

---

# The Culex mosquito microbiome and its implications on arbovirus transmission

A Data Management Plan created using DMPonline.be

**Creators:** Lander De Coninck  <https://orcid.org/0000-0001-6847-2379>, First Name Surname

**Affiliation:** KU Leuven (KUL)

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

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** First Name Surname

**Grant number / URL:** 11L1323N

**ID:** 194301

**Start date:** 01-11-2022

**End date:** 31-10-2026

## Project abstract:

Viruses transmitted by mosquitoes (arboviruses) have a major impact on global human and animal health. West Nile virus (WNV) and Usutu virus (USUV), two emerging arboviruses in Europe, are mainly transmitted by mosquitoes of the Culex genus. Culex pipiens, the northern house mosquito, is the most prevalent mosquito across the European continent and is regarded as the most important vector for WNV and USUV. Little is known about the microbiome (bacteriome and virome) of Culex mosquitoes and even less is known about how this microbiome influences the ability of these mosquitoes to transmit arboviruses to vertebrates. This PhD project will characterize the Culex mosquito microbiome using field-collected mosquitoes from different regions worldwide. Secondly, in vivo infection experiments will be performed to assess associations between the mosquito microbiome and vector competence. We will attempt to confirm such associations by recolonizing axenic mosquitoes with a selection of relevant microbes and evaluate the vector competence of the recolonized mosquitoes.

**Last modified:** 19-01-2023

# The Culex mosquito microbiome and its implications on arbovirus transmission

## 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)**

This project will generate high quality (Illumina) virome and 16S bacteriome sequencing data that will be deposited in the Sequence Read Archive (SRA) of NCBI. From our analyses, we expect to find a myriad of (near-)complete viral genomes that will also be made available in public databases (eg. Genbank) to aid the research community in unravelling the viral diversity. Our efforts should eventually lead to several published manuscripts in high impact scientific journals.

**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)**

KU Leuven policy is to store relevant research data for at least 10 years after the project has ended.

A data management plan will be constructed online using a dedicated tool: <https://dmponline.kuleuven.be/>.

a. Prof. Jelle Matthijssens will be responsible for the data storage.

b. Data will be preserved during the research on a shared storage server, easy accessible to the researcher and with several backups per day. When the project is finished, the data will move to a large volume storage server and to the newly developed Research Data Repository (RDR) of KU Leuven, where it will be stored according to KU Leuven's policy.

**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)**

NA

**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)**

NA

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

NA

## The Culex mosquito microbiome and its implications on arbovirus transmission

### DPIA

---

#### DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

## The Culex mosquito microbiome and its implications on arbovirus transmission

### GDPR

---

#### GDPR

Have you registered personal data processing activities for this project?

- Not applicable

# The Culex mosquito microbiome and its implications on arbovirus transmission

## 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
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>Generate new data</li> <li>Reuse existing data</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>Digital</li> <li>Physical</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>Observational</li> <li>Experimental</li> <li>Compiled/aggregated data</li> <li>Simulation data</li> <li>Software</li> <li>Other</li> <li>NA</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>.por, .xml, .tab, .cvs, .pdf, .txt, .rtf, .dwg, .gml, ...</li> <li>NA</li> </ul>	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <li>&lt;100MB</li> <li>&lt;1GB</li> <li>&lt;100GB</li> <li>&lt;1TB</li> <li>&lt;5TB</li> <li>&lt;10TB</li> <li>&lt;50TB</li> <li>&gt;50TB</li> <li>NA</li> </ul>	
Worldwide mosquito metagenomics	NGS data from the viromes of Culex mosquito from Brazil, USA, Belgium, Cameroon, Burkina Faso, Cambodia, New Caledonia and Guadeloupe	Generate new data	Digital	Observational	fastq	<10TB	
Bacteriome data infected field mosquitoes	NGS data of the bacteriome from field-captured mosquitoes in Belgium infected with WNV and USUV	Generate new data	Digital	Observational	fastq	<10TB	
virome data infected field mosquitoes	NGS data of the virome from field-captured mosquitoes in Belgium infected with WNV and USUV	Generate new data	Digital	Observational	fastq	<10TB	

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:

Public datasets from SRA (<https://www.ncbi.nlm.nih.gov/sra>) may be used during the project.

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.

- No

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).**

Metadata (collection date, collection site, gender, etc.) on mosquito samples is gathered by the collaborators and assembled in 1 file. Metadata will be shared with the submission of NGS reads on the SRA server.

**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.**

- Yes

At least the minimum required metadata to share NGS reads, viral genomes, etc. on the NCBI servers will be published at submission of the data at these databases.

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

**Where will the data be stored?**

Data that is used during the project will be stored on the servers of the VSC, external hard drives, KU Leuven's OneDrive server and KU Leuven's shared drives. When the project is finished, the data will be archived on KU Leuven's archive drives and datasets will be made public on NCBI servers.

**How will the data be backed up?**

Data on OneDrive and KU Leuven server is automatically backed up in regular intervals.

**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

Sufficient storage & backup capacity is available on the university's central servers, as well as on the HPC, and additional storage will be purchased if necessary.

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

The current project does not deal with sensitive personal data.

However, data will only be accessible by authorized personnel through two-factor authentication on dedicated University servers.

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

The storage of raw data (and backup) will be approximately 1-2TB. This will be stored at gbiomed (large storage volume, €113,84/TB/year). The cost of approximately €900 will be covered in the overall running costs of the lab.

## 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...).**

The most relevant data generated in this project will be deposited in the public domain:

1. Raw sequence data derived from shotgun bacteriome and virome experiments as well as sequencing of bacterial and phage isolates will be deposited short read archives (SRA)-archives.
2. Consensus sequences of obtained virus, phage and bacterial genomes will be deposited in GenBank.

Furthermore, raw data generated during this project will also be stored on dedicated servers of the KU Leuven.

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

The most relevant data generated in this project will be deposited in the public domain:

1. Raw sequence data derived from shotgun bacteriome and virome experiments as well as sequencing of bacterial and phage isolates will be deposited short read archives (SRA)-archives.
2. Consensus sequences of obtained virus, phage and bacterial genomes will be deposited in GenBank.

Furthermore, raw data generated during this project will also be stored on dedicated servers of the KU Leuven.

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

After deposition of the obtained results in the public domain (SRA, GenBank), there will be no cost associated with this storage of the present project.

## 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 most relevant data generated in this project will be deposited in the public domain:

1. Raw sequence data derived from shotgun bacteriome and virome experiments as well as sequencing of bacterial and phage isolates will be deposited short read archives (SRA)-archives.
2. Consensus sequences of obtained virus, phage and bacterial genomes will be deposited in GenBank.

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

NA

**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.**

SRA  
Genbank

**When will the data be made available?**

The full dataset will be released upon publication of the data.

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

Public Domain Mark 1.0  
GPL

**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

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

Data will be shared on the NCBI servers, which are free to use, or Zenodo which is also a free service.

## 6. Responsibilities

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

Lander De Coninck

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

Lander De Coninck

**Who will manage data preservation and sharing?**

Jelle Matthijnsens

**Who will update and implement this DMP?**

Jelle Matthijnsens

-