bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Tracing the evolutionary history and global expansion of *Candida auris* using population genomic analyses

Nancy A. Chow[1,*], José F. Muñoz[2,*], Lalitha Gade[1] , Elizabeth Berkow[1], Xiao Li[2], Rory M. Welsh[1], Kaitlin Forsberg[1], Shawn R. Lockhart[1] , Rodney Adam[3], Alexandre Alanio[4,5,6], Ana Alastruey-Izquierdo[7], Sahar Althawadi[8], Ana Belén Araúz[9], Ronen Ben-Ami[10], Amrita Bharat[11], Belinda Calvo[12], Marie Desnos-Ollivier[13], Patricia Escandón[14], Dianne Gardam[15], Revathi Gunturu[3], Christopher H. Heath[16,17,18,119], Oliver Kurzai[20,21], Ronny Martin[20,21], Anastasia P. Litvintseva[1,#], and Christina A. Cuomo[2,#]

[1]Mycotic Diseases Branch, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, GA, 30329 USA; [2]Broad Institute of MIT and Harvard, Cambridge, MA, 02142 USA; [3] Department of Pathology, Aga Khan University Hospital, Nairobi, Kenya , [4]Institut Pasteur, Molecular Mycology Unit, CNRS UMR2000, National Reference Center for Invasive Mycoses and Antifungals (NRCMA), Paris, France; [5]Laboratoire de Parasitologie-Mycologie, Hôpital Saint-Louis, Groupe Hospitalier Lariboisière, Saint-Louis, Fernand Widal, Assistance Publique-Hôpitaux de Paris (AP-HP); [6]Université Paris Diderot, Université de Paris, Paris, France ; [7]Mycology Reference Laboratory, National Centre for Microbiology, Instituto de Salud Carlos III, Carretera de Majadahonda a Pozuelo Km. 2 28220, Majadahonda, Madrid, Spain; [8]Department of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; [9]Hospital Santo Tomás, Panama City, Panama; [10]Infectious Diseases Unit, Tel Aviv Sourasky Medical Center, and the Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [11]National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, R3E 3R2, Canada; [12]Department of Infectious Diseases, School of Medicine, Universidad del Zulia, Maracaibo, Venezuela; [13]Institut Pasteur, Molecular Mycology Unit, CNRS UMR2000, Paris, France; [14]Grupo de Microbiología, Instituto Nacional de Salud, Calle 26 # 51-20, Bogotá, Colombia; [15]Department of Microbiology, PathWest Laboratory medicine FSH Network, Fiona Stanley Hospital, Murdoch, Australia; [16]Department of Microbiology, PathWest Laboratory Medicine FSH Network, Fiona Stanley Hospital, Murdoch; [17]Department of Infectious Diseases, Fiona Stanley Hospital, Department of Microbiology; [18]Infectious Diseases, Royal Perth Hospital, Perth, WA, Australia; [19]Faculty of Health & Medical Sciences, University of Western Australia, Crawley, WA, Australia; [20]German National Reference Center for Invasive Fungal Infections NRZMyk, Leibniz Institute for Natural Product Research and

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Infection Biology – Hans-Knöll-Institute, Jena, Germany; [21]University of Würzburg, Institute for Hygiene and Microbiology, Würzburg, Germany

Current addresses; XL: Advanced Institute of Information Technology, Peking University, Hangzhou, ZJ, China. 311215.

Running Head: Genomic analysis of *Candida auris* worldwide

[#]Address correspondence to Anastasia P. Litvintseva frq8@cdc.gov and Christina A. Cuomo cuomo@broadinstitute.org

*N.A.C. and J.F.M. contributed equally to this work. Co-first authorship order was assigned alphabetically.

[#]A.P.L. and C.A.C. contributed equally to this work. Co-senior authorship was assigned based on the co-first authorship and the equal contributions of both groups.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Candida auris has emerged globally as a multidrug-resistant yeast that can spread via nosocomial transmission. An initial phylogenetic study of isolates from Japan, India, Pakistan, South Africa, and Venezuela revealed four populations (Clades I, II, III, and IV) corresponding to these geographic regions. Since this description, *C. auris* has been reported in over 30 additional countries. To trace this global emergence, we compared the genomes of 304 *C. auris* isolates from 19 countries on six continents. We found that four predominant clades persist across wide geographic locations. We observed phylogeographic mixing in most clades; Clade IV, with isolates mainly from South America, demonstrated the strongest phylogeographic substructure. *C. auris* isolates from two clades with opposite mating types were detected contemporaneously in a single healthcare facility in Kenya. We estimated a Bayesian molecular clock phylogeny and dated the origin of each clade within the last 339 years; outbreak-causing clusters from Clades I, III, and IV originated 34-35 years ago. We observed high rates of antifungal resistance in Clade I, including four isolates resistant to all three major classes of antifungals. Mutations that contribute to resistance varied between the clades, with Y132F in *ERG11* as the most widespread mutation associated with azole resistance and S639P in *FKS1* for echinocandin resistance. Copy number variants in *ERG11* predominantly appeared in Clade III and were associated with fluconazole resistance. These results provide a global context for the phylogeography, population structure, and mechanisms associated with antifungal resistance in *C. auris.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Importance

In less than a decade, *C. auris* has emerged in healthcare settings worldwide; this species is capable of colonizing skin and causing outbreaks of invasive candidiasis. In contrast to other *Candida* species, *C. auris* is unique in its ability to spread via nosocomial transmission and its high rates of drug resistance. As part of the public health response, whole-genome sequencing has played a major role in characterizing transmission dynamics and detecting new *C. auris* introductions. Through a global collaboration, we assessed genome evolution of isolates of *C. auris* from 19 countries. Here, we described estimated timing of the expansion of each *C. auris* clade and of fluconazole resistance, characterized discrete phylogeographic population structure of each clade, and compared genome data to sensitivity measurements to describe how antifungal resistance mechanisms vary across the population. These efforts are critical for a sustained, robust public health response that effectively utilizes molecular epidemiology.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

In the last decade, *Candida auris* has emerged in healthcare settings as a multidrug-resistant organism in over 30 countries worldwide (1). Primarily a skin colonizer, this pathogenic yeast can cause bloodstream infections and other infections (2), is often resistant to multiple classes of antifungal drugs (3, 4), and can spread via nosocomial transmission causing outbreaks of invasive infections (5-12).

Initial studies suggested that *C. auris* emerged simultaneously and independently in four global regions, as phylogenetic analyses revealed four major clades of *C. auris* wherein isolates clustered geographically (13). These clades are referred to as the South Asian, East Asian, African, and South American clades or clades I, II, III, and IV, respectively (14). The isolates from these clades are genetically distinct, differing by tens to hundreds of thousands of single nucleotide polymorphisms (SNPs), with nucleotide diversity nearly 17-fold higher between clades compared to within clades (14). All the clades, except Clade II, have been linked to outbreaks of invasive infections; uniquely, Clade II appears to have a propensity for ear infections (15). The need for increased global efforts to understand the population structure of *C. auris* was recently highlighted by the discovery of the first Iranian *C. auris* case that yielded a single isolate representing a fifth major clade (16).

Molecular epidemiologic investigations of *C. auris* outbreaks generally show clusters of highly related isolates, supporting local and ongoing transmission (7, 17, 18). The analysis of outbreaks and individual cases has also revealed genetic complexity, with isolates from different clades detected in Germany (19), United Kingdom (20) and United States (21), suggesting multiple introductions into the countries, followed by local transmission. To date, each of the clades appears to have undergone clonal expansion; while *C. auris* genomes have conserved mating and meiosis genes, only one of the two fungal mating types are present in a given clade. Specifically, *MTL***a** is present in Clades I and IV and the other mating type, *MTL*α, is found in Clades II and III (14). Understanding whether mating and recombination between clades is occurring is critical, especially in those countries, where isolates from

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

different clades and opposing mating types overlap in time and space. This information could help contextualize complex epidemiologic findings or transmission dynamics.

In addition to its transmissibility, *C. auris* is concerning because of its high rates of drug resistance. Three major classes of antifungal drugs are currently approved for systemic use – azoles, polyenes, and echinocandins. Most *C. auris* isolates are resistant to fluconazole (azole) (22). Elevated minimal inhibitory concentrations (MICs) to amphotericin B (polyene) have been reported in several studies and resistance to echinocandins is emerging in some countries (22). Numerous mechanisms of antifungal resistance have been described for *C. auris*. Echinocandin resistance has been linked to a single mutation at S639P/F in *FKS1*, the gene that encodes the echinocandin target 1,3-beta-D-glucan synthase (23). Most isolates display a mutation linked to fluconazole resistance in *C. albicans*; three mutations, Y132F, K143R, and F126L, have been identified in *ERG11*, the gene that encodes the azole target lanosterol 14-α-demethylase. These mutations have been shown to associate by clade where Y132F and K143R are predominately found in Clades I and IV and F126L is exclusively in Clade III (13, 24). Additionally, there have been suggestions that increased copy number of *ERG11* may be a mechanism for fluconazole resistance in *C. auris* (14).

To better understand *C. auris* emergence and population structure, we engaged in a global collaboration involving 19 countries to produce a large dataset of *C. auris* whole-genome sequences from hundreds of cases and associated environmental samples from healthcare surfaces. Our goal was to generate a comprehensive genomic description of a global *C. auris* population to provide a population genetic framework for the molecular epidemiologic investigations.

Results

Geographic distribution of *C. auris* major clades

We first performed a phylogenetic analysis to characterize the global distribution of *C. auris*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

clades. By including isolates from previous studies (13, 18), we observed that all 304 isolates in this sample collection clustered in one of the four major *C. auris* clades (**Figure 1A**). In this collection, 126 (41%) were classified as Clade I, seven (2%) as Clade II, 51 (17%) as Clade III, and 120 (39%) as Clade IV. Globally, Clade I was the most widespread and found in ten countries (Canada, France, Germany, India, Kenya, Pakistan, Saudi Arabia, United Kingdom, United Arab Emirates, and United States); Clade II was found in Canada, Japan, South Korea, and United States; Clade III in Australia, Canada, Kenya, South Africa, Spain, and United States; and Clade IV in Colombia, Israel, Panama, United States, and Venezuela (**Figure 1B**). Multiple clades were found in Canada, Kenya, and United States.

In contrast to an initial report (13), we observed a weaker phylogeographic substructure as isolates from countries of most global regions appeared interspersed in phylogenies, although there was notable clustering by country within Clade IV (**Figure 1C**). Within Clade I, there were three predominant subclades, each including isolates from India and Pakistan. The smallest subclade included the B8441 reference genome and three other isolates. The other two subclades were more closely related to each other and included groups of highly related isolates from outbreaks in Kenya, United Kingdom, and United States. Additionally, Clade I isolates from countries in Europe (Germany, France) and the Middle East (Saudi Arabia, United Arab Emirates) appeared interspersed in the phylogeny, suggesting multiple introductions of *C. auris* into these countries. Clade II was rarely observed and consisted of seven diverse isolates from Japan, South Korea, United States, or Canada, where six were from cases involving ear infections. Other examples of phylogeographic mixing included isolates from Australia and Spain clustered with Clade III and isolates from Israel clustered with Clade IV, clades originally described as the African and South American clades, respectively.

<u>Evolutionary rate and molecular dating</u>

To better understand the emergence of this species, we next estimated the divergence times of the four major clades. We utilized collection dates for clinical isolates and associated environmental

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

samples, such as swabs from healthcare facilities, which ranged from 2004 to 2018; most (98%) were collected from 2012 to 2018 (**Figure 2A**). By analyzing Clade III isolates from a single healthcare facility in Kenya experiencing an outbreak of ongoing transmission (25), we confirmed that the divergence level of these isolates are temporally corelated, supporting use of molecular clock analyses, and calculated a mutation rate of 1.8695^{e-5} substitutions per site per year (**Supplementary Figure 1**). This rate was used for a Bayesian approach for molecular dating of a phylogeny for all four clades. Using a strict clock coalescent model, we estimated that the time to most recent common ancestor (TMRCA) for each clade occurred within the last 339 years (**Figure 2B, C**). Clade IV emerged most recently with a TMRCA of 1984 (95% HPD 31.5 - 37.1 years ago), while Clade II was the oldest with a TMRCA of 1679 (95% HPD 321.2 - 356.4 years ago). Lastly, we observed the divergence within the 19th century for the two most closely related clades, Clades I and III, 1878 (95% HPD 132.8 - 149.0 years ago) and 1844 (95% HPD, 164.2 - 187.3 years ago), respectively (**Figure 2B, C**). These dates are impacted by the inclusion of divergent isolates in both clades, which notably do not have *ERG11* resistant mutations and are often drug susceptible, one isolate from Canada in Clade III and two isolates from Pakistan and one from United States in Clade I (**Figures 1C** and **2B**). Excluding these drug susceptible outliers for Clades I and III, TMRCA estimates are for Clade I in 1984 (95% HPD, 17.5 – 21.74 years ago) and for Clade III in 1985 (95% HPD, 16.0 – 21.5 years ago); these more recent estimates are more similar to that estimated for Clade IV. The estimated dates of TMRCA for each individual clade supports the recent expansion during the ongoing outbreak. Together this suggests an older separation of the four clades and the recent diversification of each clade in the years before the detected outbreaks.

C. auris population structure

We next examined the global genomic data set for evidence of population substructure and recent admixture. Principal Component Analysis (PCA) identified four well separated populations

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

corresponding to Clades I, II, III, and IV, and the tight clustering of isolates within each clade suggests there is no recent admixture in any isolates (**Figure 3A**). To assess genetic diversity at the clade level, we compared population genetic statistics, including nucleotide diversity (π), Tajima's D (TD), fixation index (F_{ST}) and pairwise nucleotide diversity (D_{XY}) (**Figure 3B to E**). Overall, Clade I and III showed the lowest genetic diversity (π = 1.51e-5 and π = 1.42e-5, respectively); Clade IV exhibited nearly three times these levels (π = 4.23e-5) and Clade II presented the highest genetic diversity (π = 1.29e-4), nearly nine times higher than Clades I and III (**Figure 3B**). Clade II was also the only clade that exhibited positive TD (td = 1.153; **Figure 3B** and **C**), suggesting balancing selection as expected from the long branches observed in Clade II phylogeny (**Figure 1C**). In Clades I, III, and IV we observed negative TD values consistent with recent population expansions and the shorter phylogenetic branches (**Figure 3B** and **C, Figure 1C**); however, Clade IV exhibited a highly variable distribution of TD relative to Clade I and III (**Figure 3C**; **Supplementary Figure 2**), which suggests that these clades have experienced distinct evolutionary processes, such as different degrees of population bottlenecks.

Genome-wide F_{ST} analysis highlighted substantial interspecific divergence and reproductive isolation between *C. auris* clades (average genome-wide $F_{ST} > 0.94$ in all inter-clade comparisons; **Figure 3D**; **Supplementary Figure 2**). Comparison of the two most closely related clades (Clades I and III) revealed small regions with F_{ST} values close to zero; these regions of identity were distributed across the genome (5.86% of the genome; **Figure 3E**; **Supplementary Figure 2, 3**). Phylogenetic analysis revealed that these regions in isolates from Clades I and III are intermixed in a monophyletic clade (**Supplementary Figure 3**). Comparison of D_{XY} values across the genome highlighted regions of population divergence between *C. auris* clades. Even between these clades with substantial interspecific divergence, we detected large genomic tracts that exhibit either high or low D_{XY} values. For D_{XY}, we observed that all chromosomes exhibited a bimodal distribution of regions of both high and low levels of D_{XY}; scaffolds 8 and 10, which correspond to chromosomes 1 and 3, respectively (14), show only low

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

D_{XY} values (**Figure 3E**; **Supplementary Figure 2**). D_{XY} is expected to be elevated in regions of limited gene flow, which could have arisen in *C. auris* due to chromosomal rearrangements between the clades (26, 27), whereas D_{XY} is unchanged or decreased in regions under recurrent background selection or selective sweeps.

We observed a single *C. auris* mating type in each clade. Isolates in Clades I and IV had *MTL***a** and those in Clades II and III had *MTL*α (**Supplementary Figure 4**); this confirms prior findings from a smaller data set (14) in this larger global survey. Countries with multiple clades (i.e., Canada, Kenya, and United States) had isolates of opposite mating types; however, there is no evidence of hybridization between clades within these countries or even between isolates of opposite mating types that were observed contemporaneously in a single healthcare facility in Kenya based on the PCA analysis. Together, these findings suggest that the *C. auris* clades have been genetically isolated and that variation across the genome was likely impacted by karyotype variation that prevented equal chromosome mixing.

Antifungal drug resistance and mechanisms of resistance

To examine resistance levels, we performed antifungal susceptibility testing (AFST) to fluconazole, amphotericin B, and micafungin – drugs representing each of the major classes. Of the 294 isolates tested, 80% were resistant to fluconazole, 23% to amphotericin B, and 7% to micafungin (**Table 1, Figure 4**). Clade II had the greatest percentage (86%) of susceptible isolates, including only one isolate resistant to fluconazole, and Clade I had the greatest percentage of resistant isolates to fluconazole (97%) and amphotericin B (54%). Additionally, Clade I had the highest rates of multidrug-resistance (two antifungal classes; 49%) and was the only clade to have extensive drug-resistance (3%) to all three major classes of antifungals, including isolates from two geographic regions (United Arab Emirates and Kenya) that cluster together (**Table 1**; **Figure 4**). Amphotericin B resistance only appeared in Clades I and IV, and was dispersed across the phylogeny in Clade I and detected in a Clade IV cluster

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of isolates from Colombia. Clade IV also had the highest percentage (9%) resistant to micafungin, all isolates from Venezuela. Micafungin resistance appeared sporadically in the phylogenies of Clades I and III.

We next determined the genotype of specific drug mutations in the *ERG11* gene that have been associated with azole resistance (Y132F, K143R and F126L). The most widespread mutation was Y132F spanning 11 countries in 53% of isolates from Clade I and 40% of isolates from Clade IV (**Figure 4**). *ERG11* K143R was predominately found in a subclade within Clade I (43%) and one isolate from Clade IV. F126L was found only in Clade III, in nearly all isolates (96%) (**Figure 4**); all isolates with F126L also carried the adjacent mutation V125A. Nearly all of the isolates with these changes in *ERG11* were resistant to fluconazole; 99% of the isolates with Y132F or K143R and 100% of the isolates with F126L/V125A appeared resistant to fluconazole (MIC ≥32 µg/mL). We also identified polymorphisms in S639 in the hotspot1 of the *FKS1* gene in 90% of the isolates with decreased susceptibility to micafungin. The most frequent mutation was S639P in 13 isolates from Clade IV (11 resistant to micafungin), and S639F and S639Y were found in micafungin-resistant isolates from Clade I and III (**Figure 4**).

Analysis of the distribution of *ERG11* copy number variation (CNV) revealed that of 304 isolates, 18 (6%) had either two or three copies. Of those 18 isolates, all were resistant to fluconazole and 17 (94%) were from Clade III (**Supplementary Figure 5**). Isolates within Clade III with 2 and 3 copies of *ERG11* had significantly higher MICs ($P \leq 0.05$; Mann-Whitney test) to fluconazole than isolates with one copy (**Supplementary Figure 5**). Along with CNVs in *ERG11*, we found a total of six long regions (>40 kb) that showed increased copy number. Unlike CNVs in *ERG11*, these CNVs appeared in single isolates even in highly clonal clusters, with two isolates in each of Clades I, II, and IV (**Supplementary Figure 6**). While genes in these regions (between 23 and 125 genes in each region) have no direct relation with antifungal resistance, they might play a role in microevolution and *C. auris*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

adaptation to host stress. This includes genes associated with response to oxidative stress (*AOX2*, *HSP12*), iron assimilation (*FET33*, *FTR1*, *CHA1*, *FLC1*), cell wall and membrane integrity (*MNN2*, *ERG5*, *ERG24*), a transcription factor (*ZCF16*) and oligopeptide transporters associated with metabolic and morphologic adaptation and adherence (**Supplementary Table 2**). These data provide insight into the underlying molecular mechanisms of antifungal resistance and suggest that CNV could be a mechanism of strain variation in *C. auris*. Further exploration and monitoring of these traits are crucial to improve our understanding of *C. auris* diversity and control the expanding outbreak.

Discussion

In this study, we used whole-genome sequencing to describe a global collection of *C. auris* isolates collected from patients and healthcare facilities between 2004 and 2018. We found that the four predominant clades are genetically distinct with strong geographic substructure in Clade IV. Using collection dates to estimate a molecular clock, we dated the origins of the four clades and confirmed the recent emergence of *C. auris*. Furthermore, we characterized mutations associated with antifungal resistance by clade, which varied between clades and country of isolation. While the clades appear largely clonal in species phylogenies and represent a single mating type, we found that they have distinct evolutionary histories and genome-wide patterns of variation. We provided a browsable version for *C. auris* genomic epidemiology through Microreact (28) to explore phylogeny, geographic distribution, timeline and drug resistance mutations (https://microreact.org/project/wfwXVjf9G).

In contrast to previous reports, we observed more phylogeographic mixing for *C. auris* (13). While we found that isolates from additional global regions can be clearly assigned to one of the four previously reported clades, we observed that three countries – Canada, Kenya, and United States – had isolates corresponding to multiple *C. auris* clades (**Figure 1B**). Additionally, isolates from multiple clades have been previously reported in Germany (19) and the United Kingdom (20). As travel has been

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

278 previously shown to play a major role in the spread of *C. auris* (21), global travel of persons with prior
279 healthcare exposures to *C. auris* has likely contributed to the observed phylogeographic mixing. Our
280 analysis of the likely geographic origin of infections observed in new geographic regions is limited by
281 incomplete travel history for most patients in this study set. We noted that the strongest geographic
282 substructure was observed in Clade IV for isolates from Colombia, Panama and Venezuela, with
283 additional distinct clades of isolates from Israel and United States (**Figure 1C**). This finding further
284 supports evidence of rapid localized transmission in some of these countries (18, 21).

285 These results have confirmed prior findings from the analysis of a smaller data set where isolates
286 in Clades I and IV had *MTL***a** and those in Clades II and III had *MTL*α (14). Although mating between
287 *C. auris* clades has not been reported, it is concerning that the majority of countries reporting multiple *C.*
288 *auris* clades have clades of opposite mating types. This is especially concerning in Kenya, where
289 opposite mating types were observed in single healthcare facility experiencing ongoing transmission. In
290 such a situation, it could be possible to have mixed infections of opposite mating types. If mating
291 occurred, this would lead to increased genetic diversity and the possibility for enhanced virulence and
292 exchange of drug resistance alleles. Continued efforts to characterize *C. auris* infections at the genomic
293 level are essential for the rapid detection of potential *C. auris* hybrids.

294 Assessment of *C. auris* population structure by PCA and genome-wide F_{ST} analysis yielded no
295 evidence for admixture between the major clades. The close relationship of Clades I and III is
296 highlighted by the detection of regions with very low F_{ST} values, which suggest recent divergence or
297 genetic exchange between these clades. Given that these regions were short and spread across the
298 genome, we hypothesized that they are a result of incomplete lineage sorting rather than recent
299 introgression events. We also observed variation in the average divergence between clades (D_{XY}) along
300 each chromosome. This may be due to genome rearrangements between the clades, whereby genomic
301 areas exhibiting high D_{XY} levels, or low gene flow, arose in *C. auris* due to chromosomal

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rearrangements, which prevents recombination and supports high rates of genetic differentiation. Variation in chromosome number and size as measured by electrophoretic karyotyping as well as deletions, inversions and translocations detected by comparing genome assemblies of different *C. auris* clade isolates have been described in *C. auris* (14, 26, 27).

This global survey has provided a wider perspective of the mechanisms and the frequency of mutations associated with resistance to antifungal drugs. The presence of both resistant and susceptible isolates in the same populations along with the presence of genetically related isolates with different alleles of resistance genes indicate that the resistance in *C. auris* is not intrinsic and has been recently acquired. The most common mutation associated with azole resistance in Clades I and IV was *ERG11* Y132F; however, both clades also included genetically related isolates with *ERG11* K143R. In contrast, all fluconazole resistant isolates in Clade III carried *ERG11* F126L substitution. In addition to mutations in genes associated with drug resistance, we found that increase in copy number of *ERG11* is predominantly observed in Clade III again suggesting clade specific variation in mechanisms of azole resistance. All but three isolates with micafungin resistance had *FKS1* S639Y/P/F mutations. Taken together, these observations suggest recent emergence of antifungal resistance in *C. auris* populations, most likely in response to some unknown environmental change, such as increased use of azole antifungals in clinical practice, agriculture, or both.

By using a molecular clock, we estimated ages of the four clades by calculating time to the most recent common ancestor (TMRCA) of each clade. Our estimates demonstrated that Clade II was the oldest clade with TMRCA of 339 years, while Clade IV was the youngest with TMRCA of 34 years. Clade I and III isolates coalesce 140 and 175 years ago, respectively; however, in both Clade I or Clade III the clusters of isolates associated with ongoing drug resistant outbreaks worldwide, which display increased resistance to fluconazole and harbor mutations in *ERG11* associated with drug resistance, have emerged within the last 35 years. These results are consistent with the other population characteristics:

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

326 even with the smallest sample set, Clade II had the highest genetic diversity compared to the three other
327 clades and a positive TD indicative of the balancing selection, characteristics of an older population.
328 Conversely, three other clades had low genetic diversity and negative TD consistent with the rapid
329 emergence. Notably, the oldest *C. auris* isolate was collected from a patient in South Korea in 1996 (2)
330 and no other strains were identified by searching the historic *Candida* culture collections. The absence
331 of *C. auris* in culture collections prior to 1996 and a rapid emergence after 2012 suggest that this
332 organism only recently emerged as a human pathogen and likely occupied a different ecological niche.
333 Other notable fungal outbreaks also have been estimated to be of recent origin. For example, the
334 BdGPL lineage of the amphibian pathogen *Batrachochytrium dendrobatidis* was estimated to have
335 arisen only ~100 years ago (29). While our reported mutation rate of 1.8695^{e-5} substitutions per site is
336 consistent with that (5.7^{e-5}; $R^2 = 0.37$) reported in a previous study (17), the mutation rate over longer
337 time spans than we sampled is likely slower. We used collection dates spanning from 2004 to 2018 to
338 inform our estimate, and rates of molecular evolution measured over short time-scales tend to be
339 overestimated, as some sites will be removed over time by natural selection (30). Therefore, the rate is
340 more similar to a spontaneous mutation rather than an evolutionary substitution rate. If our mutation rate
341 is substantially overestimated, the exact times of *C. auris* emergence and clade divergence would be
342 older than we have estimated. We also acknowledge that utilizing only currently known isolates, which
343 are highly similar within clades, provides a limited sampling of a larger source population, which may
344 be also be undergoing sexual recombination. The identification and characterization of a wider
345 population sample of *C. auris* will provide a higher resolution view of the nodes separating these major
346 clades. However, as there is only speculation to date about potential associations or locations of such a
347 source population, we suggest that the dates reported be used as a rough estimate that will need further
348 evaluation when sources of additional diversity are identified.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Our molecular clock estimates demonstrate that nearly all outbreak-causing clusters from Clades I, III and IV originated 34-35 years ago in 1984-85. Such recent origin and nearly simultaneous detection of genetically distinct clades suggest that anthropogenic factors might have contributed to its emergence. Specifically: 1) In the 1980s, azole drugs first became widely used in clinical practice. The first azole topical antifungal drug, miconazole, was approved in 1971 followed by clotrimazole in 1972; both became widely used for treatment of superficial fungal infections in the late 1970s. In 1981, the first oral azole drug, ketoconazole, was released for treatment of systemic fungal infections (31). 2) In agriculture, the first azole fungicides, triadimefon and imazalil, were introduced in 1973, and by the early 1980s, ten different azole pesticide formulations were available. It has been demonstrated that azoles from the agricultural use can penetrate ground water and accumulate in soils. 3) Also noteworthy, our predicated emergence of *C. auris* as a human pathogen coincided with the early stages of AIDS epidemics; however, the wide use of antifungal drugs, such as fluconazole, for treatment of secondary fungal infections, had not started until the 1990s (32-34). Other anthropogenic factors might also have brought *C. auris* into contact with humans (35). Although the emergence of *C. auris* may be due to multiple factors, the coincidence between the introduction of azoles and *C. auris* emergence is intriguing and requires further investigation. Understanding processes that led to the emergence of *C. auris* in humans is important to prevent emergence of other drug resistant fungi and pathogens.

Although a recent study reported an isolate from a fifth clade isolated from a patient in Iran (16), all isolates in our collection were assigned to the previously described Clades I, II, III, and IV. This is noteworthy because isolates from neighboring Pakistan, Saudi Arabia, and United Arab Emirates were represented in the analysis. Indeed, this highlights the unique nature of the divergent Iranian *C. auris* case and advocates for increasing diagnostic capacity worldwide and continued phylogenetic studies to understand *C. auris* diversity.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

While we have included a set of diverse isolates, they likely differ from a random sample of the *C. auris* population. The isolates were obtained by conventional sampling, and, therefore, our findings do not represent country-specific characteristics of *C. auris* molecular epidemiology. Wider sampling may also change the population structure and antifungal susceptibility profiles. Notably, at the time of this analysis, Clade V had not yet been discovered, highlighting the importance of further sampling and genomic characterization. Finally, since the environmental reservoir of *C. auris* remains unknown, our analysis is based solely on the analysis of clinical isolates; higher genetic diversity, deeper divergence times and different population structure is likely to occur in the natural populations of this fungus.

In conclusion, we have provided a comprehensive genomic description of a global *C. auris* survey representing 19 countries on six continents. Given that *C. auris* is a transmissible multidrug-resistant organism causing outbreaks of invasive infections in healthcare studies, an understanding of how *C. auris* is spreading, evolving, and acquiring resistance to antifungal drugs is essential for robust public health responses. Continued efforts to characterize the *C. auris* population, additional mechanisms of antifungal resistance, and environments conducive for mating between clades is critical.

Methods

Sample collection

We performed genomic analyses on sequences from 304 *C. auris* isolates. This collection included *C. auris* from 19 countries on six continents and were sampled from both *C. auris* cases and environmental surfaces from healthcare facilities where ongoing transmission was occurring (**Supplementary Table 1**). Samples from *C. auris* cases were derived from a variety of specimen source sites including sterile sites, such as blood, and non-invasive sites, such as respiratory tract or urine. All samples were a result of convenience sampling. For four countries (Colombia, Kenya, United States, and Venezuela) where more than 50 samples were available, 50 representative samples were selected by

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

proportional random sampling: samples from each country were stratified by city, and then a subset was randomly selected proportionally from each strata.

Sample preparation and whole-genome sequencing (WGS)

The sample collection comprised both publicly available sequences generated from previous studies and newly sequenced isolates (**Supplementary Table 1**). For newly sequenced isolates, except those from France, DNA was extracted using the ZR Fungal/Bacterial DNA MiniPrep kit (Zymo Research, Irvine, CA, USA). For isolates from France, DNA was extracted using NucleoMag plant kit extraction (Macherey-Nagel, Germany) in a KingFisherTM Flex system (Thermo Fisher Scientific). Genomic libraries were constructed and barcoded using the NEBNext Ultra DNA Library Prep kit for Illumina (New England Biolabs, Ipswich, MA, USA) and were sequenced on either the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA) using the HiSeq Rapid SBS Kit v2 500-cycles or the MiSeq platform using the MiSeq Reagent Kit v2 500-cycles. For the two isolates from France, libraries were constructed using the Illumina Nextera Flex protocol and sequenced on an iSeq 100 to generate paired 150 bp reads.

Variant identification

We used FastQC v0.11.5 and PRINSEQ v0.20.3 (36) to assess read quality and perform filtering for low quality sequences using "-trim_left 15 -trim_qual_left 20 -trim_qual_right 20 -min_len 100 -min_qual_mean 25 -derep 14". Paired-end reads were aligned to the *C. auris* assembly strain B8441 (GenBank accession PEKT00000000.2; (14)) using BWA mem v0.7.12 (37). Variants were then identified using GATK v3.7 (38) using the haploid mode and GATK tools *RealignerTargetCreator*, *IndelRealigner*, *HaplotypeCaller* for both SNPs and indels, CombineGVCFs, GenotypeGVCFs, GatherVCFs, SelectVariants, and Variant Filtration. Sites were filtered with Variant Filtration using

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

"QD < 2.0 || FS > 60.0 || MQ < 40.0". Genotypes were filtered if the minimum genotype quality < 50, percent alternate allele < 0.8, or depth < 10 (https://github.com/broadinstitute/broad-fungalgroup/blob/master/scripts/SNPs/filterGatkGenotypes.py). Genomic variants were annotated and the functional effect was predicted using SnpEff v4.3T (39). The annotated VCF file was used to determine the genotype of known mutation sites in *ERG11* and *FKS1*. To determine the mating type locus (*MTL***a** and *MTL*α) in each isolate, the average read depth at the locus was computed from the aligned bam file and normalized by the total coverage depth.

Phylogenetic and phylodynamic analysis

For phylogenetic analysis, sites with an unambiguous SNP in at least 10% of the isolates (*n* = 222,619) were concatenated. Maximum likelihood phylogenies were constructed using RAxML v8.2.4 (40) using the GTRCAT nucleotide substitution model and bootstrap analysis based on 1,000 replicates. Phylogenetic analysis was also performed for each clade using subsets of the entire VCF and visualized using iTOL (41).

For phylodynamic analysis, we assessed for a temporal signal using a set of isolates from Kenya Clade III using TempEst v1.5.3 (42) to quantify and to estimate an initial mutation rate from the R-squared value. Bayesian phylogenies were generated using BEAST v1.8.4 (43) under strict molecular clock (both lognormal and exponential priors). In addition, we applied both Bayesian Skyline coalescent and Coalescent Exponential, and a GTR nucleotide substitution model. We obtained similar results using the molecular rate estimated for a *C. auris* outbreak un United Kingdom (17). Specimen collection dates (month and year) were used as sampling dates; the month of June (year midpoint) was assigned for samples where month was unknown. Bayesian Markov chain Monte Carlo (MCMC) analyses were run for 20 million steps using an unweighted pair-group method with arithmetic mean (UPGMA) tree as a starting tree and MCMC convergence was explored using Tracer v.1.7.1 (44). We generated a maximum

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

clade credibility tree with TreeAnnotator v1.8.4 after discarding 47% as burn-in, and we visualized phylogenies using FigTree v1.4.4.

Population genomic analyses

Population genomic analysis was performed using gdsfmt v1.14.1 (34), SNPRelate v1.12.2 (45), and the PopGenome v2.6.1 (46) R packages. Genome-wide nucleotide diversity (π), Tajima's D (TD), fixation index (F_{ST}) and pairwise nucleotide diversity (D_{XY}) were calculated and plotted per scaffold in 5 kb sliding windows. Genome-wide calculations are the average of all 5 kb windows for each metric. Genomic regions that exhibit copy number variation (CNV; deletions and duplications) were identified using CNVnator v0.3 (47) (genomic windows > 1 kb showing significant variation *P*-value < 0.01).

Antifungal susceptibility testing (AFST)

AFST was performed on 294/304 (97%) isolates. The majority (N=270, 90%) of isolates were tested at the U.S. Centers for Disease Control and Prevention (CDC) as outlined by Clinical and Laboratory Standards Institute guidelines. Custom prepared microdilution plates (Trek Diagnostics, Oakwood Village, OH, USA) were used for fluconazole and the echinocandin micafungin. Interpretive breakpoints for *C. auris* were defined based on a combination of those breakpoints which have been established for other closely related *Candida* species, epidemiologic cutoff values, and the biphasic distribution of minimum inhibitory concentrations (MICs) between the isolates with and without known mutations for antifungal resistance. Resistance to fluconazole was set at ≥32 µg/mL and at ≥4 µg/mL for micafungin. Amphotericin B was assessed by Etests (BioMerieux), and resistance was set at ≥2 µg/mL. For isolates not tested at the CDC, similar methods were employed and described previously (17, 19, 48). As there are no currently approved breakpoints for *C. auris*, for this manuscript breakpoints were set at ≥32 µg/mL for fluconazole, >1 µg/mL for amphotericin B, and ≥4 for micafungin. These MIC

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

values were based on a combination of the wild type distribution (those isolates with no mutations) and PK/PD analysis in a mouse model of infection (49).

Data and resource availability

All Illumina sequence generated by this project is available in the NCBI SRA under BioProjects PRJNA328792, PRJNA470683, PRJNA493622, and PRJNA595978. The phylogenetic tree has been deposited in Microreact (https://microreact.org/project/wfwXVjf9G). A set of isolates are available from the CDC and FDA Antimicrobial Resistance (AR) Isolate Bank (https://www.cdc.gov/drugresistance/resistance-bank/index.html).

Ethics

This project was reviewed by CDC IRB as part of the broader human subjects protocol for the Mycotic Diseases Branch, CDC.

Acknowledgements

The authors thank Anuradha Chowdhary, Joveria Farooqi, Nelesh Govender, Mathew Fisher, and Koichi Makimura for their continued support and sharing isolates, and Mathew Fisher and Angela Early for comments on the manuscript. We also thank Jacques Meis for coordinating isolate acquisition; Javier Peman and Olga Rivero-Menendez for support with the Spanish isolates; Erika Santiago, Jovanna Borace and Angel Cedeño from Hospital Santo Tomas; and Soraya Salcedo, Adriana Marín, Carmen Varón, Nohora Villalobos, Jairo Perez, Julian Escobar for participating in the collection of strains and data from Colombian institutions. This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under award U19AI110818 to the Broad Institute. CAC is a CIFAR fellow in the Fungal Kingdom Program. This work was also made possible through support from the Advanced Molecular Detection (AMD) initiative at CDC.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Disclaimer

The use of product names in this manuscript does not imply their endorsement by the U.S. Department of Health and Human Services. The finding and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1. Frequency of antifungal drug resistance among *Candida auris* isolates by clade.

	Clade I N=116[a]	**Clade II** N=7	**Clade III** N=51	**Clade IV** N=120	**Total** N=294
	% (n)	% (n)	% (n)	% (n)	% (n)
Susceptible	6 (7)	86 (6)	2 (1)	30 (36)	17 (50)
Fluconazole resistant	97 (112)	14 (1)	98 (50)	59 (71)	80 (234)
Amphotericin B resistant	47 (54)	0 (0)	0 (0)	11 (13)	23 (67)
Micafungin resistant	6 (7)	0 (0)	8 (4)	9 (11)	7 (22)
MDR[b]	49 (57)	0 (0)	8 (4)	10 (12)	24 (72)
XDR[c]	3 (4)	0 (0)	0 (0)	0 (0)	1 (4)

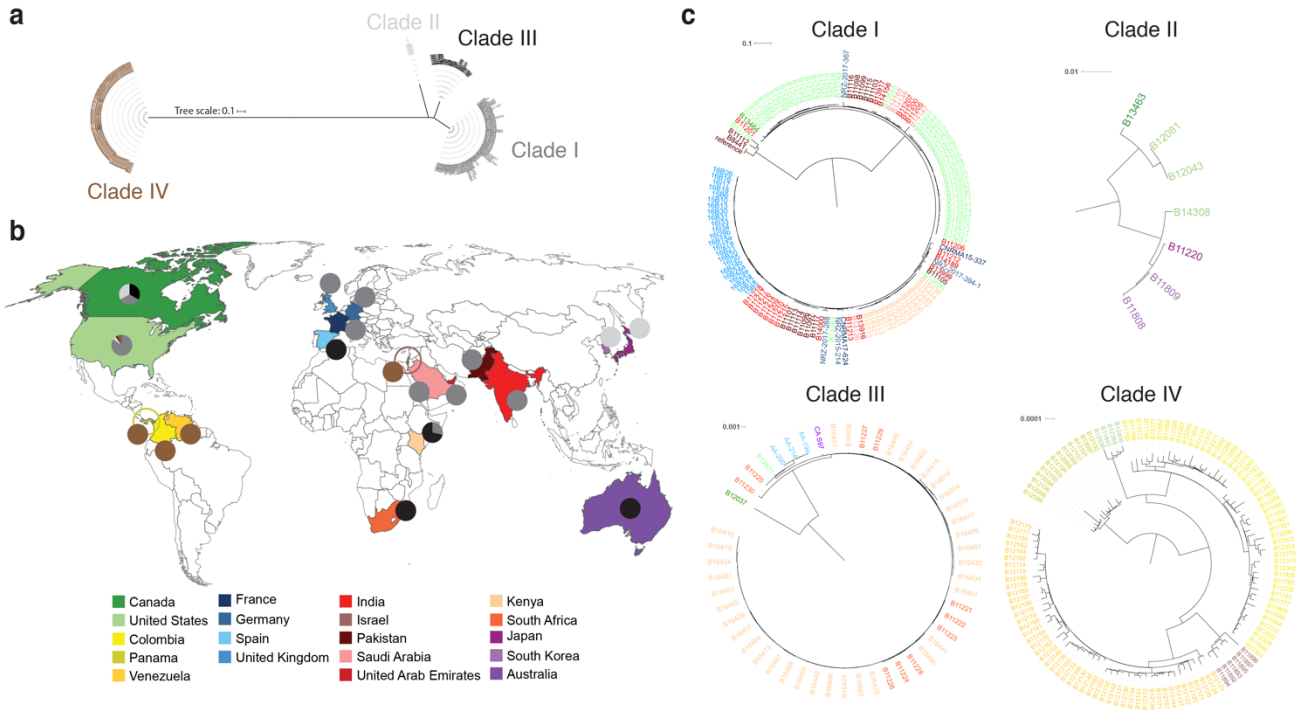
[a] Complete AFST data for 10 of the 126 (3%) Clade I isolates was missing

[b] MDR indicates multidrug-resistance to two major antifungal classes

[c] XDR indicates extensive drug-resistance to three major antifungal classes

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

506



507

508 **Figure 1. Global distribution of *Candida auris* clades. (a)** Phylogenetic tree of 304 *C. auris* whole-

509 genome sequences clustering into four major clades. Maximum likelihood phylogeny using 222,619

510 SNPs based on 1,000 bootstrap replicates. **(b)** Map detailing *C. auris* clade distribution by country (*n* =

511 19). **(c)** Phylogenetic tree of Clades I to IV; colors represent country.

512

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

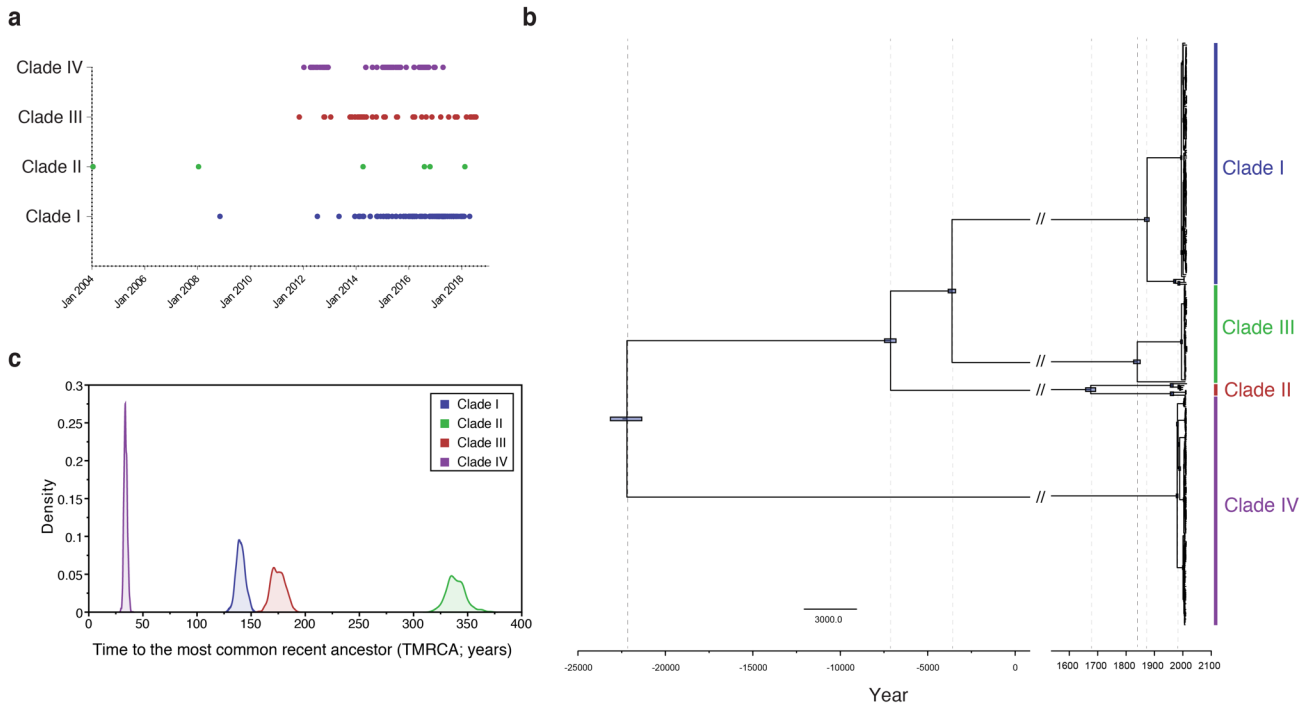


Figure 2. Dating the emergence of *Candida auris*. (a) Distribution of collection dates for source specimens by clade. **(b)** Maximum clade-credibility phylogenetic tree of *C. auris* estimated using BEAST (strict clock and coalescent model). Purple bars indicate 95% highest probability density around a node. **(c)** Marginal posterior distributions for the date of the most recent common ancestor (TMRCA) of *Candida auris* Clades I, II, III and IV. The Bayesian coalescent analysis was performed with BEAST.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

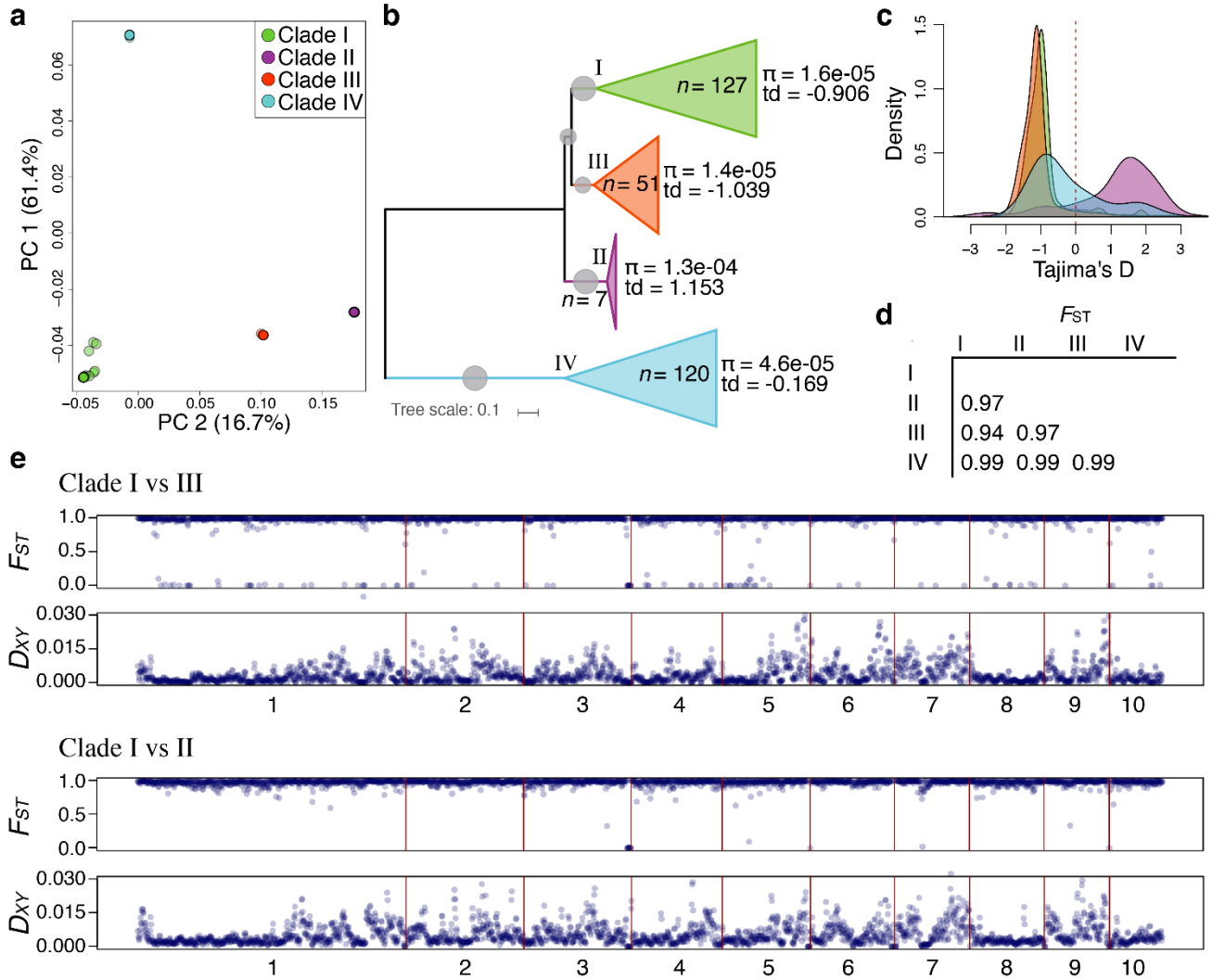


Figure 3. Population structure and genetic differentiation in *Candida auris*. (a) PCA analysis and **(b)** phylogenetic tree of 304 *C. auris* depicting genome-wide population genetic metrics of nucleotide diversity (π) and Tajima's D (TD) for each clade. **(c)** Genome-wide distribution of TD for each clade. **(d)** Average of genome-wide (5-kb windows) variation in fixation index (F_{ST}), for pairwise comparisons in each clade as designated in the first vertical and horizontal row. **(e)** Genome-wide (5-kb windows) pairwise F_{ST} and pairwise nucleotide diversity (D_{XY}) between Clade I and Clade III and Clade I and Clade II are show across the ten largest scaffolds of the B8441 reference genome. All pairwise comparisons of π, TD, F_{ST} and D_{XY} are shown in **Supplementary Figure 2**.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

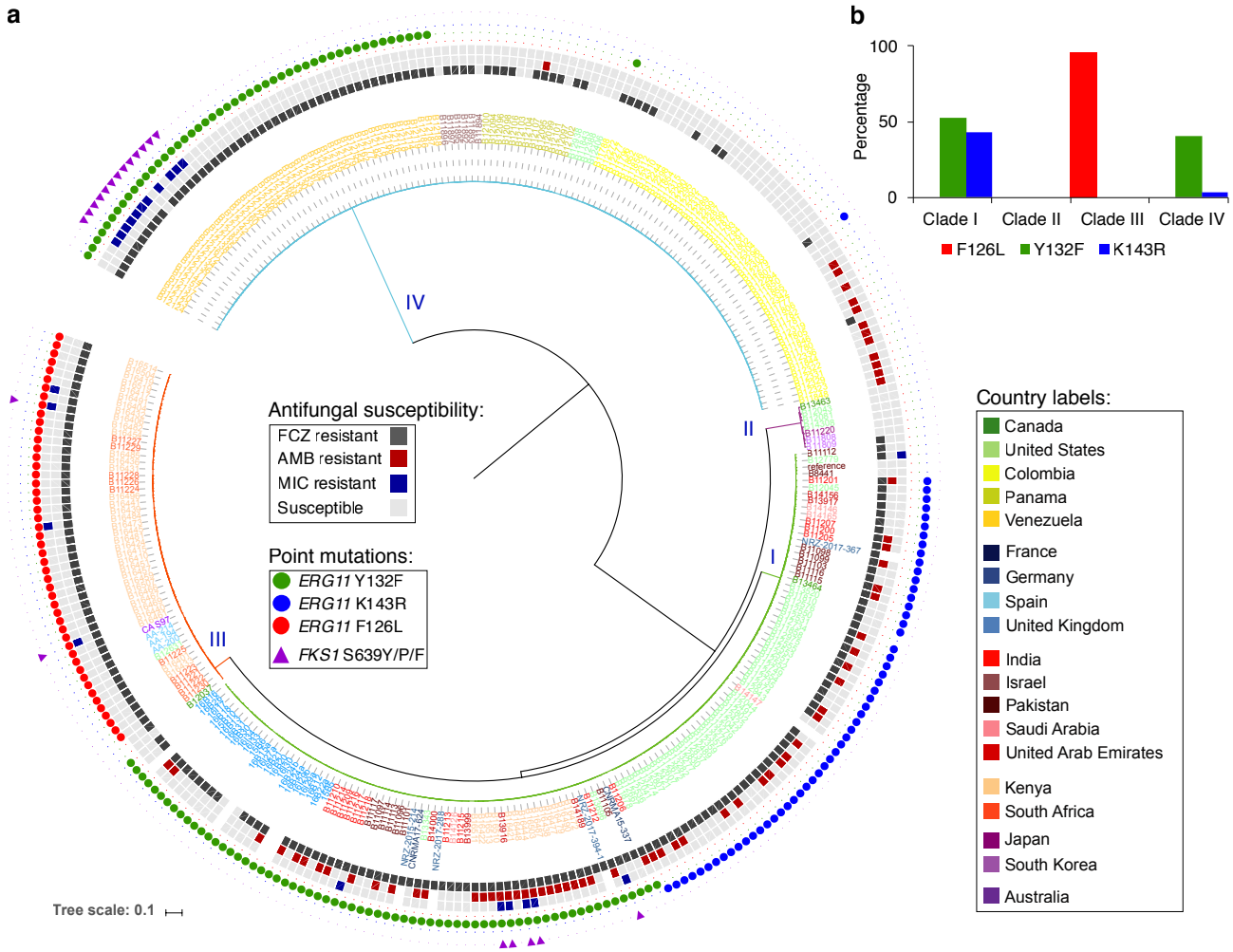


Figure 4. Antifungal susceptibility and point mutations in drug targets in *Candida auris*. (a) Phylogenetic tree detailing clade, susceptibility to fluconazole, amphotericin B, micafungin, and point mutations in *ERG11* (Y132F, K143R, F126L) and *FKS1* (S639Y/P/F) associated with resistance. **(b)** Bar plot describing frequency (%, y-axis) of Y132F, K143R, F126L point mutations in *ERG11* for each clade.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Rhodes J, Fisher MC. Global epidemiology of emerging *Candida auris.* Curr Opin Microbiol. 2019 Jul 3;52:84-9.

2. Lee WG, Shin JH, Uh Y, Kang MG, Kim SH, Park KH, et al. First three reported cases of nosocomial fungemia caused by *Candida auris*. J Clin Microbiol. 2011 Sep;49(9):3139-42.

3. Chowdhary A, Anil Kumar V, Sharma C, Prakash A, Agarwal K, Babu R, et al. Multidrug-resistant endemic clonal strain of *Candida auris* in India. Eur J Clin Microbiol Infect Dis. 2014 Jun;33(6):919-26.

4. Magobo RE, Corcoran C, Seetharam S, Govender NP. *Candida auris*-associated candidemia, South Africa. Emerg Infect Dis. 2014 Jul;20(7):1250-1.

5. Tsay S, Welsh RM, Adams EH, Chow NA, Gade L, Berkow EL, et al. Notes from the Field: Ongoing Transmission of *Candida auris* in Health Care Facilities - United States, June 2016-May 2017. MMWR Morb Mortal Wkly Rep. 2017 May 19;66(19):514-5.

6. Calvo B, Melo AS, Perozo-Mena A, Hernandez M, Francisco EC, Hagen F, et al. First report of *Candida auris* in America: Clinical and microbiological aspects of 18 episodes of candidemia. J Infect. 2016 Oct;73(4):369-74.

7. Sharma C, Kumar N, Pandey R, Meis JF, Chowdhary A. Whole genome sequencing of emerging multidrug resistant *Candida auris* isolates in India demonstrates low genetic variation. New Microbes New Infect. 2016 Sep;13:77-82.

8. Schelenz S, Hagen F, Rhodes JL, Abdolrasouli A, Chowdhary A, Hall A, et al. First hospital outbreak of the globally emerging *Candida auris* in a European hospital. Antimicrob Resist Infect Control. 2016;5:35.

9. Belkin A, Gazit Z, Keller N, Ben-Ami R, Wieder-Finesod A, Novikov A, et al. *Candida auris* Infection Leading to Nosocomial Transmission, Israel, 2017. Emerg Infect Dis. 2018 Apr;24(4):801-4.

10. Ruiz-Gaitan A, Moret AM, Tasias-Pitarch M, Aleixandre-Lopez AI, Martinez-Morel H, Calabuig E, et al. An outbreak due to *Candida auris* with prolonged colonisation and candidaemia in a tertiary care European hospital. Mycoses. 2018 Jul;61(7):498-505.

11. Sana F, Hussain W, Zaman G, Satti L, Khurshid U, Khadim MT. *Candida auris* outbreak report from Pakistan: a success story of infection control in ICUs of a tertiary care hospital. J Hosp Infect. 2019 Jun 22.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Armstrong PA, Rivera SM, Escandon P, Caceres DH, Chow N, Stuckey MJ, et al. Hospital-Associated Multicenter Outbreak of Emerging Fungus *Candida auris*, Colombia, 2016. Emerg Infect Dis. 2019 Jul;25(7).

13. Lockhart SR, Etienne KA, Vallabhaneni S, Farooqi J, Chowdhary A, Govender NP, et al. Simultaneous Emergence of Multidrug-Resistant *Candida auris* on 3 Continents Confirmed by Whole-Genome Sequencing and Epidemiological Analyses. Clin Infect Dis. 2017 Jan 15;64(2):134-40.

14. Munoz JF, Gade L, Chow NA, Loparev VN, Juieng P, Berkow EL, et al. Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species. Nat Commun. 2018 Dec 17;9(1):5346.

15. Welsh RM, Sexton DJ, Forsberg K, Vallabhaneni S, Litvintseva A. Insights into the Unique Nature of the East Asian Clade of the Emerging Pathogenic Yeast *Candida auris*. J Clin Microbiol. 2019 Apr;57(4).

16. Chow NA, de Groot T, Badali H, Abastabar M, Chiller TM, Meis JF. Potential Fifth Clade of *Candida auris*, Iran, 2018. Emerg Infect Dis. 2019 Sep 17;25(9).

17. Rhodes J, Abdolrasouli A, Farrer RA, Cuomo CA, Aanensen DM, Armstrong-James D, et al. Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. Emerg Microbes Infect. 2018 Mar 29;7(1):43.

18. Escandon P, Chow NA, Caceres DH, Gade L, Berkow EL, Armstrong P, et al. Molecular Epidemiology of *Candida auris* in Colombia Reveals a Highly Related, Countrywide Colonization With Regional Patterns in Amphotericin B Resistance. Clin Infect Dis. 2019 Jan 1;68(1):15-21.

19. Hamprecht A, Barber AE, Mellinghoff SC, Thelen P, Walther G, Yu Y, et al. *Candida auris* in Germany and Previous Exposure to Foreign Healthcare. Emerg Infect Dis. 2019 Sep 17;25(9).

20. Borman AM, Szekely A, Johnson EM. Isolates of the emerging pathogen *Candida auris* present in the UK have several geographic origins. Med Mycol. 2017 Jul 1;55(5):563-7.

21. Chow NA, Gade L, Tsay SV, Forsberg K, Greenko JA, Southwick KL, et al. Multiple introductions and subsequent transmission of multidrug-resistant *Candida auris* in the USA: a molecular epidemiological survey. Lancet Infect Dis. 2018 Dec;18(12):1377-84.

22. Lockhart SR. *Candida auris* and multidrug resistance: Defining the new normal. Fungal Genet Biol. 2019 Jun 20;131:103243.

23. Kordalewska M, Lee A, Park S, Berrio I, Chowdhary A, Zhao Y, et al. Understanding Echinocandin Resistance in the Emerging Pathogen *Candida auris*. Antimicrob Agents Chemother. 2018 Jun;62(6).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

24. Chowdhary A, Prakash A, Sharma C, Kordalewska M, Kumar A, Sarma S, et al. A multicentre study of antifungal susceptibility patterns among 350 *Candida auris* isolates (2009-17) in India: role of the ERG11 and FKS1 genes in azole and echinocandin resistance. J Antimicrob Chemother. 2018 Apr 1;73(4):891-9.

25. Adam RD, Revathi G, Okinda N, Fontaine M, Shah J, Kagotho E, et al. Analysis of *Candida auris* fungemia at a single facility in Kenya. Int J Infect Dis. 2019 Aug;85:182-7.

26. Bravo Ruiz G, Ross ZK, Holmes E, Schelenz S, Gow NAR, Lorenz A. Rapid and extensive karyotype diversification in haploid clinical *Candida auris* isolates. Curr Genet. 2019 Apr 24.

27. Munoz JF WR, Shea T, Batra D, Gade L, Litvintseva AP, Cuomo CA. Chromosomal rearrangements and loss of subtelomeric adhesins linked to clade-specific phenotypes in *Candida auris*. bioRxiv. 2019.

28. Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016 Nov;2(11):e000093.

29. O'Hanlon SJ, Rieux A, Farrer RA, Rosa GM, Waldman B, Bataille A, et al. Recent Asian origin of chytrid fungi causing global amphibian declines. Science. 2018 May 11;360(6389):621-7.

30. Ho SY, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, et al. Time-dependent rates of molecular evolution. Mol Ecol. 2011 Aug;20(15):3087-101.

31. Maertens JA. History of the development of azole derivatives. Clin Microbiol Infect. 2004 Mar;10 Suppl 1:1-10.

32. De Wit S, Urbain D, Rahir F, Weerts D, Clumeck N. Efficacy of oral fluconazole in the treatment of AIDS associated oesophageal candidiasis. Eur J Clin Microbiol Infect Dis. 1991 Jun;10(6):503-5.

33. Berry AJ, Rinaldi MG, Graybill JR. Use of high-dose fluconazole as salvage therapy for cryptococcal meningitis in patients with AIDS. Antimicrob Agents Chemother. 1992 Mar;36(3):690-2.

34. Cohen J. Treatment of systemic yeast infection with fluconazole. J Antimicrob Chemother. 1989 Feb;23(2):294-5.

35. Jackson BR, Chow N, Forsberg K, Litvintseva AP, Lockhart SR, Welsh R, et al. On the Origins of a Species: What Might Explain the Rise of *Candida auris*? J Fungi (Basel). 2019 Jul 6;5(3).

36. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011 Mar 15;27(6):863-4.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.06.896548; this version posted January 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

37. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.

38. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297-303.

39. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012 Apr-Jun;6(2):80-92.

40. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312-3.

41. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. Science. 2006 Mar 3;311(5765):1283-7.

42. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016 Jan;2(1):vew007.

43. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007 Nov 8;7:214.

44. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018 Sep 1;67(5):901-4.

45. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012 Dec 15;28(24):3326-8.

46. Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014 Jul;31(7):1929-36.

47. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011 Jun;21(6):974-84.

48. Heath CH, Dyer JR, Pang S, Coombs GW, Gardam DJ. *Candida auris* Sternal Osteomyelitis in a Man from Kenya Visiting Australia, 2015. Emerg Infect Dis. 2019 Jan;25(1):192-4.

49. Lepak AJ, Zhao M, Berkow EL, Lockhart SR, Andes DR. Pharmacodynamic Optimization for Treatment of Invasive *Candida auris* Infection. Antimicrob Agents Chemother. 2017 Aug;61(8).