nature portfolio

Corresponding author(s): Edwin C. Oh

Last updated by author(s): 02/10/2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

qPCR Analysis
- Bio-Rad CFX Opus 96 instrument with manufacturer's software (Bio-Rad Laboratories)

DNA/RNA Extraction Kits
- DNeasy PowerSoil Pro Kit protocol (Qiagen Catalog #47016)
- RNeasy PowerMicrobiome Kit protocol (Qiagen Catalog #26000-50)

Sequencing Platforms
- Illumina NextSeq 1000 platform with NextSeq 1000/2000 P1 flow cell (300 cycles) for amplicon sequencing
- Illumina NextSeq 1000 platform with NextSeq 1000/2000 P2 Reagents (300 Cycles) v3 kit (Catalog #20046813) for RNA sequencing

Mass Spectrometry
- Shimadzu MALDI-8020 mass spectrometer operated in positive linear ion mode
- SHIMADZU MALDI Solutions Data Acquisition software (Shimadzu Corporation)

Quality Control Instruments
- Agilent TapeStation system with High Sensitivity D1000 ScreenTape (Catalog #5067-5584) for library quality verification
- Thermo Fisher Qubit Flex Fluorometer with Qubit 1x dsDNA HS Assay Kit (Catalog #Q33231) for concentration measurements

Library Preparation Kits
• CleanPlex mPCR Mix (Paragon Genomics, Catalog #219011) for targeted amplification
• NEBNext Ultra II DNA Library Prep Kit for Illumina (Catalog #E7645L)
• NEBNext Multiplex Oligos for Illumina (Catalog #E7335L)
• Illumina Stranded mRNA Prep, Ligation Kit (Catalog #20040534) for RNA-seq

Primer Design Software
• PrimalScheme version 1.4.1 for amplicon panel design
• Olivar version 1.1.1 for amplicon panel design (with SADDLE algorithm)
• unikmer version 0.19.2 for species-specific marker identification

Data analysis

Amplicon Sequencing Analysis
• fastp v0.23.4 for adapter removal and quality trimming
• BWA MEM v0.7.17-r1188 for read alignment
• samtools v1.15.1 for BAM file processing
• fgbio TrimPrimers v2.4.0 for primer trimming
• ivar v1.4.3 for variant calling
• SnpEff version 5.2 for variant annotation (using custom GFF file)

Whole Genome Sequencing Analysis
• MycoSNP nextflow pipeline version 1.5
• TheiaEuk fungal genome-specific workflow v2.0.0 for species classification and subclade typing
• MUMmer version 4.0 (nucmer) for repetitive element identification
• Bedtools (makefasta) for masking repetitive elements
• FaQCs version 2.10 for quality control and preprocessing
• BWA MEM version 0.7.17-r1188 for genome alignment
• Samtools version 1.15 for SAM/BAM file processing
• Picard version 2.26.10 (MarkDuplicates and CleanSam) for alignment refinement
• GATK HaplotypeCaller version 4.2.5.0 for variant calling (haploid mode)
• GATK CombineGVCFs for merging variant calls
• SnpEff version 5.2 for variant annotation
• SnpSift version 5.2 for variant filtering and export
• snp-dists for SNP distance matrix calculation
• RAxML-NG v.1.2.2 for maximum likelihood phylogenetic tree construction
• Interactive Tree Of Life (iTOL) for tree visualization and annotation
• Python v3.10.9 data processing

RNA Sequencing Analysis
• nf-core/rnaseq pipeline version 3.13.2 for transcriptome analysis
• Trim Galore! version 0.6.7 for adapter removal
• Cutadapt version 3.4 for quality trimming
• BBTools BBSplit version 38.90 for genomic DNA screening
• SortMeRNA version 4.3.6 for ribosomal RNA screening
• STAR aligner version 2.7.9a for genome alignment
• Salmon version 1.10.1 for transcript quantification
• Samtools version 1.17 for alignment processing
• Picard version 2.26.10 (MarkDuplicates) for duplicate marking
• StringTie version 2.2.1 for transcript assembly and quantification
• Freebayes version 1.3.6 for variant calling from RNA-seq data
• nf-core/differentialabundance pipeline version 1.4.0 for differential expression analysis
• DESeq2 version 1.34.0 for differential expression analysis

Consensus Pathway Analysis (CPA)
• BLAST version 2.5.0 for gene mapping to KEGG and Gene Ontology databases
• RCPA package version 0.2.2 (R package) for FGSEA (Fast Gene Set Enrichment Analysis) consensus pathway analysis and plotting (https://github.com/tinnlab/RCPA)
• KEGG database (release 111.0) for pathway annotations
• Gene Ontology (release 2024-09-08) for gene annotations

MALDI-TOF Data Processing
• SHIMADZU MALDI Solutions software for spectra acquisition and processing (with standard baseline subtraction and peak-picking)

Reference Genomes
• C. auris strain B8441 V2 (GCA_002759435.2) for amplicon design
• C. auris B11205 (GenBank assembly GCA_016772135.1) for Clade I alignment
• C. auris B11211 (GenBank assembly GCA_002775015.1) for Clade III alignment

• C. auris B8441 (version s01-m03-r13, Candida Genome Database) for RNA-seq analysis

Visualization and Data Processing
• OmniGraffle for diagram creation
• Adobe Illustrator for figure preparation

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw sequencing data generated in this study have been deposited in the NCBI BioProject database under the following accession numbers: PRJNA1279182 for whole genome sequencing of wastewater isolates, PRJNA1279255 for wastewater amplicon sequencing, and PRJNA1279245 for RNA sequencing. Clinical isolate data collected by the Nevada State Public Health Laboratory between August 2021 and May 2024 were obtained from the existing NCBI BioProject PRJNA846332.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA - No human participants used in this study.
Reporting on race, ethnicity, or other socially relevant groupings	NA - No human participants used in this study.
Population characteristics	NA - No human participants used in this study.
Recruitment	NA - No human participants used in this study.
Ethics oversight	NA - No human participants used in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study implements wastewater-based epidemiology (WBE) to monitor Candidozyma auris infections and antifungal resistance patterns during the 2021-2024 outbreak in Southern Nevada. The study integrated longitudinal clinical surveillance data from 45 healthcare facilities (July 2021 - August 2024) with intensive wastewater monitoring at three major hospitals and two municipal wastewater treatment plants (August - December 2023). Multiple analytical approaches were employed: (1) qPCR quantification of C. auris DNA in wastewater, (2) targeted amplicon sequencing of 11 resistance-associated genes, (3) whole genome sequencing of 2,977 patient-derived assemblies (of which 2,945 were confirmed as C. auris) and 594 wastewater-derived colonies (557 confirmed as C. auris by WGS, with 443 meeting quality control thresholds for downstream analyses), (4) MALDI-TOF mass spectrometry for species and clade identification, (5) antifungal susceptibility testing across three major drug classes (azoles, echinocandins, and polyenes), and (6) RNA sequencing under multiple stress conditions including thermal stress, antifungal pressure, and antibiotic exposure. The study combined observational surveillance and genomic epidemiology with experimental laboratory components, including culture-based isolation, phenotypic susceptibility testing, and transcriptomic profiling, to provide an integrated framework for tracking fungal pathogen evolution and resistance emergence in healthcare settings.

Research sample	• Wastewater: 594 yeast colonies recovered (557 confirmed C. auris, 443 passed QC) • Clinical: 2,977 archived isolates from Nevada BioProject PRJNA846332 (2,945 confirmed C. auris)
Sampling strategy	Wastewater sampling: Weekly collection from three hospitals selected based on secure access to on-site sewer lines and their receiving municipal wastewater treatment plants (WWTP1 and WWTP4A) between August and December 2023. Hospital-specific sampling points were established upstream within each facility's sewer network to capture flows before mixing with municipal sewage. At WWTPs, grab samples of influent were collected. Approximately 100 mL of wastewater was processed per sample. Sample size determination was based on achieving sufficient pathogen detection sensitivity while maintaining feasible collection and processing workflows. Clinical isolates: All clinical isolates collected in Nevada are archived in the Nevada BioProject accession: PRJNA846332). Sequencing data from these archived clinical isolates were used for comparative genomic analyses with wastewater-derived isolates. No formal sample size calculation was performed as this was an observational surveillance study capturing all available cases during the outbreak period rather than a hypothesis-testing study requiring predetermined statistical power.
Data collection	Wastewater samples were collected by trained research personnel and processed by laboratory staff at the University of Nevada, Las Vegas. Sample metadata (date, time, location, flow conditions) were recorded at each collection point using standardized field data sheets. How data were collected: Wastewater collection - Composite samples were placed on ice in the field and stored under refrigeration until processing (hold time <36 hours). Approximately 100 mL was pelleted by centrifugation at 2,934 × g for 10 min at 4°C. DNA extraction used DNeasy PowerSoil Pro Kit. C. auris DNA quantification via qPCR (ITS2 gene, CDC assay) on Bio-Rad CFX Opus 96. Culture-based isolation involved 24-48 h pre-enrichment before colony recovery. All processing followed biosafety protocols for infectious materials. Clinical genomic data: Whole genome sequencing data for clinical isolates were obtained from the Nevada BioProject (PRJNA846332), which archives all C. auris isolates collected by the Nevada State Public Health Laboratory.
Timing and spatial scale	Sampling start/stop dates and frequency: WWTP samples are collected weekly, except for WWTP1 until August 22, 2023, varying days of the week per specific WWTP. Hospitals 1 and 2 are collected twice a week, starting in July 2023 for Hospital 1 and early September 2023 for Hospital 2. Weekly Hospital 3 sampling starts in mid-September 2023. Weekly wastewater sampling for this study was stopped in December 2023. Continuous clinical surveillance data span August 2021 to August 2024. Rationale for sampling frequency: Weekly collection provided sufficient temporal resolution to detect rapid changes in pathogen populations and resistance patterns while maintaining feasible workflows. This frequency aligns with established WBE protocols for pathogen surveillance. Gap between collection periods: Hospital 2 had a gap in collection of about 4 weeks from early October 2023 to early November 2023 due limitations in sampling site access. All Hospital samples for the weeks of Dec 04, 2023 and December 25, 2023 were not collected due to campus closure and holidays, respectively. Spatial scale: The study encompassed the greater Las Vegas metropolitan area in Southern Nevada. Three major hospitals were monitored (Hospital 1: 713 cases, Hospital 2: 89 cases, Hospital 3: 227 cases through February 22, 2024). Hospitals 1 and 2 discharge to WWTP1; Hospital 3 discharges to WWTP4A. WWTPs also receive residential wastewater. Facility-specific hospital sampling enabled comparison with diluted community-scale WWTP signals.
Data exclusions	Data were excluded based on pre-established quality control criteria. Exclusion criteria were pre-established: Yes. All QC thresholds were defined before data analysis. Whole genome sequencing exclusions: Samples failing to meet ≥70% average coverage across C. auris B8441V2 scaffolds 1-10 and 12-14 at ≥30× depth were excluded. Of 2,977 clinical colonies, 32 non-C. auris yeasts were excluded; 2,945 C. auris retained. Of 594 wastewater colonies, 37 non-C. auris yeasts were excluded; 557 C. auris identified, with 443 meeting coverage thresholds. Variant filtering excluded calls with: quality by depth ≤2.0, Fisher strand bias ≥60.0, mapping quality ≤40.0, read depth ≤30, genotype quality <50, or alternative allele frequency <80%. Amplicon sequencing exclusions: Samples with <50% positions covered at <100× depth or read depth <100 per SNP were excluded. Duplicate removal: Identical same-facility same-day wastewater genomes (0 SNP distance) were removed to avoid clonal over-representation, reducing 443 assemblies to 105 unique genomes. Rationale: All exclusions were objective quality metrics to ensure data reliability. No data excluded based on outlier status or unexpected results.
Reproducibility	All bioinformatics workflows, analysis scripts, and source code are publicly available at https://github.com/M-Moshi/UNLV-NPM-Candida-auris. The repository includes complete documentation, configuration files, software environment specifications, amplicon primer sequences, and reference files. All raw sequencing data have been deposited in NCBI BioProject databases: wastewater WGS (PRJNA1279182), wastewater amplicon sequencing (PRJNA1279255), RNA sequencing (PRJNA1279245), and clinical isolates (PRJNA846332). The Methods section provides detailed protocols with specific reagent catalog numbers, instrument settings, and quality control parameters to enable replication. Standardized analytical pipelines were used: MycoSNP version 1.5 and TheiaEuk for whole genome analysis, nf-core/rnaseq for transcriptomics, following community best practices. Quality control thresholds, variant filtering criteria, and statistical parameters are explicitly stated. For MALDI-TOF analysis, instrument settings (positive linear ion mode, m/z 2,000-20,000, laser power 80-100) and sample preparation protocols are described in detail. Antifungal susceptibility testing followed standardized gradient diffusion methods with appropriate quality control strains (C. parapsilosis ATCC 22019, C. krusei ATCC 6258, C. albicans ATCC 90028, and C. auris reference strains ATCC MYA-5002, CDC_CAU-09, CDC_CAU-27). All experiments included technical replicates where applicable (e.g., triplicate MALDI-TOF spots, biological triplicate MIC tests).
Randomization	Randomization was not relevant to this study. This was an observational surveillance study without experimental group allocation or treatment interventions.
Blinding	Blinding was not relevant to this study. As an observational epidemiological investigation focused on pathogen detection and genomic characterization, there were no treatment groups or interventions requiring blinding.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	NA - No plant material used for this study.
Novel plant genotypes	NA - No plant material used for this study.
Authentication	NA - No plant material used for this study.