

Nghiên cứu

Ứng dụng mã vạch DNA vùng ITS tự động trong định danh nấm *Candida*: Tiềm năng ứng dụng trong kiểm soát nhiễm khuẩn bệnh viện

Trần Tuấn Khôi^{1,3*}, Trần Đình Bình¹, Nguyễn Việt Tú¹, Hoàng Lê Bích Ngọc¹,
Võ Văn Minh Quân^{1,2}, Kittipan Samerpitak³

¹Khoa Kiểm soát nhiễm khuẩn, Bệnh viện Trường Đại học Y - Dược,
Trường Đại học Y - Dược, Đại học Huế

²Bộ môn Dinh dưỡng - Sức khỏe môi trường & Nghề nghiệp, Khoa Y tế công cộng,
Trường Đại học Y - Dược, Đại học Huế

³Bộ môn Vi sinh, Khoa Y, Đại học Khon Kaen, Thái Lan

*Tác giả liên hệ (Corresponding author): Trần Tuấn Khôi, email: khoitrn.med@gmail.com

Ngày nhận bài (Received): 16/04/2026; Ngày duyệt đăng (Accepted): 05/06/2026; Ngày xuất bản (Published): 18/06/2026

DOI:10.34071/jmp.2026.S-1.12

Tóm tắt

Đặt vấn đề: Định danh nhanh và chính xác nấm *Candida* đến cấp loài hoặc phức hợp loài có vai trò quan trọng trong lựa chọn kháng nấm hợp lý và kiểm soát nhiễm khuẩn bệnh viện. Công cụ mã vạch ITS tự động được xây dựng nhằm chuẩn hóa bước phân tích trình tự, giảm thiểu thời gian xử lý và tạo điều kiện thuận lợi cho tích hợp vào hệ thống xét nghiệm và giám sát kiểm soát nhiễm khuẩn.

Mục tiêu: Mô tả các tính năng chính của công cụ định danh mã vạch ITS tự động và đánh giá hiệu năng nhận dạng trên bộ trình tự chủng nấm men lâm sàng thông qua các chỉ số chẩn đoán tiêu chuẩn.

Đối tượng và phương pháp nghiên cứu: Công cụ được xây dựng trên cơ sở dữ liệu ITS tham chiếu gồm 134 loài nấm y khoa thuộc 7 chi và hoạt động theo bốn bước: K-mer indexing, so sánh khoảng cách K2P với ngưỡng hiệu chỉnh theo chi, báo cáo phân cấp (loài/phức hợp/chi) và cờ cảnh báo loài ẩn. Bộ dữ liệu đánh giá gồm 145 trình tự ITS, trong đó 92 trình tự từ chủng nấm men phân lập lâm sàng tại Bệnh viện Srinagarind, Thái Lan, và 53 trình tự từ các chủng chuẩn ngoài cơ sở dữ liệu thu thập từ GenBank. Các chỉ số độ chính xác, độ nhạy, độ đặc hiệu, giá trị tiên đoán dương (PPV), giá trị tiên đoán âm (NPV) và thời gian xử lý được tính toán, đối chiếu với tiêu chuẩn vàng là phân tích cây phát sinh loài và so sánh hiệu năng với NCBI BLAST.

Kết quả: Trên 145 trình tự được đánh giá, công cụ ghi nhận 90 dương tính thật, 52 âm tính thật, 1 dương tính giả và 2 âm tính giả. Độ chính xác đạt 97,9%, độ nhạy 97,8%, độ đặc hiệu 98,1%, PPV 98,9% và NPV 96,3%. Thời gian xử lý trung bình 0,93 giây/trình tự. Công cụ định danh đúng hoàn toàn ở mức loài đối với *C. tropicalis* (24/24), *C. glabrata* (9/9), *C. krusei* (4/4), và *C. africana* (6/6); đồng thời báo cáo an toàn ở mức phức hợp đối với *C. albicans*, phức hợp *C. parapsilosis* và phức hợp *C. rugosa*. So sánh với NCBI BLAST, công cụ tự động đạt độ chính xác cao hơn (97,9% so với 93,5%) và tốc độ xử lý nhanh hơn khoảng 10 lần (0,93 so với 10 giây/trình tự).

Kết luận: Công cụ mã vạch ITS tự động cho thấy hiệu năng định danh cao, tốc độ xử lý nhanh và cơ chế báo cáo an toàn đối với các trường hợp khó. Giá trị nổi bật là tự động hóa phân tích trình tự, chuẩn hóa kết quả đầu ra và tiềm năng tích hợp vào quy trình quản lý kháng nấm và giám sát kiểm soát nhiễm khuẩn bệnh viện.

Từ khóa: mã vạch DNA; ITS; *Candida*; định danh nấm men; kiểm soát nhiễm khuẩn.

Automated ITS DNA barcoding pipeline for *Candida* identification: Potential applications for hospital infection control

Tran Tuan Khoi^{1,3*}, Tran Dinh Binh¹, Nguyen Viet Tu¹, Hoang Le Bích Ngọc¹,
Vo Van Minh Quan^{1,2}, Samerpitak Kittipan³

¹Department of Infection Control, Hue University of Medicine and Pharmacy Hospital,
Hue University of Medicine and Pharmacy, Hue University, Vietnam

²Department of Nutrition, Environmental and Occupational Health, Faculty of Public Health, Hue University of Medicine and
Pharmacy, Hue University, Vietnam

³Department of Microbiology, Faculty of Medicine, Khon Kaen University, Thailand

Abstract

Background: Rapid and accurate identification of *Candida* species at species or species-complex level is essential for antifungal stewardship and hospital infection control. An automated ITS DNA barcoding pipeline

was developed to standardize sequence analysis, reduce processing time, and generate structured outputs suitable for laboratory information systems and infection-control surveillance.

Objectives: To describe the key features of the automated ITS DNA barcoding tool and evaluate its identification performance on clinical yeast sequences using standard diagnostic metrics.

Materials and methods: The pipeline was built on a quality-controlled ITS reference database covering 134 medically relevant fungal species across 7 genera and comprised four major components: K-mer indexing, K2P distance comparison with genus-specific thresholds, hierarchical three-tier reporting (species/complex/genus), and cryptic-species flagging. The evaluation dataset comprised 145 ITS sequences: 92 from clinical yeast isolates obtained from Srinagarind Hospital, Thailand, and 53 reference sequences from out-of-database species retrieved from GenBank. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and processing time were calculated against phylogenetic tree-based identification as the reference standard and compared with NCBI BLAST.

Results: Among 145 evaluated sequences, the tool recorded 90 true positives (TP), 52 true negatives (TN), 1 false positive (FP), and 2 false negatives (FN). Overall accuracy was 97.9%, sensitivity 97.8%, specificity 98.1%, PPV 98.9%, and NPV 96.3%. Mean runtime was 0.93 seconds/sequence. The tool achieved complete species-level concordance for *C. tropicalis* (24/24), *C. glabrata* (9/9), *C. krusei* (4/4), and *C. africana* (6/6); and safely reported results at the complex level for the *C. albicans*, *C. parapsilosis* complex, and *C. rugosa* complex. Compared with NCBI BLAST, the automated tool demonstrated higher accuracy (97.9% vs. 93.5%) and approximately 10-fold faster processing (0.93 vs. 10 seconds/sequence).

Conclusion: The automated ITS barcoding tool demonstrated high identification accuracy, rapid processing, and safe hierarchical reporting for challenging cases. Its primary strengths are automated sequence analysis, standardized structured output, and potential utility for antifungal stewardship and hospital infection-control surveillance.

Keywords: DNA barcoding; ITS; Candida; yeast identification; infection control.

1. ĐẶT VẤN ĐỀ

Nhiễm nấm bệnh viện đặc biệt do các loài *Candida* là nguyên nhân gây bệnh hàng đầu và tử vong ở bệnh nhân nằm viện, đặc biệt tại đơn vị chăm sóc tích cực (ICU) và bệnh nhân suy giảm miễn dịch. Các pipeline tin sinh học tự động bao gồm các giải pháp dựa trên trí tuệ nhân tạo (AI) và học máy (machine learning - ML) đang được nghiên cứu ứng dụng nhằm cải tiến tốc độ và độ chính xác trong chẩn đoán, phân loại, và dự báo nguy cơ nhiễm vi nấm tại bệnh viện [1-3].

Theo ước tính, có ít nhất một tỷ người bị nhiễm nấm trên toàn thế giới, dẫn đến hơn 1,6 triệu ca tử vong hàng năm [4]. *Candida* là một trong những tác nhân gây nhiễm khuẩn bệnh viện phổ biến nhất, đặc biệt là nhiễm nấm huyết do *Candida* (candidemia). Chẩn đoán truyền thống phụ thuộc vào nuôi cấy còn hạn chế về thời gian và độ phân giải (thường mất 72 - 96 giờ), trong khi MALDI-TOF MS tuy nhanh vẫn có thể gặp khó khăn với các loài hiếm, loài ẩn và các phức hợp loài gần nhau [5,6]. Sự nổi lên của *Candida auris* - loài đa kháng thuốc, có khả năng lan truyền mạnh trong bệnh viện - đã thúc đẩy nhu cầu cấp thiết về các công cụ chẩn đoán nhanh và chính xác hơn [7].

Phương pháp mã vạch DNA (DNA barcoding) dựa trên vùng ITS (Internal Transcribed Spacer) của rDNA là tiêu chuẩn vàng trong phân loại phân tử vi nấm, giúp nhận diện hầu hết các loài nấm y học quan

trọng. Sự kết hợp giữa giải trình tự ITS và các pipeline tin sinh học tự động, được tăng cường bởi AI, đang mở ra một kỷ nguyên mới trong chẩn đoán vi nấm lâm sàng [8,9].

Nghiên cứu này được thực hiện với hai mục tiêu chính: (1) mô tả các tính năng chính của công cụ mã vạch ITS tự động ứng dụng cho định danh vi nấm; (2) đánh giá hiệu năng nhận dạng của công cụ trên bộ chủng nấm men phân lập lâm sàng thông qua các chỉ số chẩn đoán chuẩn và thời gian xử lý.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Cơ sở dữ liệu tham chiếu

Cơ sở dữ liệu (CSDL) tham chiếu được xây dựng từ 288 trình tự ITS của 134 loài nấm y khoa thuộc 7 chi, gồm *Candida*, *Cryptococcus*, *Malassezia*, *Rhodotorula*, *Trichosporon*, *Aspergillus* và *Fusarium*. Trình tự được tải xuống từ GenBank (truy cập ngày 09/06/2025) và kiểm soát chất lượng nhiều bước theo khuyến cáo của cộng đồng tin sinh học năm [10]: lọc chiều dài trình tự (400 - 900 bp cho toàn bộ vùng ITS), loại bỏ trình tự có tỷ lệ base không xác định $\geq 2\%$, đối chiếu phân loại học với MycoBank và loại bỏ trùng lặp hoàn toàn trong cùng loài.

Tiêu chí lựa chọn loài vào CSDL dựa trên: (1) có ghi nhận gây bệnh ở người theo y văn; (2) ưu tiên các loài có ý nghĩa kháng thuốc hoặc dịch tễ nổi bật; (3)

là thành viên của phức hợp loài có nguy cơ nhầm lẫn trong định danh khi sử dụng phương pháp kiểu hình hoặc MALDI-TOF MS thường quy.

Đối với chi *Candida*, CSDL bao phủ 33 loài có ý nghĩa y khoa, bao gồm các loài hiện tại và các loài trước đây từng được phân loại vào chi này (Bảng 1).

Bảng 1. Danh mục 33 loài *Candida* có trong CSDL tham chiếu

Loài phức hợp/ đồng hình	Ý nghĩa lâm sàng
<i>C. albicans</i> <i>C. africana</i> <i>C. dubliniensis</i>	Loài phân lập huyết học phổ biến nhất; đề kháng azole đang gia tăng [11-13]. <i>C. albicans</i> dễ định danh nhầm với <i>C. dubliniensis</i> trên CHROMagar [14] và <i>C. africana</i> - cần xác nhận bằng sinh học phân tử [15]
<i>C. tropicalis</i>	Tỷ lệ kháng azole cao nhất trong 5 loài <i>Candida</i> chính [16]; tử vong do nhiễm xâm lấn khoảng 55 - 60% [17]
<i>C. parapsilosis</i> <i>C. metapsilosis</i> <i>C. orthopsilosis</i>	Gây dịch nhiễm khuẩn liên quan catheter tại NICU [18]; xuất hiện chủng kháng fluconazole và echinocandin [13,18]; các loài đồng hình có hoạt tính echinocandin giảm [19]
<i>C. glabrata</i> <i>C. nivariensis</i> <i>C. bracarensis</i>	MIC fluconazole và echinocandin có xu hướng tăng dần [20]; các loài đồng hình dễ bị bỏ sót bởi cơ sở dữ liệu MALDI-TOF [21]
<i>C. krusei</i> <i>C. auris</i> <i>C. haemulonii</i>	Đề kháng nội tại với fluconazole ở $\geq 70\%$ chủng phân lập toàn cầu [13] Đa kháng thuốc, gây vụ dịch bệnh viện; tỷ lệ tử vong toàn cầu từ 40 - 60% [22]; thường bị định danh nhầm sang phức hợp <i>C. haemulonii</i> bởi hệ thống sinh hoá và MALDI-TOF MS [23]
<i>C. lusitanae</i>	Có xu hướng phát sinh kháng amphotericin B trong quá trình điều trị [24]; ghi nhận trong nhiễm nấm huyết và nhiễm trùng mắt liên quan thiết bị y tế [24,25]
<i>C. kefyr</i>	Nhiễm trùng xâm lấn ở bệnh nhân suy giảm miễn dịch [26]; một số chủng giảm nhạy cảm với echinocandin [27]
<i>C. guilliermondii</i> <i>C. famata</i> <i>C. ciferrii</i> <i>C. fermentati</i> <i>C. pararugosa</i>	Thường gây nhiễm nấm huyết liên quan đến catheter tĩnh mạch [28-32]; xu hướng tăng đề kháng với nhóm Azole [13], dễ định danh nhầm với <i>C. famata</i> và <i>C. fermentati</i> trên hệ thống sinh hóa tự động [28]
<i>C. rugosa</i> <i>C. neorugosa</i> <i>C. pseudorugosa</i>	Gây nhiễm nấm huyết liên quan đến catheter tĩnh mạch [33], giảm nhạy cảm với nhóm azole và echinocandins [34]
<i>C. inconspicua</i> <i>C. norvegensis</i>	Xu hướng gia tăng ở nhóm bệnh nhân suy giảm miễn dịch, đề kháng với các thuốc kháng nấm nhóm azole [35]
<i>C. pelliculosa</i> <i>C. utilis</i>	Gây bùng phát nhiễm khuẩn bệnh viện ở trẻ sơ sinh và bệnh nhi có lưu catheter tĩnh mạch, giảm nhạy cảm với nhóm azole [35]
<i>C. lipolytica</i>	Gây dịch nhiễm khuẩn bệnh viện liên quan đến lưu catheter tĩnh mạch và có khả năng tạo màng sinh học [35]
<i>C. palmiophila</i> <i>C. zeylanoides</i> <i>C. aaseri</i> <i>C. pseudoaaseri</i>	Gây nhiễm trùng bệnh viện; có xu hướng đề kháng đặc thù và thường bị định danh nhầm trên các hệ thống sinh hóa tự động [36-39]
<i>C. subhashii</i>	Nấm men mới nổi; ca viêm phúc mạc do nấm đầu tiên trên bệnh nhân suy thận giai đoạn cuối (2009); dữ liệu kháng thuốc hạn chế, cần định danh bằng sinh học phân tử [40]

2.2. Thiết kế và các tính năng chính của công cụ

Công cụ được phát triển bằng Python 3.9 với các thư viện mã nguồn mở (BioPython, NumPy, pandas), chạy trên hệ điều hành Ubuntu 22.04 LTS (Linux), vận

hành trên phần cứng phổ thông (bộ xử lý Intel Core i5, bộ nhớ RAM 8 GB), yêu cầu bộ nhớ dưới 250 MB cho CSDL tham chiếu hiện tại. Đầu vào là các trình tự toàn bộ vùng ITS (ITS1-5.8S-ITS2) từ giải trình

tự Sanger ở định dạng FASTA; đầu ra là kết quả định danh có cấu trúc, có thể xuất dạng CSV để tích hợp với hệ thống thông tin xét nghiệm (LIS/HIS).

Công cụ hoạt động theo bốn bước tuần tự. Bước 1 – K-mer indexing: thu hẹp nhanh không gian tìm kiếm ứng viên từ cơ sở dữ liệu tham chiếu dựa trên tần suất các đoạn ngắn đặc trưng. Bước 2 – So sánh khoảng cách K2P: Ngưỡng K2P ban đầu được thiết lập cho tất cả các chi dựa trên phân bố khoảng cách nội loài (intraspecific distance) và liên loài (interspecific distance) được ghi nhận trong các nghiên cứu mã vạch ITS trên nấm y khoa: ngưỡng nội loài = 0,015; ngưỡng gian loài = 0,026; ngưỡng phức hợp loài = 0,050; ngưỡng loại trừ ngoại chi = 0,080 [41,42]. Sau đó các ngưỡng trên được hiệu chỉnh riêng cho từng chi bằng phân tích barcoding gap dựa trên phân bố khoảng cách K2P thực tế trong CSDL tham chiếu. Các K2P áp dụng sau hiệu chỉnh cho chi *Candida*: ngưỡng phân loại mức loài là 0,024; ngưỡng phân loại mức chi là 0,035; và ngưỡng loại trừ ngoại chi 0,12. Bước 3 – Báo cáo phân cấp: trả kết quả theo mức độ tin cậy giảm dần (loài, phức hợp loài và chi) kèm điểm tin cậy tương ứng. Bước 4 – Cờ cảnh báo loài ẩn (cryptic species): kích hoạt khi trình tự ITS đơn locus không đủ khả

năng phân biệt chắc chắn giữa các loài rất gần nhau, giúp phòng tránh định danh quá mức.

2.3. Bộ dữ liệu đánh giá hiệu năng công cụ

Bộ dữ liệu đánh giá gồm 145 trình tự trong đó: (1) 92 trình tự ITS của chủng nấm men phân lập từ bệnh phẩm lâm sàng có nuôi cấy dương tính tại Bệnh viện Srinagarind, Khoa Y, Đại học Khon Kaen, Thái Lan và (2) 53 trình tự ITS của các chủng chuẩn thuộc các loài nằm ngoài CSDL tham chiếu được thu thập từ GenBank (ngày truy cập 06/05/2026).

Đồng thời, toàn bộ 92 trình tự chủng nấm men lâm sàng được phân tích song song bằng NCBI BLAST online (Megablast, CSDL nucleotide nt, tham số mặc định) nhằm so sánh đối chiếu kết quả định danh với công cụ tự động [43]. Kết quả BLAST được ghi nhận dựa trên “best hit” có điểm tương đồng (percent identity) cao nhất.

Kết quả định danh tham chiếu được xác định bằng phân tích cây phát sinh loài, dùng như tiêu chuẩn vàng để đối chiếu với kết quả của công cụ tự động và NCBI BLAST trong việc đánh giá và so sánh hiệu năng.

2.4. Định nghĩa tiêu chuẩn và chỉ số đánh giá hiệu năng

Các tiêu chuẩn đánh giá hiệu năng của công cụ tự động được định nghĩa như sau:

	Đúng	Sai
Thuộc CSDL	Định danh đúng ở mức Loài/Phức hợp loài so với tham chiếu (TP)	Định danh sai ở mức Loài/Phức hợp loài so với tham chiếu (FP)
Ngoài CSDL	Trả kết quả “Unknown” hoặc định danh đúng ở mức Chi (TN)	Định danh sai ở mức Loài/Phức hợp loài/Chi so với tham chiếu (FN)

Từ đó, các chỉ số được tính theo các công thức: Độ chính xác (Accuracy) = $(TP+TN)/(TP+TN+FP+FN)$; Độ nhạy (Sensitivity) = $TP/(TP+FN)$; Độ đặc hiệu (Specificity) = $TN/(TN+FP)$. Giá trị tiên đoán dương (Positive Predictive Value – PPV) = $TP/(TP+FP)$. Giá trị tiên đoán âm (Negative Predictive Value – NPV) = $TN/(TN+FN)$.

Thời gian xử lý được đo tự động bằng công cụ, báo cáo theo giây/trình tự và tổng thời gian cho toàn bộ lô mẫu. Kết quả được xuất dạng CSV và được tổng hợp.

Kết quả BLAST được đối chiếu với phương pháp tham chiếu và được đánh giá theo dạng nhị phân (đúng/sai) nên việc so sánh được đánh giá ở độ chính xác và thời gian xử lý.

2.5. Đạo đức nghiên cứu

Nghiên cứu được thực hiện trên các chủng nấm lâm sàng, tuân thủ các quy định quốc gia và chính sách của cơ sở liên quan, theo các nguyên tắc của Tuyên bố Helsinki. Nghiên cứu cũng đã được Hội đồng Đạo đức Đại học Khon Kaen (KKUEC) phê duyệt,

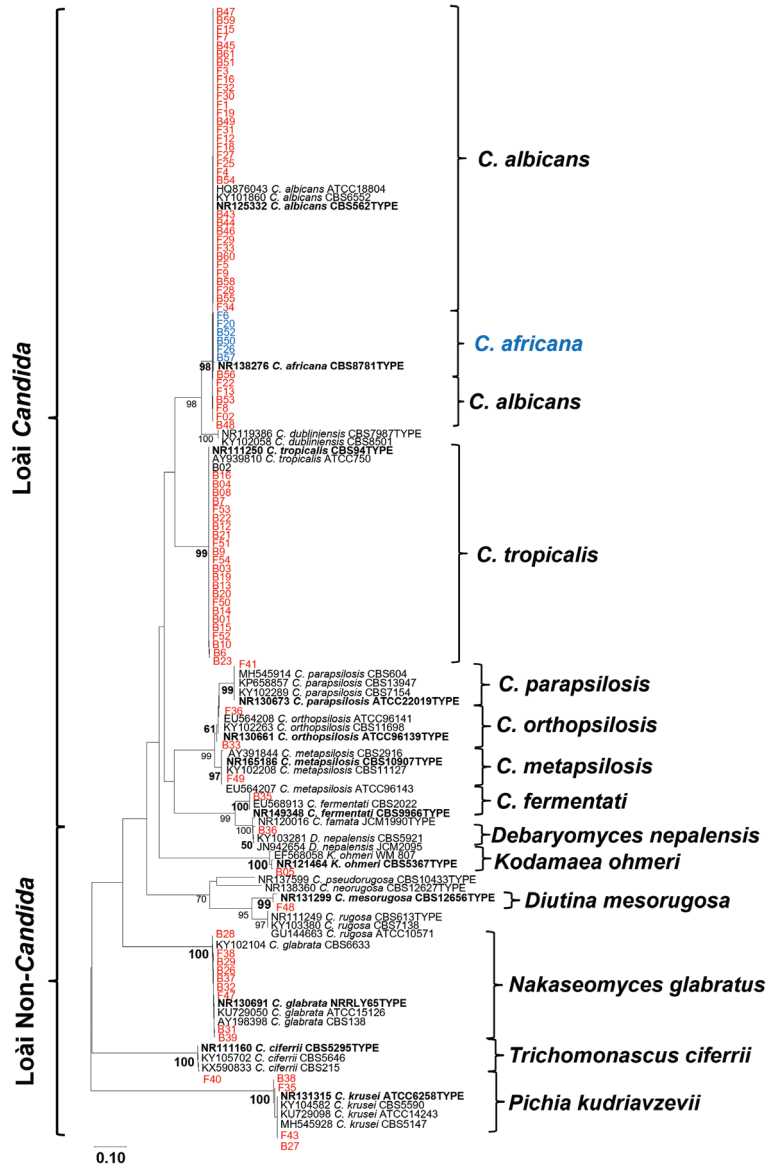
với mã số chấp thuận IRB No. HE671084.

3. KẾT QUẢ NGHIÊN CỨU

3.1. Kết quả phân tích cây phát sinh loài

Cây phát sinh loài được xây dựng bằng phương pháp Neighbor-Joining từ 135 trình tự ITS gồm 92 trình tự lâm sàng và 43 trình tự chuẩn của các loài đại diện (Hình 1).

Các nhánh chính tương ứng rõ với các loài đã biết: các cụm *C. tropicalis*, *C. glabrata*, *C. krusei* được tách biệt với giá trị bootstrap cao, hỗ trợ ranh giới loài đã xác định. Các chủng thuộc phức hợp *C. albicans* (bao gồm *C. albicans*, *C. africana* và *C. dubliniensis*) phân nhánh nằm sát nhau do khoảng cách trình tự ITS thấp. Tương tự, các nhánh *C. parapsilosis*, *C. orthopsilosis*, *C. metapsilosis* tạo thành một cụm phức hợp riêng biệt. Kết quả định danh chi tiết được trình bày tại Hình 1 và Bảng 2.



Hình 1. Cây phát sinh loài của 135 chủng xây dựng bằng phương pháp Neighbor-Joining trên MEGA12, bootstrap 1000 lần lặp, khoảng cách tiến hóa theo mô hình Tamura 3-tham số (G = 0,44), TYPE = chủng chuẩn; B = mẫu máu; F = mẫu dịch cơ thể

3.2. Kết quả định danh bằng NCBI BLAST và công cụ tự động

Bảng 2 tổng hợp kết quả định danh 92 chủng nấm men lâm sàng theo nhóm loài, đối chiếu giữa BLAST, công cụ tự động và phương pháp tham chiếu bằng cây phát sinh loài.

Bảng 2. Kết quả định danh của 92 chủng nấm men lâm sàng theo nhóm loài

Nhóm	NCBI BLAST (Megablast)	Công cụ tự động	Cây phát sinh loài	n
<i>C. albicans</i> complex	<i>C. albicans</i>	<i>C. albicans</i> cpx	<i>C. albicans</i>	40
	<i>C. albicans</i>	<i>C. africana</i>	<i>C. africana</i>	6
<i>C. tropicalis</i>	<i>C. tropicalis</i>	<i>C. tropicalis</i>	<i>C. tropicalis</i>	24
<i>C. glabrata</i>	<i>C. glabrata</i>	<i>C. glabrata</i>	<i>C. glabrata</i>	9
<i>C. krusei</i>	<i>C. krusei</i>	<i>C. krusei</i>	<i>C. krusei</i>	4

<i>C. parapsilosis</i> complex	<i>C. orthopsilosis</i>	<i>C. parapsilosis</i> cpx	<i>C. orthopsilosis</i>	1
	<i>C. metapsilosis</i>	<i>C. parapsilosis</i> cpx	<i>C. metapsilosis</i>	1
	<i>C. parapsilosis</i>	<i>C. parapsilosis</i>	<i>C. parapsilosis</i>	1
	<i>C. orthopsilosis</i>	<i>C. orthopsilosis</i>	<i>C. orthopsilosis</i>	1
<i>C. rugosa</i> complex	<i>C. mesorugosa</i>	<i>C. rugosa</i> cpx	<i>C. mesorugosa</i>	1
<i>C. cifferii</i>	<i>C. cifferii</i>	<i>C. cifferii</i>	<i>C. cifferii</i>	1
<i>C. fermentati</i>	<i>C. fermentati</i>	<i>C. fermentati</i>	<i>C. fermentati</i>	1
<i>Debaryomyces nepalensis</i>	<i>D. nepalensis</i>	<i>C. famata</i>	<i>D. nepalensis</i>	1
<i>Kodamaea ohmeri</i> (ngoài CSDL)	<i>K. ohmeri</i>	Unknown	<i>K. ohmeri</i>	1

Kết quả NCBI BLAST cho thấy có 86 trình tự được định danh đúng và 6 trình tự *C. africana* được định danh thành *C. albicans* (Bảng 2).

Trong 91 trình tự thuộc CSDL, công cụ định danh đúng 90 chủng, trong đó đạt kết quả đúng trên các nhóm *C. tropicalis* (24/24), *C. glabrata* (9/9), *C. krusei* (4/4). Đối với nhóm phức hợp *C. albicans* (n = 46): toàn bộ 40 mẫu *C. albicans* được công cụ báo cáo ở mức phức hợp loài (*C. albicans* complex); toàn bộ 6 mẫu *C. africana* được định danh đúng ở mức loài. Các trường hợp nằm trong phức hợp *C. parapsilosis* (bao gồm *C. orthopsilosis*, *C. metapsilosis*) được phân loại chính xác ở mức loài hoặc phức hợp loài. Phức hợp *C. rugosa*

cũng được phân loại chính xác ở cấp phức hợp [33,44].

Đối với 54 trình tự ITS nằm ngoài CSDL, công cụ định danh đúng 52 trình tự trong đó 18 trình tự được định danh đúng ở mức Chi, và 34 trình tự được trả về kết quả “Unknown”. Công cụ định danh sai hai trình tự, trong đó 1 trình tự *C. margitis* (ID truy cập GenBank: NR159750) được định danh thành phức hợp *C. parapsilosis* và 1 trình tự *C. hyderabadensis* (ID truy cập GenBank: NR159741) được định danh *C. parapsilosis* ở mức loài.

Chi tiết ma trận đánh giá kết quả định danh của công cụ tự động và NCBI BLAST được trình bày tại Bảng 3.

Bảng 3. Ma trận đánh giá kết quả định danh

Phương pháp		Đúng	Sai	Tổng
Công cụ tự động (n = 145)	Thuộc CSDL	90 (TP)	1 (FP)	91
	Ngoài CSDL	52 (TN)	2 (FN)	54
NCBI BLAST (n=92)		86	6	

3.3. So sánh hiệu suất giữa công cụ tự động và NCBI BLAST

Các chỉ số đánh giá hiệu suất của công cụ tự động và NCBI BLAST được trình bày chi tiết tại Bảng 4.

Toàn bộ 145 trình tự được công cụ xử lý trong 135,1 giây (tương đương thông lượng 0,93 giây/trình tự), nhanh hơn khoảng 10 lần so với NCBI BLAST và khoảng 200 - 300 lần so với quy trình phân tích thủ công dựa trên cây phát sinh loài (ước tính 3 - 5 phút/trình tự).

Bảng 4. So sánh chỉ số hiệu năng giữa hai phương pháp

Chỉ số	Công cụ tự động (n = 145)	NCBI BLAST (n = 92)	Cây phát sinh loài
Độ chính xác	97,9%	93,5%	–
Độ nhạy	97,8%	–	–
Độ đặc hiệu	98,1%	–	–
Giá trị tiên đoán dương	98,9%	–	–
Giá trị tiên đoán âm	96,3%	–	–
Thời gian xử lý trung bình/trình tự	0,93 giây	10 giây	180 - 300 giây

4. BÀN LUẬN

Trong giai đoạn 2023–2026, các nghiên cứu ứng dụng AI trong định danh vi nấm phát triển theo hai hướng chính: phân tích hình ảnh học sâu và pipeline tin sinh học phân tử tự động.

Về phân tích hình ảnh, Shankarnarayan và cộng sự (cs) cho thấy mô hình InceptionV3 đạt độ nhạy 77,1% và độ đặc hiệu 92,4% trong việc phân loại các loài *Candida* dựa trên hình ảnh từ kính hiển vi [45]; ứng dụng quang phổ Raman đơn bào kết hợp AI của

Xu và cs đã đạt độ chính xác 100% ở cấp độ loài từ bộ dữ liệu 115.129 phổ của 94 bệnh nhân; tuy nhiên, phương pháp này đòi hỏi hạ tầng thiết bị đặc thù [4]. Bên cạnh đó, ứng dụng mô hình học sâu kép (kết hợp YOLOX và MobileNet V2) của Ren và cs đã đạt độ chính xác 92,5% và độ nhạy 99,3% trong việc phát hiện sợi nấm và bào tử; tuy nhiên, phương pháp này đòi hỏi quy trình chuẩn bị mẫu chuyên biệt để chụp dưới kính hiển vi huỳnh quang thay vì soi tươi quang học thông thường [46].

Ở nhóm phương pháp phân tử, Zolotareva và cs đã phát triển hệ thống biosensor chip kết hợp mã vạch DNA (vùng ITS) và cảm biến từ điện trở, đạt độ chính xác 95% và độ nhạy 91% trong đánh giá *in silico* trên 1.239 trình tự của 58 loài nấm gây bệnh; tuy nhiên, công nghệ này hiện vẫn cần được xác thực thêm bằng các mẫu bệnh phẩm lâm sàng thực tế [47].

Công cụ định danh ITS tự động được trình bày trong nghiên cứu này thuộc nhóm pipeline phân tử, với điểm khác biệt cốt lõi là cơ sở dữ liệu tham chiếu được kiểm soát chất lượng đa bước, thuật toán so sánh khoảng cách K2P kết hợp K-mer indexing, và cơ chế báo cáo phân cấp an toàn.

Phân tích BLAST thủ công trên GenBank, dù phổ biến, nhưng vẫn có một số hạn chế hệ thống nhất định: (1) tỷ lệ ước tính khoảng 30% trình tự ITS nấm trên GenBank mang chú thích phân loại sai và tình trạng này vẫn tiếp diễn ngay cả với các trình tự được nộp gần đây, bên cạnh đó quy trình kiểm tra kết quả BLAST thủ công đòi hỏi thời gian đáng kể; một nghiên cứu phân tích cho thấy việc xác minh khoảng 20 định danh có thể mất đến hai tháng [48]; (2) nguyên tắc “best BLAST hit” không cảnh báo người dùng về tình huống các loài trong cùng phức hợp có khoảng cách ITS chồng lấp, dẫn đến nguy cơ định danh sai [10,42]. Trong nghiên cứu này, BLAST đạt độ chính xác 93,5% (86/92 trình tự), thấp hơn so với công cụ tự động, phần lớn do không phân biệt được *C. africana* khỏi *C. albicans* - phản ánh trực tiếp hạn chế của định danh dựa trên ITS đơn locus và nguyên tắc best-hit [15].

Công cụ ITS tự động giải quyết trực tiếp các hạn chế trên theo ba hướng. Về chất lượng cơ sở dữ liệu, công cụ sử dụng bộ dữ liệu tham chiếu đã qua kiểm soát chất lượng nhiều bước, phù hợp với khuyến nghị kiểm duyệt thủ công đối với dữ liệu ITS tham chiếu [10]. Về thuật toán, cơ chế K-mer indexing kết hợp so sánh khoảng cách K2P rút ngắn thời gian tìm kiếm mà vẫn đảm bảo độ phân giải liên loài. Trên toàn bộ 145 trình tự đánh giá (92 trình tự lâm sàng và 53 trình tự từ GenBank), công cụ đạt độ chính xác 97,9%, độ nhạy 97,8%, độ đặc hiệu 98,1%, PPV 98,9% và NPV 96,3%, cho thấy hiệu năng ổn định và

cân bằng giữa hai nhóm loài trong và ngoài CSDL. Về tốc độ, thời gian xử lý của công cụ nhanh hơn rõ rệt so với BLAST (10 lần) và hơn 200 lần cho phân tích cây phát sinh loài thủ công truyền thống. Về tính chuẩn hóa, kết quả ở định dạng CSV có cấu trúc nhất quán cho phép tích hợp hệ thống LIS/HIS và xây dựng giám sát chủ động.

Cơ chế báo cáo phân cấp và cờ cảnh báo loài ẩn là điểm khác biệt quan trọng nhất về mặt an toàn lâm sàng. Thay vì buộc chọn một best-hit như BLAST, công cụ trả kết quả theo mức độ tin cậy giảm dần và kích hoạt cờ cảnh báo khi trình tự truy vấn nằm trong vùng chồng lấp khoảng cách K2P liên loài. Điều này được chứng minh qua kết quả xử lý 54 trình tự ngoài CSDL: 52/54 trường hợp (96,2%) được trả về kết quả an toàn - trong đó 34 trình tự trả kết quả “Unknown” và 18 trình tự được định danh đúng ở mức Chi - thay vì buộc gán một nhãn loài sai. Đây là hành vi an toàn của công cụ đối với những loài ngoài CSDL, tránh cung cấp kết quả sai lệch có thể ảnh hưởng đến quyết định điều trị.

Ý nghĩa lâm sàng của định danh chính xác đến cấp loài nằm trực tiếp ở quyết định kháng nấm ban đầu. *C. krusei* (danh pháp mới *Pichia kudriavzevii*) có kháng fluconazole nội tại; *C. glabrata* (danh pháp mới *Nakaseomyces glabratus*) có xu hướng MIC tăng dần với echinocandin; *C. parapsilosis* và các loài đồng hình trong phức hợp có hành vi kháng echinocandin khác nhau [13]. Kết quả cho thấy công cụ nhận diện đúng toàn bộ *C. tropicalis* (24/24), *C. glabrata* (9/9) và *C. krusei* (4/4), các loài có ý nghĩa kháng thuốc hàng đầu hiện nay. Ngoài ra, nhóm phức hợp *C. parapsilosis* được xử lý an toàn: các loài đồng hình (*C. orthopsilosis*, *C. metapsilosis*) được báo cáo ở mức phức hợp loài thay vì định danh sai về loài - phù hợp với khuyến nghị lâm sàng hiện hành [49].

Trường hợp dương tính giả ghi nhận ở chủng B36 (*Debaryomyces nepalensis* bị báo là *C. famata*), nguyên nhân là do thiếu thông tin khai báo về phức hợp loài cho *C. famata* trong quá trình huấn luyện. Về di truyền, *D. nepalensis* và *C. famata* có độ tương đồng cao đối với trình tự ITS (Hình 1), khiến công cụ phân tích gán nhãn sai ở độ phân giải mức độ loài [50]. Có hai trường hợp âm tính giả (FN) ghi nhận trong nhóm 54 trình tự ngoài CSDL: (1) trình tự *C. margitis* (số truy cập GenBank: NR159750) được định danh nhầm thành phức hợp *C. parapsilosis*; (2) trình tự *C. hyderabadensis* (NR159741) được định danh nhầm thành *C. parapsilosis* ở mức loài. Cả hai loài này đều không có trong CSDL huấn luyện nhưng có trình tự ITS đủ tương đồng với *C. parapsilosis* để vượt ngưỡng K2P, dẫn đến gán nhầm nhãn thay vì trả kết quả “Unknown”. Những hạn chế này có thể được

khắc phục nếu cơ sở dữ liệu được bổ sung và khai báo đầy đủ thông tin về phức hợp loài. Trong nhóm 92 trình tự lâm sàng, không ghi nhận âm tính giả - *Kodamaea ohmeri* (B5), loài duy nhất trong nhóm này nằm ngoài CSDL, được xử lý đúng với kết quả “Unknown”.

Về tiềm năng ứng dụng trong kiểm soát nhiễm khuẩn bệnh viện, tổng quan tích hợp của El Arab và cs trên 42 nghiên cứu ghi nhận các mô hình AI — bao gồm mạng nơ-ron, decision tree, random forest và học sâu - đạt AUC > 0,80 trong phát hiện nhiễm khuẩn bệnh viện, đặc biệt thể hiện độ tin cậy cao đối với nhiễm khuẩn vết mổ và nhiễm khuẩn đường tiết niệu và hỗ trợ quyết định kiểm soát lây truyền [3]. Trong khi đó, tổng quan hệ thống và phân tích tổng hợp của Almeida và cs trên 14,620 bệnh nhân cho thấy các mô hình AI/ML đạt giá trị tiên đoán âm (NPV) > 95% trong dự báo nhiễm *Candida* xâm lấn. Đặc điểm này khiến AI trở nên phù hợp nhất cho vai trò hỗ trợ ra quyết định lâm sàng nhằm loại trừ bệnh và tối ưu hóa công tác quản lý sử dụng thuốc kháng nấm [2].

Trong bối cảnh đó, một công cụ định danh tự động với kết quả đầu ra được cấu trúc hóa - như đã trình bày trong nghiên cứu này - có cơ sở khoa học và tiềm năng để tích hợp vào mạng lưới giám sát nhiễm khuẩn bệnh viện. Khả năng tích lũy dữ liệu dọc theo thời gian của hệ thống sẽ cho phép phân tích xu hướng phân bố dịch tễ của các loài nấm và phát hiện sớm sự trỗi dậy của các chủng đa kháng. Tiềm năng này mang ý nghĩa đặc biệt đối với *C. auris* - một mầm bệnh đa kháng có khả năng bùng phát thành các vụ dịch lây nhiễm chéo tại bệnh viện, hiện đang lây lan nhanh tại châu Âu theo báo cáo giám sát của ECDC năm 2024, nhưng lại thường dễ bị định danh sai lệch bởi các hệ thống sinh hóa tự động hay thiết bị MALDI-TOF MS thông thường [51]. Tuy nhiên, để chuyển hóa những tiềm năng này thành bằng chứng ứng dụng thực tiễn, vẫn đòi hỏi các nghiên cứu tiếp theo được thiết kế chuyên biệt nhằm đánh giá tác động trực tiếp.

Nghiên cứu này tồn tại một số hạn chế cần được xem xét trong bối cảnh hiện tại. Thứ nhất, mặc dù bộ dữ liệu đánh giá đã được mở rộng lên 145 trình tự, phần trình tự lâm sàng thực tế chỉ gồm 92 chủng từ một trung tâm duy nhất (Bệnh viện Srinagarind, Thái Lan); do đó độ đặc hiệu của công cụ chưa được kiểm định trên quần thể lâm sàng đa trung tâm hay ở bối cảnh dịch tễ khác. Phần lớn nghiên cứu AI trong vi nấm gần đây đều đối mặt với hạn chế tương tự: tổng quan hệ thống 2026 theo nguyên tắc PRISMA phân tích 11 nghiên cứu ghi nhận rằng phần lớn mô hình AI chỉ được xác nhận nội bộ, rất ít nghiên cứu có đánh giá ngoài đa trung tâm [1]; kết quả đánh giá

chất lượng bằng công cụ PROBAST của Almeida và cs cũng chỉ ra nguy cơ sai số hệ thống từ mức trung bình đến cao ở hầu hết các nghiên cứu được đưa vào phân tích. Điều này càng củng cố thêm quan điểm rằng: dù các mô hình có chỉ số thống kê ấn tượng, cần phải có các thử nghiệm lâm sàng và xác thực bổ sung từ các đơn vị độc lập bên ngoài trước khi triển khai thực tế [2]. Thứ hai, đánh giá hiệu năng trong nghiên cứu này tập trung vào chi *Candida* và các loài liên quan; hiệu năng đối với các chi còn lại trong CSDL (*Cryptococcus*, *Malassezia*, *Rhodotorula*, *Trichosporon*, *Aspergillus*, *Fusarium*) chưa được đánh giá hệ thống, cần có các nghiên cứu tiếp theo để đánh giá chính xác hơn tính tổng quát của công cụ trong bối cảnh lâm sàng thực tế.

Để bổ sung bằng chứng ứng dụng lâm sàng, các nghiên cứu tiếp theo sẽ được thiết kế theo hai hướng ưu tiên: (1) xác nhận ngoài đa trung tâm trên bộ chủng lâm sàng đa dạng về địa lý và dịch tễ, nhằm đánh giá tính khái quát của hiệu năng công cụ. (2) triển khai công cụ song song với quy trình định danh thường quy của phòng xét nghiệm, cung cấp bằng chứng thực nghiệm cho giá trị ứng dụng trong quyết định kháng nấm và kiểm soát nhiễm khuẩn.

Định hướng phát triển tiếp theo của nhóm nghiên cứu là triển khai công cụ dưới dạng ứng dụng web. Giải pháp này cho phép các phòng xét nghiệm lâm sàng dễ dàng phân tích dữ liệu trình tự vùng ITS trực tiếp qua trình duyệt mà không đòi hỏi đầu tư hạ tầng hay cài đặt phần mềm. Bên cạnh đó, bộ CSDL huấn luyện sẽ được mở rộng và tích hợp thêm các locus phụ (như vùng D1/D2 LSU hoặc TEF1 α). Đặc biệt, công cụ sẽ được trang bị thêm tính năng tùy biến, cho phép người dùng tự xây dựng và tinh chỉnh CSDL huấn luyện nội bộ, giúp người dùng chủ động cập nhật CSDL để bám sát đặc điểm dịch tễ tại địa phương.

5. KẾT LUẬN

Công cụ mã vạch DNA vùng ITS tự động đạt độ chính xác 97,9%, độ nhạy 97,8%, độ đặc hiệu 98,1% và thời gian xử lý 0,93 giây/trình tự, cho thấy khả năng vận hành nhanh và ổn định. Giá trị nổi bật của hệ thống là sự kết hợp giữa tự động hóa phân tích trình tự, cơ chế báo cáo phân cấp an toàn tránh định danh quá mức và đầu ra chuẩn hóa hỗ trợ tiềm năng tích hợp LIS/HIS phục vụ quản lý kháng nấm và kiểm soát nhiễm khuẩn bệnh viện.

Lời cảm ơn

Nhóm tác giả xin chân thành cảm ơn bà Waewta Kuwatjanakul và bà Lumyai Wonglakorn, Đơn vị Vi sinh lâm sàng, Bệnh viện Srinagarind, Khoa Y, Đại học Khon Kaen, Thái Lan, đã cung cấp các chủng nấm men cho nghiên cứu này.

Tuyên bố về xung đột lợi ích

Các tác giả khẳng định không có xung đột lợi ích liên quan đến nghiên cứu, tác giả và công bố bài báo này.

TÀI LIỆU THAM KHẢO

1. Albuquerque NM, Alkahtani LM, Alharbi YA, Aljuhaymi DA, Alnufaei RD, Al Mashaihi AA, et al. Advancements in Artificial Intelligence-Based Diagnostic Tools Used to Detect Fungal Infections: A Systematic Review. *Diagnostics*. 2026 Feb 1;16(3):450.
2. Almeida H, Rodríguez-Alonso B, Alonso-Sardón M, Izquierdo I, Romero-Alegria Á, Velasco-Tirado V, et al. Can Artificial Intelligence Optimize the Early Diagnosis of Invasive Candidiasis? A Systematic Review and Meta-Analysis. *JoF*. 2026 Feb 13;12(2):138.
3. El Arab RA, Almoosa Z, Alkhunaizi M, Abuadas FH, Somerville J. Artificial intelligence in hospital infection prevention: an integrative review. *Front Public Health*. 2025 Apr 2;13:1547450.
4. Xu J, Luo Y, Wang J, Tu W, Yi X, Xu X, et al. Artificial intelligence-aided rapid and accurate identification of clinical fungal infections by single-cell Raman spectroscopy. *Front Microbiol*. 2023 Mar 22;14:1125676.
5. Tran TK, Kuwatjanakul W, Wonglakorn L, Samerpitak K. Comparative Evaluation of MALDI-TOF MS and ITS Barcoding for the Identification of Clinical and Environmental Yeast Isolates. *J Mycol Infect*. 2026 Mar; 31(1):25–38.
6. Cuevas-Tello JC, Monreal-de la Rosa A, Quistian-Navarro JL, Martínez-Gamboa A, González-Lara MF, López-García NI, et al. Deep learning for the identification of *Candida* spp. directly from blood culture gram stains from candidemia patients. *Medical Mycology*. 2025 Nov 4;63(11):myaf097.
7. Bhargava A, Klamer K, Sharma M, Ortiz D, Saravolatz L. *Candida auris*: A Continuing Threat. *Microorganisms*. 2025 Mar 13;13(3):652.
8. Gherbawy YA, Al-Harhi H, El-Dawy E, Gaber M, Pet I, Wang L, et al. The molecular revolution in fungal diagnostics: bridging gaps across clinical, agricultural, and environmental mycology. *Mycology*. 2025 Dec 3;1–28.
9. Saleh AM, Hassan RYA, Badawey AM, Marzouk HM. Toward sustainable diagnostics for *Candida albicans*: the role of artificial intelligence in analytical chemistry from data processing to Python-based blueness and redness evaluation metrics. *RSC Adv*. 2026;16(20):18389–405.
10. Irinyi L, Serena C, Garcia-Hermoso D, Arabatzis M, Desnos-Ollivier M, Vu D, et al. International Society of Human and Animal Mycology (ISHAM)-ITS reference DNA barcoding database—the quality controlled standard tool for routine identification of human and animal pathogenic fungi. *Medical Mycology*. 2015 May 1;53(4):313–37.
11. Hernández-Pabón JC, Tabares B, Gil Ó, Lugo-Sánchez C, Santana A, Barón A, et al. *Candida Non-albicans* and *Non-auris* Causing Invasive Candidiasis in a Fourth-Level Hospital in Colombia: Epidemiology, Antifungal Susceptibility, and Genetic Diversity. *Journal of Fungi*. 2024 May;10(5):5.
12. Nguyen MD, Ren P. Trends in Antifungal Resistance Among *Candida* Species: An Eight-Year Retrospective Study in the Galveston–Houston Gulf Coast Region. *J Fungi (Basel)*. 2025 Mar 19;11(3):232.
13. Calvo M, Scalia G, Trovato L. Antifungal Susceptibility Data and Epidemiological Distribution of *Candida* spp.: An In Vitro Five-Year Evaluation at University Hospital Policlinico of Catania and a Comprehensive Literature Review. *Antibiotics*. 2024 Oct;13(10):10.
14. Jan A, Bashir G, Altaf I, Fomda BA, Hamid S, Jan K. Evaluation of various phenotypic methods for differentiation of *Candida dubliniensis* from *Candida albicans*. *Journal of Microbiological Methods*. 2022 Feb;193:106400.
15. Nikmanesh B, Ahmadikia K, Getso MI, Gharehbolagh SA, Aboutalebian S, Mirhendi H, et al. *Candida africana* and *Candida dubliniensis* as causes of pediatric candiduria: A study using *HWP1* gene size polymorphism. *AIMSMICRO*. 2020;6(3):microbiol-06-03-017.
16. Lin W, Yang Q, Lin Z, Liu S, Zhang L, Jiang Y, et al. Clinical and Molecular Epidemiological Characteristics of 26 Bloodstream Infection Cases Caused by *Candida tropicalis*. *Mycopathologia*. 2025 Aug;190(4):65.
17. Keighley C, Kim HY, Kidd S, Chen SCA, Alastruey A, Dao A, et al. *Candida tropicalis*—A systematic review to inform the World Health Organization of a fungal priority pathogens list. *Med Mycol*. 2024 Jun 27;62(6):myae040.
18. Baba H, Kanamori H, Nakayama A, Sato T, Katsumi M, Chida T, et al. A cluster of *Candida parapsilosis* displaying fluconazole-trailing in a neonatal intensive care unit successfully contained by multiple infection-control interventions. *Antimicrob Steward Healthc Epidemiol*. 2024 May 16;4(1):e86.
19. Gómez-Gaviria M, García-Carnero LC, Baruch-Martínez DA, Mora-Montes HM. The Emerging Pathogen *Candida metapsilosis*: Biological Aspects, Virulence Factors, Diagnosis, and Treatment. *Infect Drug Resist*. 2024 Jan 20;17:171–85.
20. Naskar S, Prajapati A, Kaur R. Antifungal drug resistance in *Candida glabrata*: role of cellular signaling and gene regulatory networks. *FEMS Yeast Res*. 2025 May 16;25:foaf025.
21. Hou X, Xiao M, Chen SCA, Wang H, Yu SY, Fan X, et al. Identification and Antifungal Susceptibility Profiles of *Candida nivariensis* and *Candida bracarenensis* in a Multi-Center Chinese Collection of Yeasts. *Front Microbiol*. 2017 Jan 19;8:5.
22. Hsu C, Yassin M. Diagnostic Approaches for *Candida auris*: A Comprehensive Review of Screening, Identification, and Susceptibility Testing. *Microorganisms*. 2025 Jul;13(7):7.
23. Ramos LS, Barbosa PF, Lorentino CMA, Lima JC, Braga AL, Lima RV, et al. The multidrug-resistant *Candida auris*, *Candida haemulonii* complex and phylogenetic related species: Insights into antifungal resistance mechanisms. *Current Research in Microbial Sciences*. 2025;8:100354.
24. Cui Y, Huang Z, Han H, Ma Y, Li S, Hu Y, et al. Encephalopathy and Endophthalmitis Following Device-Related Invasive *Clavispora lusitanae* Infection in an Immunocompetent Patient: A Case Report. *Clin Case Rep*. 2025 Jun 16;13(6):e70550.
25. Snapiri O, Danziger CR, Sachs N, Krause I, Zvi HB, Danino D, et al. *Candida lusitanae* Fungemia in Children: A multicenter case series of emerging pathogen. *Med Mycol*.

2024 Jul 1;62(7):myae057.

26. Syed S, Afreenish Amir, Muhammad Waqas Abbas, Saima Ishtiaq, Haider Ali, Shahid Ahmad Abbasi. Systemic candidiasis by *Candida kefyr*: Case series from Pakistan. *J Med Case Rep Case Ser.* 2023 Apr 21;4(6).
27. Zareshahrabadi Z, Khaliji S, Roudbari M, Zomorodian K. Antifungal Drug Susceptibility Profile of *Candida kefyr* Isolated from Clinical Samples and Dairy Products. *Can J Infect Dis Med Microbiol.* 2024 Sep 23;2024:6594366.
28. Al-Sweih N, Ahmad S, Joseph L, Khan S, Vayalil S, Chandu R, et al. *Candida fermentati* as a Cause of Persistent Fungemia in a Preterm Neonate Successfully Treated by Combination Therapy with Amphotericin B and Caspofungin. *Journal of Clinical Microbiology.* 2015 Feb 19;53(3):1038–41.
29. McHugh JW, Bayless DR, Ranganath N, Stevens RW, Kind DR, Wengenack NL, et al. *Candida guilliermondii* fungemia: a 12-year retrospective review of antimicrobial susceptibility patterns at a reference laboratory and tertiary care center. *Journal of Clinical Microbiology.* 2024 Oct 23;62(11):e01057-24.
30. Karapetsa M, Tsolaki V, Arabatzis M, Petinaki E, Velegraki A, Zakyntinos E. Septic shock due to *Candida famata* (*Debaryomyces hansenii*) candidemia in an ICU immunocompetent trauma-patient. *Journal of Infection and Public Health.* 2019 Jul;12(4):594–7.
31. Villanueva-Lozano H, Treviño-Rangel RDJ, Hernández-Balboa CL, González GM, Martínez-Reséndez MF. An unusual case of *Candida ciferrii* fungemia in an immunocompromised patient with Crohn's and Mycobacterium bovis disease. *J Infect Dev Ctries.* 2016 Oct 31;10(10):1156–8.
32. Murata S, Mimura K, Kawamura T, Saito H, Ohno H, Tsujii E, et al. Bloodstream infection caused by *Wickerhamiella pararugosa* in a patient with intestinal obstruction: A case report. *Journal of Infection and Chemotherapy.* 2024 Sep 1;30(9):942–5.
33. Paredes K, Sutton DA, Cano J, Fothergill AW, Lawhon SD, Zhang S, et al. Molecular Identification and Antifungal Susceptibility Testing of Clinical Isolates of the *Candida rugosa* Species Complex and Proposal of the New Species *Candida neorugosa*. *J Clin Microbiol.* 2012 Jul;50(7):2397–403.
34. Pérez-Hansen A, Lass-Flörl C, Lackner M, Rare Yeast Study Group, Aigner M, Alastruey-Izquierdo A, et al. Antifungal susceptibility profiles of rare ascomycetous yeasts. *Journal of Antimicrobial Chemotherapy.* 2019 Sep 1;74(9):2649–56.
35. Kumar S, Kumar A, Roudbary M, Mohammadi R, Černáková L, Rodrigues CF. Overview on the Infections Related to Rare *Candida* Species. *Pathogens.* 2022 Aug 24;11(9):963.
36. Pfüller R, Gräser Y, Erhard M, Groenewald M. A Novel Flucytosine-Resistant Yeast Species, *Candida pseudoaaseri*, Causes Disease in a Cancer Patient. *J Clin Microbiol.* 2011 Dec;49(12):4195–202.
37. Husni R, Bou Zerdan M, Samaha N, Helou M, Mahfouz Y, Sanjour R, et al. Characterization and susceptibility of non-*albicans* *Candida* isolated from various clinical specimens in Lebanese hospitals. *Front Public Health.* 2023 Mar 10;11:1115055.
38. Wu N, Wu Y, Chu Y, Ren Z, Li H, Rong C, et al. The first rare case of *Candida palmioleophila* infection reported in China and its genomic evolution in a human host environment. *Front Microbiol.* 2023 Jul 26;14:1165721.
39. Rashak SJ, Al-Maqtoufi MY, Burghal AA. Molecular Detection of *Candida aaseri* in Oral Cavity of Immunocompromised Patients. *Asian Journal of Biology.* 2024 May 28;20(6):91–7.
40. Adam H, Groenewald M, Mohan S, Richardson S, Bunn U, Gibas CFC, et al. Identification of a new species, *Candida subhashii*, as a cause of peritonitis. *Med Mycol.* 2009 Jan;47(3):305–11.
41. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proc Natl Acad Sci USA.* 2012 Apr 17;109(16):6241–6.
42. Irinyi L, Lackner M, De Hoog GS, Meyer W. DNA barcoding of fungi causing infections in humans and animals. *Fungal Biology.* 2016 Feb;120(2):125–36.
43. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. *Bioinformatics.* 2008 Aug 15;24(16):1757–64.
44. Tavanti A, Davidson AD, Gow NAR, Maiden MCJ, Odds FC. *Candida orthopsilosis* and *Candida metapsilosis* spp. nov. to replace *Candida parapsilosis* Groups II and III. *J Clin Microbiol.* 2005 Jan;43(1):284–92.
45. Shankarnarayan SA, Charlebois DA. Machine learning to identify clinically relevant *Candida* yeast species. *Medical Mycology.* 2024 Jan 9;62(1):myad134.
46. Ren R, Tan W, Chen S, Xu X, Zhang D, Chen P, et al. Deep learning application to hyphae and spores identification in fungal fluorescence images. *Sci Rep.* 2025 Jul 26;15(1):27222.
47. Zolotareva M, Cascalheira F, Afonso R, Corado M, Martins A, Caneiras C, et al. Fungal DNA-barcoding on a chip: Magneto-resistive biosensors for yeast infection diagnosis. *VIEW.* 2025;6(6):20250015.
48. Hofstetter V, Buyck B, Eyssartier G, Schnee S, Gindro K. The unbearable lightness of sequenced-based identification. *Fungal Diversity.* 2019 May;96(1):243–84.
49. Gabaldón T. Threats from the *Candida parapsilosis* complex: the surge of multidrug resistance and a hotbed for new emerging pathogens. Heitman J, editor. *Microbiol Mol Biol Rev.* 2024 Dec 18;88(4):e00029-23.
50. Nguyen HV, Gaillardin C, Neuvéglise C. Differentiation of *Debaryomyces hansenii* and *Candida famata* by rRNA gene intergenic spacer fingerprinting and reassessment of phylogenetic relationships among *D. hansenii*, *C. famata*, *D. fabryi*, *C. flarerii* (= *D. subglobosus*) and *D. prosopidis*: description of *D. vietnamensis* sp. nov. closely related to *D. nepalensis*. *FEMS Yeast Res.* 2009 Jun 1;9(4):641–62.
51. European Centre for Disease Prevention and Control. Survey on the epidemiological situation, laboratory capacity and preparedness for *Candidozyma (Candida) auris*, 2024. [Internet]. LU: Publications Office; 2025 [cited 2026 May 7]. Available from: <https://data.europa.eu/doi/10.2900/2025052>