DMP-Kafetzopoulou-2022-FWO-Junior

Project Name My plan (FWO DMP) - DMP-Kafetzopoulou-2022-FWO-Junior

Grant Title 12X9222N

Principal Investigator / Researcher Liana Kafetzopoulou

Project Data Contact liana.kafetzopoulou@kuleuven.be

Description Lassa fever is an acute viral hemorrhagic illness caused by Lassa virus (LASV) and one of the priority diseases listed in the WHO R&D Blueprint. The high incidence in Nigeria and the annual rise of 200-300 LASV cases since 2017 have highlighted the urgent need for accelerated research and development. To address this, two major collaborative studies in both acute and surviving Lassa Fever individuals were initiated by Bernhard Nocht Institute for Tropical Medicine (BNITM, Germany) and the Irrua Specialist Teaching Hospital (ISTH, Nigeria). The two studies have been running in parallel since 2018 and have set the basis for our improved understanding of LASV viremia and persistence. Through this work a unique dataset of samples has been compiled from both the acute and convalescent phase of LASV infection. These samples are currently located at BNITM and as a long-term collaborator I am offered access to this unique comprehensive dataset. This project will use this unique opportunity to address major knowledge gaps still remaining and not previously addressed in our understanding of LASV viremia and persistence. Next generation sequencing will be used to generate virus genomic data from longitudinal samples harbouring persisting LASV and phylogenetic inference will be employed to study virus evolution during the acute infection and post-discharge. Additionally innovative phylogenetic analysis methodology developed by our team at KUL will be used to investigate viral genomic determinants of virulence and disease outcome. The objectives of this project will broaden our understanding of LASV epidemiology and its evolutionary dynamics and will for the first time allow for a comprehensive analysis of intra-host LASV variation and trait analysis to determine viral genetic factors associated with disease presentation and persistence.

Institution KU Leuven

1. General Information

Name applicant

Liana Kafetzopoulou

FWO Project Number & Title

12X9222N

Lassa virus: unravelling the within-host evolutionary dynamics of acute and persistent infections and the virulence evolution among hosts

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Work package	Type of data	Format	Volume	How created
WP1	metadata	mySQL	approx. 500GB	Data collection and transposition from xls and xlsx files
WP2	datasets consisting of sequencing reads	fastq	100-200 GB per run, estimated 10 MiSeq and 10 NextSeq	Illumina sequencing of Lassa virus from human samples
WP3	sequencing data analysis and result	fastq, csv, vcf, custom made scripts	200-400 GB per run, estimated 10 MiSeq and 10 NextSeq	Sequencing data analysis and intra-host analysis of Lassa virus sequencing reads using custom scripts and in-house developed data analysis pipelines
WP4	Data on phylogenetic inference	csv, xml	10-50 GB per run, estimated 10 MiSeq and 10 NextSeq	Phylogenetic inference study and trait identification of Lassa virus genome sequences

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

E-2021-2773

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

All of the points below apply to all researchers involved in the current project. Documents produced will be versioned and signed for traceability.

Labbook - Documentation and labbook templated have been put in place for all laboratory and bioinformatics procedures. Detailed labbook documentation is kepts for every step performed and all experiments wether lab-based or bioinformatics-based are throgouhly documented. Experimental design, sample selection and research methods are all captured as part of the labbook along with the experiment testing and results.

Protocols - All protocols will be written up to ease use by others. Tested and validated protocols will be written up as standard operating procedures and will be published in open access publications or provided on protocols.io with D.O.I identifiers.

Raw data - All raw data generated will be documented with its resepctive metadata. Files and data will be stored using a specific system for ease of data location and usage. The system will be defined prior to data generation and will be consistently used throughout the whole project

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

Metadata - All information generated, wether this is laboratory measurements or results, will be stored and collected as part of the project metadata. This informaiton will be consistenlty collected in the same way and the metadata layout will be defined prior to the start of experiments.

Database - All data and information generated will be integrated within the database that will be developed as WP1. The information will be easily interrogated and identified for use.

5. Data storage and backup during the FWO project

Where will the data be stored?

Raw data - All raw data generated will be stored along with time and name stamps on two separate locations: our research unit central storage facility and at our collaborators (BNITM) equivalent facility. The original copy of the raw data will not be touched or worked on. Copies for analysis and further investigation will be made on personal devices to protect and not corrupt the raw data.

Protocols - All protocols will be in Open Access publications or provided on Protocols.io with D.O.I identifier.

Data genenrated - All data will be stored as part of our labbook system on the L drive. Files and results genenrated will be documented and stored along with their respective documentation. For viral data we foresee local storage on desktop computers and external disks, and short term on VSC storage (Flemish Supercomputer) until acceptance of publication to further permanently store them on publicly accessible databases (eg.GenBank). Metadata and sequence will be submitted in Genomics Standards Consortium format. XML files and scripts will be permanently made available on Github repository or DataDryad. We will ensure D.O.I. 's use for full citation; public access code and tutorial will be available for BEAST analyses. All metadata and results will also be stored as part of our database being build as part of WP1.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures. We are using the L drive and investigating the possibility of a NAS system for additional backups and ease of data analysis.

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 currently have a minimum of 5TB on the L drive and can easily expandable in blocks of 5TB when required.

What are the expected costs for data storage and back up during the project? How

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

The L drive costs are the expected costs for data storage and backup during the project. These costs have been calculated within the project expenses.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data will be stored in the university's secure environment.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All data will be stored for at least 10 years.

Where will the data be archived (= stored for the longer term)?

The data will be stored on central servers at our institutions for at least 10 years. For long term storage (archiving), the data will be placed on external archiving-drives and kept in an optimal environment. For viral data we foresee local storage on desktop computers and external disks, and short term on VSC storage (Flemish Supercomputer) until acceptance of publication to further permanently store them on publicly accessible databases (eg.GenBank). Metadata and sequence will be submitted in Genomics Standards Consortium format. XML files and scripts will be permanently made available on Github repository or DataDryad. We will ensure D.O.I. 's use for full citation; public access code and tutorial will be available for BEAST analyses. All metadata and results will also be stored as part of our database being build as part of WP1.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Hardware for long term storage will be aquired during the project and will be used for the long term storage along with the space on the L drive on the servers of KU Leuven. Additionally the data will be stored long term also with our collaborators at the BNITM where the required infrsastructure is already in place and no additional costs are foreseen.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

All data will be made available upon publication

Where/how will the data be made available for reuse?

- In an Open Access repository
- Upon request by mail

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

Data will be securely shared with our project collaborator (BNITM). We have joint access to each others' servers, in consultation with respective ICT departments.

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

We do not expect to endure any additional costs for data sharing between collaborators. The costs of high publication processing fees in open access journals have been included in the project budget requested.

8. Responsibilities

Who will be responsible for data documentation & metadata?

I will be responsible together with our lab technician (Ine Boonen)

Who will be responsible for data storage & back up during the project?

Who will be responsible for data storage & back up during the project?
I will be responsible together with our lab technician (Ine Boonen)

Who will be responsible for ensuring data preservation and reuse ?
I will be responsible together with our lab technician (Ine Boonen)

Who bears the end responsibility for updating & implementing this DMP?
I bear the end responsibility of updating & implementing this DMP.