DMP SOFRAS DIMITRIOS

Title: Unravelling the multidrug resistance, drug tolerance, and collateral sensitivity in the emergent nosocomial fungus *C. auris* using experimental evolution and multi-omics analysis.

Principal researcher: Dimitrios Sofras

Affiliation: KU Leuven, Laboratory of Molecular Cell Biology

Template: KU Leuven DMP

Project administrator: Patrick Van Dijck

Data manager: Nico Vangoethem

Project abstract:
Antifungal resistance is rising on an unprecedented scale. Invasive, multidrug-resistant (MDR) fungal infections are threatening public health and are further complicated by the availability of only three major antifungal drug classes. The novel species *C. auris* embodies this antifungal resistance crisis. Since its first identification in 2009, this opportunistic pathogen has emerged worldwide in various hospital outbreaks, and it has shown resistance to an extent that has not been observed in other fungi. Despite being the first fungus to be considered an urgent threat by the CDC, an understanding of the molecular mechanisms underlying its MDR is still missing. Traditionally, antifungal resistance has been studied by wholegenome sequencing of clinical isolates. This is however challenging and genome association studies of *C. auris* are not able to resolve the extensive MDR seen in the clinic. In this project, we will overcome these limitations by performing in vitro and in vivo experimental evolution followed by multi-omics analysis and recently optimized gene-editing tools. Additionally, we will assess the role of tolerance in the development of drug resistance, and we aim to validate whether collateral sensitivity exists in *C. auris*. Eventually, this project will result in the elucidation of the molecular mechanisms which govern (multi) drug resistance in *C. auris* and possibly provide indications towards new treatment strategies and druggable targets.

ID: 99380

Start date: 01-11-2021

End date: 31-10-2025

Grant number / URL: 11J8122N

DATA COLLECTION

What data will you collect or create? Fill out the table below and/or describe.

Type of data	Format	Volume	How created?
e.g. Observations, questionaires, DNA samples	e.g. Excel, JPeg, ..	e.g. 200MB, 1GB	Computer task, observations, blood sample, ...
Digital images	.jpeg, .tif, .ai, .pdf	max. 50GB	Gel images, microscopy images, plate images illustrations, figures
MIC measurements	.xlsx	max. 10GB	Synergy H1 instrument, absorbance measurements
Growth curves	.xlsx	max. 10GB	Multiskan instrument, absorbance measurements
Growth and MIC analysis data	.pzfx	max. 10GB	Analysis of the raw .xlsx data in Graphpad Prism
Sequences	.dna	max. 10GB	Sanger sequencing data, sequence alignments, cassette construction, plasmid sequences and plasmid construction data in Snapgene software
WGS and RNAseq data	.gz	max. 80GB (from max. 150 strains)	Compressed raw data as provided by IDT sequencing services
Results of WGS and RNAseq	.xlsx	max. 10GB	Analysis through SNP discovery pipelines and R scripts, final storage in excel files
Proteomics data	.raw, .mzML	max.50GB	Raw proteomics data as provided by the mass spectrometer

Physical data:
Plasmids will be kept in a stock freezer at -20C.
Bacterial strains containing the plasmids will be kept in glycerol (25% v/v) in a freezer at -80C.
C. auris strains generated by directed microevolution and by genetic manipulation will be kept in glycerol (25% v/v) in a freezer at -80C.

Do you intend to reuse existing data?

Yes. Strains generated by other researchers, plasmid sequences utilised by other groups and by members of the lab will be used to generate mutant strains. These data are free of use, as long as the corresponding authors of the publications agree to share them with us.

DATA QUALITY, DOCUMENTATION AND METADATA

Describe the documentation that will be created for the data.

1. Digital images (gel scans and plate scans): for this type of data the format YYYYMMDD_experimentname_WP* is used as the file name and the storage is done at the appropriate directories, first on the personal server and subsequently they are transferred to the shared server.

2. Plasmid sequences: for this type of data the name of the plasmid in its original publication is used as a file name, and in the case of a novel plasmid, the accession number it takes from the responsible lab technicians becomes the file name.

3. Mass collection of data: for data collected by instruments in a batch mode (i.e. data from Multiskan and from Synergy H1) are named as YYYYMMDD_experiment_name and the specifics of each experiment can be found as metadata after the analysis of the raw data is completed.

4. Data that are generated by companies, such as WGS, RNAseq and proteomics data will be named according to their standard procedures and they will include the strain number in the file name.

5. Physical strains that come out of evolution experiments will be named as following: [letter][number], where letter will be assigned to different parental strains and number will be the day when the sample was collected and stored. Mutant strains generated in this project will be named by the lab technicians following the accession numbers of the lab's database. When referred to in publications, mutants will be given easily readable names, such as *mutated_gene* as is already being done in literature and the linkage between all the names will be done by the IT responsible or the lab technicians in the Filemaker database of the lab where there is a repository of all plasmids, strains, primers and chemicals used in the lab.

Describe the metadata for the data.

All data captured by measurements of a physicochemical property in a batch mode will be manually curated to create meaningful metadata. The processing of the raw data is carried out in Graphpad prism and by creating graphs the data become meaningful to others.
All omics data are processed in pipelines for SNP discovery, differential expression and differential protein content and graphs are created in R studio for the meaningful representation of the data. Optionally, open-source software can be used to make meaningful representations of e.g. gene ontology metadata, such as Cytoscape in which the principal researcher has been previously trained.

How will the data quality be guaranteed?

The quality of the raw data is the responsibility of the principal researcher. The quality of the processed data and metadata generated by appropriate analysis is reviewed by the project administrator, as well by other members of the lab during biannual presentations. The metadata are subsequently undergoing further quality control when they are peer reviewed after being submitted for publication.

ETHICAL, LEGAL AND PRIVACY ISSUES

Are the collected data considered to be "personal data" and are all the requirements about the collection of "personal data" met?

No

Are there any ethical issues concerning creating, sharing and use of the data?

Yes. However, the application to the ethical committee for using live murine models has not been submitted yet.

Did you consider all issues about copyrights and IPR?

Yes

DATA STORAGE AND BACKUP DURING RESEARCH

How and where will the data be stored during research?

The data generated are stored in a directory at the laboratory's server which is shared by all the collaborators including other students working on parts of the project. The shared directory is divided in different thematics according to the WP and the goals of each subdivision of the project.

Which back-up procedures are in place?

The shared directory mentioned above is routinely backed-up to the large volume storage server of the laboratory where only the data manager has write access to prevent accidental loss of the data.

Describe the data security procedures and who has access to the data?

Lab protocols, oligonucleotide sequences and Sanger sequences are available to all members of the lab via the server, to which only lab technicians have write access. The data generated by experiments of the project are only shared with close collaborators, i.e., other PhD candidates working on similar projects and the master students of the PI for the entire duration of their stay in the laboratory. KU Leuven encryption tools are used to grant access to the shared directories and to protect against malicious actions.

DATA STORAGE AND PRESERVATION AFTER RESEARCH

Which data will have long time value for the research and will be preserved?

Digital data:
All sequencing data, as well as omics data that are used in publications will be preserved for future reference and for possible follow up projects, for at least 5 years after the project is completed. Plasmid sequences constructed for this project will also be preserved indefinitely.
Physical data:
All strains that can be possibly inquired from other researchers as well as plasmids, will be kept for at least 5 years after the completion of the project. These include strains that are evolved based on the first WP of the project, and that carry important/interesting mutations and/or showcase new phenomena as is the goal of the project. Moreover, mutant strains generated by genetic engineering will also be included in the preserved strains.

Where and how will the data be stored?

Digital data are stored in different servers provided by the laboratory. The servers are namely a personal, a shared and a large storage server. The costs for preparing and storing the data are already included in the budget of the laboratory.
Physical data are stored in freezers dedicated for that purpose. These freezers are located at the premises of the laboratory and are categorised in -20C and -80C. Temporary storage of physical data can be done at 4C fridges and at the personal space of the primary researcher of a -20C freezer.

DATA SHARING

Are there any restrictions for sharing the data?

There are no restrictions for sharing the data of the project. Upon publication of the results, the data can be freely available to others.

How will the data be shared?

The collaborators of the project utilise the shared directories of the laboratory's server to exchange data. Upon request after publication, the data can be shared with data transfer tools via e-mail.

Will the data be made available on request?

- Yes

Upon publication of the results, all data generated for the project will be available upon request.

RESPONSABILITIES AND RESOURCES

Who is responsible for Data Management during the project?

The principal researcher (Dimitrios Sofras) and the IT responsible of the lab (Nico Vangoethem) are responsible for Data Management during the project.
The principal researcher is responsible for storing all data generated in the shared directories of the shared volume storage and to keep all newly generated data on his personal storage space on the laboratory's server until he transfers it to the aforementioned shared server.
The IT responsible of the lab will make sure that all shared folders are backed up into a large volume storage server at least once a month.

Which additional resources are needed for the execution of the Data Management Plan?

No additional specialist expertise is required, and the software provided by KU Leuven and by the lab is sufficient for the completion of the project.
Additional training of the principal researcher in data analysis and illustration techniques will be covered by following appropriate courses offered by KU Leuven as mentioned in his doctoral school blueprint.

Did you read the KU Leuven Data Management Policy? Link in guidance.

- Yes