Comparative analysis of resistance-fitness evolution in multidrug resistant Candida species: a role for fitness trade-off compensation and hypermutators?

A Data Management Plan created using DMPonline.be

Creator: Stef Jacobs

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Stef Jacobs

Grant number / **URL:** 11PRR24N

ID: 206196

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Invasive, multidrug-resistant (MDR) fungal infections pose a growing threat to public health, yet the evolutionary dynamics and mechanisms of resistance in fungi have been largely neglected. Nearly not investigated so far are fitness-resistance trade-offs, fitness trade-off compensatory mechanisms and drug/species specific mutation rates. In this project, we want to conduct a comparative study of antifungal drug resistance evolution in four clinically important Candida species, covering all major clades of species with MDR potential: C. auris, C. lusitaniae, C. glabrata and C. krusei. We will map the distribution of fitness-resistance associations, basal and drug/species specific mutation rates and the role these play in resistance evolution. The host lab has optimized and used a high throughput pipeline for experimental evolution and fitness-resistance testing, followed by genomic analysis and validation of mutations of interest via a novel CRISPR-cas9 gene editing system in C. glabrata and C. auris. We will use and further extend this pipeline with C. lusitaniae and C. krusei, Distribution of Fitness Effects (DFE) quantification, mutation rate analyses and transcriptomic analyses, to perform the first broad scale comparative dissection of antifungal drug resistance evolution in pathogenic fungi. Based on our analysis we will know the different resistance and fitness compensation mechanisms for the eight drugs that we will use in our assays for all four species considered.

Last modified: 15-04-2024

Comparative analysis of resistance-fitness evolution in multidrug resistant Candida species: a role for fitness trade-off compensation and hypermutators?
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

We will generate a large number of mutant strains through experimental evolution and genetic modification (CRISPR/Cas9). Part of this library of strains (experimentally evolved strains of C. auris and C. glabrata) has already been constructed in an ongoing project of the host lab and will be complemented with newly evolved strains of C. lusitaniae and C. krusei. We will also generate a large amount of sequencing data from whole-genome sequencing and RNAseq analysis (.dna, .gz). Next to these two main datatypes, we will also generate digital images (.jpeg, .tif) and workbooks with data from susceptibility and growth testing (.xlsx).

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

For the storage of all the non-physical data, Frans Panggih Purwoko is responsible (informatician of the lab). During our research, all data will be stored on KU Leuven's 'Luna Network File Storage' which consists of both a personal drive and a shared drive. After the project, the data will be transferred to an archive server for Large Volume Storage. All NGS data will be stored on both the lab servers and the servers of the Institute for Research in Biomedicine (Gabaldón lab). The generated strains will be stored in triplicate (2 in main freezer, 1 in back-up) at -80°C for long-term preservation (at least 10 years, after which regrowing and consecutive storage can be performed).

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Not applicable.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Not applicable.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Not applicable.

Comparative analysis of resistance-fitness evolution in multidrug resistant Candida species: a role for fitness trade-off compensation and hypermutators?

FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Type of data	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
Digital images	Gel images, microscopy images, plate images illustrations, figures.	Generate new data	Digital	Observational, Experimental and Simulation data	.jpeg, .tif, .ai, .pdf	<100GB	
Susceptibility data	MIC measurements - with antifungal drugs or stressors. Made with Synergy H1 instrument, absorbance measurements .	Generate new data	Digital	Experimental	.xlsx	<1GB	
Fitness data	Growth curve measurements - with and without added stressors. Made with Multiskan instrument, absorbance measurements .	Generate new data	Digital	Experimental	.xlsx	<1GB	
Cell counts for relative fitness or mutation rate analyses	For both competition experiments and fluctuation assay, cell count is used as a readout. These might be obtained by colony counting or by flow cytometry.	Generate new data	Digital	Experimental	.xlsx, .tif, .fcs, .pzfx	<10GB	
Growth and MIC analysis data	Analysis of the raw .xlsx data in Graphpad Prism.	Generate new data	Digital	Experimental	.pzfx	<10GB	
Sequences	Sanger sequencing data of the genes of strains of interest. Sequence alignments, cassette construction, plasmid sequences and plasmid construction data in CLC software.	Generate new data	Digital	Experimental	.dna	<100GB	
WGS and RNAseq data	Whole genome sequencing and RNAseq data to detect unknown resistance or fitness compensation mechanisms. Compressed raw data as provided by IDT sequencing services.	Generate new data	Digital	Experimental	.gz	<100GB (from max. 120 strains)	
Results of WGS and RNAseq	Analysis through SNP discovery pipelines and R scripts, final storage in excel files.	Generate new data	Digital	Experimental	.xlsx	<10GB	
Plasmids for the gene editing	These plasmids contain all elements to do the CRISPR allele editing.	Generate new data	Physical	Experimental			<100 vials in -20°C

Bacterial strains containing the plasmids	Bacteria are used to generate the necessary plamids for the gene editing.	Generate new data	Physical	Experimental			<100 vials in glycerol (25% v/v) in a freezer at -80°C
Collection of evolved resistant Candida strains	Obtain resistant isolates for an array of antifungals via experimental evolution. These strains will be used for further fitness and resistance testing.	Generate new data & partially reuse strains of the existing collection of the lab	Physical	Experimental			<1000 vials in glycerol (25% v/v) in a freezer at -80°C
Collection of designed resistant Candida strains	Interesting mutations will be validated by designing strains via a newly developed CRISPR system. These strains will be used to confirm the phenotypes.	Generate new data	Physical	Experimental			<1000 vials in glycerol (25% v/v) in a freezer at -80°C

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

We will reuse already evolved and designed strains of the existing strain collection of the lab. These strains were obtained in the context of these publications and pre-prints: 10.1128/mBio.03333-20, 10.3390/jof7090754, https://doi.org/10.21203/rs.3.rs-3621420/v1, https://doi.org/10.21203/rs.3.rs-3463071/v1.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, animal data

Workpackage 4, Task 4.3: we will use mice for in vivo fitness/virulence testing. Ethical approval will be requested for this project, but requests for similar experiments in Candida species have been granted in the past for our lab. After experiments with different treatment schemes, we will harvest the organs of the mice to determined the colonisation by Candida. The data we will obtain here are cell counts after selective plating of the strains. These will be analysed with Excel and graphs will be made using Graphad Prism.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Prior to any experimental work, thorough preparations are conducted, with every step meticulously documented. Lab-specific standard operating procedures (SOPs) are outlined and securely stored alongside experimental data within designated folders, facilitating seamless retrieval of metadata. We follow this lab guide with standardized protocols implemented in our laboratory. Furthermore, the results will be published, with detailed descriptions of materials and methods to facilitate replication by experienced researchers. All generated data is securely stored on the KU Leuven server, which includes protected project directories accessible only to involved researchers. Each researcher also has a personal directory on the KU Leuven server for safe data storage, along with dedicated directories on OneDrive and Teams. Data is never stored on personal or work devices to mitigate the risk of data loss due to technical failures.
We also have weakly lab meetings to discuss procedures and techniques so that everyone is always up to date.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

All raw datasets are deposited in the designated online directories, alongside processed data and final results (e.g., figures published in papers), within a dedicated directory for raw data files as described above.

3. Data storage & back-up during the research project

Where will the data be stored?

All data is stored on the KU Leuven shared server and additional online directories (e.g. personal directory on KU Leuven server, Teams, OneDrive).

How will the data be backed up?

Standard back-up provided by KU Leuven ICTS.

**Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.**

- Yes

We have large storage capacity accessible on both the I and J drives. Additionally, if needed, we can expand our storage space upon payment.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be securely stored within the university's protected environment.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The anticipated expense is €500 per terabyte per year. We don't anticipate exceeding this cost. We incorporate such expenses into our consumable budget for any grant applications.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

Both digital and physical materials will be retained for a period of 10 years. Physical materials, including evolved and engineered strains, will be preserved in a freezer set to -80°C.

Where will these data be archived (stored and curated for the long-term)?

The data will be archived on the large volume storage and the shared network drive (J-drive).

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Considering the anticipated size of the database (likely less than 1 TB), the initial setup cost is estimated to be below €500, with an annual support fee of €50.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

The physical materials can be requested and will be sent to the other lab. Digital information will also be accessible upon request.

If access is restricted, please specify who will be able to access the data and under what conditions.

Acces is not restricted.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The data will be disseminated through publications and presentations at international conferences. Interested researchers can trace back a strain and request it from the lab based on information provided in the publications. All digital data will be accessible via the Research Data Repository.

When will the data be made available?

Upon publication (or patenting) of the research results.

Which data usage licenses are you going to provide? If none, please explain why.

Data from the project that can be shared will be released under a Creative Commons Attribution license (CC-BY 4.0). This ensures that users must provide credit to the original data creators when using the data.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

All publications will receive a DOI. A PID will be added upon deposit of data in a repository.

What are the expected costs for data sharing? How will these costs be covered?

Sharing data via NCBI incurs no charges. However, the inclusion of data in manuscripts will involve publication costs, which can range upwards of €7000, particularly for papers in the Nature group (currently €10500 for Nature Microbiology). If other researchers request strains, we typically request them to provide a DHL or FedEx number for shipping.
Costs related to a patent application (€10000) will be covered, with the intention of seeking a company to obtain a license for the patent.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Patrick Van Dijck (PI, KU Leuven); Stef Jacobs (PhD Student, KU Leuven); Hans Carolus (Postdoc, KU Leuven)

Who will manage data storage and backup during the research project?

Luc Grauwels (ICT support, KU Leuven)

Who will manage data preservation and sharing?

Patrick Van Dijck (PI, KU Leuven); Hans Carolus (Postdoc, KU Leuven)

Who will update and implement this DMP?

Patrick Van Dijck (PI, KU Leuven)