

---

## Factors driving gut virome composition and stability in zoo-housed and wild bonobos (*Pan paniscus*)

*A Data Management Plan created using DMPonline.be*

**Creator:** Kim Vermeulen

**Affiliation:** KU Leuven (KUL)

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

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

**Grant number / URL:** 11Q6124N

**ID:** 206225

**Start date:** 01-11-2023

**End date:** 31-10-2027

### Project abstract:

The human gut is inhabited by a substantial amount of viruses, which have only recently started to get attention in microbiome research. The viruses in the gut are collectively referred as the gut virome and seem to play an important role in regulating our health and wellbeing. Current evidence suggests that in humans, the composition of the gut virome is highly individual-specific and stays relatively stable over time, but the factors driving these patterns of composition are still poorly understood. Progress is hindered by a lack of standardized methods in virome research, and ethical and practical considerations that accompany experimental work and longitudinal sampling in humans. Animal studies, and especially in primates, offer a unique potential for larger experimental control and provide insight into the evolutionary processes that shaped the human gut virome. In this project, I will study the factors driving gut virome composition and stability in zoo-housed and wild bonobos (*Pan paniscus*). I will investigate to what extent virome profiles are shaped by host-specific factors like sex, age or genetics, or rather by external factors like geography, diet and sociality using both a cross-sectional and a longitudinal sampling approach. I will also study the effects of antibiotic-use on the ability of the virome to increase the emergence of antibiotic-resistance genes through interaction with the gut bacteria, which could have major implications for human treatment protocols.

**Last modified:** 07-04-2024

# Factors driving gut virome composition and stability in zoo-housed and wild bonobos (*Pan paniscus*)

## 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, .csv, .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>	
Zoo-housed bonobo fecal samples	Fecal samples of the entire European zoo-housed bonobo population	Reuse existing data	Physical	-	-	-	Approximately 3 boxes of 160 aliquots in 2 ml CryoTubes
Behavioral observations of zoo-housed bonobos	Behavioral observations of sociality and dietary intake for the European zoo-housed bonobo population	Reuse existing data	Digital	Observational	.XML .xlsx	< 100 GB	-
Wild bonobo fecal samples	Fecal samples collected from two wild bonobo communities in the DRC	Generate new data	Physical	-	-	-	Approximately 4 boxes of 224 aliquots in 2 ml CryoTubes
Bonobo gut virome	Metagenomic sequencing data of viral DNA and RNA in fecal matter of zoo-housed and wild bonobos	Generate new data	Digital	Observational	FASTQ	< 1 TB	-
Behavioral observations of wild bonobos	Behavioral observations of sociality and dietary intake for two wild bonobo communities	Generate new data	Digital	Observational	.CTX .xlsx	< 100 GB	-
Statistical analyses	Code for statistical analyses of virome data	Generate new data	Digital	Aggregated data	-	< 100 MB	-

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:

Fecal samples and metadata on the European zoo-housed population of bonobos were collected for the purpose of an ongoing PhD project on the bacterial component of the gut microbiome (PhD student Jonas Torfs, University of Antwerp, supervised by Dr. Nicky Staes). Several aliquots of the fecal samples were stored at Antwerp Zoo and can be used for virome analysis in my project. As the PhD is still ongoing, sequencing results on the bacterial gut microbiome and metadata are not yet publicly available, but can be accessed through my co-promoter Dr. Staes.

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

I will collect fecal samples and behavioral observations of bonobos, both in the wild and in European zoos. The methodology outlined in our project is strictly non-contact and non-invasive, following IUCN guidelines. It also fits the requirements of the *Institut Congolaise pour la Conservation de la Nature* (ICCN), the national conservation authority granting us permission to conduct research on the wild bonobos in the Democratic Republic of Congo. I have also received approval for conducting research at the European zoos from the ethical committee of the Royal Zoological Society of Antwerp (ref: EC-15/DC(10-08-2023)).

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).

**Behavioral data:** Behavioral observations will be collected using the CyberTracker and the Observer XT (Noldus). Files of these programs will be exported as Excel tables, which will contain a GUIDE tab explaining all columns, variables and abbreviations used in the file.

**Metagenomic sequencing data:** Cleaned data on the viral content of the fecal samples (e.g. viral richness, diversity) will be made available in Excel files. Explanatory comments and a GUIDE tab will be included to explain how each parameter was calculated.

**Statistical analysis:** R files will contain comments on how the models were made and what the different steps in the model specifically do using # inside the scripts or with R.markdown files.

**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

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

**Where will the data be stored?**

Original files of raw reads will be stored on the K-drive of KU Leuven and physical hard drives at the lab of promotor Prof. Matthijnsens. Observational data of behavior will be stored on a personal CRC OneDrive and physical hard drives at Antwerp Zoo.

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

The KU Leuven servers and KU Leuven OneDrive have automatic back-up services. The observational data of behavior will also be backed up automatically on the online cloud storage of the CRC.

**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

The KU Leuven and CRC OneDrive provides 2TB and 1TB storage for non-confidential data, respectively, which should be enough for the project. Additional physical 5TB hard drives for back up can be purchased by the research groups if necessary.

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

The KU Leuven and CRC servers are password-protected. Access to the data as well as the access level will be limited on a project need and individual basis.

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

All KU Leuven and CRC personnel has access to 2TB and 1TB of data storage on OneDrive, respectively. We do not plan to generate massive amounts of data, so these online storages should provide sufficient storage. If not, additional physical hard drives (5TB) can be bought for 150 euros. Costs will be covered by the research groups.

The physical samples will be stored within available freezers in our laboratory. The costs of their storage is covered by the lab of Prof. Matthijnsens.

#### 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...).

All data will be preserved for 10 years according to KU Leuven RDM policy.

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

Metagenomic sequencing data will be stored on the long-term on physical hard drives at the lab of promoter Prof. Matthijnsens. Behavioral data will be stored on physical hard drives at the CRC.

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

There are no substantial additional costs for data storage and backup. The host institutes of my promoter Prof. Matthijnsens and co-promoter Dr. Staes will provide funding for the external hard drives and will provide access to an online server.

#### 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 a restricted access repository (after approval, institutional access only, ...)
- Yes, in an Open Access repository

R scripts and metagenomic data will be made available on the GitHub page of promoter Prof. Matthijnsens <https://github.com/Matthijnsenslab>. Observational data of behavior can be made available upon reasonable request after publication of research results.

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

Observational data of behavior can be made available after approval by co-promoter Dr. Staes and host institute the CRC.

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.

R scripts and metagenomic data will be made available on the GitHub page of promoter Prof. Matthijnsens <https://github.com/Matthijnsenslab>. Observational data of behavior can be made available upon reasonable request after publication of research results.

When will the data be made available?

Upon publication of research results.

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

Data will be made available under a creative commons attribution license (CC-BY 4.0).

**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?**

There are no substantial additional costs for data sharing. Data will be made available after results are published. According to FWO requirements, publications will open access and publication costs will be covered by the bench fee.

## **6. Responsibilities**

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

Kim Vermeulen, Jelle Matthijnsens and Nicky Staes

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

Kim Vermeulen, Jelle Matthijnsens and Nicky Staes

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

Kim Vermeulen, Jelle Matthijnsens and Nicky Staes

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

Kim Vermeulen, Jelle Matthijnsens and Nicky Staes