Short-Term Evolution and Dispersal Patterns of Fluconazole-Resistance in *Candida auris* Clade III

Irving Cancino-Muñoz (1,2*); Juan Vicente Mulet-Bayona (3*); Carme Salvador-García (3); Nuria Tormo-Palop (3); Remedios Guna (3); Concepción Gimeno-Cardona (3$); Fernando González-Candelas (1,2,5$)

(1) Unidad Mixta Infección y Salud Pública FISABIO-Universidad de Valencia, Valencia, España
(2) Instituto de Biología Integrativa de Sistemas, I2SysBio (CSIC-UV), Valencia, España
(3) Servicio de Microbiología y Parasitología, Consorcio Hospital General Universitario de Valencia, Valencia, España
(4) Departamento de Microbiología y Ecología, Universidad de Valencia, Valencia, España
(5) CIBER en Epidemiología y Salud Pública, ISCIII, Madrid, España

(*) Equal contribution

($) Authors for correspondence: concepcion.gimeno@uv.es ; fernando.gonzalez@uv.es

ABSTRACT

The rapid increase of infections caused by the emerging fungal pathogen *Candida auris* is of global concern, and understanding its expansion is a priority. The phylogenetic diversity of the yeast is clustered in five major clades, among which clade III is particularly relevant, as most of its strains exhibit resistance to fluconazole, reducing the therapeutic alternatives and provoking outbreaks that are difficult to control. In this study, we have investigated the phylogenetic structure of clade III by analyzing a global collection of 566 genomes. We have identified three subgroups within clade III, among which two are genetically most closely related. Moreover, we have estimated the evolutionary rate of clade III to be 2.25e-7 substitutions/site/year (2.87 changes per year). We found that one of these subgroups shows intrinsic resistance to fluconazole and is responsible for the majority of cases within this clade globally. We inferred that this subgroup may have originated around December 2010 (95% CI: April 2010 - June 2011), and since then it has spread across continents, generating multiple large outbreaks, each with a unique pattern of transmission and dissemination. These results highlight the remarkable ability of the pathogen to adapt to its environment and its rapid global spread, underscoring the urgent need to effectively address this epidemiological challenge.

Keywords: Fungal pathogen; *Candida auris*; Fluconazole resistance; phylogenetic structure; Dispersion.

INTRODUCTION

Infections caused by the emerging pathogen *Candida auris* have increased dramatically worldwide (Du *et al.*, 2020; Chakrabarti and Sood, 2021; Akinbobola *et al.*, 2023). This fungus is characterized by causing large, difficult-to-control nosocomial outbreaks (Chow *et al.*, 2020; Tian *et al.*, 2021; Mulet Bayona *et al.*, 2023), as well as its rapid acquisition of resistance to antifungals (Muñoz *et al.*, 2018; Du *et al.*, 2020). As a result of this epidemiological situation, the World Health Organization has recently included it in its fungal priority pathogens list (WHO, 2022).

The phylogenetic structure of *C. auris* consists of 5 major clades with different genomic and phenotypic characteristics. For instance, clade I, originated in South Asia, is characterized by its rapid dissemination in clinical settings and is responsible for outbreaks that are difficult to eradicate worldwide (Chow *et al.*, 2018; Thatchanamoorthy *et al.*, 2022; Codda *et al.*, 2023). In contrast, strains belonging to clade V have only been detected in individuals from Iran (Spruijtenburg *et al.*, 2022). Most of the strains classified into clade III are resistant to fluconazole, an antifungal agent used as a first-line treatment against *Candida* infections (Chybowska, Childers and Farrer, 2020).

Global phylodynamic studies suggest that the oldest clade is clade II, while the most recent ones are clades I and III (Chow *et al.*, 2020). Estimates of the mutation rate of *C. auris* vary between 1.34e-5 - 5.74e-5 substitutions per site per year (s/s/y), depending on the clade analyzed and the epidemiological context of each study (Rhodes *et al.*, 2018; Chow *et al.*, 2020; Kekana *et al.*, 2023).

Since 2016, the number of *C. auris*-associated infections in Europe has tripled, with the Mediterranean countries being the most affected. In the latest ECDC report, Spain, Italy and Greece were responsible for 96.3% of the cases in this continent. Of these, Spain is the largest contributor with 50.5% of the total (Kohlenberg *et al.*, 2022). National genomic surveillance studies in Italy have identified the emergence of multidrug-resistant strains in a large hospital in Liguria since 2019, all belonging to clade I (Codda *et al.*, 2023). The situation in Spain is similar. Monitoring studies in different medical centers have detected *C. auris* since 2016. In all cases, the strains have been classified into clade III, and were resistant to fluconazole, while only a few have been assigned as multidrug-resistant (Ruiz-Gaitán *et al.*, 2018; Mulet Bayona *et al.*, 2023).

The origin of clade III is thought to be South Africa. There have been cases reported in 2011 and 2012 in Kenya and South Africa (Chow *et al.*, 2020; Naicker *et al.*, 2021). Nevertheless, there are reports describing cases belonging to this clade in Canada in 2012, so the origin of this clade remains controversial (De Luca *et al.*, 2021). Genomic epidemiology studies of nosocomial outbreaks in other continents have shown that the

strains causing these outbreaks were more closely genetically related to African than to North American strains.

In this study, we have analyzed the evolution and dispersal of clade III of *C. auris*, focusing on two different epidemiological contexts, one regional, analyzing the initial stages of a long-term nosocomial outbreak in Spain, and the other global. Using information from a large collection of genomes, we describe 3 genetically distant subgroups within clade III, one of which is intrinsically resistant to fluconazole. The mutation rate obtained as well as the most likely inferred ancestral dates provide new insights into understanding how this clade has evolved and novel information for the genomic surveillance of this emerging pathogen.

MATERIALS AND METHODS

Sample collection and whole genome sequencing

This study focuses on the evolutionary analysis of global dispersion of clade III *C.auris.* Hence, the analyzed dataset is mainly enriched with samples from this clade; nevertheless, we included samples from the other *C. auris* clades to determine the genotype of the unknown isolates. The search criteria for public samples in the NCBI database were as follows (latest access on 2023/06/14): records from clinical and/or environmental isolates, generated with short-read sequencing technology, coming from all continents, and containing "clade III" in their genotype description, when possible. Only records containing at least the year of diagnosis in the variable "collection date" were selected. Importantly, all samples belonging to three previously characterized clade III nosocomial outbreaks in the UK (Eyre *et al.*, 2018), China (Tian *et al.*, 2021), and Kenya (Chow *et al.*, 2020) were downloaded. Additionally, we included 35 representative samples of a nosocomial outbreak in the Comunitat Valenciana (CV) region, Spain (Mulet-Bayona *et al.*, 2024). In cases with more than one sample per patient (e.g., serial samples from the same individual), we selected the first diagnosed sample from each case, when possible. Records with less than a minimum depth of 20x and minimum horizontal coverage of 90% were discarded. Thus, our final dataset consisted of 689 *C. auris* genomes (clade III: n=567; other clades: n=122, **Supplementary Table 1**).

Genomic libraries of isolates from the CV region outbreak were prepared using the Illumina DNA Prep kit (Illumina, San Diego, CA, USA) and sequenced on a Nextseq instrument generating 2x150bp paired-end sequencing reads following the manufacturer's instructions.

Bioinformatic analysis

All raw sequencing reads were trimmed and filtered using fastp (Chen *et al.*, 2018) with the following parameters “--cut_tail, --cut_window_size=10, --cut_mean_quality=20, --length_required=50, --correction, and --dedup”. These filtered reads were used to detect mutations associated with antifungal resistance as well as

for phylogenetic purposes as described below. In the case of the reference strains, we first downloaded their respective assembly file, and then, short-reads were simulated using the ART software (Huang *et al.*, 2012) with the following parameters "-l 150 -ss MSv3 -f 100 -m 350 -s 10". The reference assembly samples were "GCA_002759435.2" (clade I), "GCA_003013715.2" (clade II), "GCF_002775015.1" (clade III), and "GCA_003014415.1" (clade IV).
The Valencian region raw sequence reads were deposited in the European Nucleotide Archive public server under project PRJEB70513. Additional data and accession numbers of the sequenced reads can be found in **Supplementary Table 1**.

Variant calling and identification of antifungal resistance mutations
We used two reference genomes for the identification of variants (indels and SNPs) depending on the goal of each analysis. The clade III reference genome (strain B12221, GCF_002775015.1) was used for evolutionary and phylogenetic analyses, whereas the fully susceptible clade I reference (strain B8441, GCA_003014415.1) was chosen for screening of mutations associated with antifungal resistance. In both approaches, the snippy v4.6.0 software (Seemann, 2015) was employed with parameters "--mincov 10 --minfrac 0.9 --minqual 60" for variant calling. For evolutionary analyses, polymorphisms detected in duplicated and repetitive regions were filtered out. Duplicated regions were identified with MUMmer 4.0.0 (Marçais *et al.*, 2018) while repeat regions were identified using RepeatModeler2 (Flynn *et al.*, 2020). In total, 3842 regions totaling 409,448 bp ~ 3.21% of the genome were removed from posterior analyses. The presence of resistance-associated mutations in *erg11* and *fsk1* genes were determined by annotation of the variants with snpEFF (Cingolani *et al.*, 2012).

Phylogenetic/Evolutionary analyses
Three different multiple sequence SNP-based alignments (MSAs) were constructed. An initial MSA was obtained including all the complete genomes (n=689) previously described, to place all the samples in a global phylogenetic context. The second MSA was generated using only the genomes belonging to clade III. Finally, a third MSA was constructed for the dating analysis, which included samples belonging to subgroups 2 and 3 of clade III (a detailed description of this analysis is given below).

All the alignments were reconstructed from the SNPs identified by mapping against the reference genome of clade III. In all cases, the phylogenies were generated with Iqtree2 (Minh *et al.*, 2020) using the GTR model with 1000 replicates and with the *-fconst* option added for each case. Finally, pairwise SNP distances were calculated for all MSAs with the R ape package.

tMCRA inference and Bayesian skyline analysis
For the dating analysis, 380 global genomes with precise date of sampling (at least month and year) were used. This dataset included samples corresponding to four

globally previously described nosocomial outbreaks, one from the UK, one from China and one from Kenya, as well as the CV outbreak. First, root-to-tip distances were calculated with Tempest v1.5.3 (Rambaut *et al.*, 2016) to determine whether sufficient temporal signal was present and to obtain an initial estimate of the mutation rate. Once we were confident that the temporal signal was strong, a Bayesian-based phylogeny was constructed using the collection date of each isolate in Beast2 (Bouckaert *et al.*, 2019) For this, we used the GTR nucleotide substitution model, with a Coalescent Exponential Population tree as a prior, and under a strict molecular clock with a LogNormal distribution prior. Furthermore, sampling bias correction was performed following the ascertainment correction recommendation (https://www.beast2.org/2019/07/18/ascertainment-correction.html). Three independent MCMC runs of 50 million generations were conducted. All the runs and trees were combined and annotated with LogCombiner and TreeAnnotator v2.7.1, respectively. The maximum credibility tree was visualized with FigTree v1.4.4.

Once the mutation rate of clade III was obtained, a Coalescent Bayesian Skyline analysis was performed to determine the dynamics of the effective population size of each outbreak analyzed. The estimates and parameters used were the same as for the tMCRA of clade III using Beast2. The Coalescent Bayesian Skyline analysis was performed with Tracer (Rambaut *et al.*, 2018).

RESULTS

Phylogenetic structure of *Candida auris* clade III.

From a global dataset containing 676 genomes, 22 genomes that did not meet our quality control criteria were filtered out. In addition, 35 genomes belonging to a nosocomial outbreak in the CV were added. For the reconstruction of the initial phylogenetic tree, we chose to exclude polymorphisms detected in duplicated and repetitive genomic regions. In total, 3,842 regions spanning 409,448 base pairs (~3.21% of the total genome) were filtered out from further analyses. Thus, our initial dataset comprised 689 genomes and an alignment of 12,586,076 bp with 194,389 SNPs.

The initial ML phylogenetic tree was constructed using the global dataset described above, which included isolates from all *C. auris* clades. We found that 566 genomes belonged to clade III (**Supplementary Figure 1**). Because most isolates belonging to this clade are fluconazole-resistant and the cause of outbreaks that are difficult to control in Europe and globally, we decided to investigate this clade in more detail.

In order to have a higher resolution of the phylogenetic structure of clade III, we inferred a ML phylogenetic tree using only the genomes from this clade (n=566, MSA of 1,071 SNPs). The resulting topology showed that clade III is composed of 3 well-defined monophyletic subgroups (denoted as SG1, SG2, and SG3 in this study) (**Figure 1A**). Globally, subgroup SG3 is the most widespread, with 562 isolates from

9 countries in all the continents. The country with most isolates in SG3 was the USA with 279 isolates, followed by South Africa and the United Kingdom with 107 and 51 cases each, suggesting that this subgroup has spread worldwide. In contrast, subgroups SG1 and SG2 have only 2 isolates each. Within the former, both were sampled in Canada, while those in SG2 are from South Africa (**Figure 1A**). Examining the global phylogeny of clade III, we observed that isolates in subgroups SG2 and SG3 are genetically most closely related (mean distance between SG3 and SG2 of 63 SNPs), while SG1 is the most distant (mean distance from SG3 and SG2 of 778 SNPs) (**Figure 1B**).

Within the SG3 subgroup, there are two major clusters, one exclusive to the United States with 191 cases (189/191 isolates from California), and another, with a wider distribution, which included 372 samples from nosocomial outbreaks in Spain, UK, China and Kenya, as well as two small clusters from the United States with the majority of the cases coming from Florida (**Figure 2**). The average genetic distance between isolates from the US cluster (California State) was 9 SNPs, while those of the outbreaks from Spain, UK, China, and Kenya were 2, 3, 6, and 10 mutations, respectively, confirming recent transmission events (**Figure 3**).

Notably, we observed that almost all isolates belonging to SG3 (except strain "B19006") possess the ERG11:VF125AL mutation associated with fluconazole resistance. In contrast, isolates classified in SG1 and SG2 possess the wildtype allele and are phenotypically susceptible to fluconazole. As for echinocandin resistance, only 9 isolates contained any echinocandin-related mutation. The most common mutation was FSK1:S639P with 3 cases, followed by FSK1:S639F, FSK1:S639Y, and FSK1:S639dup with two cases each.

Local and global spread of clade III

Given the closer phylogenetic relationship between the SG2 and SG3 subgroups, we decided to focus our analysis on these subgroups to examine their global dispersion. By considering the phylogenetic structure obtained through Bayesian analysis and the origin of each isolate, we found that SG3 originated in Africa, resulting in clinical cases in Kenya and South Africa between 2011 and 2014 (**Figure 2**). Since then, cases have been reported on other continents. First in Oceania (Australia) and then in Europe (United Kingdom) in 2015. A year later, in 2016, the first cases of nosocomial infections were described in China and Spain. In 2017, the first case was reported in the Americas: a patient in the United States in Indiana (strain "B12631"). However, this case may be considered as a single case, given that it did not lead to outbreaks. It was not until 2019 when multiple cases of infection were reported in the states of California and Florida, the latter causing large nosocomial outbreaks in those locations. Finally, in 2020, isolated cases were described in Connecticut and New York. Importantly, since the first cases were reported in each country, the pathogen has been identified afterwards. For instance, in Australia, new cases were reported in 2018, three years after the first described case.

When examining the temporal dynamics of the populations in the nosocomial outbreaks (Kenya, UK, China, Spain and the large US cluster), an increase in the population size of the pathogen was observed in every case, starting from the report of the first case in each locality. In the Spanish outbreak, a plateau phase was reached three years after the first reported case, while in the US cluster this occurred two years later, and in the China outbreak it was attained only one year after the first case. In contrast, the dynamics of outbreak expansion in Kenya and the UK were different. In both cases, an increase in population size was observed at the beginning of the outbreak, but this decreases slightly one year later.

In the case of the UK outbreak, it has been reported that preventive measures were implemented to control the spread of the pathogen, which could explain the dynamics in this population (**Supplementary Figure 2**).

Dating analyses of clusters and outbreaks

Next, we analyzed subgroups SG2 and SG3 to infer the probable dates of origin of both lineages. For these analyses we used a dataset with samples having at least year/month in their available diagnosis date (n=380, MSA of 194,389 SNPs) (see Methods). Once we ascertained that we had a strong temporal signal (R^2=0.69) (**Supplementary Figure 3**), we decided to calculate the rate of evolution of both subgroups, as well as the most likely date of their most recent common ancestor (tMRCA). The inferred evolution rate was 2.25e-7 s/s/y (95% CI, 1.92e-7 - 2.46e-7) which translates as 2.87 changes per year (95% CI, 2.45 - 3.13). The tMRCA of the SG2 and SG3 subgroups was inferred 7/2003 (95% HDP, 3/2001 - 10/2005), while the origin of the fluconazole-resistant SG3 subgroup was 12/2010 (95% HDP, 4/2010 - 6/2011) (**Supplementary Figure 4**).

Additionally, we obtained the most likely date for the MRCA of the nosocomial outbreaks as well as that of the US cluster. The dates of the Spain, UK, China and Kenya outbreaks were 9/2015 (95% HDP, 2/2015 - 3/2016), 11/2013 (95% HDP, 5/2013 - 4/2014), 9/2015 (95% HDP, 4/2015 - 1/2016), and 7/2011 (95% HDP, 4/2011 - 8/2011), respectively, coinciding in all cases with the earliest reported cases in each corresponding setting (**Supplementary Figure 4**). The ancestor of the large US cluster likely originated in 7/2017 (95% HDP, 2/2017 - 5/2018). The states where this subgroup has been reported are California with 189 cases, followed by Florida with 75 cases, New York with 4 isolates, and Connecticut and Indiana with 1 case each. Looking at the phylogeny of clade III we may suggest that there were at least two introductions in the US, one in New York and one in California. The latter is the cause of a large outbreak since 2017. This geographic restriction suggests that the clades diversified locally after their ancestral strains were introduced into each region (**Supplementary Figure 5**).

DISCUSSION

In this work we have analyzed the evolution of clade III of *Candida auris* and described it could have spread and established itself in the world causing nosocomial resistant outbreaks in all continents. To the best of our knowledge, this is the first study in which the evolution of this clade is examined in depth. We found that this clade is divided into three subgroups (SG1, SG2 and SG3). Among them, the SG2 and SG3 subgroups are more genetically related, with an average genetic distance of 62 SNPs between them.

The strains that belong to the SG3 subgroup are resistant to fluconazole and carry the VF125AL mutation in the ERG11 protein. As a result, the use of other classes of antifungal drugs is essential and, thus, the emergence of multidrug-resistant isolates is highly probable. However, the frequency of known resistance-related mutations to these antifungals is relatively low. We observed only 9 cases (1.6% of the total) with mutations in the *FSK1* gene associated with resistance to echinocandins. This might be due to the fact that there are unknown mutations associated with these compounds, or that many isolates lack this phenotypic information, or that there are other mechanisms of resistance that remain unknown. As a practical case in this study, the strain "B19006" is resistant to fluconazole but has no mutation VF125AL:ERG11. This isolate probably carries other variants (SNPs, deletions or insertions), or alterations in the transcription of yet unknown genes that may be associated with resistance. Recently, new mutations or target genes related to resistance have been described (Carolus *et al.*, 2021).

Based on the ancestral dates inferred from the evolutionary analysis, we were able to determine the origin, timing and introduction of clade III at each setting. By using available genomes and the dates of different previously reported outbreaks, we have established that the likely origin of SG3 was in Africa in December 2010 (95% confidence interval: April 2010 - June 2011), from which an outbreak originated in a facility in Kenya in mid-2011 (July 2011) (Adam *et al.*, 2019). In 2013, a major outbreak was reported in a hospital in the United Kingdom (Eyre *et al.*, 2018), coinciding with our obtained estimates.

In the second half of 2015, the first reported cases were described simultaneously in a hospital in Shenyang province, China, as well as in Valencia, Spain. These dates match those reported in both areas (Ruiz-Gaitán *et al.*, 2018; Tian *et al.*, 2021). Finally, the SG3 subgroup arrived to the United States in 2017 (July 2017) through two introductions, one in New York and another in California. The New York isolates are more closely related to some described in South Africa, suggesting ancestry from Africa, while those from California are more restricted to that area and their most recent ancestry is unknown. The first cases reported in California are from late 2017 (Price *et al.*, 2021).

Fluconazole resistance and the origin of clade III have been widely debated. In this study, we identified two genetically closely related subgroups (SG2 and SG3), one susceptible and the other surprisingly resistant to the aforementioned antifungal drug. Considering the inferred dates of the possible ancestry of both subgroups and SG3 (7/2003 and 12/2010, respectively), we suggest that the acquisition of resistance in SG3 occurred during this period. We hypothesize that the increased use of antifungals in both clinical and other ecological niches has exerted selective pressure for the emergence and fixation of resistant strains. In fact, increased use of antifungals has been reported in South Africa between 2004-2006 (Chakrabarti and Sood, 2021). This, along with the fact that there are many reported cases of fungal infections caused by other *Candida* and non-*Candida* species in South Africa, which have been treated with fluconazole (Arastehfar *et al.*, 2020), suggests that this resistant phylogenetic subgroup likely emerged in South Africa and extended worldwide.

A contrasting result is the evolutionary rate estimated. In this study the estimated mutation rate at 2.25e-7 changes per genome per year (95% CI, 1.92e-7 - 2.46e-7), while the rate reported in other studies has an average value of 3.02e-5 (Rhodes *et al.*, 2018; Chow *et al.*, 2020; Kekana *et al.*, 2023). This discrepancy may be due to different factors regarding how this value was obtained. The main difference is the number of samples involved. In this study we have used 380 complete genomes from clade III, ensuring to include as much diversity as possible, while in previous studies smaller datasets were used, mainly including genomes from a transmission outbreak to calibrate estimates for a whole clade. Since more genomes are now available, the inferred rate of evolution is more accurate and closer to reality. Furthermore, the r2 value we report is higher than in other studies (0.69 vs. 0.35 and 0.55), giving a higher confidence to the inferred rate. Another important difference that may affect the evolutionary estimate is the reference genome used. In other studies, the reference genome of clade I (strain B8441, GCA_003014415.1) was used, whereas in this analysis, we selected the reference genome of clade III (strain B12221, GCF_002775015.1). The impact of the reference genome used when performing this kind of studies has been discussed elsewhere, from the percentage of reads mapped to the number of variants called (Valiente-Mullor *et al.*, 2021). In light of these previous comments, we consider our approach appropriate and robust.

The limitations of this study stem from the number of samples analyzed. Although many epidemiological studies report isolates belonging to clade III (Chow *et al.*, 2018, 2020; Naicker *et al.*, 2021; Biran *et al.*, 2023), there are other clades that are more successful, for example clade IV. We suggest that this may be due to the fact that the latter have arrived first and have rapidly colonized these regions (Chow *et al.*, 2020).

In conclusion, we have analyzed complete genomes of *Candida auris* clade III to provide new insights into the origin of a fluconazole-resistant subgroup, and its dispersal dynamics over the world. These results open new discussions and extend

the knowledge to implement epidemiological measures to control the pathogen, especially in healthcare settings.

ACKNOWLEDGEMENTS

Irving Cancino-Muñoz is a beneficiary of the European Union-funded NextGenerationUE (code UP2021-044) grant competition for the requalification of the Spanish university system of the Ministry of Universities of the Government of Spain. Project funded by MICIN (PID2021-127010OB-I00) and Generalitat Valenciana (CIPROM2021-053).

REFERENCES

Adam, R.D. *et al.* (2019) 'Analysis of *Candida auris* fungemia at a single facility in Kenya', *International Journal of Infectious Diseases*, 85, pp. 182–187.

Akinbobola, A.B. *et al.* (2023) 'Environmental reservoirs of the drug-resistant pathogenic yeast *Candida auris'*, *PLoS pathogens*, 19(4), p. e1011268.

Arastehfar, A. *et al.* (2020) 'Drug-Resistant Fungi: An Emerging Challenge Threatening Our Limited Antifungal Armamentarium', *Antibiotics (Basel, Switzerland)*, 9(12). Available at: https://doi.org/10.3390/antibiotics9120877.

Biran, R. *et al.* (2023) 'Nationwide Outbreak of Candida auris Infections Driven by COVID-19 Hospitalizations, Israel, 2021-2022', *Emerging infectious diseases*, 29(7), pp. 1297–1301.

Bouckaert, R. *et al.* (2019) 'BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis', *PLoS computational biology*, 15(4), p. e1006650.

Carolus, H. *et al.* (2021) 'Genome-Wide Analysis of Experimentally Evolved *Candida auris* Reveals Multiple Novel Mechanisms of Multidrug Resistance', *mBio*, 12(2). Available at: https://doi.org/10.1128/mBio.03333-20.

Chakrabarti, A. and Sood, P. (2021) 'On the emergence, spread and resistance of *Candida auris*: host, pathogen and environmental tipping points', *Journal of medical microbiology*, 70(3), p. 001318.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics* , 34(17), pp. i884–i890.

Chow, N.A. *et al.* (2018) 'Multiple introductions and subsequent transmission of multidrug-resistant *Candida auris* in the USA: a molecular epidemiological survey', *The Lancet infectious diseases*, 18(12), pp. 1377–1384.

Chow, N.A. *et al.* (2020) 'Tracing the Evolutionary History and Global Expansion of *Candida auris* Using Population Genomic Analyses', *mBio*, 11(2). Available at: https://doi.org/10.1128/mBio.03364-19.

Chybowska, A.D., Childers, D.S. and Farrer, R.A. (2020) 'Nine Things Genomics Can Tell Us About Candida auris', *Frontiers in genetics*, 11, p. 526662.

Cingolani, P. *et al.* (2012) '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*, 6(2), pp. 80–92.

Codda, G. *et al.* (2023) 'In vivo evolution to echinocandin resistance and increasing clonal heterogeneity in *Candida auris* during a difficult-to-control hospital outbreak, Italy, 2019 to 2022', *Eurosurveillance*, 28(14), p. 2300161.

De Luca, D.G. *et al.* (2021) 'Four genomic clades of *Candida auris* identified in Canada, 2012–2019', *Medical Mycology*, 60(1), p. myab079.

Du, H. *et al.* (2020) ‘*Candida auris*: Epidemiology, biology, antifungal resistance, and virulence’, *PLoS pathogens*, 16(10), p. e1008921.

Eyre, D.W. *et al.* (2018) ‘A *Candida auris* Outbreak and Its Control in an Intensive Care Setting’, *The New England journal of medicine*, 379(14), pp. 1322–1331.

Flynn, J.M. *et al.* (2020) ‘RepeatModeler2 for automated genomic discovery of transposable element families’, *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), pp. 9451–9457.

Huang, W. *et al.* (2012) ‘ART: a next-generation sequencing read simulator’, *Bioinformatics* , 28(4), pp. 593–594.

Kekana, D. *et al.* (2023) ‘*Candida auris* Clinical Isolates Associated with Outbreak in Neonatal Unit of Tertiary Academic Hospital, South Africa - Volume 29, Number 10—October 2023 - Emerging Infectious Diseases journal - CDC’, *Emerging Infectious Disease journal* [Preprint]. Available at: https://doi.org/10.3201/eid2910.230181.

Kohlenberg, A. *et al.* (2022) ‘Increasing number of cases and outbreaks caused by *Candida auris* in the EU/EEA, 2020 to 2021’, *Eurosurveillance*, 27(46), p. 2200846.

Marçais, G. *et al.* (2018) ‘MUMmer4: A fast and versatile genome alignment system’, *PLoS computational biology*, 14(1), p. e1005944.

Minh, B.Q. *et al.* (2020) ‘IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era’, *Molecular biology and evolution*, 37(5), pp. 1530–1534.

Mulet Bayona, J.V. *et al.* (2023) ‘*Candida auris* from colonisation to candidemia: A four-year study’, *Mycoses*, 66(10), pp. 882–890.

Mulet-Bayona, J.V. *et al.* (2024) ‘Genotypic and phenotypic characterization of a nosocomial outbreak of *Candida auris* in Spain during five years’, *bioRxiv*. Available at: https://doi.org/10.1101/2024.03.01.582904.

Muñoz, J.F. *et al.* (2018) ‘Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species’, *Nature communications*, 9(1), pp. 1–13.

Naicker, S.D. *et al.* (2021) ‘Clade distribution of *Candida auris* in South Africa using whole genome sequencing of clinical and environmental isolates’, *Emerging microbes & infections* [Preprint]. Available at: https://doi.org/10.1080/22221751.2021.1944323.

Price, T.K. *et al.* (2021) ‘Genomic Characterizations of Clade III Lineage of *Candida auris*, California, USA’, *Emerging infectious diseases*, 27(4), pp. 1223–1227.

Rambaut, A. *et al.* (2016) ‘Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen)’, *Virus Evolution*, 2(1), p. vew007.

Rambaut, A. *et al.* (2018) ‘Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7’, *Systematic biology*, 67(5), pp. 901–904.

Rhodes, J. *et al.* (2018) ‘Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris’*, *Emerging microbes & infections*, 7(1), p. 43.

Ruiz-Gaitán, A. *et al.* (2018) ‘An outbreak due to *Candida auris* with prolonged colonisation and candidaemia in a tertiary care European hospital’, *Mycoses*, 61(7), pp. 498–505.

Seemann, T. (2015) *snippy: fast bacterial variant calling from NGS reads*, *GitHub*. Available at: https://github.com/tseemann/snippy (Accessed: 6 March 2023).

Spruijtenburg, B. *et al.* (2022) ‘Confirmation of fifth *Candida auris* clade by whole genome sequencing’, *Emerging microbes & infections* [Preprint]. Available at: https://doi.org/10.1080/22221751.2022.2125349.

Thatchanamoorthy, N. *et al.* (2022) ‘*Candida auris*: A Mini Review on Epidemiology in Healthcare Facilities in Asia’, *Journal of Fungi*, 8(11), p. 1126.

Tian, S. *et al.* (2021) ‘Genomic epidemiology of *Candida auris* in a general hospital in Shenyang, China: a three-year surveillance study’, *Emerging microbes & infections* [Preprint]. Available at: https://doi.org/10.1080/22221751.2021.1934557.

Valiente-Mullor, C. *et al.* (2021) ‘One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads’, *PLoS computational biology*, 17(1), p. e1008678.

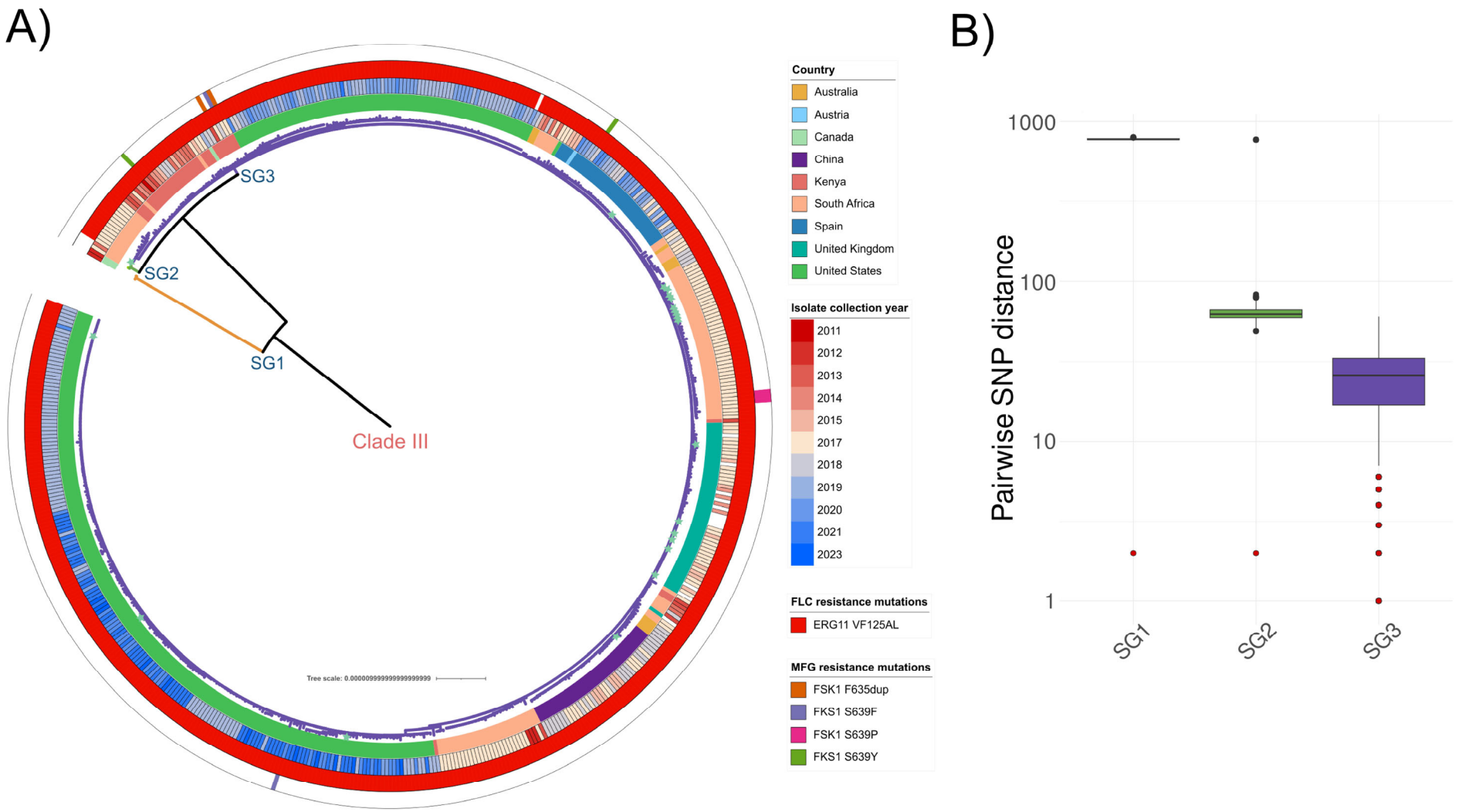
WHO (2022) *WHO fungal priority pathogens list to guide research, development and public health action*. World Health Organization. Available at: https://www.who.int/publications/i/item/9789240060241 (Accessed: 21 December 2023).

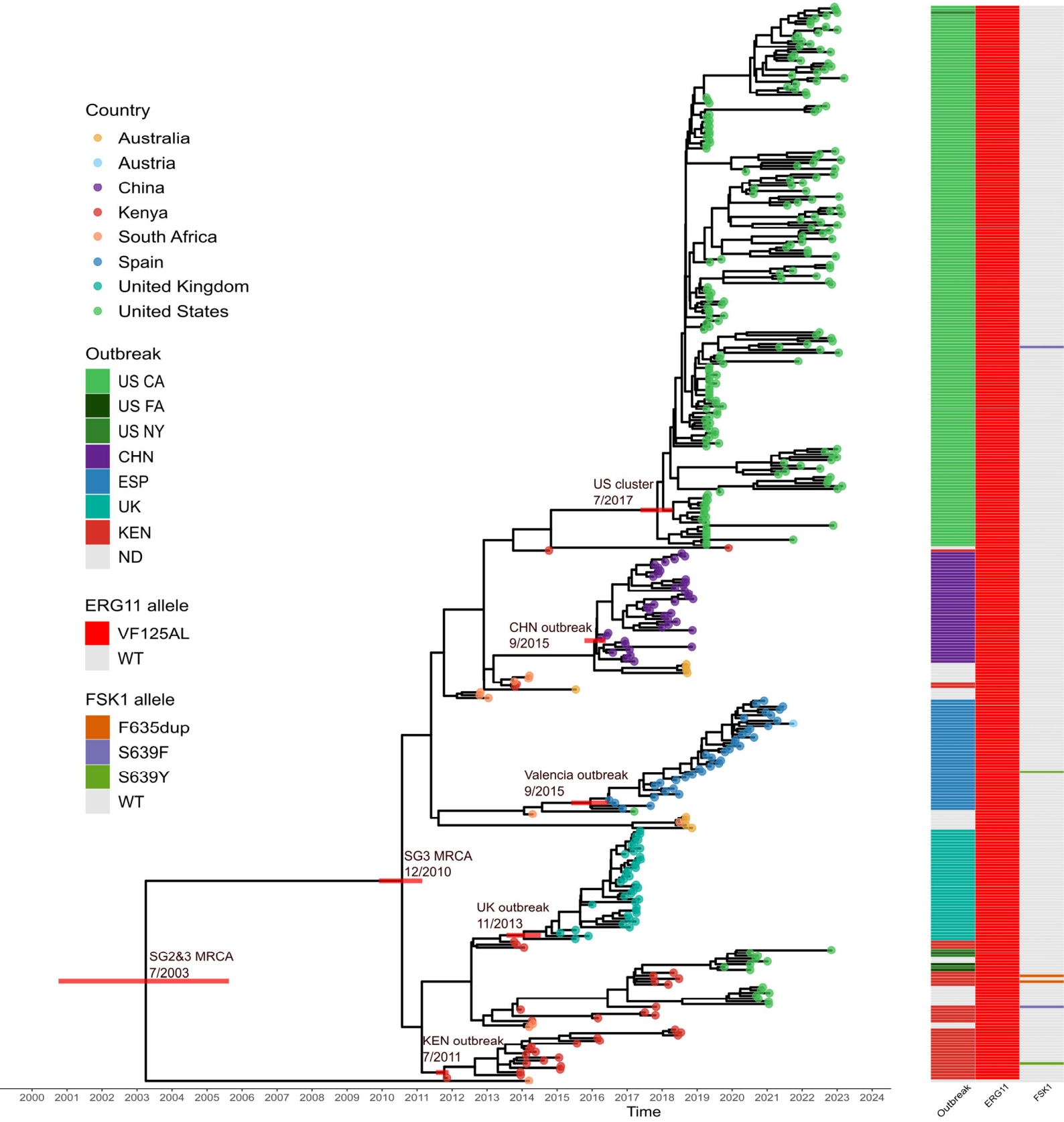
Figure legends

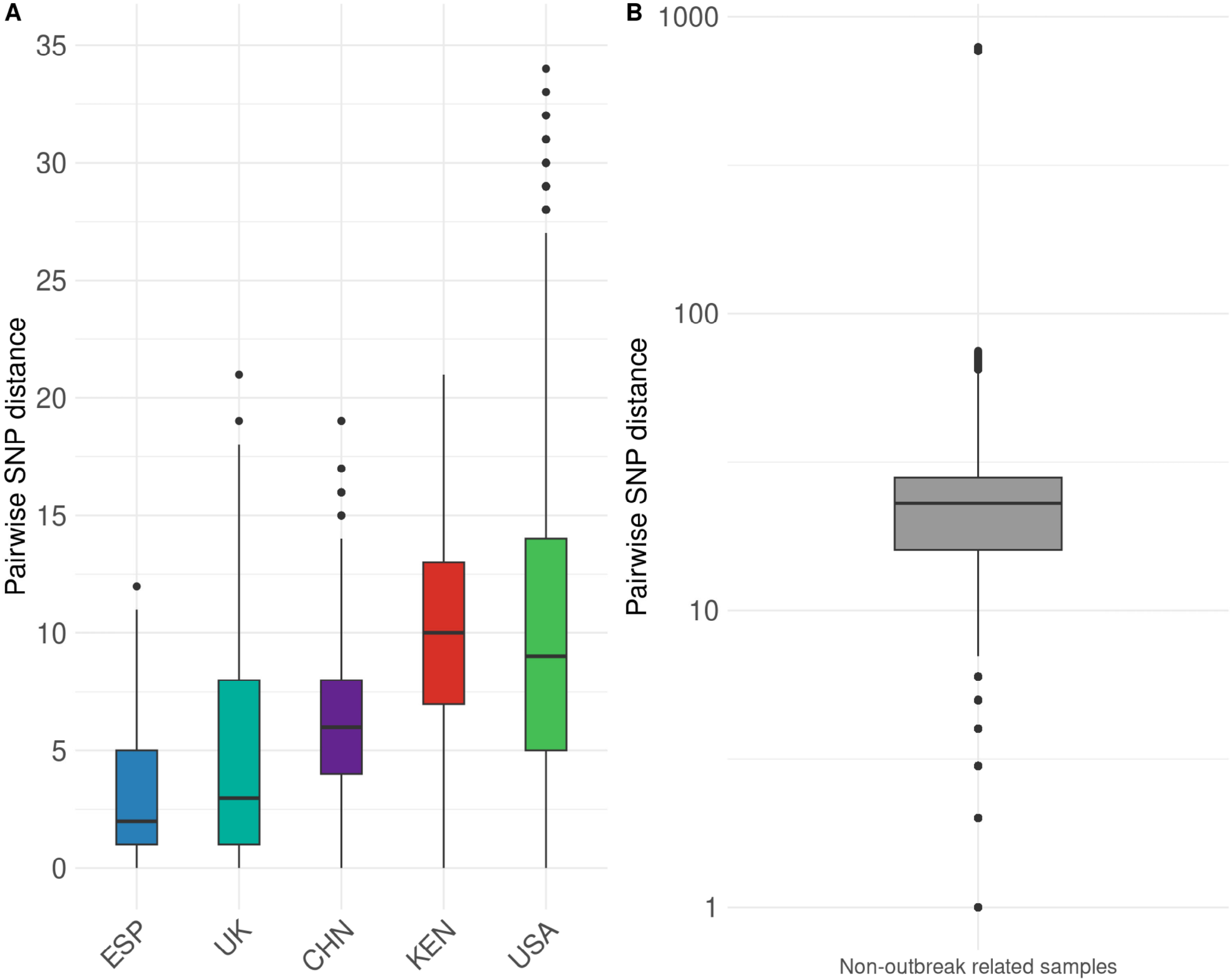
Figure 1. Phylogenetic structure of clade III of *Candida auris*. A) Global phylogeny of clade III of *Candida auris* (n=566). The phylogeny was constructed from 2,152 SNPs. The color of the branches indicates the subgroup (referred as SG label) of each sample. The first ring (from inner to outer) shows the country of origin of each isolate, the second ring shows the year of culture of each isolate, the third ring indicates mutations related to fluconazole resistance, and the fourth ring shows mutations related to echinocandin resistance, and the last ring Blank spaces mean that the isolates harbor the wildtype allele. The star shape in the branch tips indicate isolate environmental source. **B)** Pairwise genetic distance (in SNPs) of each subgroup. The red points represent the genetic distance within each subgroup.

Figure 2. *Candida auris* clade III subgroups SG2&3 phylogeny (n=380). The bayesian-based tree was built from 1,071 SNPs. The branch tip color indicates the origin country of each sample analyzed. The first column indicates the nosocomial outbreak that corresponds to each sample analyzed. The US cluster is divided into three groups according to the US origin state. The second column shows mutations associated with fluconazole resistance (ERG11 allele), while the third column indicates mutations associated with echinocandins(FSK1 allele). WT means that the isolates harbor the wildtype allele. ND means "Not Detected".

Figure 3. Pairwise genetic distance of *Candida auris* clade III subgroups SG2&3. A) Genetic distances between isolates of each outbreak. B) Genetic distances between samples that do not belong to an outbreak.


A)
SG3
SG2
SG1
Clade III
Tree scale: 0.00000999999999999999
Country
Australia
Austria
Canada
China
Kenya
South Africa
Spain
United Kingdom
United States
Isolate collection year
2011
2012
2013
2014
2015
2017
2018
2019
2020
2021
2023
FLC resistance mutations
ERG11 VF125AL
MFG resistance mutations
FSK1 F635dup
FKS1 S639F
FSK1 S639P
FKS1 S639Y
B)
Pairwise SNP distance
1000
100
10
1
SG1
SG2
SG3


Country
Australia
Austria
China
Kenya
South Africa
Spain
United Kingdom
United States
Outbreak
US CA
US FA
US NY
CHN
ESP
UK
KEN
ND
ERG11 allele
VF125AL
WT
FSK1 allele
F635dup
S639F
S639Y
WT
US cluster
7/2017
CHN outbreak
9/2015
Valencia outbreak
9/2015
SG3 MRCA
12/2010
UK outbreak
11/2013
SG2&3 MRCA
7/2003
KEN outbreak
7/2011
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024
Time
Outbreak
ERG11
FSK1


A
Pairwise SNP distance
35
30
25
20
15
10
5
0
ESP
UK
CHN
KEN
USA
B
Pairwise SNP distance
1000
100
10
1
Non-outbreak related samples