# Nanopore Genome Sequencing and Variant Analysis of the Susceptible *Candida auris* Strain L1537/2020, Salvador, Brazil

**Abstract** *Candida auris* has been reported worldwide, but only in December 2020, the first strain from a COVID-19 patient in Brazil was isolated. Here, we describe the genome sequence of this susceptible *C. auris* strain and performed variant analysis of the genetic relatedness with strains from other geographic localities.

*Candida auris* is an emerging multidrug-resistant pathogen able to cause invasive infection outbreaks in tertiary care hospitals worldwide. *C. auris* has the unique ability to persistently colonize the skin and medical surfaces, making it easily transmissible [1, 2]. Currently, 4 major populations have been described, representing geographic linked clades: South East Asia (clade I), Eastern Asia (clade II), Southern Africa (clade III), South America (clade IV), and a minor fifth lineage (clade V) represented by a single Iranian isolate [2, 3].

*Candida auris* was first reported from Brazil in December 2020 where it was isolated from a COVID-19 patient hospitalized in a reference center at Salvador city. Microsatellite typing showed that the strain (L1537/2020) from the index patient (case one

in [4]) is part of the South Asia clade (clade I). Remarkably, the strain was susceptible for all antifungal classes [4]. Whole genome sequencing (WGS) was performed on strain L1537/2020 to enable in-depth analysis of its genetic relatedness with publicly available clade I strains. Additionally, WGS of this strain provides genome data of an antifungal susceptible *C. auris* strain [4, 5].

The strain was cultured in 10 ml peptone glucose broth (LP0040; Oxoid, Basingstoke, UK) and incubated (125 rpm) at 25 °C for 3 days. High-quality genomic DNA was extracted as described before [6]. However, the final step with chloroform/isoamyl-alcohol was replaced by column-based purification using the Fungi/Yeast Genomic DNA kit (catalog nr. 27300; Norgen Biotek, Thorold, ON, Canada) and DNA was eluted in 50 μl molecular grade 1 × IDTE buffer (pH 8.0; IDT, Coralville, IA, USA).

Sequencing was performed on the Oxford Nanopore MinION platform (Oxford Nanopore Technologies, Oxfordshire, UK) using the ligation sequencing kit (SQK-LSK109) and native barcoding kit (EXP-NBD104). The sample was sequenced on the MK1B MinION (MIN-101B) device with a FLO-MIN106 (SpotON R9.4) flowcell using MinKNOWN release 20.10.3.

Raw nanopore reads were basecalled using Guppy (v4.4.1 + 1c81d623j; ONT) with the parameters --flowcell FLO-MIN106 --kit SQK-LSK109 --barcode_kits EXP-NBD104 --device CUDA:0. De novo
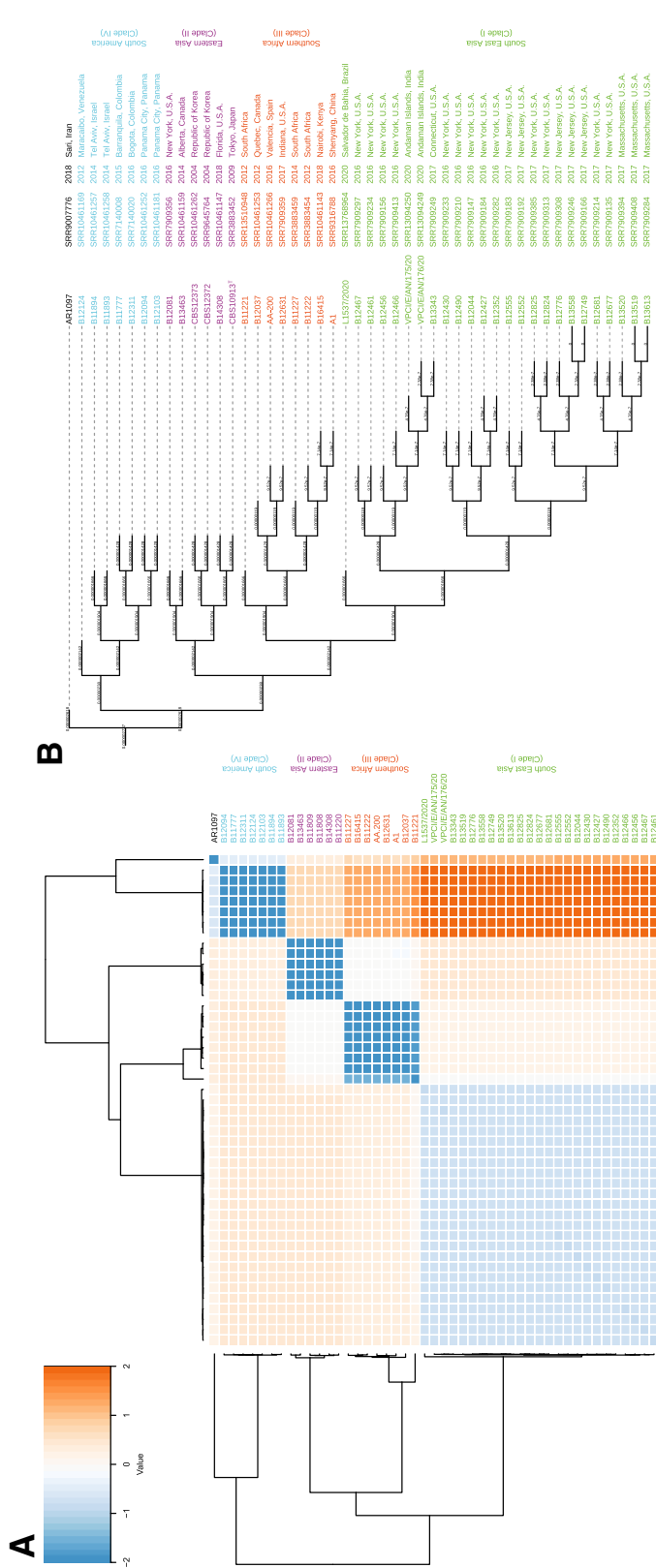
**Fig. 1** **a** Heatmap visualization of the pairwise SNP-number comparison based on sequences grouped by their z-scores (generated in R package, https://tvpham.github.io/ion/). **b** IQ-TREE generated maximum likelihood phylogenetic analysis using 420,164 SNPs, rooted with strain AR1097 (*C. auris* clade V). Numbers above branches represent branch lengths interpreted as the numbers of nucleotide substitutions per nucleotide site. The heatmap z-scores and the IQ-Tree branch lengths show that clade I strains were very similar to each other, however, strain L1537/2020 had a longer branch length and a high number of SNPs (1893–2089) compared to other strains in clade I

genome assembly was performed using Flye v2.8.2-b1689 (https://github.com/fenderglass/Flye; [7]) with the parameters --nano-raw < fastq > --out-dir < directory > --genome-size 12.5 m.

The assembly resulted in 15 fragments with a total length of 12,687,478 bp ($N_{50}$ of 2,134,410 bp; largest fragment 4,519,230 bp) with a mean coverage of 300X. However, when the 7 smaller contigs (range 490–4970 bp) were omitted, the mean coverage was 387X. The 8 large contigs (range 27,834–4,519,230 bp; total length 12,675,277 bp) included the circular mitochondrial genome (27,834 bp; 1000X coverage), the remaining nuclear contigs had a coverage of $\sim$ 300X.

Variant calling was performed using a subset of 47 published genomes representing the five *C. auris* clades, including a benchmark set for clade I, (Fig. 1; [3, 8–12]). Sequencing reads were aligned with reference genome B8441 (NCBI accession SRS1558430) using minimap2 for nanopore and bwa v0.7.17-r1188 for Illumina data [13, 14]. Sam-files were sorted and indexed using samtools v1.9 [15]. For nanopore data, longshot was performed on the bam-file to obtain variants of the reference genome [16]. Indels and SNPs with a quality < 20 were removed using vcftools v0.1.15 [17]. For Illumina data, picard (http://broadinstitute.github.io/picard/) was performed to mark duplicates. Variants were then identified with gatk HaplotypeCaller, and SNPs were selected with gatk HaplotypeCaller and filtered with the settings "QD < 2.0||MQ < 40.0||FS > 60.0||SOR > 3.0||MQRankSum < − 12.5   ||ReadPosRankSum < − 8.0" [18]. SNP-files were merged using vcf-merge to produce a fasta-alignment file for all samples with R-package SNPRelate [19]. Different nucleotides within the fasta-alignment file were counted by pairwise SNP-number comparison (Fig. 1, panel A). IQ-TREE v1.6.1 was performed to build a maximum-likelihood tree from the alignment file that was visualized with itol (https://itol.embl.de/) (Fig. 1, panel B; [20]). As a result, 420,164 SNPs were counted. Strains in clade I had $\sim$ 1200–1900 SNPs, while much higher SNP-counts were observed for clade II ($\sim$ 45 K), clade III ($\sim$ 65 K), clades IV and V ($\sim$ 170 K). Pairwise SNP-number comparison showed that all clade I strains were very much similar with only 230–1011 nucleotide differences, except for strain L1537/2020 that had 1893–2089 SNPs compared to the other clade I strains (Fig. 1).

Although strain L1537/2020 belongs to clade I, it is distantly related to all other representatives of that clade, and in contrast it is susceptible to all common antifungals (Fig. 1; [4]). Several mutations are reported to play a role in *C. auris*' antifungal resistance, viz. *CIS2* (A27T), *ERG3* (W182*, L207I), *ERG11* (Y132F, K143R), *FKS1* (S639P), *MEC3* (A272V), *PEA2* (D367V), *TAC1B* (FS191S, F214S, R495G, S611P), and *UPC2* (M365) [3, 9, 21–23]. None of these mutations are present in the genome of L1537/2020.

Nearly all publicly available *C. auris* genome data was generated by short-read Illumina sequencing [3, 9, 11, 12]. Nonetheless, the relative distant relation of the nanopore-based genome of strain L1537/2020 to other Illumina-sequenced clade I strains cannot be explained by the differences in sequencing technologies. The nanopore flowcell, chemistry and basecalling software used here approaches an accuracy of > 98%. This means that a SNP precision of > 99.9% can be achieved in the case of 50X genome coverage [24]. With the 300X coverage for L1537/2020 it is unlikely that an erroneous mutation was introduced in the assembly. Hence, further studies are needed to investigate the epidemiological and biological impact of the phenotypic and genotypic differences in L1537/2020 versus its multi-drug resistant siblings within clade I.

Astellas, Biotoscana, United Medical, Gilead, MSD and Pfizer. The other authors report no conflict of interest.

**Author name and affiliation**

[1,2]Auke W. de Jong ⓘD
[1,3]Elaine C. Francisco ⓘD
[3,4]João Nóbrega de Almeida Jr ⓘD
[5]Igor B. Brandão
[6]Felicidade M. Pereira ⓘD
[7]Pedro H. Presta Dias ⓘD
[8]Magda M. de Miranda Costa
[9]Regiane T. de Souza Jordão
[1]Duong Vu ⓘD
[3]Arnaldo L. Colombo ⓘD
[1,2,10]Ferry Hagen ⓘD

[1]Department of Medical Mycology, Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, NL-3584CT Utrecht, The Netherlands
e-mail: a.jong@wi.knaw.nl
e-mail: elaineperol@yahoo.com.br
e-mail: d.vu@wi.knaw.nl
e-mail: f.hagen@wi.knaw.nl
e-mail: f.hagen@gmail.com

[2]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands

[3]Laboratório Especial de Micologia, Disciplina de Infectologia, Universidade Federal de São Paulo, São Paulo, Brazil
e-mail: jnaj99@gmail.com
e-mail: colomboal@terra.com.br

[4]Hospital Israelita Albert Einstein, São Paulo, Brazil

[5]Comissão de Controle de Infecção Hospitalar, Hospital de Bahia, Salvador, Brazil
e-mail: igorbrasilbrandao@gmail.com

[6]Laboratório Central de Saúde Pública Professor Gonçalo Muniz, Salvador, Brazil
e-mail: felicidade.pereira@saude.ba.gov.br

[7]Centro de Informações Estratégicas de Vigilância em Saúde da Bahia, Salvador, Brazil
e-mail: pedropresta@hotmail.com

[8]Agência Nacional de Vigilância Sanitária, Brasília, Brazil
e-mail: magda.miranda@anvisa.gov.br

[9]Coordenação Geral de Laboratórios de Saúde Pública, Ministério da Saúde, Brasília, Brazil
e-mail: regiane.jordao@saude.gov.br

[10]Department of Dermatology, Jining No. 1 People's Hospital, Shandong, China

# References

1. Rossato L, Colombo AL. *Candida auris*: what have we learned about its mechanisms of pathogenicity? Front Microbiol. 2018;9:3081.

2. de Jong AW, Hagen F. Attack, defend and persist: how the fungal pathogen *C. auris* was able to emerge globally in healthcare environments. Mycopathologia. 2019;184(3):353–65. https://doi.org/10.1007/s11046-019-00351-w.

3. Chow NA, Muñoz JF, Gade L, Berkow EL, Li X, Welsh RM, Forsberg K, Lockhart SR, Adam R, Alanio A, Alastruey-Izquierdo A, Althawadi S, Araúz AB, Ben-Ami R, Bharat A, Calvo B, Desnos-Ollivier M, Escandón P, Gardam D, Gunturu R, Heath CH, Kurzai O, Martin R, Litvintseva AP, Cuomo CA. Tracing the evolutionary history and global expansion of *C. auris* using population genomic analyses. mBio. 2020;11(2):e03364–19. https://doi.org/10.1128/mBio.03364-19.

4. de Almeida JN, Francisco EC, Hagen F, Brandão IB, Pereira FM, Presta Dias PH, de Miranda Costa MM, de Souza Jordão RT, de Groot T, Colombo AL. Emergence of *C. auris* in Brazil in a COVID-19 intensive care unit. J Fungi. 2021;7(3):220. https://doi.org/10.3390/jof7030220.

5. Mac Aogáin M, Chaturvedi V, Chotirmall SH. MycopathologiaGENOMES: the new 'home' for the publication of fungal genomes. Mycopathologia. 2019;184:551–4. https://doi.org/10.1007/s11046-019-00366-3.

6. Navarro-Muñoz JC, de Jong AW, Gerrits van den Ende B, Haas PJ, Then ER, Mohd Tap R, Collemare J, Hagen F. The high-quality complete genome sequence of the opportunistic fungal pathogen *Candida vulturna* CBS 14366(T). Mycopathologia. 2019;184(6):731–4. https://doi.org/10.1007/s11046-019-00404-0.

7. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6. https://doi.org/10.1038/s41587-019-0072-8.

8. Arora P, Singh P, Wang Y, Yadav A, Pawar K, Singh A, Padmavati G, Xu J, Chowdhary A. Environmental isolation of *C. auris* from the coastal wetlands of Andaman Islands, India. mBio. 2021;12(2):e03181-e3220. https://doi.org/10.1128/mBio.03181-20.

9. Lockhart SR, Etienne KA, Vallabhaneni S, Farooqi J, Chowdhary A, Govender NP, Colombo AL, Calvo B, Cuomo CA, Desjardins CA, Berkow EL, Castanheira M, Magobo RE, Jabeen K, Asghar RJ, Meis JF, Jackson B, Chiller T, Litvintseva AP. Simultaneous emergence of multidrug-resistant *C. auris* on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. Clin Infect Dis. 2017;64(2):134–40. https://doi.org/10.1093/cid/ciw691.

10. Muñoz JF, Gade L, Chow NA, Loparev VN, Juieng P, Berkow EL, Farrer RA, Litvintseva AP, Cuomo CA. Genomic insights into multidrug-resistance, mating and virulence in *C. auris* and related emerging species. Nat Commun. 2018;9(1):5346. https://doi.org/10.1038/s41467-018-07779-6.

11. Tian S, Bing J, Chu Y, Chen J, Cheng S, Wang Q, Zhang J, Ma X, Zhou B, Liu L, Huang G, Shang H. Genomic epidemiology of *C. auris* in a general hospital in Shenyang, China: a three-year surveillance study. Emerg Microbes Infect. 2021;10(1):1088–96. https://doi.org/10.1080/22221751.2021.1934557.

12. Welsh RM, Misas E, Forsberg K, Lyman M, Chow NA. *Candida auris* whole-genome sequence benchmark dataset for phylogenomic pipelines. J Fungi. 2021;7(3):214. https://doi.org/10.3390/jof7030214.

13. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191.

14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;24:1754–60.

15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.

16. Edge P, Bansal V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. Nat Commun. 2019;10:4660. https://doi.org/10.1038/s41467-019-12493-y.

17. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 1000 genomes project analysis group. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8. https://doi.org/10.1093/bioinformatics/btr330.

18. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media. https://www.oreilly.com/library/view/genomics-in-the/9781491975183/.

19. 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;28:3326–8. https://doi.org/10.1093/bioinformatics/bts606.

20. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300.

21. Carolus H, Pierson S, Muñoz JF, Subotić A, Cruz RB, Cuomo CA, Van Dijck P. Genome-wide analysis of experimentally evolved *C. auris* reveals multiple novel mechanisms of multidrug resistance. mBio. 2021;12(2):e03333–20. https://doi.org/10.1128/mBio.03333-20.

22. Li J, Coste AT, Liechti M, Bachmann D, Sanglard D, Lamoth F. Novel *ERG11* and *TAC1b* mutations associated with azole resistance in *C. auris*. Antimicrob Agents Chemother. 2021;65(5):e02663-e2720. https://doi.org/10.1128/AAC.02663-20.

23. Rybak JM, Muñoz JF, Barker KS, Parker JE, Esquivel BD, Berkow EL, Lockhart SR, Gade L, Palmer GE, White TC, Kelly SL, Cuomo CA, Rogers PD. Mutations in *TAC1B*: a novel genetic determinant of clinical fluconazole resistance in *C. auris*. mBio. 2020;11(3):e00365–20. https://doi.org/10.1128/mBio.00365-20.

24. Applications: Nanopore sequencing accuracy. https://www.nanopore.com/accuracy Accessed September 9, 2021.