UC Irvine
UC Irvine Previously Published Works

Title
Skin Metagenomic Sequence Analysis of Early Candida auris Outbreaks in U.S. Nursing Homes

Permalink
https://escholarship.org/uc/item/6mg158xj

Journal
mSphere, 6(4)

ISSN
1556-6811

Authors
Huang, Xin
Welsh, Rory M
Deming, Clay
et al.

Publication Date
2021-08-25

DOI
10.1128/msphere.00287-21

Peer reviewed

Skin Metagenomic Sequence Analysis of Early *Candida auris* Outbreaks in U.S. Nursing Homes

Xin Huang,[a] Rory M. Welsh,[b] Clay Deming,[a] Diana M. Proctor,[a] Pamela J. Thomas,[c] NISC Comparative Sequencing Program,[c] Gabrielle M. Gussin,[d] Susan S. Huang,[d,e] Heidi H. Kong,[f] Meghan L. Bentz,[b] Snigdha Vallabhaneni,[b,g] Tom Chiller,[b] Brendan R. Jackson,[b] Kaitlin Forsberg,[b] Sean Conlan,[a] Anastasia P. Litvintseva,[b] Julia A. Segre[a]

[a]Microbial Genomics Section, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA
[b]Mycotic Diseases Branch, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia, USA
[c]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Maryland, USA
[d]Division of Infectious Diseases, Department of Medicine, University of California Irvine School of Medicine, Irvine, California, USA
[e]Health Policy Research Institute, University of California Irvine School of Medicine, Irvine, California, USA
[f]Dermatology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA
[g]Division of Healthcare Quality Promotion, National Center for Emerging and Zoonotic Infectious Diseases, CDC, Atlanta, Georgia, USA

Xin Huang and Rory M. Welsh contributed equally to this work. Author order was determined alphabetically.

ABSTRACT *Candida auris* is a human fungal pathogen classified as an urgent threat to the delivery of health care due to its extensive antimicrobial resistance and the high mortality rates associated with invasive infections. Global outbreaks have occurred in health care facilities, particularly, long-term care hospitals and nursing homes. Skin is the primary site of colonization for *C. auris*. To accelerate research studies, we developed microbiome sequencing protocols, including amplicon and metagenomic sequencing, directly from patient samples at health care facilities with ongoing *C. auris* outbreaks. We characterized the skin mycobiome with a database optimized to classify *Candida* species and *C. auris* to the clade level. While *Malassezia* species were the predominant skin-associated fungi, nursing home residents also harbored *Candida* species, including *C. albicans*, and *C. parapsilosis*. Amplicon sequencing was concordant with culturing studies to identify *C. auris*-colonized patients and provided further resolution that distinct clades of *C. auris* are colonizing facilities in New York and Illinois. Shotgun metagenomic sequencing from a clinical sample with a high fungal bioburden generated a skin-associated profile of the *C. auris* genome. Future larger scale clinical studies are warranted to more systematically investigate the effects of commensal microbes and patient risk factors on the colonization and transmission of *C. auris*.

IMPORTANCE *Candida auris* is a human pathogen of high concern due to its extensive antifungal drug resistance and high mortality rates associated with invasive infections. *Candida auris* skin colonization and persistence on environmental surfaces make this pathogen difficult to control once it enters a health care facility. Residents in long-term care hospitals and nursing homes are especially vulnerable. In this study, we developed microbiome sequencing protocols directly from surveillance samples, including amplicon and metagenomic sequencing, demonstrating concordance between sequencing results and culturing.

KEYWORDS *Candida auris*, fungal pathogen, mycobiome, medical mycology, ITS1

*C*andida auris has emerged as a serious threat because of its high rates of evolved antimicrobial resistance, high mortality associated with bloodstream infection, and high transmission rates within health care facilities (1, 2). *C. auris* is one of only five

microbial pathogens, and the only fungal species, rated by the U.S. Centers for Disease Control and Prevention (CDC) at the highest level of urgent threat in the 2019 antibiotic resistance report (3). Over one million invasive fungal infections occur globally, increasingly from multidrug-resistant *C. auris* (4).

C. auris was first identified in 2009 from a clinical culture of a Japanese patient's external ear canal (5), and in the ensuing decade, four main independent clades of *C. auris* have been identified from four distinct regions, clade I from South Asia, clade II from East Asia, clade III from Africa, and clade IV from South America (6–10). *C. auris* isolates are commonly azole resistant, with some strains resistant to all three classes of antifungals (azoles, amphotericin B, and echinocandins) (2, 11). As the global emergence of invasive infections was recognized, both the CDC (12) and Public Health England (13) issued multiple clinical alerts.

Approximately half of the *C. auris* clinical cases in the United States have been in the bloodstream, but skin sites (axilla [Ax] and groin [G]) and anterior nares (N) are considered the primary sites of asymptomatic colonization (14). *C. auris* colonizes and is shed from human skin, persists on environmental surfaces (15, 16), and is resistant to certain standard cleaning products, enabling its spread through health care settings. Identifying *C. auris*-colonized patients is a crucial component of infection control since these patients are at increased risk for developing a subsequent bloodstream infection, and they pose a risk for transmission to other patients in the ward.

Outbreaks reported in multiple countries, including American hospitals and nursing homes (7, 17–19), have proven difficult to control. Long-term acute care facilities and nursing homes with ventilator beds are particularly vulnerable to *C. auris* outbreaks since residents stay for long periods of time and have complex care needs. Furthermore, these facilities often do not have the infection control resources or structural protections (e.g., single-occupancy rooms) that are available in short-stay acute care hospitals. These factors underlie the increased risk to residents of health care-associated microbial infections, often by multidrug-resistant pathogens (20).

Human skin of young adult healthy volunteers is inhabited by a diverse community of bacteria (21), fungi (22), and viruses (23). As the first line of defense, the skin harbors trillions of commensal microbes essential for colonization resistance (24, 25). Disruption of the commensal microbial community might provide opportunities for pathogens to colonize and infect the host (26). Amplicon and shotgun metagenomic sequencing have enabled the characterization of microbial communities. Specifically, amplifying and sequencing the bacterial 16S rRNA gene or the fungal internal transcribed spacer 1 (ITS1) region directly from a clinical or environmental sample enables the characterization of these communities, typically at the genus level. Achieving higher level resolution from the 16S rRNA gene or ITS1 region is dependent on the microbial genus and the sequencing strategy but requires custom analysis with curated reference genomes (22, 27). Shotgun metagenomic sequencing enables further investigation of a community's functional potential and may achieve species or strain level resolution, but analysis is limited for rare constituents and for uncultivated strains lacking reference genomes.

A collaboration between the CDC and the National Institutes of Health (NIH) aimed to develop microbiome sequencing protocols, including amplicon and metagenomic sequencing, directly from patient samples at health care facilities with ongoing *C. auris* outbreaks. First, we modified laboratory protocols to obtain skin-associated microbiome DNA directly from surveillance samples, obtained as part of an ongoing point prevalence study. After DNA sequencing, we implemented a strategy to classify *Candida* ITS1 sequences to the species level. In addition, by leveraging high-confidence differences in the ITS1 region, we accurately classified the clade of *C. auris* directly from amplicon sequencing data. The skin ITS1-based mycobiome of patients from multiple health care facilities, including early *C. auris* outbreaks, showed elevated skin colonization by *Candida* species, including *C. auris* and *C. albicans*, concordant with *Candida* culturing results. We also demonstrated the possibility of metagenomic

sequencing directly from surveillance samples to characterize *C. auris* colonizing the skin of a patient with high fungal abundance.

RESULTS

When patients colonized with *Candida auris* were identified in Illinois and New York health care facilities in 2017, local health departments launched point prevalence surveillance surveys to identify any additional residents colonized by *C. auris*. *C. auris* surveillance was a combined axilla/groin swab, but anterior nares was also considered a possible reservoir for these studies. These clinical samples were cultured on selective media, and individual colonies were confirmed as *C. auris* by matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (15). To generalize our study on the skin microbiome of patients colonized with *C. auris*, we sequenced eight or nine patient samples from each of two ventilator-capable skilled nursing facilities (vSNF). These facilities provide enhanced care, including mechanical ventilation, to patients receiving treatment for acute and chronic health conditions. Facilities in New York (region A) and Illinois (region B) had *C. auris* positive colonization rates of 87.5% and 11.1%, respectively. We obtained similar clinical samples from 10 control patients in health care facilities in California serving a similar demographic, but with no history of *C. auris*.

To classify the ITS1 reads derived from amplicon sequencing, we first improved a custom ITS1 database by incorporating ITS1 sequences and their annotations from ITSoneDB (22, 28). Next, we downloaded the complete ITS1 sequences of 14 common and clinically relevant *Candida* species from public databases. Single nucleotide variants (SNVs) within the ITS1 sequences were used to distinguish *C. auris* clades I/III, II, and IV (Fig. 1A).

The skin microbiomes of nursing home residents taken as controls from the unaffected California facility were dominated by *Malassezia* species. Overall, *Malassezia* spp. constituted >80% of the fungal relative abundance in 22 of 25 samples; 7 of 25 samples harbored >5% *Candida* spp., and 5 of these 7 were samples from the groin (Fig. 1B). Three groin samples had high representation of non-*Malassezia* fungi: two were predominated by *C. albicans* and one with *Diutina mesorugosa* (previously classified as *Candida rugosa*). No *C. auris* positive samples or ITS1 reads mapping to *C. auris* were identified from these samples, consistent with surveillance culturing.

In contrast, the skin of residents in region A was dominated by *Candida* species, with four samples dominated by *C. albicans*, two samples by *C. auris*, one sample by a combination of *C. albicans* and *C. auris*, and one sample by *C. dubliniensis* (Fig. 1C). Reads matching *C. auris* were obtained in seven of the eight residents sampled, ranging from less than 1% to greater than 50% relative abundance (Fig. 1C and Table 1). In all cases, sequencing results were consistent with culturing results. Residents in region B with a low *C. auris* positivity rate exhibited similar skin fungal profiles as the control facility, with six out of nine samples dominated by *Malassezia* spp., two samples by *C. albicans*, and one sample by *C. parapsilosis* (Fig. 1C). Again, sequencing results recapitulated culturing results, identifying one of the nine residents as colonized with *C. auris* (Fig. 1C and Table 1). Clade identity derived from ITS1 sequencing confirmed that residents in region A carried clade I, while residents in region B carried clade IV, matching genomic sequencing of isolates from these areas (Table 2) (14).

The bacterial communities of control samples from axilla and nares were dominated by skin commensals, predominantly *Corynebacterium*, *Cutibacterium*, and *Staphylococcus*, and five out of nine groin samples were dominated by *Proteobacteria* (Fig. 2A). Similarly, bacterial communities on the skin of region B residents were dominated by *Corynebacterium* and *Staphylococcus*, with some representation by *Proteobacteria* (Fig. 2B). However, the skin of region A residents with high *C. auris* positivity was colonized by species more commonly considered health care-associated pathogens, including *Acinetobacter*, *Escherichia/Shigella*, *Klebsiella*, *Morganella*, *Proteus*, and *Pseudomonas* (Fig. 2B).

A

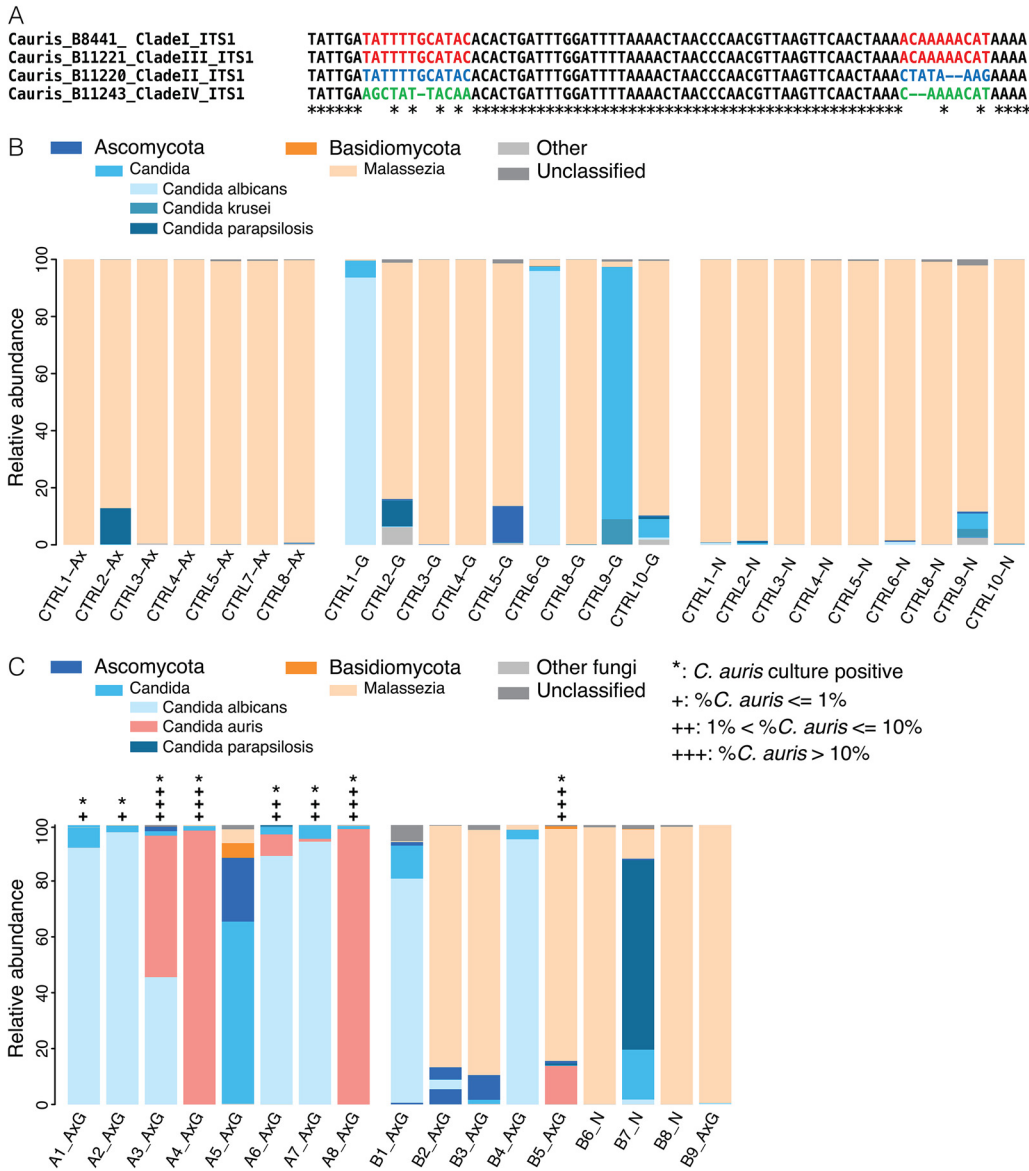|	
Cauris_B8441_ CladeI_ITS1	TATTGATATTTTGCATACACACTGATTTGGATTTTAAAACTAACCCAACGTTAAGTTCAACTAAAACAAAAACATAAAA
Cauris_B11221_CladeIII_ITS1	TATTGATATTTTGCATACACACTGATTTGGATTTTAAAACTAACCCAACGTTAAGTTCAACTAAAACAAAAACATAAAA
Cauris_B11220_CladeII_ITS1	TATTGATATTTTGCATACACACTGATTTGGATTTTAAAACTAACCCAACGTTAAGTTCAACTAAACTATA--AAGAAAA
Cauris_B11243_CladeIV_ITS1	TATTGAAGCTAT-TACAAACACTGATTTGGATTTTAAAACTAACCCAACGTTAAGTTCAACTAAAC--AAAACATAAAA
	****** * * * * ** * * ****

FIG 1 *Candida auris* detection from clinical samples based on ITS1 sequencing. (A) Genetic variation in ITS1 sequences of the four clades of *C. auris*. Asterisks below the alignment indicate consensus positions among all four clades. (B) Relative fungal abundance plot for samples from control (CTRL) residents in a region with no *C. auris* detected. Ax, axilla; G, groin; N, anterior nares. CTRL1–Ax is a sample from the axilla from control subject 1. (C) Relative fungal abundance plot for samples from facilities in regions A and B with *C. auris* outbreaks. Asterisks indicate *C. auris* culture-positive samples, and plus signs indicate the relative abundance of *C. auris* in the samples from ITS1 sequencing. AxG, axilla/groin; N, anterior nares. A1_AxG is a sample from the axilla/groin from subject 1 in region A.

In order to explore the possibility of performing metagenomic sequencing directly from surveillance samples and also to investigate potential adaptation to human skin by *C. auris*, we performed shotgun metagenomic sequencing of four additional samples, all PCR positive for *C. auris*, including samples from additional skilled nursing homes in regions A and B. The microbiomes of all four metagenomic samples were highly dysbiotic, dominated by hospital-associated bacterial species like *Acinetobacter baumannii*, *Klebsiella pneumoniae*, and *Enterococcus faecalis*, similar to what was seen in region A samples by V1-V3 sequencing (data not shown).

For healthy volunteers, fungi typically make up <5% of the microbiome at the sampled body sites (22). Similarly, for these samples, fungi represented less than 5% of the microbial community, with a *C. auris* relative abundance of 0.25 to 0.73% with one

TABLE 1 *Candida* species detected by culturing or MALDI-TOF MS and ITS1 sequencing

Source[a]	*Candida* species detected by culture	*Candida* species detected by sequencing
A1_Ax/G	*C. auris, C. albicans, C. glabrata*	*C. auris, C. albicans, C. glabrata*
A2_Ax/G	*C. auris, C. albicans*	*C. auris, C. albicans*
A3_Ax/G	*C. auris, C. albicans*	*C. auris, C. albicans, C. tropicalis*
A4_Ax/G	*C. auris*	*C. auris*
A5_Ax/G		*C. albicans, C. dubliniensis, C. tropicalis*
A6_Ax/G	*C. auris, C. albicans*	*C. auris, C. albicans, C. parapsilosis*
A7_Ax/G	*C. auris, C. albicans*	*C. auris, C. albicans, C. parapsilosis*
A8_Ax/G	*C. auris*	*C. auris*
B1_Ax/G	*C. albicans*	*C. albicans, C. tropicalis*
B2_Ax/G		*C. albicans, C. tropicalis*
B3_Ax/G		*C. tropicalis*
B4_Ax/G	*C. albicans*	*C. albicans*
B5_Ax/G	*C. auris*	*C. auris, C. parapsilosis*
B6_N		
B7_N	*C. parapsilosis*	*C. albicans, C. dubliniensis, C. glabrata, C. parapsilosis*
B8_N		
B9_Ax/G	*C. albicans, C. glabrata*	*C. albicans, C. tropicalis*

[a]AxG, axilla/groin; N, anterior nares. A1_AxG is a sample from the axilla/groin from resident 1 in region A.

exception. One sample (B12_Ax/G; SRA accession no. SRX5795138) had a high fungal burden and high overall *C. auris* relative abundance (14.24% of reads). We mapped these reads to a high-quality reference genome derived form a 2019 Illinois region clade IV isolate (ACUZR; BioSample accession no. SAMN18141223), which was reported to be susceptible to the commonly used antifungal agents fluconazole, amphotericin B, and micafungin (29). Read coverage of the reference was ~70×, allowing us to confidently call variants between the patient metagenome and our reference. We detected 49 variants, 42 of which were predicted to be in noncoding regions. The seven variants in coding regions (six missense and one synonymous) were all found in genes encoding hypothetical proteins.

DISCUSSION

Over the last decade, *Candida auris* has rapidly emerged as a serious threat for vulnerable patients with immunodeficiency and/or other preexisting conditions, especially in long-term care facilities, such as long-term post-acute care hospitals and nursing homes, with recent clusters also occurring in acute care facilities (30, 31). Surveillance for *C. auris*

TABLE 2 *C. auris* ITS1 read clade classification

Source	% ITS1 reads[b]		
	Clade I/III[a]	Clade II	Clade IV
A1_Ax/G	0.04	nd	nd
A2_Ax/G	0.02	nd	nd
A3_Ax/G	51.24	nd	nd
A4_Ax/G	98.10	nd	nd
A5_Ax/G	nd	nd	nd
A6_Ax/G	8.36	nd	nd
A7_Ax/G	1.00	nd	nd
A8_Ax/G	98.86	nd	nd
B1_Ax/G	nd	nd	nd
B2_Ax/G	nd	nd	nd
B3_Ax/G	nd	nd	nd
B4_Ax/G	nd	nd	nd
B5_Ax/G	nd	nd	13.23
B6_N	nd	nd	nd
B7_N	nd	nd	nd
B8_N	nd	nd	nd
B9_Ax/G	nd	nd	nd

[a]Clades I and III cannot be distinguished by ITS1 sequence.
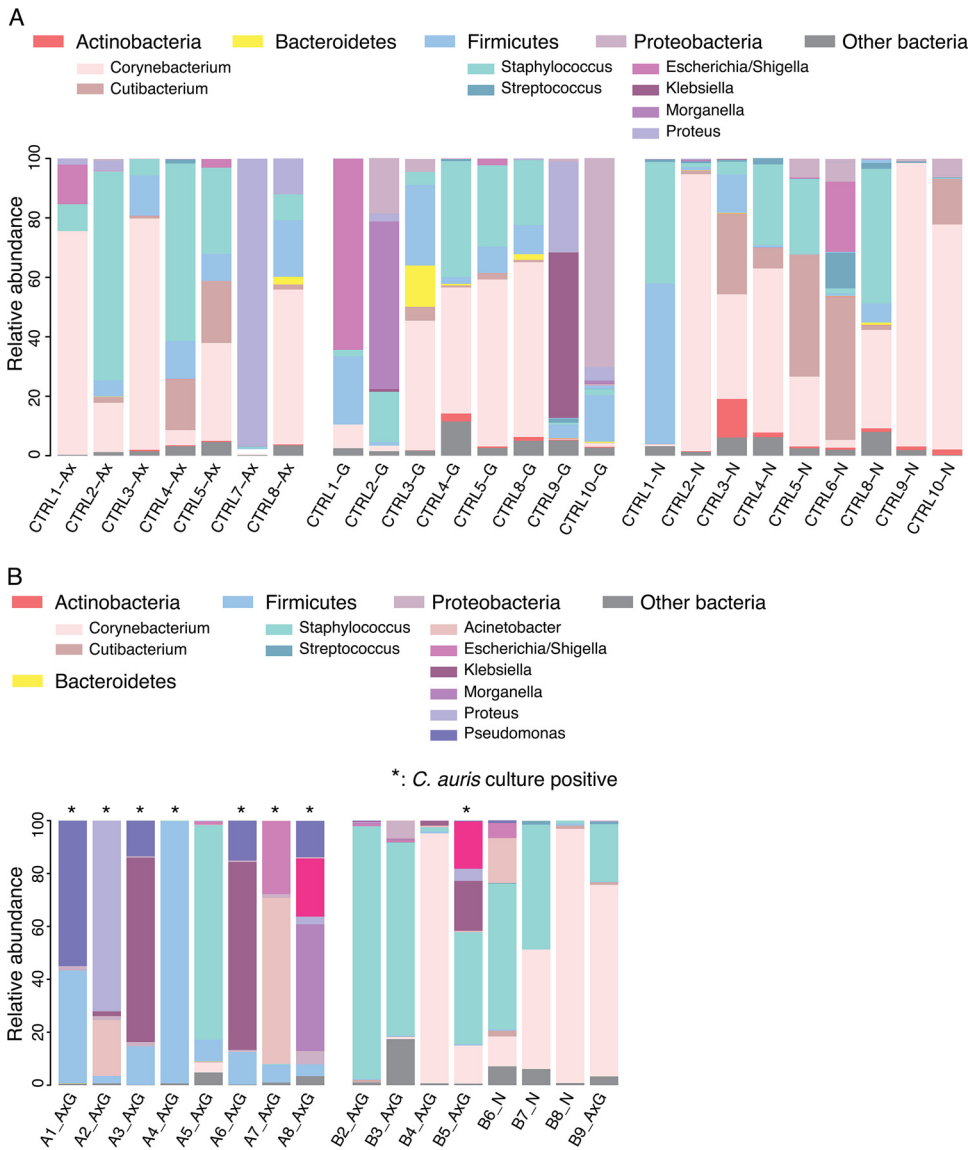[b]nd, none detected.

FIG 2 Matching bacterial community profiles for clinical samples obtained for *C. auris* surveillance. (A) Relative bacterial abundance plot for samples from the control region without *C. auris*. Ax, axilla; G, groin; N, anterior nares. CTRL1–Ax is a sample from the axilla from control subject 1. (B) Relative bacterial abundance plot for samples from facilities in regions A and B with *C. auris* outbreaks. Asterisks indicate *C. auris* culture-positive samples. AxG, axilla/groin; N, anterior nares. A1_AxG is a sample from the axilla/groin from subject 1 in region A.

in these health care facilities plays a large role in assessing and controlling outbreaks. We have demonstrated the utility of next generation sequencing technologies, including amplicon and metagenomic sequencing, to analyze surveillance samples to reveal the microbial composition of these samples, supplement culture results, and elucidate potential adaptation of this pathogen to the human host.

Skin fungal communities of healthy young adults are dominated by *Malassezia* (22), with little if any *Candida* spp. However, patients in long-term post-acute care hospitals and skilled nursing facilities tend to harbor more *Candida* spp., especially in facilities with *C. auris* outbreaks (e.g., region A), raising the possibilities that cocolonization by multiple *Candida* species may be an underlying risk factor for *C. auris* colonization or that a skin microbiome disruption leading to *Candida* dominance may precede *C. auris* colonization. A larger study of surveillance samples from facilities with *C. auris* in the NY region also cultured multiple *Candida* species from skin, including *C. albicans*, *C. glabrata*, *C. tropicalis*,

and *C. parapsilosis* (19). Skin bacterial communities of healthy adults are dominated by *Corynebacterium*, *Propionibacterium*, *Staphylococcus*, and *Streptococcus* (23), with very little if any *Proteobacteria*. However, patients in these ventilated skilled nursing facilities tend to harbor more *Proteobacteria* species that are potential pathogens associated with nosocomial transmission, especially in the facility with more severe *C. auris* outbreaks (e.g., region A).

A limitation of this study is that samples provided for sequencing were unlinked to individual clinical metadata. As such, we are unable to generate or test hypotheses linking antimicrobial receipt, underlying disease state, or other factors to *C. auris* burden. Exposure to antimicrobials was associated with an increased odds of *C. auris* colonization in a recent study of long-term care patients (32).

Amplicon sequencing provides cross validation for culturing and insight into the skin fungal and bacterial community profiles that cannot be easily generated by culturing. Amplicon sequencing may be more comprehensive in detecting the full constituency of the fungal community rather than pursuing multiple culture conditions (Table 1). We were able to classify human skin-associated ITS1 reads to the genus level for most fungi, to the species level for most *Candida* reads, and even to the clade level for *C. auris*. The potential clade V *C. auris*, identified from Iran (33), has an ITS1 sequence that is distinguishable from other clades, upholding the potential for ITS1 sequencing to classify *C. auris* reads to the clade level.

Metagenomic sequencing opens up another level of possibility for integrating surveillance data with host microbial profiles and functional relevance of commensal microbes. We are also aware of the challenges posed by the low biomass of skin microbiome samples, so recovery of sufficient number of fungal reads may pose future challenges. From the clinical sample with sufficient *C. auris* bioburden, 49 high-coverage sequence variants were identified. Most variants (42/49) were predicted to be in noncoding regions of the genome. The remaining seven variants were in genes encoding hypothetical proteins, highlighting the need for improved fungal annotation. Future larger scale clinical studies are warranted to more systematically investigate the effects of commensal microbes and patient risk factors on the colonization and transmission of this global public health threat.

MATERIALS AND METHODS

Point prevalence survey of the health care facilities. The ventilator-capable skilled nursing facilities in regions A and B had already reported positive cases of *Candida auris* when the surveys were performed. Samples used in this study were collected as part of a point prevalence study. Sharing leftover clinical samples with NIH colleagues for protocol development and microbial analysis was reviewed and approved by the Centers for Disease Control and Prevention Institutional Review Board. Samples were collected in BD ESwabs (BD Biosciences) in Amies medium, and an aliquot was stored at −80°C. Body sites sampled include combined axilla and inguinal crease (groin) and anterior nares.

Nursing homes in a region with no reported cases of *C. auris* served as a control. Swabs of axilla, inguinal crease (groin), and nares were collected for skin microbiome sequencing by a standard method (23), and samples were stored at −80°C before processing.

DNA extraction, library preparation, and amplicon and metagenomic sequencing. (i) DNA extraction. For samples collected with BD ESwabs, 100 μl of each specimen in Amies medium was transferred to a 2.0-ml tube containing 1-mm glass beads (MPBio Matrix C), 300 μl yeast cell lysis buffer (Lucigen), and 10,000 (10K) units of Readylyse (Lucigen). For the samples collected directly in yeast cell lysis buffer by the standard method, 1-mm glass beads and 10K units of Readylyse were added. After incubation at 37°C for 1 h, samples were mechanically disrupted using a Tissuelyser (Qiagen) for 2 min at 30 Hz, cooled for 2 min, and disrupted for an additional 2 min at 30 Hz. Samples were then incubated for 30 min at 65°C for complete lysis and MasterPure Complete reagent (Lucigen) was added. The resulting supernatant was processed using the Purelink Genomic DNA kit (Life Technologies). DNA was eluted in DNA-free PCR water (Qiagen).

(ii) Library preparation and amplicon sequencing. The ITS1 region was amplified with flanking primers 18S-F and 5.8S-1R using a dual indexing strategy (22, 34, 35). Briefly, composite primers consisting of an adapter, index, heterogeneity spacer and primer were synthesized: forward primer, 5′-CAAGCAGAAGAC GGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT<index><spacer>GTAAAAGTCGTAACAA GGTTTC; reverse primer, 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC T<index><spacer>GTTCAAAGAYTCGATGATTCAC. Index and spacer sequences, as well as the dual-index plate layouts are described in additional files 1 and 2 of Fadrosh et al. (34). The following PCR conditions were used: 2.5 μl of 10× PCR buffer, 4 μl deoxynucleoside triphosphate (dNTP) mix, 0.25 μl TaKaRa LA *Taq*

polymerase (Clontech), 1 μl 18S-F (10 μM), 1 μl 5.8S-1R (10 μM), 13.75 μl PCR water (MoBio), and 2.5 μl DNA. Reactions were performed in duplicates for 30 cycles, combined, purified using Agencourt AmpureXP (Beckman Coulter), and quantified using the Promega Glomax (Promega). Equivalent amounts of amplicons were pooled together, purified using a MinElute PCR purification kit (Qiagen), and sequenced on the Illumina MiSeq platform with paired-end 300-bp read length.

The 16S rDNA V1-3 region was amplified with universal primers 27F and 534R (22, 35). These were synthesized as composite primers: forward primer, 5′-CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCT<index><spacer>AGAGTTTGATCCTGGCTCAG; reverse primer, 5′-AATGATA CGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT<index><spacer>ATTACCGCGGC-TGCTGG. The same PCR conditions described above for ITS1 amplification were used.

Custom fungal ITS1 reference database update. All fungal ITS1 sequences from ITSoneDB (version 1.131) were downloaded, with both ENA (European Nucleotide Archive) and HMM (hidden Markov model) annotations (28). Additional *C. auris* sequences were added to our curated ITS1 database (22), to include all four clades of *C. auris*. The curated ITS1 database was subjected to a search against the ITSoneDB sequences with the basic local alignment tool (BLAST), and the entries that had matches (alignment length of ≥50 bp and percent identity of ≥70%) to both the ENA and HMM annotations of the same record were kept ($n = 20,168$). The rationale was that, if our curated ITS1 sequence matches both the ENA and HMM annotations of the same ITSoneDB record, this sequence is most likely an authentic ITS1 sequence. ITS1 entries that had a match (alignment length of ≥50 bp and percent identity of ≥70%) to an ITSoneDB ENA annotation with the same genus name were also kept ($n = 337$), taking into account teleomorphic and anamorphic names of the same genus. A total of 20,482 out of 23,458 sequences (87%) from the curated Findley et al. ITS1 database (22) were kept. Most of the dropped sequences represented the ITS2 region. In addition, ITS1 sequences of >1 kb were removed and replaced with the longer sequence of ITSoneDB ENA and HMM annotations of ≤1 kb. The remaining ITSoneDB sequences ($n = 35,188$) were clustered with CD-HIT (36) at 95% identity, resulting in 5,547 sequences, which were added to the final database. Taxonomy strings from the NCBI GenBank annotations of the ITSoneDB entries were parsed to match the laboratory ITS1 sequence format. Finally, laboratory *Malassezia* sequences from several species derived from Sanger sequencing were clustered at 95% identity and added to the final database, which included 25,312 sequences.

***Candida* species database and clade level variation for *C. auris*.** Sequences containing complete ITS1 regions from 14 common and clinically relevant *Candida* species were downloaded from NCBI: *Candida albicans*, *Candida auris*, *Candida dubliniensis*, *Candida glabrata*, *Candida haemulonii* (*Candida haemulonis*), *Candida krusei* (*Pichia kudriavzevii/Issatchenkia orientalis*), *Candida lusitaniae* (*Clavispora lusitaniae*), *Candida metapsilosis*, *Candida orthopsilosis*, *Candida parapsilosis*, *Candida pseudohaemulonii*, *Candida rugosa*, *Candida sake*, and *Candida tropicalis*. *Candida duobushaemulonii* has the same ITS1 sequences as *Candida haemulonii* and is therefore not included as an independent species in the database. Strict ITS1 sequences were derived from the ENA and HMM annotations of the 14 species in ITSoneDB. NCBI sequences were subjected to a search with BLAST against the strict ITS1 sequences, and flanking sequences were trimmed. The strict ITS1 sequences were clustered with CD-HIT at 100% and combined into the final *Candida* species database. To validate the *Candida* species database, first reference sequences in the database were mapped to the database itself. Then, additional testing was performed by introducing random mutations throughout the ITS1 sequences (1% error rate) to simulate sequencing errors and mapped against the original database. In both scenarios, classification accuracy was 100%. A fungal mock community (FMC, courtesy of G. Giannoukos, Broad Institute) was also used to verify the classification accuracy of the *Candida* species database. The three *Candida* species (*C. albicans*, *C. tropicalis*, and *C. lusitaniae*) in the FMC were correctly identified using the database across 13 FMC samples. SNVs within the ITS1 sequences were used to distinguish *C. auris* clades I/III, II, and IV (Fig. 1A). See "Data availability" below for access to database files and scripts.

Metagenomic sequencing. Libraries were generated from genomic DNA with Nextera XT library preparation kit (Illumina) per the manufacturer's recommendations. Sequencing was performed as previously described (23). Paired-end 151-bp reads with a target of 15 million to 50 million clusters per sample were sequenced on an Illumina NovaSeq 6000 instrument.

Amplicon and metagenomic sequencing data analysis. ITS1 MiSeq sequences were demultiplexed, trimmed of barcodes and primers, trimmed to 200 bp, and subsampled to 5,000 reads per sample. Samples without at least 1,000 reads were not included in the analysis. Reads were first subjected to a search using BLAST against the updated fungal ITS1 database described above (E-value cutoff, 1e−6), and reads classified in the *Candida* genus were further subjected to a BLAST search against the *Candida* species database (evalue cutoff, 1e−6). At the genus level, reads that had a match with at least 60% coverage and 70% identity against a reference were classified. At the species level, reads that had a match with at least 50 bp and 95% identity against a reference were classified. At the *C. auris* clade level, reads that had a match with at least 70 bp and 95% identity were classified.

16S V1-3 MiSeq sequences were processed using the mothur pipeline (37), version 1.39.1, as previously described (38, 39). Briefly, 16S sequences were preprocessed to remove primers and barcodes, aligned to the Silva bacteria rRNA database (version 128) and subsampled to 5,000 sequences per sample. Samples without at least 5,000 reads were not included in the analysis. Sequences were preclustered, and chimeras from PCR artifacts were identified and removed using VSEARCH in mothur (40). Remaining sequences were classified to the genus level using RDP training set 16.

For the metagenomic sequencing data, adapters were trimmed using cutadapt (https://github.com/marcelm/cutadapt), low-quality reads were removed using prinseq-lite (-lc_method entropy -lc_threshold 70 -min_len 50 -min_qual_mean 20 -ns_max_n 5 -min_gc 10 -max_gc 90) (41) and host (human; GRCh38)

reads were removed. Reads were classified using the Clinical Pathoscope pipeline (42) with default parameters. Total fungal reads and *C. auris* reads were summed for each patient sample to obtain the fungal and *C. auris* relative abundance, respectively. SNVs were called against a haploid 12.5 Mb *C. auris* clade IV genome (ACUZR; BioSample accession no. SAMN18141223) using Snippy, version 4.4.1 (with the BWA-MEM algorithm, version 0.7.17, and FreeBayes software, version 1.3.1-dirty; https://github.com/tseemann/snippy). Paired-end reads were used by Snippy with default parameters, except –minfrac 0.01. Gene annotation was transferred from the annotated *Candida auris* strain B11243 (PYGM00000000.1) to ACUZR. Briefly, the B11243 assembled contigs were individually aligned to the ACUZR finished chromosomes, and annotation was transferred using the Import Features function in the sequence alignment window of the GenBank submission software Sequin (version 15.10). Annotation was validated in Sequin and manually inspected and corrected where applicable. A custom Perl script yielded the genetic consequences of the identified variants.

Data availability. Database files (ITS and *Candida*-specific), scripts and a description of their usage can be found at https://github.com/skinmicrobiome/fungal_toolkit_XH. Data are available under NCBI BioProject accession no. PRJNA657014. For metagenomic reads, the human reads were removed (mapped against the GRCh38 human genome build), and only reads that positively map to a microbial genome in the reference genomes (as in reference 39) were retained for analysis and deposited in the SRA: SRX5795136 to SRX5795139.

REFERENCES

1. Centers for Disease Control and Prevention. 2021. Candida auris. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/fungal/candida-auris. Accessed 9 August 2020.

2. 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. 2017. Simultaneous emergence of multidrug-resistant Candida auris on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. Clin Infect Dis 64:134–140. https://doi.org/10.1093/cid/ciw691.

3. Centers for Disease Control and Prevention. 2019. Antibiotic resistance threats in the United States, 2019. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf.

4. Lionakis MS, Hohl TM. 2020. Call to action: how to tackle emerging nosocomial fungal infections. Cell Host Microbe 27:859–862. https://doi.org/10.1016/j.chom.2020.04.011.

5. Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. 2009. Candida auris, a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a Japanese hospital. Microbiol Immunol 53:41–44. https://doi.org/10.1111/j.1348-0421.2008.00083.x.

6. Calvo B, Melo AS, Perozo-Mena A, Hernandez M, Francisco EC, Hagen F, Meis JF, Colombo AL. 2016. First report of Candida auris in America: clinical and microbiological aspects of 18 episodes of candidemia. J Infect 73:369–374. https://doi.org/10.1016/j.jinf.2016.07.008.

7. Chow NA, Gade L, Tsay SV, Forsberg K, Greenko JA, Southwick KL, Barrett PM, Kerins JL, Lockhart SR, Chiller TM, Litvintseva AP, US Candida auris Investigation Team. 2018. Multiple introductions and subsequent transmission of multidrug-resistant Candida auris in the USA: a molecular epidemiological survey. Lancet Infect Dis 18:1377–1384. https://doi.org/10.1016/S1473-3099(18)30597-8.

8. Chowdhary A, Sharma C, Duggal S, Agarwal K, Prakash A, Singh PK, Jain S, Kathuria S, Randhawa HS, Hagen F, Meis JF. 2013. New clonal strain of Candida auris, Delhi, India. Emerg Infect Dis 19:1670–1673. https://doi.org/10.3201/eid1910.130393.

9. Lee WG, Shin JH, Uh Y, Kang MG, Kim SH, Park KH, Jang HC. 2011. First three reported cases of nosocomial fungemia caused by Candida auris. J Clin Microbiol 49:3139–3142. https://doi.org/10.1128/JCM.00319-11.

10. Magobo RE, Corcoran C, Seetharam S, Govender NP. 2014. Candida auris-associated candidemia, South Africa. Emerg Infect Dis 20:1250–1251. https://doi.org/10.3201/eid2007.131765.

11. Ostrowsky B, Greenko J, Adams E, Quinn M, O'Brien B, Chaturvedi V, Berkow E, Vallabhaneni S, Forsberg K, Chaturvedi S, Lutterloh E, Blog D, C. auris Investigation Work Group. 2020. Candida auris isolates resistant to three classes of antifungal medications - New York, 2019. MMWR Morb Mortal Wkly Rep 69:6–9. https://doi.org/10.15585/mmwr.mm6901a2.

12. Centers for Disease Control and Prevention. 2016. Clinical alert to U.S. healthcare facilities - June 2016. Centers for Disease Control and Prevention, Atlanta, GA. https://www.cdc.gov/fungal/candida-auris/candida-auris-alert.html. Accessed 8 July 2021.

13. Public Health England. 2017. Candida auris identified in England. Public Health England, London, United Kingdom. https://www.gov.uk/government/publications/candida-auris-emergence-in-england/candida-auris-identified-in-england. Accessed 8 July 2021.

14. Vallabhaneni S, Kallen A, Tsay S, Chow N, Welsh R, Kerins J, Kemble SK, Pacilli M, Black SR, Landon E, Ridgway J, Palmore TN, Zelzany A, Adams EH, Quinn M, Chaturvedi S, Greenko J, Fernandez R, Southwick K, Furuya EY, Calfee DP, Hamula C, Patel G, Barrett P, Lafaro P, Berkow EL, Moulton-Meissner H, Noble-Wang J, Fagan RP, Jackson BR, Lockhart SR, Litvintseva AP, Chiller TM. 2016. Investigation of the first seven reported cases of Candida auris, a globally emerging invasive, multidrug-resistant fungus — United States, May 2013−August 2016. MMWR Morb Mortal Wkly Rep 65:1234–1237. https://doi.org/10.15585/mmwr.mm6544e1.

15. Welsh RM, Bentz ML, Shams A, Houston H, Lyons A, Rose LJ, Litvintseva AP. 2017. Survival, persistence, and isolation of the emerging multidrug-resistant pathogenic yeast Candida auris on a plastic health care surface. J Clin Microbiol 55:2996–3005. https://doi.org/10.1128/JCM.00921-17.

16. Piedrahita CT, Cadnum JL, Jencson AL, Shaikh AA, Ghannoum MA, Donskey CJ. 2017. Environmental surfaces in healthcare facilities are a potential source for transmission of Candida auris and other Candida species. Infect Control Hosp Epidemiol 38:1107–1109. https://doi.org/10.1017/ice.2017.127.

17. Adams E, Quinn M, Tsay S, Poirot E, Chaturvedi S, Southwick K, Greenko J, Fernandez R, Kallen A, Vallabhaneni S, Haley V, Hutton B, Blog D, Lutterloh E, Zucker H, Candida auris Investigation Workgroup. 2018. Candida auris in healthcare facilities, New York, USA, 2013–2017. Emerg Infect Dis 24:1816–1824. https://doi.org/10.3201/eid2410.180649.

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

19. Zhu Y, O'Brien B, Leach L, Clarke A, Bates M, Adams E, Ostrowsky B, Quinn M, Dufort E, Southwick K, Erazo R, Haley VB, Bucher C, Chaturvedi V, Limberger RJ, Blog D, Lutterloh E, Chaturvedi S. 2020. Laboratory analysis of an outbreak of *Candida auris* in New York from 2016 to 2018: impact and lessons learned. J Clin Microbiol 58:e01503-19. https://doi.org/10.1128/JCM.01503-19.

20. Snitkin ES, Won S, Pirani A, Lapp Z, Weinstein RA, Lolans K, Hayden MK. 2017. Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak. Sci Transl Med 9:eaan0093. https://doi.org/10.1126/scitranslmed.aan0093.

21. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA. 2009. Topographical and temporal diversity of the human skin microbiome. Science 324:1190–1192. https://doi.org/10.1126/science.1171700.

22. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, NIH Intramural Sequencing Center Comparative Sequencing Program, Kong HH, Segre JA. 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498:367–370. https://doi.org/10.1038/nature12171.

23. Oh J, Byrd AL, Deming C, Conlan S, NISC Comparative Sequencing Program, Kong HH, Segre JA. 2014. Biogeography and individuality shape function in the human skin metagenome. Nature 514:59–64. https://doi.org/10.1038/nature13786.

24. Byrd AL, Belkaid Y, Segre JA. 2018. The human skin microbiome. Nat Rev Microbiol 16:143–155. https://doi.org/10.1038/nrmicro.2017.157.

25. Sanford JA, Gallo RL. 2013. Functions of the skin microbiota in health and disease. Semin Immunol 25:370–377. https://doi.org/10.1016/j.smim.2013.09.005.

26. SanMiguel AJ, Meisel JS, Horwinski J, Zheng Q, Grice EA. 2017. Topical antimicrobial treatments can elicit shifts to resident skin bacterial communities and reduce colonization by Staphylococcus aureus competitors. Antimicrob Agents Chemother 61:e00774-17. https://doi.org/10.1128/AAC.00774-17.

27. Conlan S, Kong HH, Segre JA. 2012. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. PLoS One 7:e47075. https://doi.org/10.1371/journal.pone.0047075.

28. Santamaria M, Fosso B, Licciulli F, Balech B, Larini I, Grillo G, De Caro G, Liuni S, Pesole G. 2018. ITSoneDB: a comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences. Nucleic Acids Res 46:D127–D132. https://doi.org/10.1093/nar/gkx855.

29. Proctor DM, Dangana T, Sexton DJ, Fukuda C, Yelin RD, Stanley M, Bell PB, Baskaran S, Deming C, Chen Q, Conlan S, Park M, NISC Comparative Sequencing Program, Welsh RM, Vallabhaneni S, Chiller T, Forsberg K, Black SR, Pacilli M, Kong HH, Lin MY, Schoeny ME, Litvintseva AP, Segre JA, Hayden MK. 2021. Integrated genomic, epidemiologic investigation of *Candida auris* skin colonization in a skilled nursing facility. Nat Med https://doi.org/10.1038/s41591-021-01383-w.

30. Chowdhary A, Tarai B, Singh A, Sharma A. 2020. Multidrug-resistant Candida auris infections in critically ill coronavirus disease patients, India, April-July 2020. Emerg Infect Dis 26:2694–2696. https://doi.org/10.3201/eid2611.203504.

31. Rodriguez JY, Le Pape P, Lopez O, Esquea K, Labiosa AL, Alvarez-Moreno C. 2020. Candida auris: a latent threat to critically ill patients with COVID-19. Clin Infect Dis https://doi.org/10.1093/cid/ciaa1595.

32. Rossow J, Ostrowsky B, Adams E, Greenko J, McDonald R, Vallabhaneni S, Forsberg K, Perez S, Lucas T, Alroy K, Slifka KJ, Walters M, Jackson BR, Quinn M, Chaturvedi S, Blog D, New York Candida auris Investigation Workgroup. 2021. Factors associated with Candida auris colonization and transmission in skilled nursing facilities with ventilator units, New York, 2016–2018. Clin Infect Dis 72:e753–e760. https://doi.org/10.1093/cid/ciaa1462.

33. Chow NA, de Groot T, Badali H, Abastabar M, Chiller TM, Meis JF. 2019. Potential fifth clade of Candida auris, Iran, 2018. Emerg Infect Dis 25:1780–1781. https://doi.org/10.3201/eid2509.190686.

34. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2:6. https://doi.org/10.1186/2049-2618-2-6.

35. Jo JH, Deming C, Kennedy EA, Conlan S, Polley EC, Ng WI, NISC Comparative Sequencing Program, Segre JA, Kong HH. 2016. Diverse human skin fungal communities in children converge in adulthood. J Invest Dermatol 136:2356–2363. https://doi.org/10.1016/j.jid.2016.05.130.

36. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

37. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.

38. Oh J, Freeman AF, NISC Comparative Sequencing Program, Park M, Sokolic R, Candotti F, Holland SM, Segre JA, Kong HH. 2013. The altered landscape of the human skin microbiome in patients with primary immunodeficiencies. Genome Res 23:2103–2114. https://doi.org/10.1101/gr.159467.113.

39. Tirosh O, Conlan S, Deming C, Lee-Lin SQ, Huang X, NISC Comparative Sequencing Program, Su HC, Freeman AF, Segre JA, Kong HH. 2018. Expanded skin virome in DOCK8-deficient patients. Nat Med 24:1815–1821. https://doi.org/10.1038/s41591-018-0211-7.

40. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. https://doi.org/10.7717/peerj.2584.

41. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.

42. Byrd AL, Perez-Rogers JF, Manimaran S, Castro-Nallar E, Toma I, McCaffrey T, Siegel M, Benson G, Crandall KA, Johnson WE. 2014. Clinical PathoScope: rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. BMC Bioinformatics 15:262. https://doi.org/10.1186/1471-2105-15-262.