genomic reference datasets (Schaeffer et al., 2017). To address this, there is a computational pipeline called mSWEEP to analyze sweep metagenomic data with high accuracy and speed. mSWEEP was designed to work with the

large and expanding sets of reference genomes available for bacterial pathogens and other species that can be cultured (Sankar et al., 2016). The main contribution is a workflow that combines three steps: it first clusters reference genomes into groups that reflect biological relatedness, then pseudoaligns sequencing reads quickly, and finally estimates group abundances with a probabilistic model while applying strict controls to reduce false positives (**Figure 5**) (Maklin et al., 2021).

Compared with approaches based on pseudoalignment, with probabilistic models, mSWEEP methods improve abundance estimation by clustering reads within large reference collections (Schaeffer et al., 2017). This pipeline supports high-resolution analysis of plate sweep experiments and sets a clear framework for this type of work. Using existing detailed genomic maps of pathogen populations, mSWEEP goes beyond species identification and supports studies of within-host variation, patterns of transmission, and how ecological factors shape microbial diversity (Fisher et al., 2017). mSWEEP is a method for analyzing sweep metagenomics data to address biological questions in microbial ecology and epidemiology (Maklin et al., 2021).

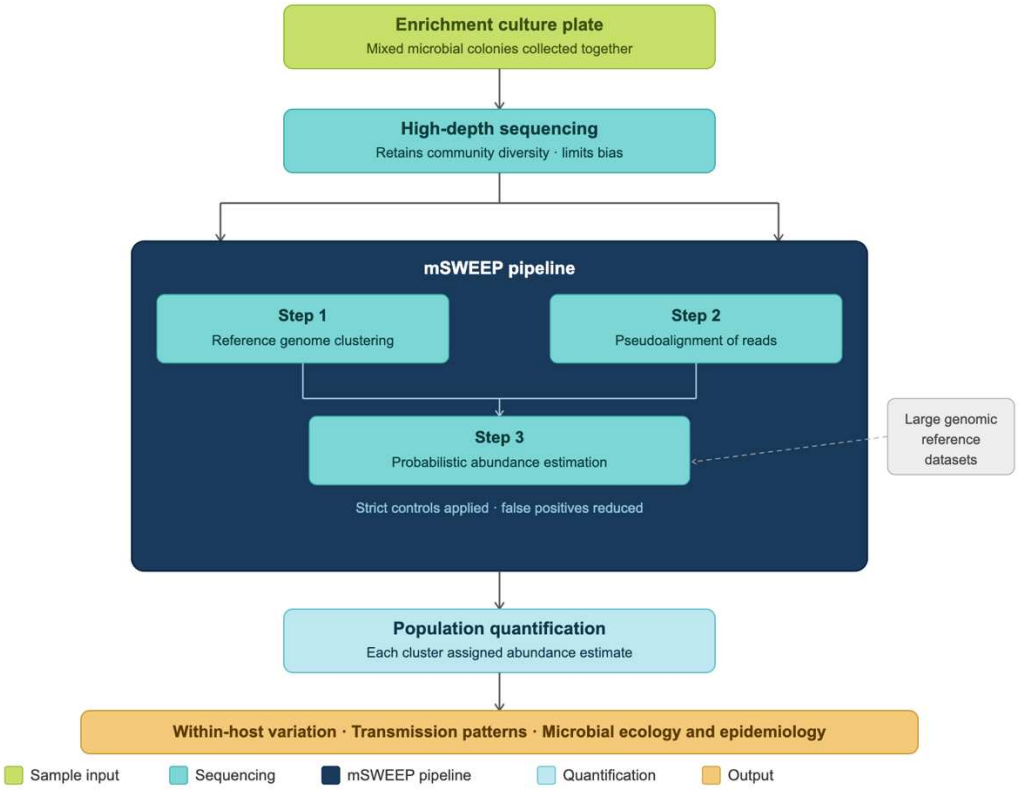

Figure 5: Flowchart of the mSWEEP pipeline describing a typical workflow for relative abundance estimation (Maklin et al., 2021).

1.15.3 Advantages of sweep approach

The sweep method provides a clear strategic benefit by linking single-isolate sequencing with culture-independent metagenomic analysis (Maklin et al., 2021). This method collects a mixed set of colonies directly from a primary culture plate so that the sample reflects the full cultured community grown on that specific medium.This position offers clear advantages like

(1) Economic and logistical efficiency in large-scale surveillance is improved because this approach avoids the slow, labor-intensive work of isolating and processing individual colonies. This reduces processing time and effort while supporting higher sample throughput. Processing hundreds or thousands of samples in a high-throughput workflow reduces reagent use, time, and sequencing library preparation steps compared with single-colony sequencing (Bisso et al., 2025).

(2) Compared with isolating a single colony, which reflects only a limited subset of the within-host community, sweep sampling collects many colonies at once and gives a broader view of population diversity (McMurray et al., 2026). This is important in monitoring the emergence of antimicrobial resistance during treatment.

(3) Sweep sequencing focuses mostly on the reads of organisms of interest. This yields higher-depth data for those targets and supports higher-resolution analysis when existing genomic reference data are available. This makes them well suited for research on pathogens that are sequenced often. Moreover, this method lets researchers draw on large reference genome collections to study new questions about within-host variation, transmission, and distribution with high precision (Maklin et al., 2021).

(4) Sweep sequencing supports high-throughput screening approaches by serving as an initial pass that filters samples before more detailed follow-up analyses. In large studies, it can quickly identify samples that have mixed infections, new resistance genes, or newly emerging strains (Ayon et al., 2023). Overall, the sweep approach is way more than to reduce costs. It supports larger-scale sampling and a more realistic view of diversity, and when it is combined with other

bioinformatics tools it allows deeper analysis of microbial populations for epidemiology and evolutionary research (Nam et al., 2023).

AIM OF THE WORK

Recovering microbial strain at single nucleotide polymorphism-level resolution, at the same time as capturing the microbial diversity within a clinical population can be extremely useful for surveillance and diagnosis but can be very challenging. Using metagenomic sweep approaches allow the capture of the full microbial population within a sample, including low-frequency variants and mixed infections (Maklin et al., 2021). The overall aim of this study was to evaluate the utility of the sweep metagenomic approach combined with the mSWEEP probabilistic classification pipeline for characterizing *Candida Spp.* colonisation at strain-level resolution at San Matteo Hospital, Pavia, Italy during the first COVID-19 wave. Specifically this study aimed to:-

1. Assess whether deep sequencing after culturing, combined with mSWEEP probabilistic classification pipeline could reliably identify and quantify *Candida* species and their genomic cluster from nasal, rectal and respiratory samples.

2. Characterise the diversity of *Candida* species that were colonising patients across different body sites during the first COVID-19 wave at San Matteo Hospital, Pavia, Italy.

3. Identify the strain-level population structure of the dominant *Candida* species by identifying circulating clusters and their distribution across sample types and ward categories.

4. And lastly, to investigate whether the colonising strains detected could be linked to the same persistent strain previously characterised at the same hospital, in order to assess the potential of this approach for nosocomial outbreak detection and monitoring.

MATERIALS AND METHODS

3.1 Species Detection from the datasets

To check relative species present in the sample, we used Kraken for the detection of the species. After isolation, the samples were sequenced with Illumina technology and species identification was assessed through Kraken version 1.1.1 (Wood & Salzberg, 2014). Kraken is a metagenomic sequence classification tool that rapidly and accurately assigns taxonomic labels to DNA reads by using exact k-mer alignments against a precomputed database. It works by extracting all k-mers (default length 31 bp) from each read, looking up the lowest common ancestor (LCA) of all genomes containing each k-mer, and then scoring root-to-leaf paths in a classification tree built from those LCA assignments to determine the most specific taxon possible. The approach achieves speed over 4 million reads per minute—hundreds of times faster than BLAST—while maintaining genus-level precision near 100% and sensitivity (Wood & Salzberg, 2014). We obtained Candida species/strains: (32), Other fungal species: (34) and Bacterial species: (58) different species out of which only *Candida* Species was extracted for further analysis.

3.2 Database Construction

We built reference databases for use with the mSWEEP–mGEMS (Mäklin et al., 2023) pipeline; which consisted of detailed databases of the pathogen species of interest. To build the broad reference, we used Kraken version 1.1.1 (Wood & Salzberg, 2014) to identify the species present within the sequencing read sets. We identified 25 species but 3 of the major species causing infection were taken into account which were (*C. parapsilosis, C. albicans, C. glabrata*) and downloaded their genomes from the National Center for Biotechnology Information and subjected them to quality control. For each pathogen species of interest we used filters for genome length of 50 and obtained detailed sets of genome assemblies from curated collections to enable sequence cluster-level (SCs) identification.

3.3 Demix Pipeline for Strain-Level Resolution

To resolve strain mixtures within dominant species, the demix_check pipeline was used (Thorpe et al., 2024). This pipeline assesses the demixed binned reads from an mGEMS

analysis to help with interpreting the assigned clusters. This is an important step when running mGEMS on complex mixtures, when there is possible contamination, or when the species being analysed does not have a comprehensive reference set. In these cases, the assigned clusters may represent the closest available sequences from the reference set, but the reads may actually originate from an unknown cluster absent from the reference set.

Prior to demix_check validation, sequencing reads were aligned to the reference database using Bowtie2, which performed rapid and accurate short-read alignment against the indexed reference genomes of the target species. The resulting alignments were subsequently used as input for abundance estimation. Themisto was first applied to pseudoalign the reads and assign them to reference clusters, after which mSWEEP estimated the relative abundance of each cluster within the sample by modelling the distribution of pseudoalignment results across the reference panel.

To address potential mis-assignments, demix_check calculates the genetic distances between reference isolates, which are used to build within-cluster and between-cluster distance distributions. Genetic distances are then calculated between the demixed binned reads and the reference isolates, and the query-reference distances are compared to the reference-reference distances to determine whether the binned reads originate from the assigned cluster or from an uncharacterised lineage (https://github.com/harry-thorpe/demix_check). This step helped to identify the presence and approximate abundance of specific strains within the sample and this method was executed in three distinct modes:

--mode_setup	This was used to set up one or more reference datasets
--mode_run	Then this step was performed to run mGEMS and then checked the results
--mode_check	This is used to check the results of an existing mGEMS analysis

3.4 Data Filtering for High-Quality Analysis

To ensure robust downstream comparisons, a strict filtering step was applied to all samples. Samples were retained only if they met all of the following criteria:

1. Total read count > 1,000,000.
2. High genome-wide coverage and
3. Mean coverage with a quality score of 4 (on a scale where 4 represents the lowest confidence).

A summary table was generated detailing the number of samples retained after applying each filter sequentially. Cluster assignment was not considered during this initial filtering stage.

3.5 Single Nucleotide Polymorphism (SNP) Analysis

Reads mapping and variant calling pipeline proceeds with genomic analysis; the alignment of the sequence to a reference genome is the essential first step, especially if the ultimate goal is to identify variants (Saada et al., 2024).

Reads were trimmed to remove adapter sequences and filter low-quality reads using FastP v.0.23.2 (Chen, 2023), an ultrafast all-in-one quality control tool for FASTQ data. Once trimmed, the reads were aligned to the reference genomes for both Coca and Single colony using the bowtie2 aligner tool, an highly-efficient and memory-saving tool for rapid alignment of read on a known reference (Langmead & Salzberg, 2012). The reference genomes were indexed before the alignment to organize the genome into a structured format that simplifies the search for specific regions (Cenzato & Lipták, 2024; Langmead & Salzberg, 2012). A SAM file, for each isolate, is obtained from the alignment and converted to sorted BAM format using Picard SortSam v.3.1.1 (https://broadinstitute.github.io/picard/). To reduce systematic errors in base quality scores and PCR-based errors, such as PCR-duplicates, BAM files were identified and marked using Picard MarkDuplicate. Genome Analysis Toolkit (GATK) v.4.4.2 HaplotypeCaller was used for variant calling (single nucleotide polymorphism (SNPs) and indels) with default parameters and specifying samples diploidy and the option -ERC GVCF (Poplin, et al., 2017). Sample diploid state was specified by the --sample-ploidy parameter and a genomic VCF intermediate file (GVCF) was generated as output. Differently from the VCF, the GVCF file contains information of both variant and non-variant sites and contains information about the confidence level in determining if the site matches the reference genome (https://gatk.broadinstitute.org/hc/en-us/articles/360035531812-GVCF-Genomic-Variant-Call-Format). The GVCFs were subsequently imported into a GATK GenomicsDB to merge GVCFs from multiple samples into a combined VCF (one for each Candida species) with the GenotypeGVCFs GATK function, which facilitates the joint genotyping on one or more samples pre-called with HaplotypeCaller.

Variants from the preliminary merged VCF were filtered through GATK VariantFiltration using the following parameters: DP<=20, QD<2.0, MQ <40.0, FS>60.0, MQRankSum < -12.5 and ReadPosRankSum<-8.0. SNPs were selected using GATK SelectVariants with --select-type-to-include SNP option. SNPs that had a less than minimum genome quality of 50 and sample depth of 30 were removed.

3.6 Phylogenetic Analysis of *C. parapsilosis* Clusters

For the *Candida parapsilosis* cluster C3, a maximum-likelihood phylogenetic tree was constructed. High-quality SNPs from both sweep samples and monocolony isolates were used. The tree was built with IQ-TREE 2 (Minh et al., 2020) using the MFP-GTR model and 1000 bootstrap replicates. Maximum likelihood is a statistical method that selects the best-fitting phylogenetic tree from a set of possible trees, based on the data and a substitution model (King & Van Doorslaer 2018). This analysis aimed to contextualize the CoCa (C3) strains from metagenomic sweeps within the genetic diversity of isolated colonies.

3.7 Software and Statistical Environment

All computational analyses described in this study were performed on a Linux environment. Software tools were installed and managed primarily via conda (Anaconda Inc.) using the conda-forge and bioconda channels, ensuring reproducibility and version control across the pipeline. The full analytical workflow spans metagenomic classification, reference database construction, read pseudoalignment, species and cluster abundance estimation, read binning, quality control, short-read alignment, variant calling, phylogenetic inference, within-sample diversity profiling, and statistical data visualisation. Each tool and environment is described below in the order in which it was applied (Software and Tools Used.pdf).

RESULTS

4.1 Kraken-based species detection and Candida selection

We investigated the epidemiology of candida colonisation at IRCCS Policlinico San Matteo Hospital, Italy, during the first wave of the COVID-19 pandemic, following the observation of a higher-than-expected incidence of these infections. The culture-based deep sequencing approach was applied to a total of 601 clinical samples, including nasal, rectal, and respiratory specimens, collected from 144 inpatients during the first COVID wave, from April to May 2020. Based on an initial screening of the species present in the samples, performed with Kraken, a broad range of microbial species was detected across the sequencing reads. 32 *Candida* species, 34 non-*Candida* fungal species, and 58 bacterial species were identified (**Figure 6**). The most notable taxa included *C. albicans*, *C. glabrata*, *C. krusei*, *Candida lusitaniae*, *C. parapsilosis*, *Acinetobacter baumannii*, and *Acinetobacter nosocomialis*, among others.

The detection of multiple *Candida* species across different sample types and wards confirms the complex and heterogeneous nature of fungal colonization in COVID-19 patients. The co-occurrence of multiple species in the same clinical environment marks the importance of broad surveillance strategies, particularly in immunocompromised and critically ill patients who are at elevated risk for invasive candidiasis.

Microbial Species Distribution Using Kraken

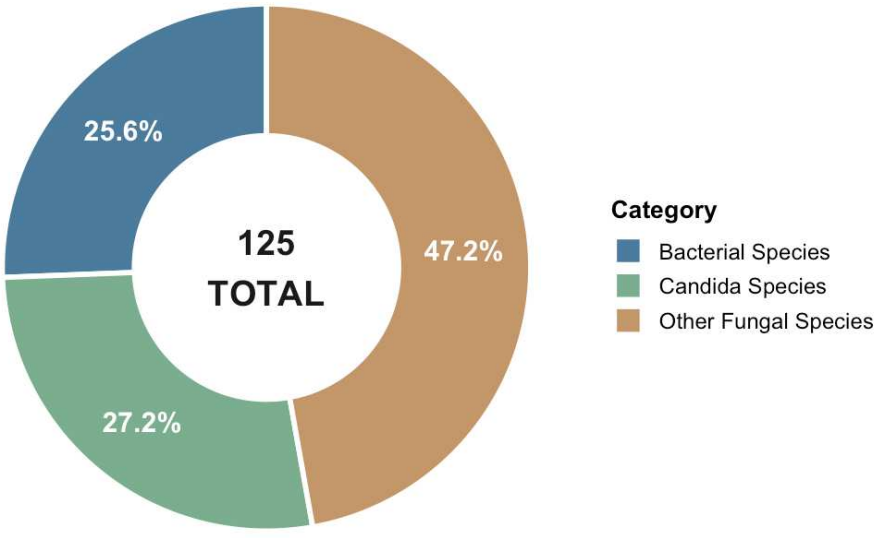

Figure 6: Distribution of microbial species identified using Kraken. Bacterial species constitute the largest proportion (n = 59), followed by other fungal species (n = 34) and *Candida* species (n = 32).

4.2 Sub-Clustering of Selected Species Using FastBAPS

C. albicans, C. parapsilosis and *C. glabrata* were chosen as focal species because of their frequent detection at San Matteo hospital and their involvement in multiple cases of candidemia between 2015 and 2020. Global population structure within each of these three species was thus investigated at a finer scale through sub-clustering analysis using FastBAPS on a dataset of published genomes (Tonkin-Hill et al., 2019).

Maximum-likelihood phylogenetic trees were constructed for each species from published whole-genome sequence data, and FastBAPS was subsequently applied to infer population structure. The clustering algorithm partitioned each phylogeny into distinct sub-clusters (designated C1-C10, depending on the species) that represent the major lineages circulating within the hospital.

For each of the three species, the reference genome database was sub-clustered using FastBAPS, which grouped publicly available reference genomes into distinct genetic lineages based on hierarchical Bayesian partitioning. For *C. parapsilosis*, this yielded six reference clusters (C1–C6); for *C. albicans* two clusters (C1-C2) and for *C. glabrata* nine clusters (C1-C10) were identified. These cluster labels were then incorporated into the demix reference database. The 601 clinical samples were subsequently analysed using the demix pipeline. mSWEEP for species-level abundance estimation and mGEMS for read binning which assigned each sample's sequencing reads to the appropriate cluster based on their match to the pre-labelled reference genomes.

4.3 Distribution of *Candida* and other species across hospital wards and sample types

4.3.1 Distribution of *Candida* and other microbial species across sample types

The analysis of the 601 clinical samples across rectal, nasal, respiratory swabs, revealed rectal samples as the one yielding the highest number of colonization, followed by nasal and

respiratory samples **(Figure 7).** *C. parapsilosis* was the most abundant species in both nasal and rectal samples, while *C. albicans* was consistently the second most detected species across all three sample types. *C. glabrata* showed a notably higher presence in rectal samples compared to nasal and respiratory samples, consistent with its known gastrointestinal niche (Beardsley et al., 2024). Among the bacterial species, *Acinetobacter baumannii* and *Acinetobacter nosocomialis* were prominently detected, particularly in nasal samples, reflecting their known role as nosocomial respiratory pathogens. Beyond these dominant species, a diverse range of additional fungal and bacterial species were detected across all sample types, indicating that the microbial landscape of these hospitalised COVID-19 patients extended well beyond the primary species of interest.

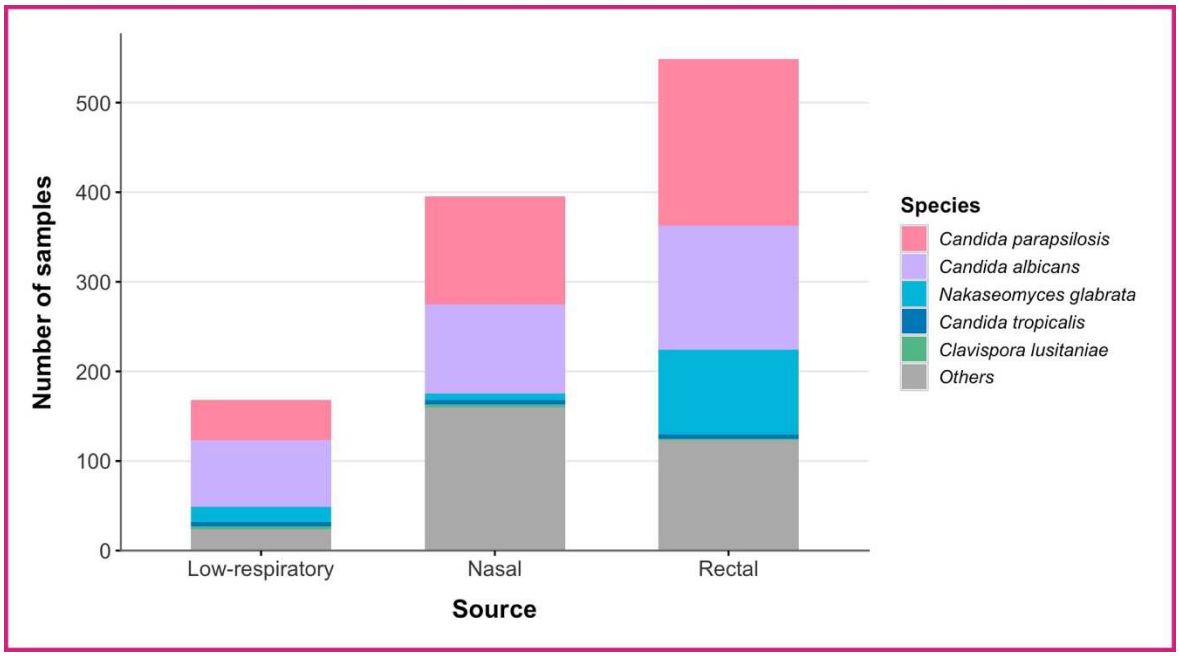

Figure 7: Barplot showing the number of colonizations for the most frequent microbial species characterized across the three different sample types, collected from 144 hospitalized patients. Species with fewer than 30 cases were grouped as "Others".

4.3.2 Distribution of *Candida* and other microbial species across wards

ICU samples accounted for the large majority of total detections (ICU- 2342 & Other wards-774) across all species (**Figure 8**). *C. parapsilosis* emerged as the most abundant species in the ICU, followed by *C. albicans*. In ordinary wards, *C. albicans* constituted the majority of the isolates, while *C. parapsilosis* represented a lower proportion than that observed in the ICU.

C. glabrata, *C. tropicalis* and a broader collection of other fungal and bacterial species were detected in both ward types but were disproportionately more abundant in the ICU setting. This pattern is consistent with the known vulnerability of ICU patients. These distribution patterns highlight the complex and diverse microbial burden in ICU patients and mark the ecological dominance of *C. parapsilosis* in the most critically ill individuals.

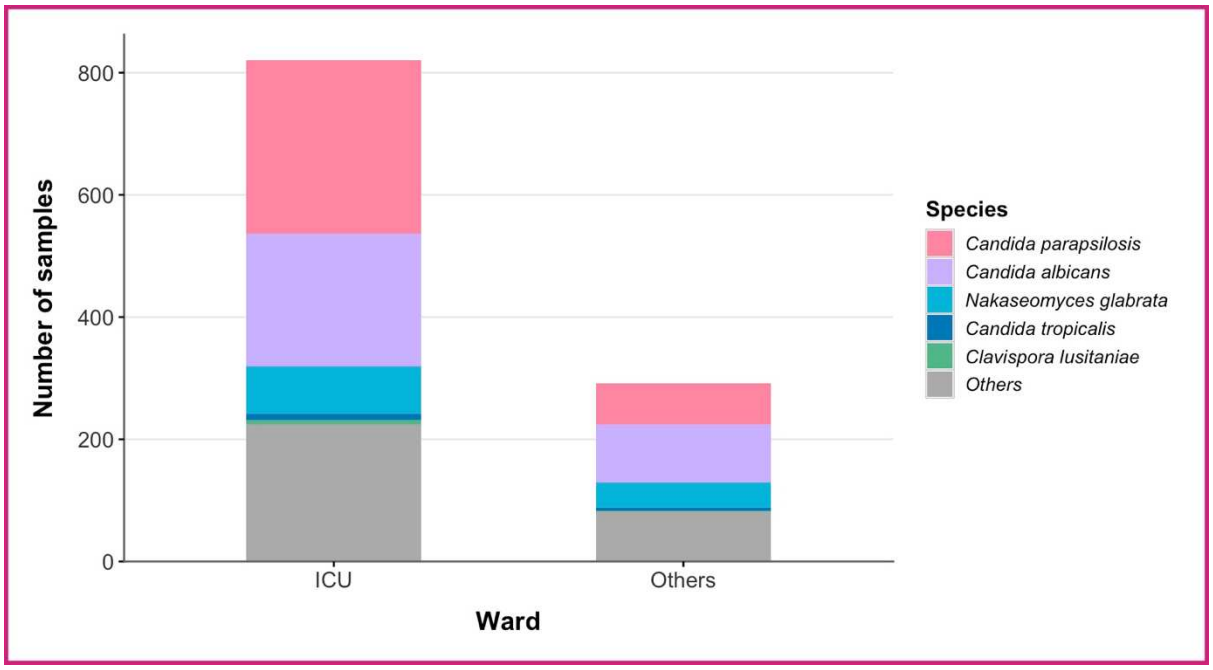

Figure 8: Barplot showing the number of detections for each microbial species across ICU and ordinary wards collected from 144 hospitalised patients. Species with fewer than 30 cases were grouped as "Others".

4.3.3 *Nakaseomyces glabrata (C. glabrata)* cluster distribution across sample types

Across sample types, *C. glabrata* showed a strongly rectal-biased distribution, with the majority of detections occurring in rectal samples (~117), compared to nasal (~22) and respiratory (~28) samples (**Figure 10**). All nine clusters were detected in rectal samples, with C9 being the most abundant, followed by C8, C6, C3, C5, C7, C2, C10, and C1. Nasal and respiratory samples showed much lower total counts and a similarly diverse but compressed cluster distribution. This highly uneven distribution across sample types with rectal samples dominating confirms that *C. glabrata* predominantly occupies the gastrointestinal tract in these patients, with minimal presence in the upper respiratory tract (Beardsley et al., 2024). The consistent multi-cluster pattern across all three sample types rules out a single transmitted strain and instead points to diverse endogenous gut colonisation as the main source.

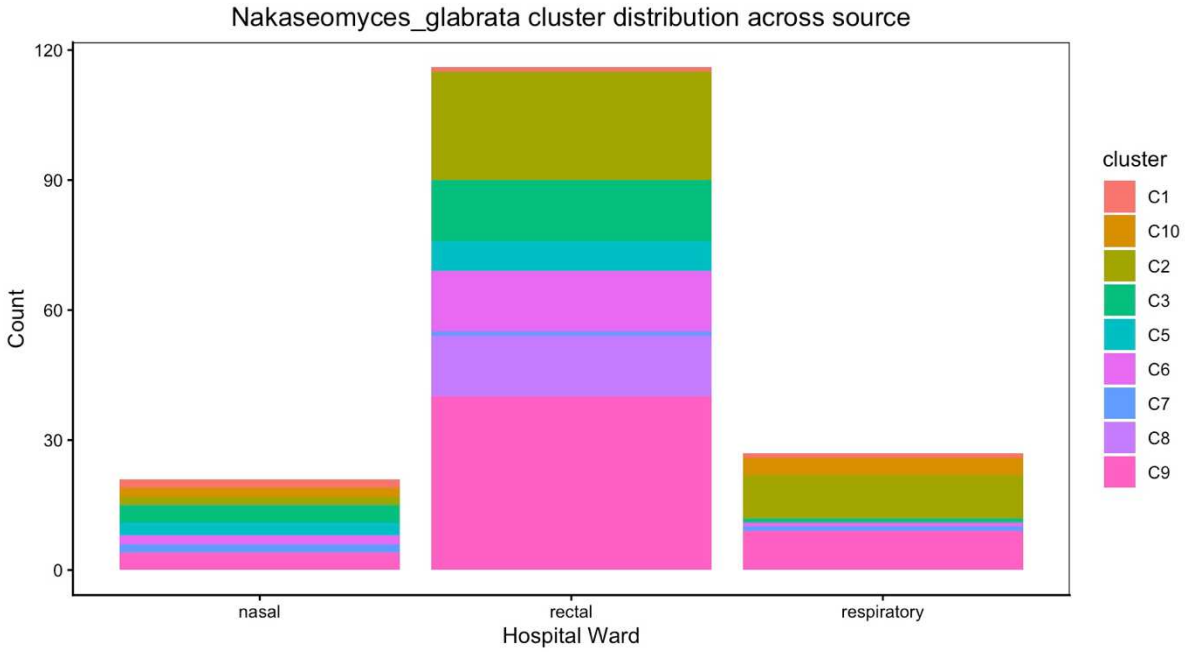

Figure 10: *Candida glabrata* cluster distribution across clinical sample types. Rectal samples overwhelmingly dominate (~117 total), with all nine clusters (C1, C2, C3, C5, C6, C7, C8, C9, C10) represented, and C9 as the most abundant. Nasal (~22) and respiratory (~28) samples show substantially lower counts with a similarly diverse but compressed cluster representation. This striking rectal predominance confirms the gastrointestinal tract as the primary ecological niche of *C. glabrata* in this patient cohort.

4.3.4 *Nakaseomyces glabrata (C. glabrata)* cluster distribution across hospital wards

The distribution of *C. glabrata* across hospital wards showed higher detections in ICU (90) compared to ordinary wards (70) (**Figure 9**). Importantly, nine distinct clusters were identified as C1-C10 and all nine were present in both ward types. No single cluster dominated. Instead, multiple clusters co-circulated in roughly similar proportions across both ICU and ordinary wards. Cluster C9 appeared as the largest contributor in the ICU, but no cluster came close to the near-complete dominance. Samples from C4 do not appear because they were excluded during the filtering step (see Materials and Methods). This broad cluster diversity in both wards confirms that *C. glabrata* is not undergoing clonal nosocomial spread, but rather reflects diverse strains colonising patients independently.

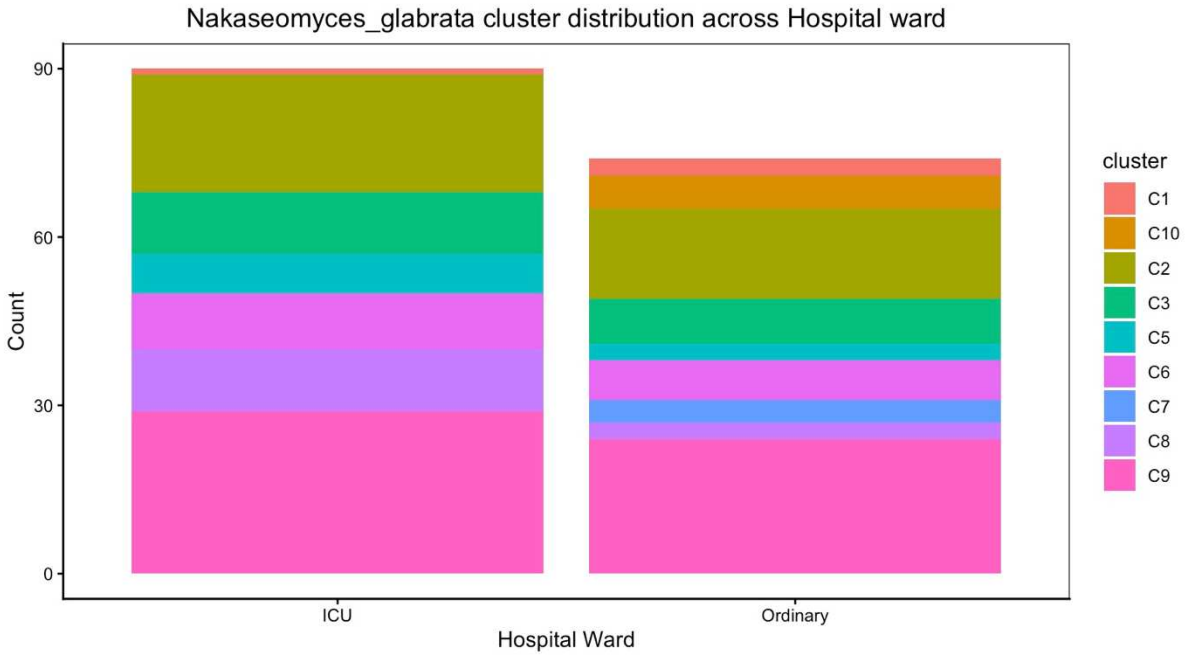

Figure 9: *Candida glabrata* cluster distribution across ICU and ordinary wards. Nine co-circulating clusters are detected in both ward types, with ICU samples showing slightly higher total counts (~90) compared to ordinary wards (~70). Unlike *C. parapsilosis*, no single cluster dominates; C9 (pink) is the largest contributor in the ICU but accounts for only a modest fraction of total detections. The broadly similar multi-cluster composition across both wards reflects diverse, endogenous gut-reservoir colonisation rather than clonal hospital transmission.

4.3.5 *Candida albicans* cluster distribution across sample types

Across sample types, *C. albicans* was detected in nasal (~165), rectal (~195), and respiratory (~115) samples, with rectal samples showing the highest count (**Figure 11**). Both clusters C1 and C2 were present across all three sample types in broadly similar proportions, with C2 consistently making up a substantial portion of detections at every site. The relatively even distribution of both clusters across nasal, rectal, and respiratory samples is consistent with *C. albicans* being a pan-mucosal commensal which is naturally present across multiple body sites simultaneously. Unlike *C. glabrata*, which is strongly restricted to the gut, and unlike *C. parapsilosis*, as shown later, which is dominated by a single clone, *C. albicans* shows broad anatomical spread with no site-specific or cluster-specific pattern. This distribution is consistent with the opportunistic colonization behaviour of *C. albicans* and suggests the absence of a single dominant lineage associated with a particular anatomical niche.

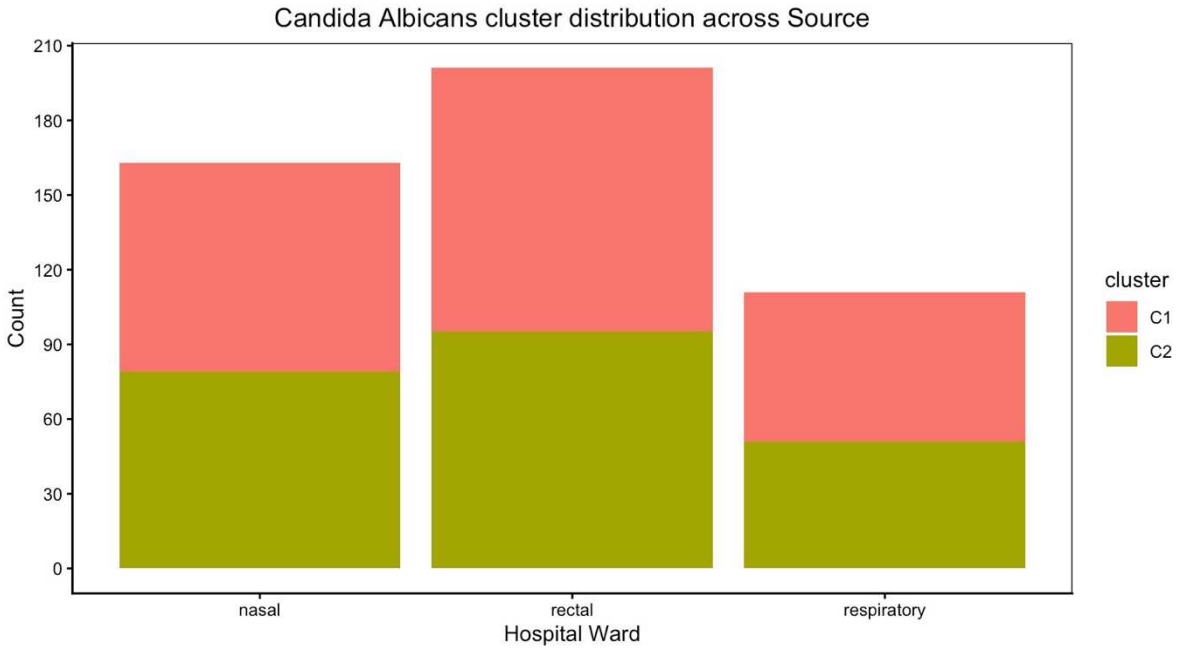

Figure 11: *Candida albicans* cluster distribution across clinical sample types. Two co-circulating clusters C1 (pink) and C2 (olive green) are detected across all three sample types, with rectal samples showing the highest total count (~195), followed by nasal (~165) and respiratory (~115). Both clusters are present in broadly similar proportions in each anatomical site, with no single cluster dominating, reflecting the pan-mucosal and genetically diverse nature of *C. albicans* colonisation in this patient cohort.

4.3.6 *Candida albicans* cluster distribution across hospital wards

Candida albicans was detected across both ward types, with higher counts in the ICU (~320) compared to ordinary wards (~140) (**Figure 12**). Two clusters were identified C1 and C2, both present in similar proportions in both ward types. Neither cluster dominated exclusively. C2 made up roughly half the detections in each ward, with C1 accounting for the remaining portion. The broadly similar cluster proportions between ICU and ordinary wards suggest that *C. albicans* colonisation is not driven by a single hospital-adapted clone spreading between patients, but rather by diverse strains colonising patients from their own endogenous sources.

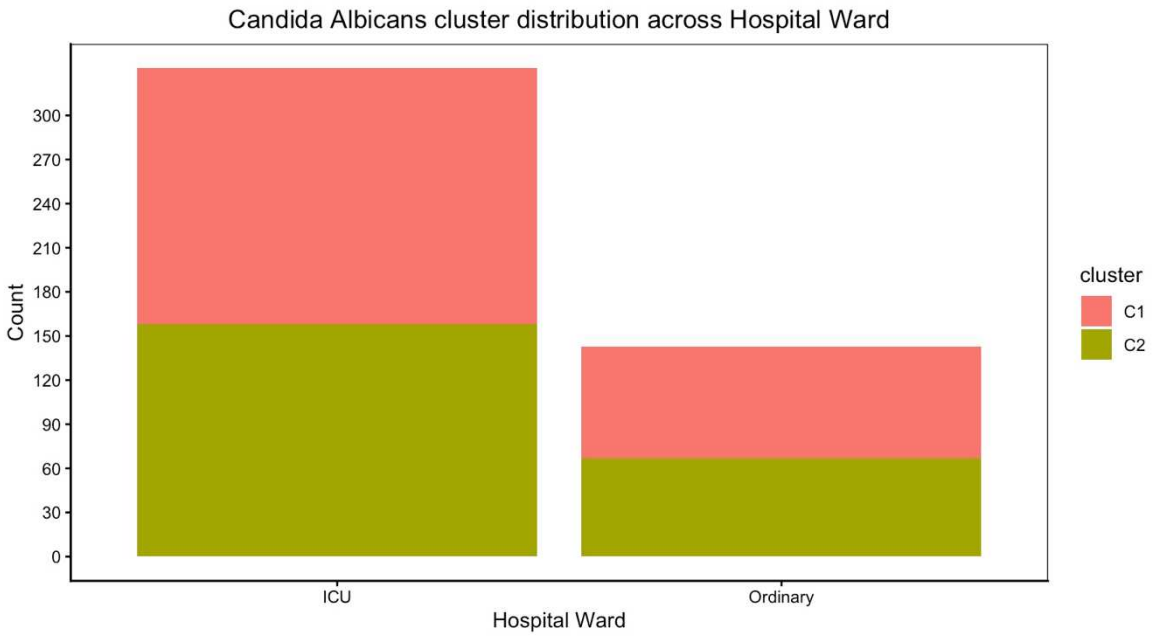

Figure 12: *Candida albicans* cluster distribution across ICU and ordinary wards. Both clusters C1 and C2 are present in comparable proportions in both ward types, with ICU samples showing a higher total count (~320) compared to ordinary wards (~140). The absence of a dominant single cluster in either ward type distinguishes *C. albicans* from the striking C3-dominated pattern observed for *C. parapsilosis*, and is consistent with diverse endogenous colonisation rather than clonal nosocomial spread.

4.3.7 *Candida parapsilosis* cluster distribution across sample types

The distribution of *C. parapsilosis* clusters was examined across the three sample types. As shown in (**Figure 13**), the distribution of *C. parapsilosis* was characterized by a marked predominance of cluster C3, which represents the main circulating lineage, while the remaining lineages are only detected at low frequency. This pattern differs from that observed in *C. albicans* and *C. glabrata*, which showed a more even representation of all identified clusters across samples.

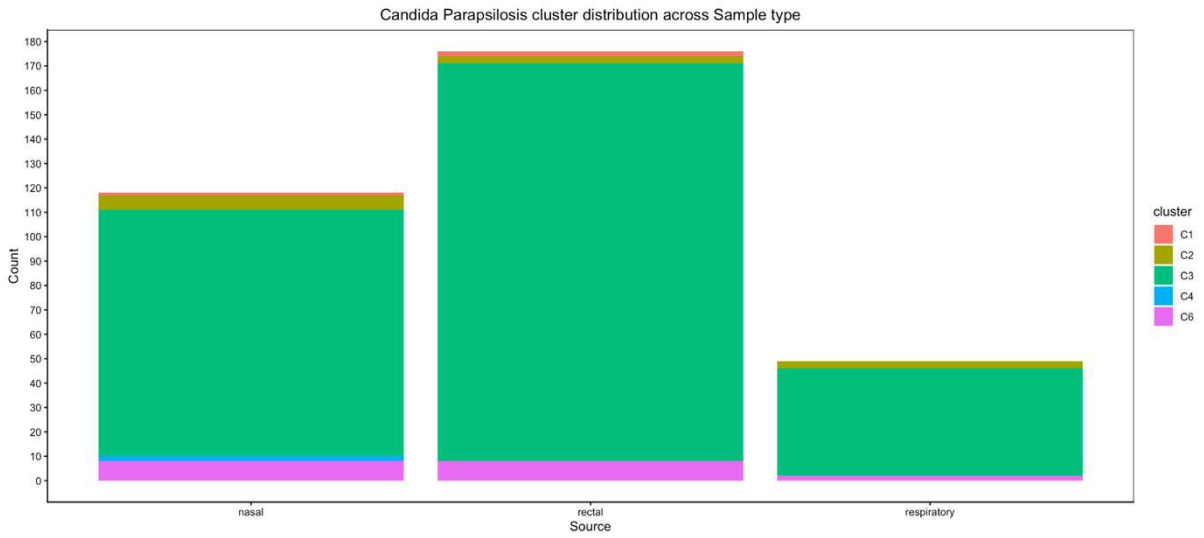

Figure 13: *Candida parapsilosis* cluster distribution across different sample types. Rectal samples show the highest total count (175), followed by nasal (118) and respiratory (49) samples. Cluster C3 (green) dominates across all three sources. Minor contributions from clusters C1, C2, and C6 are visible at the top of each bar.

Rectal samples exhibited the highest total count of *C. parapsilosis* detections (175), followed by nasal samples (118) and respiratory samples (49). The predominance of *C. parapsilosis* in rectal samples is consistent with gastrointestinal colonization as a primary reservoir.

The widespread presence of cluster C3 across all three anatomical sites supports the idea of a single, highly successful clone driving *C. parapsilosis* colonization in the patient population. Minor clusters (C1, C2, C4 and C6) were detected at low frequencies across sample types, indicating that although C3 is the prevailing lineage, complete clonal replacement has not occurred and multiple lineages continue to co-circulating at low abundance, potentially representing independent colonization events.

4.3.8 *Candida parapsilosis* Cluster Distribution Across Hospital Wards

The distinctive distribution of *Candida parapsilosis* C3 was also evident across hospital wards, with a strong prevalence of C3 in both ordinary and ICU, as shown in (**Figure 14)**.

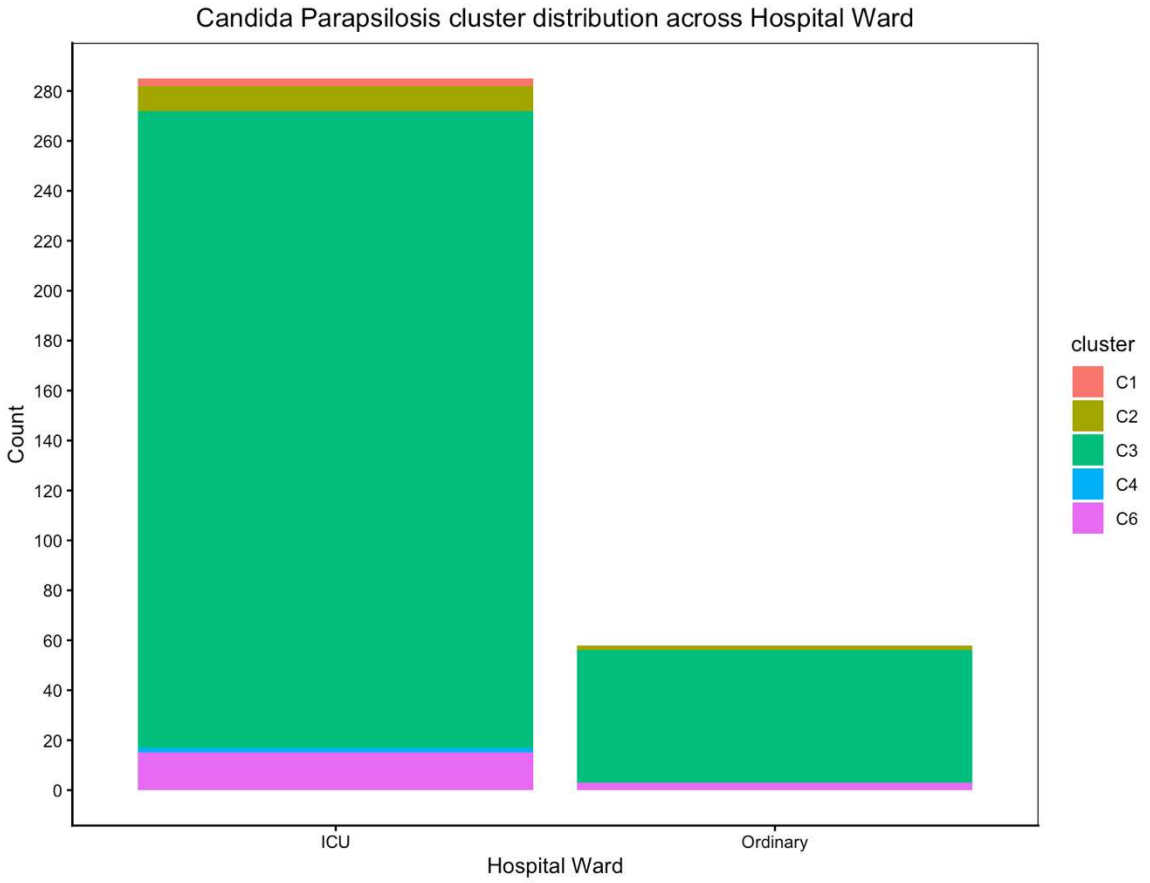

Figure 14: *Candida parapsilosis* cluster distribution across hospital wards (grouped by ward type). Cluster C3 (green) overwhelmingly dominates in the ICU (285 total), with C6 and C2 present at very low frequencies. The ordinary ward shows a similar C3-dominant pattern at a lower overall count (59).

The dominance of cluster C3 in **(Figure 14)** is remarkable: it accounts for the vast majority of all *C. parapsilosis* detections across both ward types, with only minor contributions from C1, C2, C4, and C6. This pattern strongly suggests the expansion of a single successful lineage of *C. parapsilosis* within the hospital environment.

The biological interpretation of these findings is noteworthy. *C. parapsilosis* is distinguished among major *Candida* species by its ability to form robust biofilms on abiotic surfaces, which can contribute to persistence in hospital environments despite relatively low MICs to echinocandins in planktonic cells (Govrins & Lass-Flörl, 2024). The intensive care settings, characterized by frequent use of invasive devices and close patient contact, may therefore facilitate the persistence and selection of a single and successful lineage. The presence of C3

across both ward types in **(Figure 14)** raises the possibility that this lineage was already circulating as a colonizing strain prior to the COVID-19 pandemic. This observation is consistent with the presence of an azole-resistant and persistent strain previously reported at San Matteo Hospital since 2018.

4.4 Phylogenetic analysis of *C. parapsilosis* C3 and its relationship to candidemia-causing strains

To assess whether the *C. parapsilosis* strains associated with colonization in hospitalized patients during the study period corresponds to the azole-resistant and persistent clone responsible for multiple candidemia cases between 2018 and 2020, a phylogenetic analysis was performed combining isolates from the present study with those from the previous published dataset (Vumbaca et al; 2024). The phylogenetic tree built from whole-genome SNP data showed that the colonizing *C. parapsilosis* C3 (labeled as CoCa) isolates form a large and genomically identical group, represented by the pink arc that runs around most of the circle (**Figure 15**). Importantly, the CoCa samples in this study represent colonisation isolates collected from rectal, nasal and respiratory swabs from hospitalised patients and are distinct from the candidemia-causing isolates, which are represented by Group B, Group A, and the sporadic cases in the phylogeny.

The key finding of this analysis is that samples that cause colonisation, clusterize within Group B samples in the phylogenetic tree (**Figure 15**). Group B represents the main nosocomial outbreak strain responsible for candidemia at San Matteo Hospital between 2018 and 2020, as documented by Vumbaca et al. 2024. The clustering of CoCa colonisation samples within this group, rather than forming a distinct branch, provides evidence that the *C. parapsilosis* C3 lineage is the same persistent, azole-resistant strain detected and circulating in the hospital since 2018.

Looking more closely at the three candidemia groups in the phylogeny, each tells a different story. Group B samples clusterize closely together, indicating very low genetic divergence among them. This pattern is consistent with transmission of a single strain between patients, potentially mediated by healthcare workers or contaminated equipment, which are recognized routes of nosocomial spread for *C. parapsilosis* in hospital settings (Franconi et al., 2023). Group A isolates also form a small, compact cluster, suggesting a separate but less successful

transmission event. Lastly, sporadic cases did not form a monophyletic group, suggesting multiple independent emergence events occurred within the hospital over time.

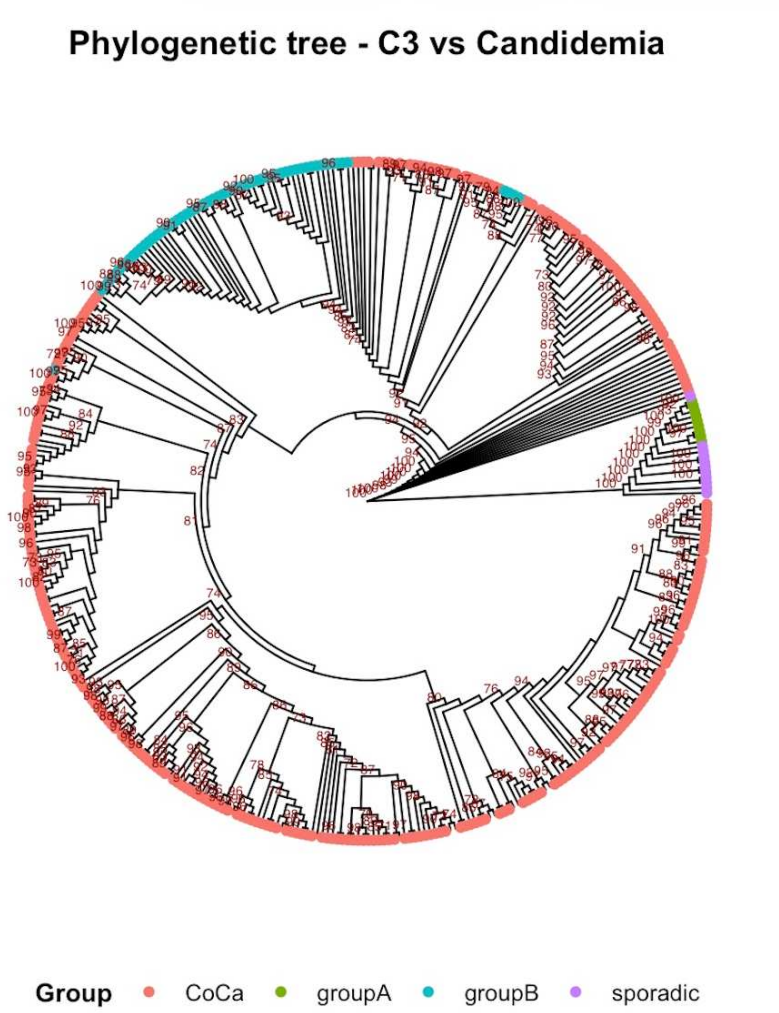

Figure 15: Phylogenetic tree showing relationship between *C. parapsilosis* C3 (CoCa) colonising isolates and candidemia-causing strains. The tree was inferred from whole-genome SNP data. Isolates are colour-coded by group: CoCa (pink), Group A (olive green), Group B (cyan), and sporadic candidemia (purple). Bootstrap support values ≥ 70 are displayed at internal nodes.

As previously described, the clustering of CoCa colonisation samples together with Group B isolates supports their relatedness. This finding provides evidence the strain colonizing patients is closely related to the clone previously associated with candidemia, suggesting that this lineage is still circulating within the hospital. This has direct implications for infection control: the fact that this clone continues to colonise patients means that standard measures for treating candidemia after it occurs are insufficient. Identifying colonised high-risk clones early is

essential to interrupt the chain from silent colonisation to bloodstream infection. Stopping cluster-related spread requires strict hand hygiene and decontamination of shared equipment, while reducing sporadic cases requires identifying which patients are already colonised with *C. parapsilosis* and considering early antifungal intervention (Douglas et al., 2023).

Overall, this phylogenetic analysis provides clear genomic evidence that the *C. parapsilosis* C3 clone occupying the mucosal niches of hospitalized patients during the first COVID wave at San Matteo Hospital is the same persistent, azole-resistant strain responsible for the nosocomial candidemia outbreak of 2018–2020.

DISCUSSION AND CONCLUSION

This study applied a metagenomic sweep approach combined with the demix_check hierarchical pipeline to characterise *Candida* spp. colonisation at strain-level resolution across 601 clinical samples collected during the first COVID-19 wave at the San Matteo Hospital, Pavia. The obtained results show how the novel sweep metagenomics approach can resolve fungal population structure at the level of sequence clusters. Samples collected by plate sweeps were sequenced with Illumina and then reads were analyzed using the demix_check pipeline to investigate *Candida* population structure at strain level in complex clinical samples. Among the detected species, *C. parapsilosis*, *C. albicans*, and *C. glabrata,* emerged as the most frequent in rectal, nasal and respiratory samples from hospitalized patients, consistent with previously described colonization patterns in clinical populations.

C. glabrata showed a strongly rectal-biased distribution (~117 detections in rectal samples versus ~22 nasal and ~28 respiratory), consistent with its established gastrointestinal niche (Beardsley et al., 2024). Importantly, the nine detected clusters were identified across both ICU and ordinary wards, without one cluster being more prevalent than the others. C9 was the most abundant in the ICU, but accounted for only a modest fraction of total detections. The distribution of multiple *C. glabrata* clusters across ICU and ordinary ward is inconsistent with the widespread transmission of a single hospital-associated lineage. Rather, the observed population structure suggests independent colonization events arising from genetically diverse strains carried by individual patients. This interpretation is supported by the ecological characteristics of *C. glabrata*. As a common gastrointestinal commensal, the species is frequently acquired from endogenous reservoirs, which may explain the presence of multiple genetically distinct lineages among hospitalized patients (Romo et al., 2020).

C. albicans was detected across all three sample types: nasal (~165), rectal (~195), and respiratory (~115), with two co-circulating clusters (C1 and C2) present in similar proportions in the three different body sites examined in the study. Neither cluster dominated, and the cluster composition was comparable between ICU (~320) and ordinary wards (~140). This even distribution across hospital wards and anatomical sites is consistent with *C. albicans* behaving as a widespread mucosal commensal. The observed genomic diversity indicates that *C. albicans* circulates within the hospital as a *C. albicans* is the most common commensal

fungal species in the human microbiota, known for its broad colonisation of mucosal surfaces, and the genomic diversity observed here confirms that within this hospital setting it circulates as a heterogeneous population without the sign of nosocomial outbreak dynamics. The higher counts in ICU samples reflect the greater vulnerability of critically ill patients to fungal overgrowth under immunosuppression and antibiotic pressure, rather than hospital-adapted clonal spread. This result validates the pipeline's ability to characterise a diffuse commensal species and distinguishes its signal from that of a nosocomial pathogen.

5.1 *Candida parapsilosis*: A Persistent Clone and Its Hidden Reach

The most significant and clinically urgent finding of this study concerns *C. parapsilosis*. Among the three major species, *C. parapsilosis* stood apart in a way that was immediately visually apparent in the distribution data and progressively more alarming as the analysis deepened. A single genomic cluster (cluster C3) accounted for the overwhelming majority of *C. parapsilosis* detections across all three sample types and both ward categories. Across nasal, rectal, and respiratory specimens, across ICU and ordinary wards, cluster C3 represented approximately 90–95% of observed detections. This is not the pattern of a commensal organism passively residing in its ecological niche. It is the signature of a successful, hospital-adapted clone actively expanding across a patient population.

What makes this finding more than an epidemiological curiosity is the phylogenetic context. When the C3 sweep-derived sequences were placed on a maximum-likelihood phylogenetic tree alongside single-colony isolates from candidemia patients, the result was clear. The colonizing strains and the bloodstream infection-causing isolates are part of the same persistent clone. This is the same azole-resistant lineage that has been circulating at San Matteo Hospital since at least 2018 (Vumbaca et al., 2024). The COVID-19 pandemic did not introduce a new pathogen to this hospital but it amplified the conditions in which an already established, already resistant, and already dangerous clone could expand its reach. The extraordinary physiological stress of severe COVID-19 systemic inflammation, prolonged mechanical ventilation, high-dose corticosteroids, disrupted gut microbiomes, and compromised epithelial barriers created the environment that allowed a fungus usually living harmlessly on surfaces to enter the bloodstream.

The pan-mucosal colonization achieved by C3 is itself a remarkable biological capability. Which means a single lineage colonizes nasal passages, gastrointestinal mucosa, and respiratory tract and has the ability to adhere, survive, and compete across physiologically

distinct environments or the disruption of normal anatomical barriers during ICU care. Either explanation has serious implications: this clone does not occupy a single ecological niche that can be targeted by a specific intervention, it inhabits the patient as a whole.

The minor clusters (C1, C2, C4, C6) detected at low frequencies across sample types represent another important observation. Their persistence at the margins of the dominant C3 population indicates that the hospital has not undergone complete clonal replacement which means genetic diversity persists, including lineages that may carry different resistance profiles or virulence characteristics. Monitoring these minority populations over time is clinically important: the history of antimicrobial resistance teaches us that today's minority variant can become tomorrow's dominant strain under the selective pressure of antifungal treatment.

5.2 Looking Forward

The findings of this thesis point toward several concrete directions for future research. Longitudinal surveillance studies using this pipeline, tracking colonization and resistance evolution in high-risk patients from ICU admission through discharge, would transform the current cross-sectional snapshot into a dynamic map of nosocomial *Candida* transmission. Integration of antifungal susceptibility data obtained from both clinical isolates and from the sweep-derived genomic data through resistance gene characterisation would add the clinical management dimension that this genomic epidemiology study.Beyond *Candida*, this study contributes to a growing body of evidence that sweep metagenomics combined with demix_check probabilistic approach represents a powerful tool for the epidemiology of any pathogen population capable of hospital-adapted clonal expansion. The approach is not species-specific as it is a general approach to evaluate more carefully the microbial communities that inhabit patients and hospitals.

REFERENCES

1. Abdel-Hamid, R. M., El-Mahallawy, H. A., Abdelfattah, N. E., & Wassef, M. A. (2023). The impact of increasing non-albicans Candida trends on diagnostics in immunocompromised patients. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology], 54(4), 2879–2892.
2. Alves, R., Barata-Antunes, C., Casal, M., Brown, A. J. P., ... & Paiva, S. (2020). Adapting to survive: How Candida overcomes host-imposed constraints during human colonization. PLoS pathogens, 16(5).
3. Amann, V., Kissmann, A. K., Firacative, C., & Rosenau, F. (2025). Biofilm-Associated Candidiasis: Pathogenesis, Prevalence, Challenges and Therapeutic Options. Pharmaceuticals (Basel, Switzerland), 18(4), 460.
4. Anderson, F. M., Visser, N., Amses, K., Hodgins-Davis, A., ... & O'Meara, T. R. (2022). Human commensal Candida albicans strains demonstrate substantial within-host diversity and retained pathogenic potential. bioRxiv, 2022-09.
5. Anderson, M. Z., & Dietz, S. M. (2024). Evolution and strain diversity advance exploration of Candida albicans biology. mSphere, 9(8), e0064123.
6. Andrzejuk, P., Tokarska-Rodak, M., Dyrda, A., & Zarębska, M. (2024). Phenotypic and genotypic characterization of Candida species from the oral cavity of healthy individuals in Lublin province, Poland. Journal of oral microbiology, 17(1), 2437335.
7. Angoulvant, A., Guitard, J., & Hennequin, C. (2016). Old and new pathogenic Nakaseomyces species: epidemiology, biology, identification, pathogenicity and antifungal resistance. *FEMS Yeast Research*, *16*(2), fov114.
8. Arafa, S. H., Elbanna, K., Osman, G. E. H., & Abulreesh, H. H. (2023). Candida diagnostic techniques: a review. Journal of Umm Al-Qura University for Applied Sciences, 9(3), 360–377.
9. Arastehfar, A., Carvalho, A., Nguyen, M. H., Hedayati, M. T., … & Hoenigl, M. (2020). COVID-19-associated candidiasis (CAC): an underestimated complication in the absence of immunological predispositions?. *Journal of Fungi*, *6*(4), 211.
10. Ayon, N. J. (2023). High-throughput screening of natural product and synthetic molecule libraries for antibacterial drug discovery. Metabolites, 13(5), 625.
11. Bayat, A., Gaëta, B., Ignjatovic, A., & Parameswaran, S. (2017). Improved VCF normalization for accurate VCF comparison. Bioinformatics , 33(7), 964–970.

12. Beardsley, J., Kim, H. Y., Dao, A., Kidd, S., ... & Alffenaar, J. W. (2024). Candida glabrata (Nakaseomyces glabrata): A systematic review of clinical and microbiological data from 2011 to 2021 to inform the World Health Organization Fungal Priority Pathogens List. Medical mycology, 62(6), myae041.
13. Bennett, R. J., Forche, A., & Berman, J. (2014). Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harbor perspectives in medicine*, *4*(10), a019604.
14. Bhattacharya, S., Sae-Tia, S., & Fries, B. C. (2020). Candidiasis and Mechanisms of Antifungal Resistance. Antibiotics (Basel, Switzerland), 9(6), 312.
15. Bisso, Y. I., Duron, O., Plantard, O., Prevost, G., ... & Boyer, P. (2025). High throughput sequencing in the diagnosis and surveillance of tick-borne diseases: A narrative review. One Health, 101312.
16. Borman, A. M., & Johnson, E. M. (2023). Changes in fungal taxonomy: mycological rationale and clinical implications. Clinical microbiology reviews, 36(4), e0009922.
17. Branco, J., Miranda, I. M., & Rodrigues, A. G. (2023). Candida parapsilosis virulence and antifungal resistance mechanisms: a comprehensive review of key determinants. *Journal of Fungi* , *9* (1), 80.
18. Butler, G., Rasmussen, M. D., Lin, M. F., Santos, M. A., ... & Cuomo, C. A. (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature*, *459*(7247), 657-662.
19. Caetano, C. F., Gaspar, C., Martinez-de-Oliveira, J., Palmeira-de-Oliveira, A., & Rolo, J. (2023). The role of yeasts in human health: a review. Life, 13(4), 924.
20. Carolus, H., Pierson, S., Lagrou, K., & Van Dijck, P. (2020). Amphotericin B and other polyenes—discovery, clinical use, mode of action and drug resistance. Journal of Fungi, 6(4), 321.
21. Cavalheiro, M., & Teixeira, M. C. (2018). Biofilms: Threats, Challenges, and Promising Strategies. Frontiers of Medicine, 5, 28.
22. Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta, 2(2), e107.
23. Chen, X.-F., Hou, X., Xiao, M., Zhang, L., … & Hsueh, P.-R. (2021). Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) Analysis for the Identification of Pathogenic Microorganisms: A Review. Microorganisms, 9(7), 1536.

24. Chow, E. W. L., Pang, L. M., & Wang, Y. (2021). From Jekyll to Hyde: The Yeast–Hyphal Transition of Candida albicans. Pathogens, 10(7), 859.
25. Ciurea, C. N., Kosovski, I. B., Mare, A. D., Toma, F., …& Man, A. (2020). Candida and Candidiasis-Opportunism Versus Pathogenicity: A Review of the Virulence Traits. Microorganisms, 8(6), 857.
26. Corbu, V. M., Gheorghe-Barbu, I., Dumbravă, A. Ș., Vrâncianu, C. O., & Șesan, T. E. (2023). Current Insights in Fungal Importance-A Comprehensive Review. Microorganisms, 11(6), 1384.
27. Cortegiani, A., Misseri, G., Fasciana, T., Giammanco, A., … & Chowdhary, A. (2018). Epidemiology, clinical characteristics, resistance, and treatment of infections by Candida auris. Journal of Intensive Care, 6(1), 1-13.
28. Corzo-Leon, D. E., Peacock, M., Rodriguez-Zulueta, P., Salazar-Tamayo, G. J., & MacCallum, D. M. (2021). General hospital outbreak of invasive candidiasis due to azole-resistant Candida parapsilosis associated with an Erg11 Y132F mutation. *Medical mycology*, *59*(7), 664-671.
29. Crotty, M. P., Akins, R., Nguyen, A., Slika, R., Rahmanzadeh, K., Wilson, M. H., & Dominguez, E. A. (2020). Investigation of subsequent and co-infections associated with SARS-CoV-2 (COVID-19) in hospitalized patients. *medRxiv*, 2020-05.
30. Czajka, K. M., Venkataraman, K., Brabant-Kirwan, D., Santi, S. A., … & Tharmalingam, S. (2023). Molecular Mechanisms Associated with Antifungal Resistance in Pathogenic Candida Species. Cells, 12(22), 2655.
31. De-la-Pinta, I., Marcos-Arias, C., Sevillano, E., Eraso, E., & Quindós, G. (2026). Emergent Candida Species on Healthcare Surfaces: Abiotic Reservoirs as a Source of Invasive Candidiasis. Microorganisms , 14 (2), 367.
32. d'Enfert, C., Kaune, A. K., Alaban, L. R., Chakraborty, S., … & Brown, A. J. P. (2021). The impact of the Fungus-Host-Microbiota interplay upon Candida albicans infections: current knowledge and new perspectives. FEMS microbiology reviews, 45(3), fuaa060.
33. Deorukhkar, S. C. (2017). Virulence traits contributing to pathogenicity of candida species. Journal of Microbiology & Experimentation, 5(1).
34. Dire, O., Ahmad, A., Duze, S., & Patel, M. (2023). Survival of Candida auris on environmental surface materials and low-level resistance to disinfectant. Journal of Hospital Infection, 137, 17-23.

35. Dos Santos, M. M., & Ishida, K. (2023). We need to talk about Candida tropicalis: Virulence factors and survival mechanisms. Medical mycology, 61(8), myad075.
36. Douglas, A. P., Stewart, A. G., Halliday, C. L., & Chen, S. C. A. (2023). Outbreaks of fungal infections in hospitals: epidemiology, detection, and management. Journal of Fungi, 9(11), 1059.
37. Ene, I. V., & Bennett, R. J. (2014). The cryptic sexual strategies of human fungal pathogens. *Nature Reviews Microbiology*, *12*(4), 239-251.
38. Farhan, M. S., Abdullah, B. A., Mamdwooh, A. E., & Numan, R. S. (2024). Review of Virulence Factors in Candida. Journal for Research in Applied Sciences and Biotechnology, 3(2), 75–82.
39. Figueroa-Bossi, N., Balbontín, R., & Bossi, L. (2022). Basic bacteriological routines. Cold Spring Harbor Protocols, 2022(10), pdb-prot107849.
40. Fischer, M., Strauch, B., & Renard, B. Y. (2017). Abundance estimation and differential testing on strain level in metagenomics data. Bioinformatics, 33(14), i124-i132.
41. Fisher, M. C., Gurr, S. J., Cuomo, C. A., Blehert, D. S., ... & Cowen, L. E. (2020). Threats posed by the fungal kingdom to humans, wildlife, and agriculture. MBio , 11 (3), 10-1128.
42. Franconi, I., Rizzato, C., Poma, N., Tavanti, A., & Lupetti, A. (2023). Candida parapsilosis sensu stricto antifungal resistance mechanisms and associated epidemiology. Journal of Fungi , 9 (8), 798.
43. Gabaldón, T., & Carrete, L. (2016). The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in Candida glabrata. *FEMS yeast research*, *16*(2), fov110.
44. Gauthier G. M. (2017). Fungal Dimorphism and Virulence: Molecular Mechanisms for Temperature Adaptation, Immune Evasion, and In Vivo Survival. Mediators of inflammation, 2017, 8491383.
45. Gerós-Mesquita, Â., Carvalho-Pereira, J., Franco-Duarte, R., Alves, A., Gerós, H., … & Sampaio, P. (2020). Oral Candida albicans colonization in healthy individuals: prevalence, genotypic diversity, stability along time and transmissibility. Journal of oral microbiology, 12(1), 1820292.
46. Ghasemi, R., Lotfali, E., Rezaei, K., Madinehzad, S. A., ... & Fattahi, M. (2022). M eyerozyma guilliermondii species complex: review of current epidemiology,

antifungal resistance, and mechanisms. *Brazilian Journal of Microbiology*, *53*(4), 1761-1779.

47. Gómez-Gaviria, M., Ramírez-Sotelo, U., & Mora-Montes, H. M. (2022). Non-albicans Candida Species: Immune Response, Evasion Mechanisms, and New Plant-Derived Alternative Therapies. Journal of fungi (Basel, Switzerland), 9(1), 11.

48. Gong, J., Bing, J., Guan, G., Nobile, C. J., & Huang, G. (2020). The Als3 cell wall adhesin plays a critical role in human serum amyloid A1-induced cell death and aggregation in Candida albicans. Antimicrobial agents and chemotherapy, 64(6), 10-1128.

49. Govrins, M., & Lass-Flörl, C. (2024). Candida parapsilosis complex in the clinical setting. *Nature Reviews Microbiology*, *22*(1), 46-59.

50. Guinea, J. (2014). Global trends in the distribution of Candida species causing candidemia. *Clinical Microbiology and Infection*, *20*, 5-10.

51. Guitard, J., Angoulvant, A., Letscher-Bru, V., L'Ollivier, C., ... & Hennequin, C. (2013). Invasive infections due to Candida norvegensis and Candida inconspicua: report of 12 cases and review of the literature. *Medical Mycology*, 51(8), 795-799.

52. Gulati, M., & Nobile, C. J. (2016). Candida albicans biofilms: development, regulation, and molecular mechanisms. Microbes and Infection / Institut Pasteur, 18(5), 310–321.

53. Hassan, Y., Chew, S. Y., & Than, L. T. L. (2021). Candida glabrata: Pathogenicity and Resistance Mechanisms for Adaptation and Survival. Journal of fungi (Basel, Switzerland), 7(8), 667.

54. Hadrich, I., & Ranque, S. (2015). Typing of fungi in an outbreak setting: lessons learned. *Current Fungal Infection Reports*, *9*(4), 314-323.

55. Hartfield, M., & Bataillon, T. (2020). Selective sweeps under dominance and inbreeding. *G3: Genes, Genomes, Genetics*, *10*(3), 1063-1075.

56. Hori, Y., & Shibuya, K. (2018). Role of FKS gene in the susceptibility of pathogenic fungi to echinocandins. Medical mycology journal, 59(2), E31-E40.

57. Houlihan, C. F., & Whitworth, J. A. (2019). Outbreak science: recent progress in the detection and response to outbreaks of infectious diseases. *Clinical Medicine*, *19*(2), 140-144.

58. Howell, S. A., Hazen, K. C., & Brandt, M. E. (2015). Candida, Cryptococcus, and other yeasts of medical importance. Manual of clinical microbiology, 1984-2014.
59. Hsu, C., & Yassin, M. (2025). Diagnostic Approaches for Candida auris: A Comprehensive Review of Screening, Identification, and Susceptibility Testing. Microorganisms, 13(7), 1461.
60. Kadosh, D., & Mundodi, V. (2020). A re-evaluation of the relationship between morphology and pathogenicity in Candida species. Journal of Fungi (Basel, Switzerland), 6(1), 13.
61. Kamli, M. R., Sabir, J. S., Malik, M. A., & Ahmad, A. (2022). Characterization of the secretome of pathogenic Candida glabrata and their effectiveness against systemic candidiasis in BALB/c mice for vaccine development. *Pharmaceutics* , *14* (10), 1989.
62. Karakoyun, A. S., Spruijtenburg, B., Unal, N., Meijer, E. F., ... & Ilkit, M. (2024). Molecular typing and antifungal susceptibility profile of Candida krusei bloodstream isolates from Türkiye. *Medical Mycology*, 62(2), myae005.
63. Keighley, C., Kim, H. Y., Kidd, S., Chen, S. C., … & Alffenaar, J. W. (2024). Candida tropicalis-A systematic review to inform the World Health Organization of a fungal priority pathogens list. Medical mycology, 62(6), myae040.
64. Kernien, J. F., Johnson, C. J., & Nett, J. E. (2017). Conserved Inhibition of Neutrophil Extracellular Trap Release by Clinical Biofilms. Journal of Fungi (Basel, Switzerland), 3(3).
65. Kidd, S. E., Chen, S. C.-A., Meyer, W., & Halliday, C. L. (2019). A New Age in Molecular Diagnostics for Invasive Fungal Disease: Are We Ready? Frontiers in Microbiology, 10, 2903.
66. Kim, T. S., Kim, T. S., Park, H., & Park, J. H. (2024). A Case of Invasive Candidiasis Caused by Candida norvegensis in a Liver Transplant Patient. *Laboratory Medicine Online*, 14(1), 60-65.
67. King, K. M., & Van Doorslaer, K. (2018). Building (viral) phylogenetic trees using a maximum likelihood approach. *Current Protocols in Microbiology*, *51*(1), e63.
68. Kollef, M. H., Torres, A., Shorr, A. F., Martin-Loeches, I., & Micek, S. T. (2021). Nosocomial infection. Critical care medicine, 49(2), 169-187.
69. Koundal, S., & Cojandaraj, L. (2020). Candida species-morphology, medical aspects and pathogenic spectrum. European Journal of Molecular & Clinical Medicine , 7 (07), 2020.

70. Kucukates, E., Gultekin, N. N., Alisan, Z., Hondur, N., & Ozturk, R. (2016). Identification of Candida species and susceptibility testing with Sensititre YeastOne microdilution panel to 9 antifungal agents. Saudi medical journal, 37(7), 750–757.
71. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359.
72. Lass-Flörl, C., Kanj, S. S., Govender, N. P., Thompson III, G. R., … & Govrins, M. A. (2024). Invasive candidiasis. Nature Reviews Disease Primers, 10(1), 20.
73. Leite-Jr, D.P., Vivi-Oliveira, V.K., Maia, Macioni, M.B., Oliboni, G.M., & De Oliveira, I.D. (2023). The Candida genus complex: Biology, evolution, pathogenicity virulence and one health aspects, beyond the Candida albicans paradigm. A comprehensive review. *Virology Immunol. J*, 7, 1-38.
74. Leon, C., Ruiz-Santana, S., Saavedra, P., Castro, C., ... & Gómez, J. (2018). Contribution of Candida biomarkers and DNA detection for the diagnosis of invasive candidiasis in ICU patients with severe abdominal conditions. Critical Care, 20(1), 1-12.
75. Lewis, W. H., Tahon, G., Geesink, P., Sousa, D. Z., & Ettema, T. J. G. (2020). Innovations to culturing the uncultured microbial majority. Nature Reviews Microbiology, 19(4), 225-240.
76. Li, C., He, Q., Qian, H., & Liu, J. (2021). Overview of the pathogenesis of COVID-19. *Experimental and therapeutic medicine*, 22(3), 1011.
77. Li, Y., Hind, C., Furner-Pardoe, J., Sutton, J. M., & Rahman, K. M. (2025). Understanding the mechanisms of resistance to azole antifungals in Candida species. JAC-Antimicrobial Resistance , 7 (3), dlaf106.
78. Liu, J. Y., & Dickter, J. K. (2020). Nosocomial infections: a history of hospital-acquired infections. Gastrointestinal Endoscopy Clinics , 30 (4), 637-652.
79. Liu, Y., Shen, X., Gong, Y., Liu, Y., … & Zeng, X. (2023). Sequence Alignment/Map format: a comprehensive review of approaches and applications. Briefings in Bioinformatics, 24(5).
80. Long, R., Hernaez, M., Ochoa, I., & Weissman, T. (2017). GeneComp, a new reference-based compressor for SAM files. Proceedings. Data Compression Conference, 2017, 330–339.
81. Lopes, J. P., & Lionakis, M. S. (2022). Pathogenesis and virulence of Candida albicans. Virulence, 13(1), 89-121.

82. Malinovská, Z., Čonková, E., & Váczi, P. (2023). Biofilm Formation in Medically Important Candida Species. Journal of fungi (Basel, Switzerland), 9(10), 955.
83. Mallick, D. C., Kaushik, N., Goyal, L., Mallick, L., & Singh, P. (2025). A Comprehensive Review of Candidemia and Invasive Candidiasis in Adults: Focus on the Emerging Multidrug-Resistant Fungus Candida auris. Diseases, 13(4), 93.
84. Mäklin, T., Kallonen, T., David, S., Boinett, C. J., … & Honkela, A. (2021). High-resolution sweep metagenomics using fast probabilistic inference. Wellcome open research, 5, 14.
85. Mayer, F. L., Wilson, D., & Hube, B. (2013). Candida albicans pathogenicity mechanisms. *Virulence* , *4* (2), 119-128.
86. McCarty, T. P., & Pappas, P. G. (2016). Invasive candidiasis. *Infectious Disease Clinics*, *30*(1), 103-124.
87. McMurray-Jones, A., Spann, K., Yarlagadda, P. K., Fernando, J., & Roberts, L. W. (2026). Environmental surveillance of bacteria in a new intensive care unit using plate sweeps. Microbial Genomics , 12 (1), 001597.
88. Michel, C., Martin, C., Smeesters, P., Goffard, J.-C., … & Hallin, M. (2023). Next-generation sequencing: what are the needs in routine clinical microbiology? A survey among clinicians involved in infectious diseases practice. Frontiers of Medicine, 10, 1225408.
89. Minh, B.Q. et al. (2020). IQ-TREE 2: New models and methods for phylogenetic inference. Molecular Biology and Evolution, 37(5), 1530–1534.
90. Misas, E., Witt, L.S., Farley, M.M., Thomas, S., ... & Lyman, M. (2024). Molecular and epidemiological investigation of fluconazole-resistant Candida parapsilosis—Georgia, United States, 2021. In *Open Forum Infectious Diseases* (Vol. 11, No. 6, p. ofae264). US: Oxford University Press.
91. Moorhouse, A. J., Moreno-Lopez, R., Gow, N. A., & Hijazi, K. (2021). Clonal evolution of Candida albicans, Candida glabrata and Candida dubliniensis at oral niche level in health and disease. *Journal of Oral Microbiology*, 13(1), 1894047.
92. Motaung, T. E., Albertyn, J., Pohl, C. H., & Köhler, G. (2015). Candida albicans mutant construction and characterization of selected virulence determinants. *Journal of microbiological methods*, 115, 153-165.
93. Nagarathnamma, T., Chunchanur, S. K., Rudramurthy, S. M., Vineetha, K. R., … & Ambica, R. (2017). Outbreak of Pichia kudriavzevii fungaemia in a neonatal intensive care unit. *Journal of Medical Microbiology*, 66(12), 1759-1764.

94. Nam, N. N., Do, H. D. K., Loan Trinh, K. T., & Lee, N. Y. (2023). Metagenomics: an effective approach for exploring microbial diversity and functions. Foods , 12 (11), 2140.

95. Neto Junior, J. M., Dias, V. C., de Andrad Bastos, V. Q., de Andrade Bastos, L. Q., … & Diniz, C. G. (2024). Clinical and epidemiological aspects of Candida yeast infections and rational use of antifungals. Future microbiology, 19(7), 577–584.

96. Nguyen, T. A., Kim, H. Y., Stocker, S., Kidd, S., … & Alffenaar, J.-W. (2024). Pichia kudriavzevii (Candida krusei): A systematic review to inform the World Health Organisation priority list of fungal pathogens. Medical Mycology: Official Publication of the International Society for Human and Animal Mycology, 62(6).

97. Nielsen, H., Stenderup, J., Bruun, B., & Ladefoged, J. (1990). Candida norvegensis peritonitis and invasive disease in a patient on continuous ambulatory peritoneal dialysis. *Journal of clinical microbiology*, 28(7), 1664-1665.

98. Ostrowsky, B., Greenko, J., Adams, E., Quinn, M., … & C. auris Investigation Work Group (2020). Candida auris Isolates Resistant to Three Classes of Antifungal Medications - New York, 2019. MMWR. Morbidity and mortality weekly report, 69(1), 6–9.

99. Pais, P., Galocha, M., Viana, R., Cavalheiro, M., … & Teixeira, M. C. (2019). Microevolution of the pathogenic yeasts Candida albicans and Candida glabrata during antifungal therapy and host infection. Microbial cell (Graz, Austria), 6(3), 142–159.

100. Pappas, P. G., Lionakis, M. S., Arendrup, M. C., Ostrosky-Zeichner, L., & Kullberg, B. J. (2018). Invasive candidiasis. Nature Reviews Disease Primers, 4(1), 1-20.

101. Pemán, J., Ruiz-Gaitán, A., García-Vidal, C., Salavert, M., Ramírez, P., Puchades, F., ... & Quindós, G. (2020). Fungal co-infection in COVID-19 patients: Should we be concerned?. *Revista iberoamericano de micologia*, 37(2), 41-46.

102. Pérez-García, L. A., Macías-Pérez, J. R., León-Buitimea, Á., Alvarado-Sánchez, B., … & Rodríguez-Reyes, S. C. (2017). Candida and candidiasis. In *Current progress in medical mycology* (pp. 91-118). Cham: Springer International Publishing.

103. Pfaller, M. A., & Diekema, D. (2007). Epidemiology of invasive candidiasis: a persistent public health problem. *Clinical microbiology reviews* , *20* (1), 133-163.

104. Pfaller, M. A., Diekema, D. J., Gibbs, D. L., Newell, V. A., ... & Ling, T. A. (2010). Results from the ARTEMIS DISK Global Antifungal Surveillance Study, 1997 to 2007: a 10.5-year analysis of susceptibilities of Candida species to fluconazole and voriconazole as determined by CLSI standardized disk diffusion. *Journal of clinical microbiology*, 48(4), 1366-1377.

105. Piñeiro, C., & Pichel, J. C. (2022). BigSeqKit: a parallel Big Data toolkit to process FASTA and FASTQ files at scale. GigaScience, 12.

106. Pinho, S., Miranda, I. M., & Costa-de-Oliveira, S. (2024). Global epidemiology of invasive infections by uncommon Candida species: A systematic review. Journal of Fungi, 10(8), 558.

107. Polke, M., Hube, B., & Jacobsen, I. D. (2015). Candida survival strategies. *Advances in Applied Microbiology*.

108. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., ... & Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *biorxiv* , 201178.

109. Rhodes, J., & Fisher, M. C. (2019). Global epidemiology of emerging Candida auris. Current Opinion in Microbiology, 52, 84-89.

110. Richardson, J. P., Ho, J., & Naglik, J. R. (2018). Candida–epithelial interactions. Journal of fungi, 4(1), 22.

111. Romo, J. A., & Kumamoto, C. A. (2020). On Commensalism of Candida. Journal of fungi (Basel, Switzerland), 6(1), 16.

112. Ruiz, G. B., & Lorenz, A. (2021). What do we know about the biology of the emerging fungal pathogen of humans Candida auris?. *Microbiological Research*, *242*, 126621.

113. Saada, B., Zhang, T., Siga, E., Zhang, J., & Magalhães Muniz, M. M. (2024). Whole-genome alignment: methods, challenges, and future directions. *Applied Sciences*, 14(11), 4837.

114. Saeed, N. K., Almusawi, S., & Al-Beltagi, M. (2024). Candidemia chronicles: Retrospective analysis of candidemia epidemiology, species distribution, and antifungal susceptibility patterns in Bahrain. World journal of virology, 13(4), 98839.

115. Sanders E. R. (2012). Aseptic laboratory techniques: plating methods. Journal of visualized experiments : JoVE, (63), e3064.

116. Sankar, A., Malone, B., Bayliss, S. C., Pascoe, B., ... & Honkela, A. (2016). Bayesian identification of bacterial strains from sequencing data. Microbial genomics, 2(8), e000075.

117. Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., … & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. Biology, 12(7).

118. Schaeffer, L., Pimentel, H., Bray, N., Melsted, P., & Pachter, L. (2017). Pseudoalignment for metagenomic read assignment. Bioinformatics, 33(14), 2082-2088.

119. Schbath, S., Martin, V., Zytnicki, M., Fayolle, J., … & Gibrat, J.-F. (2012). Mapping reads on a genomic sequence: an algorithmic overview and a practical comparative analysis. Journal of Computational Biology: A Journal of Computational Molecular Cell Biology, 19(6), 796–813.

120. Schelenz, S., Hagen, F., Rhodes, J. L., Abdolrasouli, A., ... & Armstrong-James, D. (2016). First hospital outbreak of the globally emerging Candida auris in a European hospital. Antimicrobial Resistance & Infection Control, 5(1), 1-7.

121. Sherry, N. L., Gorrie, C. L., Kwong, J. C., Higgs, C., ... & Howden, B. P. (2022). Multi-site implementation of whole genome sequencing for hospital infection control: a prospective genomic epidemiological analysis. *The Lancet Regional Health–Western Pacific*, *23*.

122. Silva, I., Miranda, I. M., & Costa-de-Oliveira, S. (2024). Potential environmental reservoirs of Candida auris: a systematic review. Journal of Fungi, 10(5), 336.

123. Soliman S. S. M. (2023). Editorial: Candida spp.-transmission, pathogenesis, host-pathogen interaction, prevention, and treatment. Frontiers in microbiology, 14, 1258837.

124. Soll, D. R. (2014). The role of phenotypic switching in the basic biology and pathogenesis of Candida albicans. Journal of Oral Microbiology, 6.

125. Soriano, A., Honore, P. M., Puerta-Alcalde, P., Garcia-Vidal, C., … & Verweij, P. E. (2023). Invasive candidiasis: current clinical challenges and unmet needs in adult populations. The Journal of antimicrobial chemotherapy, 78(7), 1569–1585.

126. Stajich, J. E. (2017). Fungal genomes and insights into the evolution of the kingdom. *Microbiology spectrum*, 5(4), 10-1128.

127. Strickland, A. B., & Shi, M. (2021). Mechanisms of fungal dissemination. Cellular and Molecular Life Sciences: CMLS, 78(7), 3219–3238.

128. Swaney, R., Jokomo-Nyakabau, R., Nguyen, A. A., Kenny, D., ... & Velagapudi, M. (2023). Diagnosis and Outcomes of Fungal Co-Infections in COVID-19 Infections: A Retrospective Study. *Microorganisms*, 11(9), 2326.

129. Szekely, J., Rakchang, W., Rattanaphan, P., & Kositpantawong, N. (2023). Fluconazole and echinocandin resistance of Candida species in invasive candidiasis at a university hospital during pre-COVID-19 and the COVID-19 outbreak. Epidemiology and infection, 151, e146.

130. Talapko, J., Juzbašić, M., Matijević, T., Pustijanac, E., … & Škrlec, I. (2021). Candida albicans-The Virulence Factors and Clinical Manifestations of Infection. Journal of fungi (Basel, Switzerland), 7(2), 79.

131. Tang, J., Cai, L., He, H., Li, X., ... & Pan, W. (2025). Candidemia caused by Meyerozyma guilliermondii species complex in patients with cancer: A systematic review of case reports. *Medical Mycology*, *63*(10), myaf087.

132. Tavanti, A., Davidson, A. D., Gow, N. A., Maiden, M. C., & Odds, F. C. (2005). Candida orthopsilosis and Candida metapsilosis spp. Nov. to replace Candida parapsilosis groups II and III. *Journal of clinical microbiology* , *43* (1), 284-292.

133. Thomas-Rüddel, D. O., Schlattmann, P., Pletz, M., Kurzai, O., & Bloos, F. (2022). Risk Factors for Invasive Candida Infection in Critically Ill Patients: A Systematic Review and Meta-analysis. Chest, 161(2), 345–355.

134. Thorpe, H. A., Pesonen, M., Corbella, M., Pesonen, H., ... & Corander, J. (2024). Pan-pathogen deep sequencing of nosocomial bacterial pathogens in Italy in spring 2020: a prospective cohort study. *The Lancet Microbe* , *5* (10).

135. Thompson, D. S., Carlisle, P. L., & Kadosh, D. (2011). Coevolution of morphology and virulence in Candida species. *Eukaryotic cell*, 10(9), 1173-1182.

136. Todd, R. T., Wikoff, T. D., Forche, A., & Selmecki, A. (2019). Genome plasticity in Candida albicans is driven by long repeat sequences. eLife, 8, e45954.

137. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D., & Corander, J. (2019). Fast hierarchical Bayesian analysis of population structure. Nucleic acids research , 47 (11), 5539-5549.

138. Tönnies, H., Prior, K., Harmsen, D., & Mellmann, A. (2021). Establishment and evaluation of a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of Pseudomonas aeruginosa. *Journal of clinical microbiology*, 59(3), 10-1128.

139. Tsai, C. S., Lee, S. S. J., Chen, W. C., Tseng, C. H., ... & Hung, Y. P. (2023). COVID-19-associated candidiasis and the emerging concern of Candida auris infections. *Journal of Microbiology, Immunology and Infection*, 56(4), 672-679.

140. Turner, S. A., & Butler, G. (2014). The Candida pathogenic species complex. *Cold Spring Harbor perspectives in medicine*, 4(9), a019778.

141. Usher, J. (2019). The mechanisms of mating in pathogenic fungi—a plastic trait. *Genes*, 10(10), 831.

142. Vazquez, J. A., Whitaker, L., & Zubovskaia, A. (2025). Invasive Candidiasis in the Intensive Care Unit: Where Are We Now? Journal of Fungi, 11(4), 258.

143. Vitiello, A., Ferrara, F., Boccellino, M., Ponzo, A., … & Sabbatucci, M. (2023). Antifungal Drug Resistance: An Emergent Health Threat. Biomedicines, 11(4), 1063.

144. Vumbaca, M., Biffignandi, G. B., Cavanna, C., Bellinzona, G., ... & Sassera, D. (2024). Genomic characterization of a persistent, azole-resistant C. parapsilosis strain responsible for a hospital outbreak during the first COVID-19 wave. *bioRxiv*, 2024-11.

145. Wall, G., Montelongo-Jauregui, D., Vidal Bonifacio, B., Lopez-Ribot, J. L., & Uppuluri, P. (2019). Candida albicans biofilm growth and dispersal: contributions to pathogenesis. Current Opinion in Microbiology, 52, 1–6.

146. Wang, Y., Lo, W. C., & Chou, C. S. (2017). A modeling study of budding yeast colony formation and its relationship to budding pattern and aging. *PLOS computational biology*, 13(11), e1005843.

147. Wiederhold, N. P. (2022). Emerging fungal infections: new species, new names, and antifungal resistance. *Clinical chemistry*, *68*(1), 83-90.

148. Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, 15(3), R46.

149. Xu, J. (2021). Is natural population of Candida tropicalis sexual, parasexual, and/or asexual?. *Frontiers in Cellular and Infection Microbiology*, 11, 751676.

150. Yadav, V., Sun, S., & Heitman, J. (2023). On the evolution of variation in sexual reproduction through the prism of eukaryotic microbes. *Proceedings of the National Academy of Sciences*, 120(11), e2219120120.