

DMP Heleen Maetens

Project Name DMP Heleen Maetens - DMP Heleen Maetens

Principal Investigator / Researcher Heleen Maetens

Institution KU Leuven

1. General Information

Name applicant

Heleen Maetens

FWO Project Number & Title

1128222N

Swimming in the shadow of cichlids: the enigmatic species richness and evolutionary history of the African small barbs.

Affiliation

- KU Leuven
- Other

Royal Museum for Central Africa

Royal Belgian Institute of Natural Sciences

2. Data description

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

- Generate new data
- Reuse existing 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).

Generate new data

General new data

Type of data	Format	Volume	How created
Manuscripts	.doc .pdf	several MB	Manuscripts of articles and publications
Powerpoint presentations	.ppt .pdf	Several GB	Presentations for conferences, seminars, meetings
Pictures from the field	.jpeg	several GB	Pictures from voucher specimens taken on the field
Database	.xlsx .accdb	several MB	Information (tags, identification, locality of catch, date of catch...) of all specimens used during the PhD

-

WP1: genetics, phylogenetics and phylogeography

We will use a multi-locus approach using two mitochondrial (COI, Cytb,) and five nuclear (RAG1, rhodopsin, intron 2 of GH, EGR2B and IRBP) genes. These results will be used to study the geographical distribution of sister groups.

Type of data	Format	Volume	How created
DNA sequences of different genes (COI, Cytb, RAG1, rhodopsin, GH intron 2, EGR2B, IRBP)	.ab1 (raw sequence files) .fas (alignments of sequences)	Several MB	DNA barcoding from fin clips
Gene trees	.fas .mtsx .xml .txt .t	Several MB	Using MEGA, BEAUTi-BEAST, PhyML and Mr Bayes software
Molecular clock calculations	.xml .txt	Several MB	Using BEAUTi and BEAST software and visualisation in FigTree
Maps and distribution ranges	.qgz	Several MB	Created in QGIS projects
Dispersal-Vicariance Analysis and Dispersal-Extinction Cladogenesis	tree file format	Severa MB	Using S-DIVA and S-DEC
Ancestral state reconstructions	tree file format	Several MB	Using RASP

WP2: morphology and morphometrics

We will use traditional morphometrics, geometric morphometrics both on voucher specimens and on 3D μ CT scan images and the information found in literature about some specific characters.

Type of data	Format	Volume	How created
Numeric data of traditional morphology	.xlsx .csv	Several MB	Measurements and meristics are taken on voucher specimens using a dial caliper
Scripts for data analysis	.R	Several MB	Multivariate analyses are executed using R
Pictures of specimens to use for geometric morphometrics	.jpeg	Several GB	Pictures taken in a standardised way including a scale
Landmark coordinates	.TPS	Several MB	Landmarks are placed using tpsDIG
3D μ CT scan images	.tif	Several GB	Pictures will be cleaned using GOM inspect
Data on characteristics	.xlsx	Several MB	Data from observations and literature that will be used to reconstruct character states
Character mapped phylogenetic trees	Tree file format	Several GB	Traits will be reconstructed and mapped on phylogenetic trees

WP3: morphology and morphometrics

We will perform stable isotopes of ^{13}C and ^{15}N on the samples from the Lake Edward system and the N.-E. Congo.

Type of data	Format	Volume	How created
Numeric data of stable isotopes	.xlsx .csv	Several MB	Data will be processed in the lab of Prof. S. Bouillon
Scripts for data analysis	.R	Several MB	Multivariate analyses are executed using R

WP4: morphology and morphometrics

We will use a whole genome scanning approach on the Lake Edward system and Congo basin. Genomic data will be analysed and stored on the VSC. Reuse existing data

Reuse of data

The RMCA harbours the largest collection of African freshwater fishes, including several thousands of *Enteromius* specimens. Therefore, most material for morphological analyses is already available.

Type of data	Format	Volume	How created
Voucher specimens		> 3000 specimens	Caught during different expeditions in the past
Genetic samples (fin clips from fish specimens)		> 450 samples	During the field trips, a fin clip was taken from different individuals and stored in 100% EtOH
DNA sequences from GenBank and Bold	.fas	About 1315 sequences of <i>Enteromius</i> are available now	Collecting DNA sequences from public databases in order to compare our data with existing data

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)

- No

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?

For each WP the following documentation will be provided (if applicable):

- lab protocols (labbook, .doc files)
- raw data (.xlsx files)
- data analyses and used parameters (labbook, .R files)
- automatically generated data from pictures
- databases with information about the used samples (voucher specimens and fin clips)

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.

- No

No metadata standard will be used. We will create metadata files for each WP along the way. We will follow the rules of the institute regarding the organisation and storage of the data.

5. Data storage and backup during the FWO project

Where will the data be stored?

The data will be stored at my personal work computer.

The genomic data will be stored at the VSC.

Preserved specimens are in the collections of the Royal Museum for Central Africa (RMCA) and newly acquired specimens will be registered and curated in the RMCA. Raw data (2D and 3D images, μ CT data) will be stored at the KU Leuven and the RMCA. μ CT data will be uploaded on OpenSource depositories, such as MorphoSource. Sequence data will be uploaded on FishBOL and GenBank. Genomic data will be uploaded on the NCBI Sequence Read Archive.

How is backup of the data provided?

Regular backups will be done on an external hard drive and on OneDrive.

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

I expect that the capacity of my own personal computer will be enough to store the data, excluding the pictures. Pictures will be stored separately on an external hard drive.

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

I do not expect additional costs for data storage.

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

The data will be protected by a password that will change regularly.

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 retained.

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

Miguel Parrent, fish collection manager at the RMCA, Tervuren will be responsible for the curation of the fish collections. Preserved museum specimens that are put at our disposal for the study will return afterwards to the RMCA collections. Newly collected specimens will be deposited at the RMCA for curation.

At the end of the project, the promoters will be the designated responsible persons for the data. Data storage capacities are available to preserve the data during at least 5 years after the end of the research.

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

No extra costs are expected for the storage of the data.

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 accessible on demand

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 accessible through publications and OpenSource depositories as GenBank and Bold. Data will be accesible upon request by mail.

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

There are no expected costs for data sharing.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Heleen Maetens

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

Heleen Maetens

Who will be responsible for ensuring data preservation and reuse ?

Heleen Maetens

Who bears the end responsibility for updating & implementing this DMP?

The PI, Jos Snoeks, bears the end responsibility of updating & implementing this DMP.