

1 Discovery of the sixth *Candida auris* clade in Singapore

2
3 Chayaporn Suphavitai¹, Karrie Kwan Ki Ko^{1,2,3,4,5}, Kar Mun Lim¹, Mei Gie Tan², Patipan
4 Boonsimma¹, Joash Jun Keat Chu¹, Sui Sin Goh², Prevena Rajandran², Lai Chee Lee⁶, Kwee
5 Yuen Tan⁶, Bushra Binte Shaik Ismail⁶, May Kyawt Aung⁶, Yong Yang^{4,6}, Jean Xiang Ying
6 Sim^{4,6,7}, Indumathi Venkatachalam^{4,6,7}, Benjamin Pei Zhi Cherng^{4,7}, Bram Spruijtenburg^{8,9}, Kian
7 Sing Chan^{3,4}, Lynette Lin Ean Oon^{3,4}, Ai Ling Tan^{2,4}, Yen Ee Tan^{2,4}, Limin Wijaya^{4,7}, Ban Hock
8 Tan^{4,7}, Moi Lin Ling^{4,6}, Tse Hsien Koh^{2,4}, Jacques F. Meis^{8,9,10}, Clement Kin Ming Tsui^{11,12,13},
9 Niranjana Nagarajan^{1,5}

10
11 [^]Contributed equally to this work

12 ^{*}Corresponding authors

- 13
14 1. Genome Institute of Singapore, Agency for Science, Technology and Research
15 (A*STAR), Singapore
16 2. Department of Microbiology, Singapore General Hospital, Singapore
17 3. Department of Molecular Pathology, Singapore General Hospital, Singapore
18 4. Duke-NUS Medical School, Singapore
19 5. Yong Loo Lin School of Medicine, National University of Singapore, Singapore
20 6. Department of Infection Prevention and Epidemiology, Singapore General Hospital,
21 Singapore
22 7. Department of Infectious Diseases, Singapore General Hospital, Singapore
23 8. Department of Medical Microbiology and Infectious Diseases, Canisius-Wilhelmina
24 Hospital, Nijmegen, The Netherlands
25 9. Center of Expertise in Mycology Radboud University Medical Center/Canisius-
26 Wilhelmina Hospital, 6532 SZ Nijmegen, The Netherlands
27 10. Department I of Internal Medicine, University Hospital Cologne, Excellence Center for
28 Medical Mycology, Cologne, Germany
29 11. Infectious Diseases Research Laboratory, National Centre for Infectious Diseases, Tan
30 Tock Seng Hospital, Singapore
31 12. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
32 13. Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

33 Summary

34 Background

35 The emerging fungal pathogen *Candida auris* poses a serious threat to global public health due
36 to its worldwide distribution, multidrug-resistance, high transmissibility, propensity to cause
37 outbreaks and high mortality rates. We report three *C. auris* isolates detected in Singapore,
38 which are genetically distinct from all known clades (I-V) and represent a new clade (Clade VI).

39 Methods

40 Three epidemiologically unlinked clinical isolates belonging to the potential new *C. auris* clade
41 were whole-genome sequenced and phenotypically characterized. The complete genomes of
42 these isolates were compared to representative genomes of all known clades. To provide a
43 global context, 3,651 international whole-genome sequences (WGS) from the NCBI database
44 were included in the high-resolution single nucleotide polymorphism (SNP) analysis. Antifungal

45 resistance genes, mating type locus, and chromosomal rearrangements were characterized
46 from the WGS data of the Clade VI isolates. We further implemented Bayesian logistic
47 regression models to simulate the automatic detection of Clade V and VI as their WGS data
48 became available.

49 Findings

50 The three Clade VI isolates were separated by >36,000 SNPs from all existing *C. auris* clades.
51 These isolates had opposite mating type allele and different chromosomal rearrangements
52 when compared to their closest Clade IV relatives. As a proof-of-concept, our classification
53 model was able to flag these outlier genomes as a potential new clade. Furthermore, an
54 independent WGS submission from Bangladesh was found to belong to this new clade.

55 Interpretation

56 The discovery of a new *C. auris* clade in Singapore and Bangladesh, showing close relationship
57 to Clade IV members in South America, highlights the unknown genetic diversity and origin of *C.*
58 *auris*, particularly in under-resourced regions. Active surveillance in clinical settings, along with
59 effective sequencing strategies and downstream analysis, will be essential in the identification of
60 novel strains, tracking of transmission, and containment of adverse clinical impacts caused by
61 *C. auris* infections.

62 Funding

63 This work was supported by the Singapore National Medical Research Council (NMRC)
64 research training fellowship (MOH-FLWSHP19may-0005), the NCRS Duke-NUS Academic
65 Medical Center Academic Clinical Program grant (09/FY2022/P1/17-A32, GRDUKP003401),
66 and the Genedant-GIS Innovation Program grant.

67 Research in context

68 Evidence before this study

69 We searched PubMed using the search terms “*Candida auris*” AND “clade”, for papers
70 published between Jan 1, 2009, and July 1, 2023. This search retrieved 115 publications. 60
71 relevant publications were identified. 28 studies analyzed and discussed the molecular
72 epidemiology of *C. auris*, including the description of *C. auris* clades, either in outbreak or
73 surveillance settings. There were 11 case reports of *C. auris* clinical cases that included clade
74 determination. Two studies focused on the detection and clade determination of *C. auris* from
75 non-healthcare environments. Clade-specific characteristics were described or analyzed in 14
76 studies. One study applied machine learning to *C. auris* drug resistance analysis, but not for
77 clade determination. Four studies focused on the description of potentially new *C. auris*
78 lineages, subclades, or clades. All publications described isolates that belong to one of the five
79 known *C. auris* clades (I-V). All publications found that strains from different clades differed by
80 more than 35,000 SNPs, and that there are clade-specific differences in geographical
81 distribution, phenotypic characteristics, antifungal susceptibility profile, outbreak potential, and
82 clinical manifestations.

83
84 The NCBI Pathogen Detection system contained 4,506 *C. auris* genomes on July 1, 2023.
85 There were ten (0.22%) submissions from Southeast Asian countries and 92 (2.04%)

86 submissions from South Asia and the Indian subcontinent, which are parts of the Indomalayan
87 biogeographic realm.

88 Added value of this study

89 To the best of our knowledge, we are the first group to perform hybrid assemblies on three
90 representative isolates in a new *C. auris* clade, which is separated from all other existing clades
91 (I-V) by >36,000 SNPs. Whole-genome SNP analysis and phenotypic characterization of these
92 epidemiologically unlinked isolates detected in Singapore suggest that they represent a
93 previously unreported sixth major clade. High-resolution SNP analysis of 3,651 international
94 whole-genome sequences from the NCBI database, which generated the final dataset
95 consisting of more than 6.6 million genome pairs, revealed six distinct genetic clusters
96 representing the five known clades and the new sixth clade (Indomalayan). In addition, we
97 demonstrate that a machine learning approach can be used to flag these outlier genomes for
98 further investigations as soon as they become available, thus providing the earliest possible
99 alert for potential new public health threats.

100 Implications of all the available evidence

101 Despite the high antimicrobial resistance burden in Southeast Asia and South Asia, these
102 regions are disproportionately underrepresented in terms of genomic surveillance of *C. auris*, a
103 multidrug-resistant fungal pathogen. The detection of three epidemiologically unlinked *C. auris*
104 isolates in Singapore belonging to a new *C. auris* clade suggests that yet-to-be-reported strains
105 may be circulating in the region. Given the propensity for multidrug resistance, healthcare-
106 associated infection outbreaks, and the associated high mortality, active surveillance and
107 continued vigilance is necessary.

108 Introduction

109 *Candida auris* is an emerging multidrug-resistant (MDR) fungal pathogen that has been
110 implicated in a large number of healthcare-associated infection outbreaks¹⁻⁸. Invasive infections
111 caused by MDR *C. auris* are associated with high mortality rates in vulnerable patient groups<sup>2-
112 10</sup>. Within healthcare settings, *C. auris* appears to be highly transmissible and is known to
113 persist in the environment^{4-6,11}. Given its worldwide distribution, MDR characteristics, high
114 transmissibility, and high mortality rates, *C. auris* represents a serious threat to global public
115 health. The World Health Organization highlighted this threat by classifying *C. auris* as a 'critical
116 priority' pathogen last year¹², while the U.S. Centers for Disease Control and Prevention
117 declared *C. auris* as an 'urgent' antimicrobial resistance threat¹³, further emphasizing the need
118 to understand and mitigate this public health threat.

119
120 Five geographically distinct *C. auris* clades have been described to date. This includes Clade I
121 (South Asian), Clade II (East Asian), Clade III (African), Clade IV (South American), and Clade
122 V (Iranian)^{10,14-16}. These clades were recognized based on whole-genome sequence analyses,
123 where strains from different clades differed by more than 35,000 single nucleotide
124 polymorphisms (SNPs)^{14,17}. Furthermore, these clades differed in terms of their antifungal
125 susceptibility profiles, outbreak potential, and clinical manifestations¹⁸⁻²⁰.

126
127 Here we report three epidemiologically unlinked *C. auris* isolates detected in Singapore, which
128 appear to represent members of a new clade. The index isolate (A) was first detected by
129 polymerase chain reaction (PCR) from a routine *C. auris* screening swab and subsequently

130 cultured in April 2023. WGS analysis of isolate A showed that the strain was separated from all
131 known clades by more than 36,000 SNPs. Retrospective analysis of our laboratory's archived *C.*
132 *auris* isolates subsequently discovered two additional isolates (B and C) with similar genomic
133 and phenotypic characteristics. Isolate C was previously short-read sequenced and reported in
134 2019 (SRR10102336)²¹.

135
136 To understand the global context, we analyzed 3,651 *C. auris* whole-genome sequences in the
137 NCBI database together with the sequence data generated in this study. The final dataset
138 consisted of >6.6 million genome pairs, and six distinct genetic clusters were recognized,
139 consisting of the five known clades and the new Indomalayan clade. From the NCBI database,
140 we identified one additional independently submitted genome from Bangladesh
141 (SRR24877249), which belongs to this new clade. The previously submitted WGS data of
142 isolate C from Singapore (SRR10102336)²¹, and the independent upload from Bangladesh
143 (SRR24877249) were neither identified nor reported as representatives of a potential new clade
144 at the time of writing. The presence of potentially novel, but unrecognized genome data present
145 in the NCBI database demonstrates a gap in novel pathogen or genotype detection despite the
146 availability and accessibility of genomic surveillance data. Given the vast body of microbial
147 genomic data available in the post-COVID-19 pandemic era, it is evident that the challenge lies
148 not only in data availability but also in the ability to fully exploit the data for timely public health
149 alerts and decision-making. To address this gap, we implemented a machine learning approach
150 to demonstrate that potentially novel genomes can be flagged automatically for further
151 investigations, as soon as the genome data become available.

152
153 This study reports the hybrid assemblies, genomic analysis, and phenotypic characteristics of
154 three representatives of the sixth *C. auris* clade. We further simulated the automatic detection of
155 the new clade, demonstrating that microbial genomic surveillance can be further augmented to
156 provide early alerts for potential emergent threats.

157 Methods

158 *Candida auris* microbiological and epidemiological analysis

159 Three *C. auris* isolates from three epidemiologically unlinked patients in Singapore were
160 included in this study. The full description of the isolates is summarized in appendix 1 (p3).
161 Details of phenotypic identification and antifungal susceptibility testing are in appendix 1 (p4).
162 Four ATCC *C. auris* isolates from Clade I-IV were included as clade-specific controls in this
163 study (appendix 2). The sample and patient identifiers used throughout the manuscript have
164 been modified from the original identifiers, and the identifiers used in the manuscript and
165 supplementary materials cannot reveal the identity of the subjects.

166
167 This study used pre-existing retrospective collections of isolates and our analyses led to no
168 clinical intervention. Epidemiological data collection was previously performed as part of routine
169 surveillance and infection prevention measures and hence constituted a non-research infection
170 control surveillance activity. Institutional review board exemption was granted by the SingHealth
171 Centralised Institutional Review Board (Reference number 2017-2576).

172 DNA extraction, WGS and sequence data analysis

173 Details of DNA extraction and WGS methods are in appendix 1 (p4). Methods used for
174 sequence data analysis, including sequence read quality control, genome assembly, variant

175 calling, genome relatedness analysis, phylogenetic analysis, antibiotic resistance gene analysis,
176 mating type loci analysis, chromosomal rearrangement analysis, *in silico* short tandem repeat
177 (STR) analysis, and divergence time estimation are available in appendix 1 (pp 4-7). Quality
178 statistics for the sequence data and genome assemblies generated in this study are
179 summarized in appendix 2. In addition to the new sequence data generated in this study, 4,475
180 publicly available genomes were initially included, of which 3,651 passed quality control and
181 were included in downstream analysis. The list of publicly available genomes included, and their
182 quality statistics are in appendix 3.

183 Machine-learning models for detecting potential new clade

184 Current methods of clade determination rely on phylogenetic analysis or STR genotyping, the
185 interpretation of which can be challenging, resulting in clade assignments that may be uncertain,
186 as evidenced by the previous report on Isolate C, which was labeled as part of the Clade IV
187 phylogenetic cluster²¹. There is currently no established threshold for SNP distance to flag an
188 unusual *C. auris* genome as a potential new clade. To address this, we implemented a
189 Bayesian logistic regression model to determine whether a pair of genomes are from the same
190 clade based on SNP distances and previously reported clades. This threshold was then used to
191 determine the relationship between genomes (an edge links genomes from the same clade) in a
192 graph, which captured clusters (connected components) representing existing and potential new
193 clades. We implemented the model using 3,651 publicly available whole-genome sequences,
194 and the Clade VI whole-genome sequences generated in this study (appendix 3). 1,132 (31.0%)
195 of the publicly available sequence data had previously reported clade information. All WGS data
196 were grouped into 19 quarterly datasets based on their date of public release (2019 - June
197 2023). The number of clusters in the graph captured the number of clades present at each time
198 point. Detailed methods regarding the algorithm and model evaluation are in appendix 1 (pp 8-
199 11).

200 Role of the funding source

201 The funders of the study had no role in study design, data collection, data analysis, data
202 interpretation, or writing of the report.

203 Results

204 All three Clade VI isolates (A, B, C) were phenotypically identified to be *C. auris*. Details for
205 phenotypic identification results are in appendix 1 (p13). The colony morphology of all Clade VI
206 isolates and the lactophenol cotton blue slide mount of isolate A at 600× magnification are in
207 appendix 1 (p16). Antifungal susceptibility testing results are in **Table 1**. Further results and
208 discussions on antifungal resistance testing are in appendix 1 (pp 22-24).

209
210 The index isolate from Patient A, isolate A (LIMS ID F3485, Biosample SAMN36753178) was
211 cultured from a combined screening swab (nasal, axillae, and groin) in April 2023. The
212 screening swab was taken within 24 hours of Patient A's admission as part of a *C. auris* active
213 surveillance program in the hospital. Patient A received treatment in the Department of Internal
214 Medicine and reported no travel history in the preceding two years. Since *C. auris* is not known
215 to be endemic in Singapore, epidemiological, microbiological, and molecular investigations were
216 initiated. Retrospective analysis of our laboratory's archived *C. auris* isolates found two
217 additional isolates (B and C) to be similarly genetically distant to the five known clades, out of 43
218 isolates analyzed. Isolate B (LIMS ID F1580, Biosample SAMN36753179) was cultured from an

219 intraoperative tissue specimen from Patient B, in April 2022. Patient B received treatment in the
220 orthopedic unit and reported no recent travel history in the preceding two months, but had
221 outpatient visits at hospitals in Indonesia in 2021. Isolate C (LIMS ID F0083, Biosample
222 SAMN36753180) was cultured from the positive blood culture of Patient C in January 2018.
223 Patient C was admitted directly from Bangladesh and the blood culture that was taken within 48
224 hours of admission grew *C. auris*. Isolate C was previously sequenced in 2019 and is one of the
225 ten publicly available *C. auris* genomes submitted from Singapore (SRR10102336)²¹.

226
227 Extensive epidemiological investigations determined that Patients A, B, and C were separated
228 spatially and temporally, with no overlap in physical space, clinical unit, and care facilities
229 utilization in the large 1700-bed tertiary hospital. In accordance with the hospital's infection
230 prevention policy, every positive case of *C. auris* triggered epidemiological investigations and
231 contact tracing. All inpatients who shared the same ward or room with the positive cases were
232 defined as contacts and they were screened for *C. auris* colonization as part of the infection
233 prevention investigations. A total of 68 contacts were identified for Patients A and B. Patient C
234 was isolated on admission and had no known contacts. None of the contacts screened positive
235 for *C. auris* by routine diagnostic *C. auris* in-house validated PCR test or cultures. Furthermore,
236 a total of 51 environmental samples were cultured to look for *C. auris* during the investigations,
237 none of which yielded *C. auris*.

238
239 All three Clade VI *C. auris* isolates were whole-genome sequenced using two sequencing
240 platforms: Illumina (short-read technology) and Oxford Nanopore (long-read technology), to
241 obtain high-quality hybrid assemblies of the genomes. *De novo* hybrid genome assembly of the
242 *C. auris* genomes was performed for the Singapore Clade VI isolates, and each consisted of
243 eight to nine contigs, corresponding to near-complete genomes. The total assembly size was
244 12.4 to 12.5 Mbp. Quality statistics for Clade VI isolate sequence data and assemblies are
245 available in appendix 2.

246
247 A comprehensive genome relatedness analysis using 3,651 *C. auris* WGS data in the NCBI
248 database and sequence data generated in this study was performed (appendix 3). The final
249 dataset included a total of 3,654 whole-genome sequences after quality control checks
250 (appendix 1 p5), and >6.6 million genome pairs were analyzed. We found six distinct genetic
251 clusters in this large-scale analysis. One additional *C. auris* genome (SRR24877249),
252 independently submitted from Bangladesh, was found to belong to the same cluster as the
253 Singaporean Clade VI isolates.

254
255 Since complete chromosomal sequences allow us to perform both whole genome phylogenetic
256 and chromosomal rearrangement analyses, 14 representative genomes from known clades (I-V)
257 with complete sequences of chromosomes were selected for in-depth downstream analysis.
258 Phylogenetic analysis of the Clade VI genomes with these representative genomes showed that
259 the Clade VI genomes were genetically distinct from the five known clades (**Figure 1A**). The
260 four newly assembled Clade VI genomes clustered together in the new clade, which was more
261 closely related to Clade IV but differed from all five known clades by at least 36,000 SNPs. The
262 SNP distances between clades are shown in **Figure 1B**. The number of SNP differences within
263 each clade is in appendix 1 (p17). *In silico* STR analysis was performed based on a previously
264 published scheme, and the topology of the dendrogram also demonstrated clustering of Clade
265 VI isolates based on STR profiles, congruent with the observation of a sixth clade using WGS
266 ^{22,23} (appendix 1 p18). Additional Bayesian phylogenetic analysis suggests that the estimated
267 divergence time of Clade VI from the most closely related Clade IV was at 4831 BCE (HPD
268 272.4 – 311.6 years ago). Details of the molecular clock analysis are in appendix 1 (p19).

269

270 We further investigated whether the Clade VI isolates have clade-specific features that
271 distinguish them from their closest relatives in Clade IV. Previous studies observed clade-
272 specific chromosomal rearrangements in *C. auris*^{24,25}. We reconstructed complete genomes to
273 investigate if similar clade-specific large chromosomal rearrangements exist in the Singaporean
274 Clade VI isolates (**Figure 2** and appendix 1 p20). Compared to Clade IV, an inversion in
275 chromosome 5 was not observed in the Singaporean Clade VI genomes relative to Clade I.
276 Furthermore, even though the new Clade VI isolates were most closely related to Clade IV, we
277 found that they were of opposite mating types (**Figure 1C**). Clade IV isolates carry the *MTLa*
278 allele at the *MAT* locus²⁴, while all Clade VI isolates belong to mating type α with amino acid
279 identity of homologs above 95% based on BLAST matches (appendix 1 p21).

280
281 Given the marked difference among whole genomes of *C. auris* from different clades, we
282 exploited the SNP distances between the genomes to develop a system to monitor and alert for
283 the emergence of potentially new clades. Using publicly available WGS data of 3,651 isolates
284 and the three representative genomes from Clade VI, with a final dataset of >6.6 million genome
285 pairs, we simulated the detection of Clade V and Clade VI at two time points, when the WGS
286 data became available (**Figure 3**). Although the amount of WGS data increased from 554
287 genomes in 2019 Q1 to 3,654 genomes in 2023 Q3 (appendix 1 pp25-26), the SNP distance
288 thresholds inferred by the Bayesian logistic regression model were consistently between 19,000
289 – 23,000 SNPs across the time points (**Figure 3A**). The model detected the emergence of
290 Clades V (2019 Q3) and VI (2023 Q2) as it monitored the SNP distance amongst the genomes
291 over time (**Figure 3B**). As an illustrative example, in 2023 Q3, there were six connected
292 components and each component consisted of genomes from the same clade (**Figure 3C**).
293 Detailed results on the automatic detection of new clades can be found in appendix 1 pp 25-26.

294 Discussion

295 We report the discovery of the sixth major clade of *C. auris*, which includes three
296 epidemiologically unlinked isolates detected in Singapore, and one publicly available WGS data
297 of an isolate reported from Bangladesh (SRR24877249). This conclusion is based upon
298 extensive genomic analysis demonstrating the clustering of these four isolates, and large
299 genetic divergence from the five known clades. We obtained high-quality hybrid genome
300 assemblies of the Singaporean Clade VI isolates, and analyzed these genomes together with
301 3,651 publicly available *C. auris* whole-genome sequences.

302
303 Although the samples that yielded these isolates were collected in Singapore and Bangladesh,
304 the Clade VI isolates were most closely related to representatives of Clade IV (divergence of
305 ~36,000 SNPs), which were mostly from South America. The genomic distance observed
306 between Clades IV and VI is greater than that observed between Clades I and III (~35,000
307 SNPs). Within Clade VI, the three Singaporean isolates were distinct in at least 70 SNPs. This
308 within-clade difference is similar to what was previously reported among Iranian Clade V
309 isolates, which were cultured from patients with community-onset otomycosis¹⁵. The
310 independent detection of Clade VI isolates from Singapore and Bangladesh suggests that Clade
311 VI strains might have been circulating undetected in Southeast Asia and South Asia. Given the
312 estimated divergence time from Clade IV, we speculate that there may exist an endemic zone or
313 a special ecological habitat for Clade VI *C. auris* in the Indomalaya biogeographic realm. Further
314 discussion on the divergence time is available in appendix 1 p25.

315
316 The Clade VI isolates in Singapore were distinct from their closest relatives in Clade IV based
317 on phenotypic investigations and genomic analysis. Firstly, the susceptibility to antifungals, and

318 lack of known AMR resistance-conferring mutations in the Clade VI isolates (**Table 1** and
319 appendix 1 p22-24) are unlike that of Clade IV, where approximately half of the isolates are
320 resistant to fluconazole^{10,14,26,27}. Secondly, Clade VI isolates lack an inversion observed in Clade
321 IV (**Figure 2** and appendix p20). Thirdly, Clade VI isolates carry the *MTL α* allele, a mating-type
322 allele opposite to the *MTL α* observed in Clade IV. As isolates of opposite mating types are yet to
323 be found within the same clade, taken together with other findings presented in this study, the
324 Clade VI isolates are likely representatives of a new major clade rather than a sub-lineage of an
325 existing clade. Further discussion on mating type can be found in appendix 1 p25.

326
327 Unlike the five known *C. auris* clades that are named after their geographic location, these
328 Clade VI isolates detected in Singapore have no known specific geographic associations yet.
329 Due to the retrospective nature of this study, extensive travel history could not be obtained from
330 the patients. We postulate that isolate C was related to Bangladesh geographically, as Patient C
331 was airlifted from Bangladesh to Singapore and isolate C was grown from the blood culture
332 taken within 48 hours of his admission. The independent detection of another Clade VI isolate in
333 Bangladesh (SRR24877249, sample collected in 2021) suggests that Clade VI strains may be
334 circulating undetected in Bangladesh, with the first documented Clade VI isolate being detected
335 in Singapore (isolate C, F0083). Patient B was known to have epidemiological links to Indonesia
336 and Singapore, but not to Bangladesh or South Asia. Patient A was not known to have left
337 Singapore in the preceding two years and has no known epidemiological link to South Asia. The
338 identification of more Clade VI isolates may help us to pinpoint a geographic origin for this
339 emergent clade. Further discussion on the detection and containment of *C. auris* in Singapore is
340 in appendix 1 p25.

341
342 While the origin of Clade VI cannot be resolved, the cryptic circulation of Clade VI *C. auris* in
343 Southeast Asian and South Asian countries is evident because all four patients had
344 epidemiological links to Bangladesh, Indonesia, and Singapore. Due to the paucity of *C. auris*
345 data in the region, it is unclear whether Clade VI is widespread or expanding from its presumed
346 endemic zone. Its potential for causing outbreaks also remains unknown. Clades I, III, and IV
347 are known to cause prolonged outbreaks in healthcare settings^{3,6,14,27-31}. As a result, the
348 acquisition of Clades I, III, and IV *C. auris* is often traced to healthcare settings. On the contrary,
349 Clades II and V are not known to be healthcare-associated^{10,15,32,33}. Although the generally wild-
350 type antifungal susceptibility profile of Clade VI resembles that of Clades II and V, two of the
351 three Clade VI isolates were detected in invasive clinical infections (blood and intraoperative
352 tissue). This is unlike Clades II and V, which are most commonly associated with community-
353 onset otomycosis^{10,15,32,33}. As the virulence and outbreak potential of Clade VI *C. auris* is yet to
354 be established, it is imperative to ensure its early detection and containment for the purpose of
355 patient safety. Since we have yet to identify the source, the risk factor(s) for acquisition, and the
356 specific geographical associations of Clade VI *C. auris*, active surveillance strategies and
357 priorities may need to be reassessed by individual healthcare institutions, to balance possible
358 emergent threat containment with optimal healthcare resource utilization.

359
360 The Clade VI isolates were detected as part of a microbial genomic surveillance program within
361 a Singaporean academic medical center. Although laboratories are increasingly able to obtain
362 microbial genome sequences, it is evident that ongoing, large-scale genomic analysis to provide
363 relevant public health alerts continues to be challenging. We aimed to augment our ability to
364 detect outlier genomes, such as the Clade VI genomes reported in this study, by demonstrating
365 a proof-of-concept machine learning approach that can automatically identify outlier genomes
366 (detection of Clades V and VI) for enhancing genomic surveillance (**Figure 3**). Further
367 discussion on the results and limitations of this machine learning approach can be found in
368 appendix 1 pp25-26.

369
370 This study has several limitations. Due to the retrospective nature of this study, full travel and
371 exposure history could not be traced. The sample size of Clade VI isolates was small, and
372 therefore the reported genomes may not be representative of the entire clade. The antifungal
373 susceptibility test was not performed with a reference broth dilution method but was performed
374 with Sensititre™ YeastOne™ plates. Stringent quality control of publicly available WGS data
375 discarded 18.4% (824/4475) of available genome data, including isolate C's sequence
376 (SRR10102336) reported from Singapore in 2019²¹. This significant data loss from suboptimal
377 quality suggests that when applied in practice, our mathematical modelling approach to detect
378 new clades will face additional challenges, and the framework presented in this study therefore
379 represents a proof-of-concept that requires further evaluations and fine-tuning.

380
381 Worsening spread of *C. auris* has recently been reported in Europe and in the United States^{34–}
382 ³⁶. It is evident that *C. auris* continues to be a threat to public health notwithstanding the
383 resources available for surveillance and containment in high-income countries. Given the
384 resource limitation in parts of Indomalaya, innovative solutions are necessary to augment
385 regional microbial genomic surveillance. Cost-effective integration of existing routine infectious
386 disease diagnostics with genomic surveillance at sentinel sites^{37–39}, together with effective
387 sequence data analysis, are essential components for pandemic preparedness. Our study
388 provides evidence for the discovery of the sixth major *C. auris* clade in Singapore and
389 underscores the need to enhance genomic surveillance in strategic sentinel sites.

390

391 Figure legends

392 **Figure 1.** (A) Genetic relationships among representative *Candida auris* genomes from five
393 known clades and the newly identified sixth clade. Six distinct genetic clusters were observed,
394 each consisting of genomes from the same clade. Genetic relationships were inferred using
395 maximum-likelihood phylogenetic analysis, with 1,000 bootstrap replicates. The publicly
396 available B8441 genome was used as a reference to identify small nucleotide polymorphisms
397 (SNPs) in other isolates. Hybrid assemblies of the Singaporean *C. auris* isolates A (F3485), B
398 (F1580), and C (F0083), were used in this analysis. An additional publicly available whole-
399 genome sequence independently submitted from Bangladesh (SRR24877249) was found to
400 belong to the same cluster as the Singaporean Clade VI isolates. The Clade VI genomes are
401 labeled in blue bolded fonts. Yellow circle denotes the bootstrap value. The scale bar represents
402 the mean number of nucleotide substitutions per site. (B) SNP difference among representative
403 *C. auris* genomes. The numbers within each box indicate the number of SNP differences
404 observed between genomes in distinct clades (in thousands). Isolates between different clades
405 were divergent in at least 35,000 positions genome-wide (e.g. between Clades I and III). Clade
406 VI genomes were divergent in at least 36,000 SNPs from all known clades (I-V). Purple boxes
407 within the heatmap represent genomes that are divergent in $\leq 5,000$ SNPs from each other. (C)
408 Summary of mating type predictions for each clade, determined based on the MTL α / α alleles
409 present at the MAT locus.

410
411 **Figure 2.** Chromosomal rearrangements in *Candida auris*. Chord diagrams showing
412 translocations and inversions in representative genomes for clades II, III, IV, V, and VI, using
413 CA-AM1 (Clade I) as a reference. Only rearrangement events of size $>10,000$ bp are displayed.
414 Based on whole-genome alignments, Clade VI has an inversion in the long arm of chromosome
415 1, which is also observed in the representative genomes from Clades II, III, IV, and V. We also
416 identified a unique translocation event between chromosomes 2 and 4 in Clade VI.
417 Representative isolate genomes used in this figure: Clade I, CA-AM1 (GCA_014673535); Clade
418 II, B13463 (GCA_016495665); Clade III, A1 (GCA_014217455); Clade IV, B11245
419 (GCA_008275145); Clade V, B18474 (SRR13269545); Clade VI, isolate A (F3485).

420
421 **Figure 3.** Detection of potentially new *Candida auris* clades using a Bayesian analysis
422 framework. **(A)** Line plot illustrates small nucleotide polymorphism (SNP) distance thresholds
423 learned across time points using Bayesian logistic regression. The Y-axis represents the SNP
424 distance between genome pairs. The pink dashed line represents the mode of the SNP distance
425 threshold (d), and the shaded region represents the 95% high-density interval (HDI) bounded by
426 d_{low} and d_{high} . The purple solid line marks d_{high} . When the SNP distance is greater than d_{high} , the
427 input genome pair is predicted to be from different clades. The X-axis represents time in
428 quarterly units, where *C. auris* whole genome sequence (WGS) data released within the same
429 time period were added incrementally into the training dataset. **(B)** Line plot illustrates the
430 number of connected components (Y-axis) detected over time (X-axis in quarterly units). The
431 emergence of Clade V and Clade VI were detected by the rise in the number of connected
432 components (CCs) (black arrows). **(C)** Graph visualization of the relationships between
433 genomes in 2023 Q3. Each node represents a *C. auris* genome, and node color denotes clades.
434 Each color represents genomes with known clade information. Gray nodes represent genomes
435 without clade information in the literature. For ease of visualization, we subsampled the
436 genomes in the components which contain more than 200 genomes. An edge links two nodes
437 when $s < d_{high}$, indicating that the genomes are predicted to belong to the same clade. Six
438 connected components were observed across 3,654 genomes.

439 **Tables**

440 **Table 1.** Minimum inhibitory concentration (MIC) values of Singapore Clade VI *Candida auris*
 441 isolates. Antifungal susceptibility testing was performed using the Sensititre™ YeastOne™
 442 ITAMYUCC Plate (Thermo Scientific, Cleveland, OH, USA) according to the manufacturer’s
 443 instructions, in a College of American Pathologists-accredited clinical laboratory. *5-Flucytosin
 444 MIC was determined using the Sensititre™ YeastOne™ YO10 Plate (International version)
 445 (Thermo Scientific, Cleveland, OH, USA). Isolates A and B exhibited wild-type MICs for all
 446 tested antifungals^{40,41}.
 447

	Minimum inhibitory concentration (mg/L) of		
	Isolate A (F3485)	Isolate B (F1580)	Isolate C (F0083)
Anidulafungin	0.12	0.12	0.12
Micafungin	0.12	0.06	0.06
Caspofungin	0.06	0.03	0.06
5-Flucytosine*	≤ 0.06	≤ 0.06	≤ 0.06
Isavuconazole	0.03	≤ 0.008	0.06
Posaconazole	0.03	≤ 0.008	0.06
Voriconazole	0.03	≤ 0.008	0.06
Itraconazole	0.06	0.03	0.12
Fluconazole	2.00	1.00	4.00
Amphotericin B	1.00	0.50	2.00

448

449 Contributors

450 CS and KK conceptualized the study and CS, KK, and CT prepared the draft manuscript. CS,
451 KK, CT, KML, and MGT performed the literature search. KML, CT, and NN contributed to the
452 conceptualization and design of the study. MGT, ALT, and YET contributed isolates for analyses
453 and provided expert advice on clinical mycology. CS, KK, KML, PB, SSG, PR, BS and JJKC
454 contributed to the data acquisition and analysis process. KSC and LO contributed samples for
455 analysis and provided expert advice on clinical molecular microbiology. LCL, MLL, BBSI, MKA,
456 YY, JXYS, and IV provided epidemiological data and expert infection control and prevention
457 advice. BC and LW provided specialist infectious disease inputs and clinical data. JFM provided
458 expert advice and supported STR and phylogenetic timing analysis. BS performed STR and
459 phylogenetic timing analysis. BHT and THK critically evaluated the clinical, microbiological, and
460 epidemiological data and analysis. NN critically evaluated the genomic data analysis. All authors
461 contributed to the manuscript, and edited and reviewed the final manuscript. CS and KK
462 accessed and verified the raw data. CS and KK contributed equally to this work. All authors had
463 full access to all the data in the study and had final responsibility for the decision to submit for
464 publication.

465 Data sharing

466 The code used to perform WGS analyses and train machine-learning models in this study is
467 available at https://github.com/CSB5/Candida_auris_CladeVI. Reads and genome assemblies
468 from this study have been deposited in the National Centre for Biotechnology Information
469 (NCBI) Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>) under
470 BioProject accession number PRJNA1000034.

471 Declaration of interests

472 All authors declare no competing interests.

473 Supplementary Material

474 Supplementary Appendix 1
475 Supplementary Appendix 2
476 Supplementary Appendix 3

477 Acknowledgements

478 This work was supported by the NCRS Duke-NUS Academic Medical Center Academic Clinical
479 Program grant (09/FY2022/P1/17-A32, GRDUKP003401), the Genedant-GIS Innovation
480 Program grant, resources from the Genome Institute of Singapore (GIS), and resources from
481 Singapore General Hospital (SGH). Karrie K. K. Ko is supported by the Singapore National
482 Medical Research Council (NMRC) research training fellowship (MOH-FLWSHP19may-0005).
483 The work was supported in part through computational resources (Sockeye) and services
484 provided by Advanced Research Computing at the University of British Columbia, Vancouver,
485 Canada. We thank members of the Diagnostic Bacteriology Laboratory and the Molecular
486 Laboratory, SGH, for their technical assistance. We thank members of the Department of
487 Infection Prevention and Epidemiology, SGH, for their assistance with epidemiologic data. We
488 thank the Department of Infectious Diseases, SGH, for their assistance with clinical data. We

489 thank members of the Laboratory of Metagenomic Technologies and Microbial Systems, GIS,
490 for their helpful comments on genomic analysis.

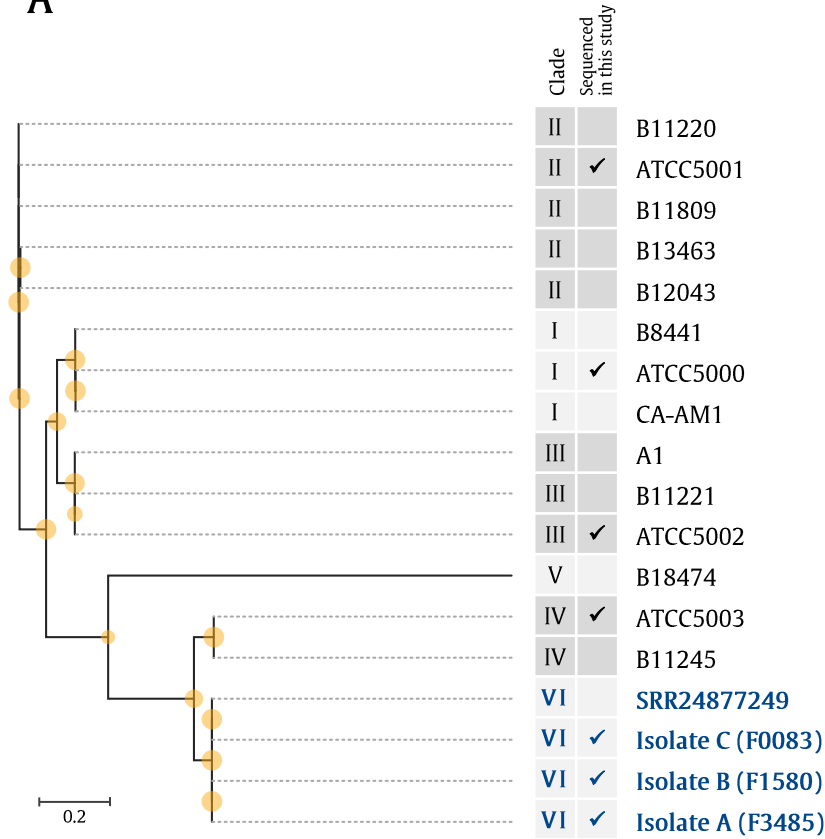
491 References

- 492 1 Chowdhary A, Jain K, Chauhan N. *Candida auris* Genetics and Emergence. *Annu Rev*
493 *Microbiol* 2023; published online July 5. DOI:10.1146/annurev-micro-032521-015858.
- 494 2 Morales-López SE, Parra-Giraldo CM, Ceballos-Garzón A, *et al.* Invasive Infections with
495 Multidrug-Resistant Yeast *Candida auris*, Colombia. *Emerg Infect Dis* 2017; **23**: 162–4.
- 496 3 Ruiz-Gaitán A, Moret AM, Tasiás-Pitarch M, *et al.* An outbreak due to *Candida auris* with
497 prolonged colonisation and candidaemia in a tertiary care European hospital. *Mycoses*
498 2018; **61**: 498–505.
- 499 4 Adams E, Quinn M, Tsay S, *et al.* *Candida auris* in Healthcare Facilities, New York, USA,
500 2013–2017. *Emerg Infect Dis* 2018; **24**: 1816–24.
- 501 5 Eyre DW, Sheppard AE, Madder H, *et al.* A *Candida auris* Outbreak and Its Control in an
502 Intensive Care Setting. *N Engl J Med* 2018; **379**: 1322–31.
- 503 6 Schelenz S, Hagen F, Rhodes JL, *et al.* First hospital outbreak of the globally emerging
504 *Candida auris* in a European hospital. *Antimicrob Resist Infect Control* 2016; **5**: 35.
- 505 7 Chowdhary A, Sharma C, Duggal S, *et al.* New clonal strain of *Candida auris*, Delhi, India.
506 *Emerg Infect Dis* 2013; **19**: 1670–3.
- 507 8 Theodoropoulos NM, Bolstorff B, Bozorgzadeh A, *et al.* *Candida auris* outbreak involving
508 liver transplant recipients in a surgical intensive care unit. *Am J Transplant* 2020; **20**: 3673–
509 9.
- 510 9 Calvo B, Melo ASA, Perozo-Mena A, *et al.* First report of *Candida auris* in America: Clinical
511 and microbiological aspects of 18 episodes of candidemia. *J Infect* 2016; **73**: 369–74.
- 512 10 Chow NA, de Groot T, Badali H, Abastabar M, Chiller TM, Meis JF. Potential Fifth Clade of
513 *Candida auris*, Iran, 2018. *Emerg Infect Dis* 2019; **25**: 1780–1.
- 514 11 Welsh RM, Bentz ML, Shams A, *et al.* Survival, Persistence, and Isolation of the Emerging
515 Multidrug-Resistant Pathogenic Yeast *Candida auris* on a Plastic Health Care Surface. *J*
516 *Clin Microbiol* 2017; **55**: 2996–3005.
- 517 12 Fisher MC, Denning DW. The WHO fungal priority pathogens list as a game-changer. *Nat*
518 *Rev Microbiol* 2023; **21**: 211–2.
- 519 13 Centers for Disease Control and Prevention (U.S.) National Center for Emerging Zoonotic
520 and Infectious Diseases (U.S.). Division of Healthcare Quality Promotion. Antibiotic
521 Resistance Coordination and Strategy Unit. Antibiotic resistance threats in the United
522 States, 2019. DOI:10.15620/cdc:82532.
- 523 14 Lockhart SR, Etienne KA, Vallabhaneni S, *et al.* Simultaneous Emergence of Multidrug-
524 Resistant *Candida auris* on 3 Continents Confirmed by Whole-Genome Sequencing and
525 Epidemiological Analyses. *Clin Infect Dis* 2017; **64**: 134–40.

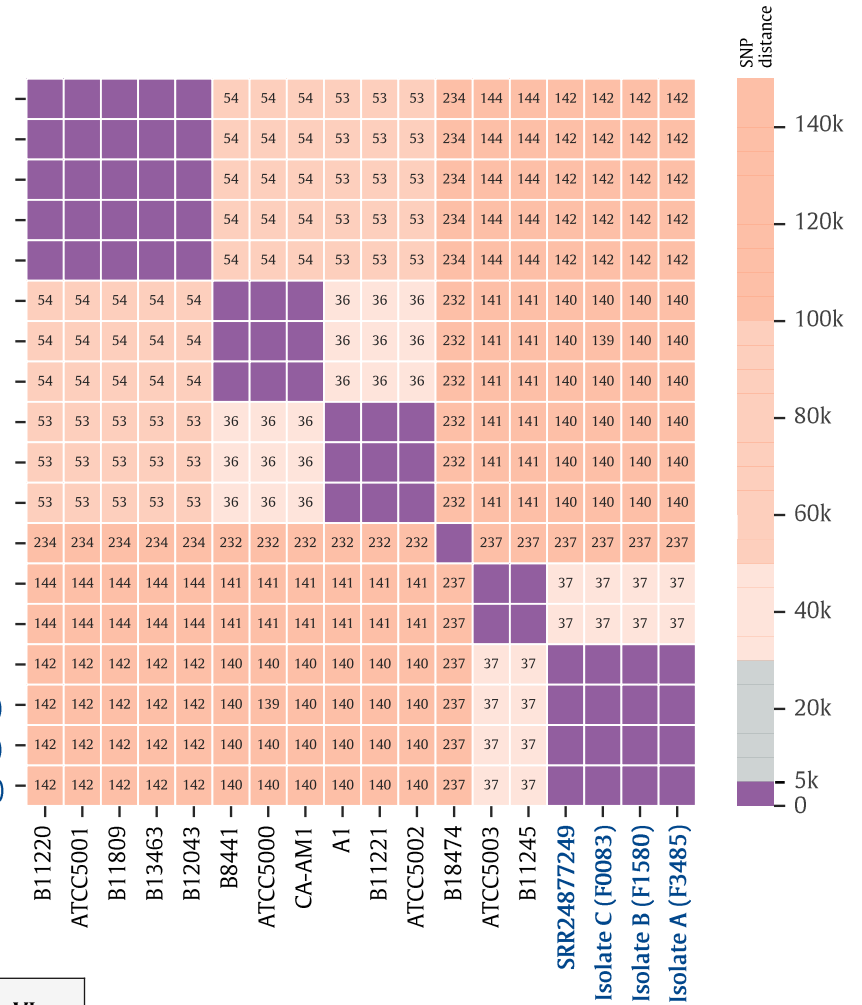
- 526 15 Spruijtenburg B, Badali H, Abastabar M, *et al.* Confirmation of fifth *Candida auris* clade by
527 whole genome sequencing. *Emerg Microbes Infect* 2022; **11**: 2405–11.
- 528 16 Abastabar M, Haghani I, Ahangarkani F, *et al.* *Candida auris* otomycosis in Iran and review
529 of recent literature. *Mycoses* 2019; **62**: 101–5.
- 530 17 Jeffery-Smith A, Taori SK, Schelenz S, *et al.* *Candida auris*: a Review of the Literature. *Clin*
531 *Microbiol Rev* 2018; **31**. DOI:10.1128/CMR.00029-17.
- 532 18 Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. *Candida auris* sp.
533 nov., a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a
534 Japanese hospital. *Microbiol Immunol* 2009; **53**: 41–4.
- 535 19 Rhodes J, Fisher MC. Global epidemiology of emerging *Candida auris*. *Curr Opin Microbiol*
536 2019; **52**: 84–9.
- 537 20 Iguchi S, Itakura Y, Yoshida A, *et al.* *Candida auris*: A pathogen difficult to identify, treat,
538 and eradicate and its characteristics in Japanese strains. *J Infect Chemother* 2019; **25**:
539 743–9.
- 540 21 Tan YE, Teo JQ-M, Rahman NBA, *et al.* *Candida auris* in Singapore: Genomic
541 epidemiology, antifungal drug resistance, and identification using the updated 8.01
542 VITEK[®]2 system. *Int J Antimicrob Agents* 2019; **54**: 709–15.
- 543 22 de Groot T, Spruijtenburg B, Parnell LA, Chow NA, Meis JF. Optimization and Validation of
544 *Candida auris* Short Tandem Repeat Analysis. *Microbiol Spectr* 2022; **10**: e0264522.
- 545 23 de Groot T, Puts Y, Berrio I, Chowdhary A, Meis JF. Development of *Candida auris* Short
546 Tandem Repeat Typing and Its Application to a Global Collection of Isolates. *MBio* 2020;
547 **11**. DOI:10.1128/mBio.02971-19.
- 548 24 Muñoz JF, Gade L, Chow NA, *et al.* Genomic insights into multidrug-resistance, mating and
549 virulence in *Candida auris* and related emerging species. *Nat Commun* 2018; **9**: 5346.
- 550 25 Muñoz JF, Welsh RM, Shea T, *et al.* Clade-specific chromosomal rearrangements and loss
551 of subtelomeric adhesins in *Candida auris*. *Genetics* 2021; **218**: iyab029.
552 DOI:10.1093/genetics/iyab029.
- 553 26 Chowdhary A, Prakash A, Sharma C, *et al.* A multicentre study of antifungal susceptibility
554 patterns among 350 *Candida auris* isolates (2009-17) in India: role of the ERG11 and FKS1
555 genes in azole and echinocandin resistance. *J Antimicrob Chemother* 2018; **73**: 891–9.
- 556 27 Chow NA, Muñoz JF, Gade L, *et al.* Tracing the Evolutionary History and Global Expansion
557 of *Candida auris* Using Population Genomic Analyses. *MBio* 2020; **11**.
558 DOI:10.1128/mBio.03364-19.
- 559 28 Mulet Bayona JV, Tormo Palop N, Salvador García C, *et al.* Characteristics and
560 Management of Candidaemia Episodes in an Established *Candida auris* Outbreak.
561 *Antibiotics (Basel)* 2020; **9**. DOI:10.3390/antibiotics9090558.
- 562 29 Salah H, Sundararaju S, Dalil L, *et al.* Genomic Epidemiology of *Candida auris* in Qatar
563 Reveals Hospital Transmission Dynamics and a South Asian Origin. *J Fungi (Basel)* 2021;

- 564 7. DOI:10.3390/jof7030240.
- 565 30 Chow NA, Gade L, Tsay SV, *et al.* Multiple introductions and subsequent transmission of
566 multidrug-resistant *Candida auris* in the USA: a molecular epidemiological survey. *Lancet*
567 *Infect Dis* 2018; **18**: 1377–84.
- 568 31 Ben Abid F, Salah H, Sundararaju S, *et al.* Molecular characterization of *Candida auris*
569 outbreak isolates in Qatar from patients with COVID-19 reveals the emergence of isolates
570 resistant to three classes of antifungal drugs. *Clin Microbiol Infect* 2023; published online
571 April 26. DOI:10.1016/j.cmi.2023.04.025.
- 572 32 Sekizuka T, Iguchi S, Umeyama T, *et al.* Clade II *Candida auris* possess genomic structural
573 variations related to an ancestral strain. *PLoS One* 2019; **14**: e0223433.
- 574 33 Kwon YJ, Shin JH, Byun SA, *et al.* *Candida auris* Clinical Isolates from South Korea:
575 Identification, Antifungal Susceptibility, and Genotyping. *J Clin Microbiol* 2019; **57**.
576 DOI:10.1128/JCM.01624-18.
- 577 34 Lyman M, Forsberg K, Sexton DJ, *et al.* Worsening Spread of *Candida auris* in the United
578 States, 2019 to 2021. *Ann Intern Med* 2023; **176**: 489–95.
- 579 35 Kilburn S, Innes G, Quinn M, *et al.* Antifungal Resistance Trends of *Candida auris* Clinical
580 Isolates in New York and New Jersey from 2016 to 2020. *Antimicrob Agents Chemother*
581 2022; **66**: e0224221.
- 582 36 Kohlenberg A, Monnet DL, Plachouras D, *Candida auris* survey collaborative group.
583 Increasing number of cases and outbreaks caused by *Candida auris* in the EU/EEA, 2020
584 to 2021. *Euro Surveill* 2022; **27**. DOI:10.2807/1560-7917.ES.2022.27.46.2200846.
- 585 37 Ko KKK, Chu JJK, Lim KM, *et al.* Clonal serotype 1c multidrug-resistant *Shigella flexneri*
586 detected in multiple institutions by sentinel-site sequencing. *Front Med* 2022; **9**: 964640.
- 587 38 Yingtaweessittikul H, Ko K, Rahman NA, Tan SYL, Nagarajan N, Suphavilai C. CalmBelt:
588 rapid SARS-CoV-2 genome characterisation for outbreak tracking. *Front Med* 2021; **8**:
589 790662. doi: 10.3389/fmed.2021.790662 .
- 590 39 Ko KKK, Chng KR, Nagarajan N. Metagenomics-enabled microbial surveillance. *Nat*
591 *Microbiol* 2022; **7**: 486–96.
- 592 40 Clinical and Laboratory Standard Institutes. Clinical and Laboratory Standard Institutes
593 (CLSI). Epidemiological Cutoff Values for Antifungal Susceptibility Testing, 4th Edition.
594 CLSI Supplement M57S. USA,2022. USA: Clinical and Laboratory Standard Institutes.
- 595 41 Antifungal Susceptibility Testing and Interpretation. 2023; published online April 5.
596 <https://www.cdc.gov/fungal/candida-auris/c-auris-antifungal.html> (accessed July 01, 2023).

A

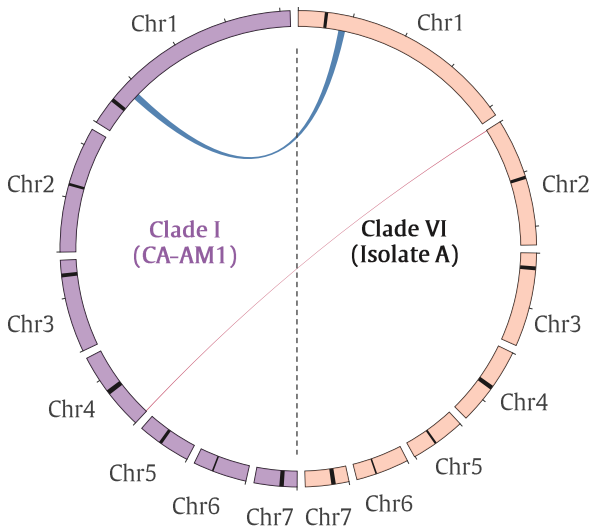
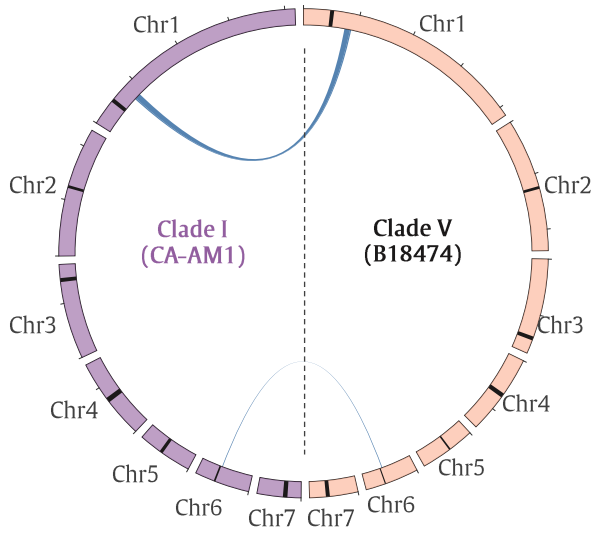
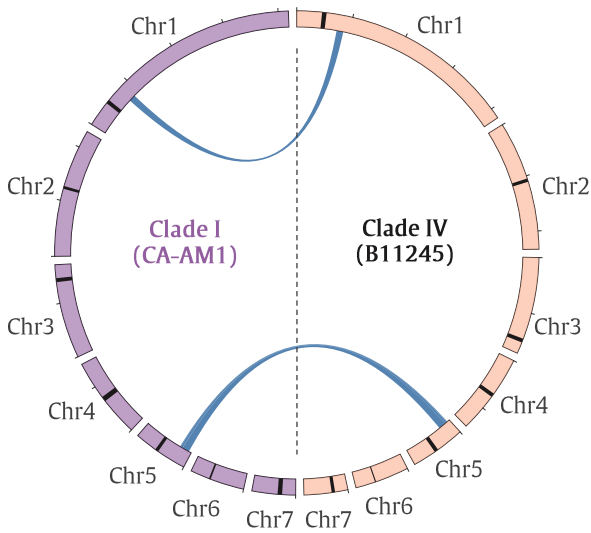
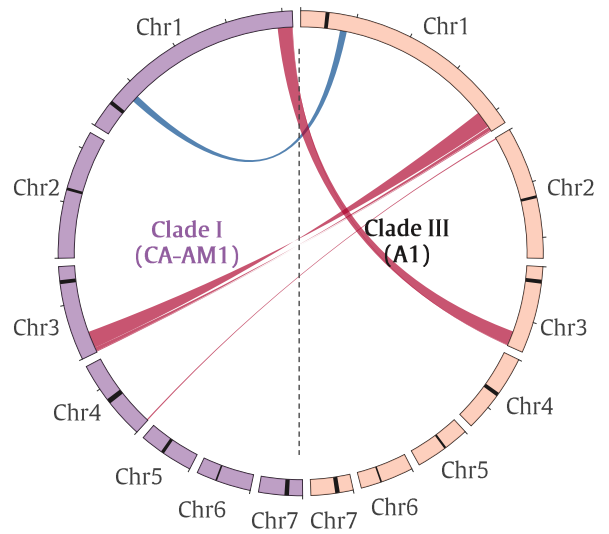
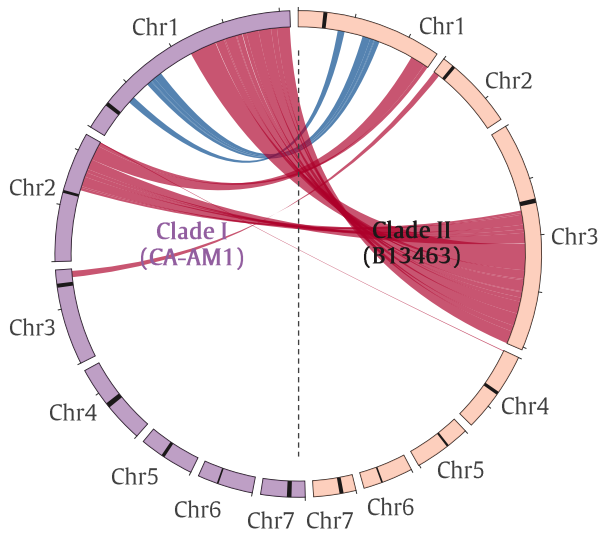


B



C

Clade	I	II	III	IV	V	VI
Mating type locus	<i>MTLa</i>	<i>MTLα</i>	<i>MTLα</i>	<i>MTLa</i>	<i>MTLa</i>	<i>MTLα</i>



Genome information

- Reference genome
- Query genome
- Centromere

Rearrangement event

- Inversion
- Translocation

