A phylogeographic, ecological and organoleptic characterization of wild coffee species from the Congo Basin

A Data Management Plan created using DMPonline.be

Creator: An-Sofie Tas

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** 1173723N

ID: 197791

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Coffee is one of the most valuable crops in the world and it is the second-most exported agronomic product of developing countries. Almost all cultivated coffee varieties descend from two wild species, Coffea arabica and C. canephora which makes coffee cultivation worldwide highly susceptible to emerging diseases and the consequences of climate change. Therefore, an assessment of the cultivation potential of other coffee species is highly needed. As a first step to achieve this, this project aims to uncover the ecology, phylogeography and organoleptic profile of a set of Central and West African wild coffee species. Therefore it will (1) Describe the phylogeny of the species with a special focus on the canephora-group within the Congo Basin, using Genotyping-by-Sequencing data; (2) Characterize the ecological niche of the Coffea species using observational and experimental approaches; (3) quantify possible future shifts in their distribution range due to climate change; and (4) Analyze the potential of the coffee species from the Congo Basin for further commercialization through a standardized organoleptic analysis. This project capitalizes on the broad expertise in studying the wild relatives of coffee in the DRC by the three involved research groups.

Last modified: 16-03-2023

A phylogeographic, ecological and organoleptic characterization of wild coffee species from the Congo Basin
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)

The following physical samples of the investigated coffee species will be collected:
- Silica dried leaf samples
- Herbarium vouchers
- Seeds

The following datasets on the investigated coffee species will be obtained:
-Genotype-by-Sequencing data
-Data on morphological traits
-Data on the ecological niche (biotic and abiotic factors)
-Occurrence data
-Climate data
-Robusta Green Grading form data
-Descriptive cupping data

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)

1. Designation of responsible person (If already designated, please fill in his/her name.)
2. Storage capacity/repository
    - during the research
    - after the research

1. Responsible person for the data management during the project: An-Sofie Tas (KU Leuven) & Olivier Honnay (KU Leuven). Responsible person after the end of the research: Olivier Honnay (KU Leuven).
2. Both during as after the end of the project the data will be stored at the KU Leuven Enterprise Box (100GB) and the Meise Botanic Garden Drive (30GB) which have more than sufficient storage capacities.

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)

The samples collected in DR Congo are subjected to the Nagoya protocol. To align with the protocol the needed documents are arranged. The project is embedded in a broader collaboration with local partners (INERA-Yangambi, UNIKIS & CSB). A MOU is signed between the Botanic Garden Meise (project partner) and UNIKIS. In order to ensure the fair and equitable utilisation of the *Coffea canephora* genetic resources of the Democratic Republic of the Congo, the Prior Informed Consent (PIC), the Mutually Agreed Terms (MAT) and Material Transfer Agreement (MTA) will be requested at the Congolese Competent Authority, in accordance to the "Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity".

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

Not applicable

A phylogeographic, ecological and organoleptic characterization of wild coffee species from the Congo Basin
DPIA

DPIA

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

- Not applicable

A phylogeographic, ecological and organoleptic characterization of wild coffee species from the Congo Basin
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

A phylogeographic, ecological and organoleptic characterization of wild coffee species from the Congo Basin
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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Silica dried leaf samples	Coffee leaf samples dried in silica gel for DNA extraction	Generate new & reuse existing data	Physical				Max. 1000 individuals
Herbaria specimens	A branches of coffee trees with leaves dried in an herbarium vouchers for morphological studies	Generate new & reuse existing data	Physical				Max. 1000 individuals
seeds	Coffee seeds for germination experiments and organoleptic analysis	Generate new data	Physical				Max. 8 kg
SNP data set of collected coffee individuals	GBS sequencing output of the DNA extractions from the leaf samples	Generate new data	Digital	Observational	.vcf	<100GB	
Climate data	Data on 19 climate variables from BioClim database	Reuse existing data	Digital	Observational	.tif	<1GB	
Occurrence data	Coördinates of coffee plant individuals	Generate new & reuse existing data	Digital	Observational	.csv	< 1GB	
Seed germination data	Germination parameters derived form germination experiment	Generate new data	Digital	Experimental	.csv	< 1GB	
Scans of leaf samples	Scans of leaf samples for morphological studies	Generate new data	Digital	Observational	.tif	< 100GB	
Data on morphological traits and ecological niche	Data on the morphological traits of coffee tree individuals (such as height, stem diameter) and habitat characteristics	Generate new data	Digital	Observational	.csv	< 1GB	
Robusta Green Grading Form data	scores of coffee bean samples to indicate the quality	Generate new data	Digital	Observational	.csv	< 1GB	
Descriptive Cupping Data	scores and descriptions of coffee bean samples to determine the flavor profile	Generate new data	Digital	Observational	.csv	< 1GB	

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:

This project will reuse coffee samples (herbarium specimens & silica dried leaf samples) stored at the Herbarium of Meise Botanic Garden (stored with accession numbers in BGBase) and the corresponding passport data.
The project will also make use of the existing bioclimate variables of the WorldClim database:
Bioclimatic variables — WorldClim 1 documentation

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.

- Yes

The use of coffee samples (herbaria specimens, silica dried leaf samples, seeds) and the derived data from these samples are subjected to the Nagoya protocol. In accordance with the Nagoya Protocol, Meise Botanic Garden and University of Kisangani (UNIKIS), both partners of the project, signed a MTA (Material Transfer Agreement). This MTA allows Meise Botanic Garden to use and analyse the samples and the associated data in accordance with the defined activities. Any use of the samples and data that is not described in the MTA has to first be approved by UNIKIS.

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).

- Physical data (herbarium specimens, silica dried leaf samples and seeds) will be stored at the herbarium of Meise Botanic Garden. The passport data corresponding to the samples will be stored in BGbase (Meise Botanic Garden).
- All data will be collected by standardized procedures. The sampling methodologies will be described in detail in a codebook.
- For raw datafiles, as previously described, a README;txt will be made containing information on the content of the dataset (description of variables etc.)

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

There are two metadata standards that will be used in this project:

1. Ecological Metadata Language (EML) will be used for community metadata, as this standard is specific for ecological research. The Morpho application will be used to create, edit and search metadata files following EML. These files can be used to exchange information within and among research groups.
2. Metadata related to genomics data will be organized and reported according to the standard developed by the Genomic Standards Consortium (GSC).

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

Where will the data be stored?

1. Physical samples (herbaria specimens, silica dried leaf samples, seeds) will be stored at the herbarium of Meise Botanic Garden
2. Datasets will be stored at the server space at Department of Biology and KU Leuven Enterprise BOX, as well as on the servers at Meise Botanic Garden. Additionally personal copies will be made on an external hard drive.

How will the data be backed up?

1. The data stored using the KU Leuven and Meise Botanic Garden services have automatic back-up procedures.
2. Additionally, back-ups from personal copies will be made using external hard drives.

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

Storage capacity of the KU Leuven Enterprise Box: 100 GB
Storage capacity of the Meise Botanic Garden drive: 30 GB
This storage and backup capacities will be sufficient to store the data of this project.

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

Both the KU Leuven Enterprise Box as the The Meise Botanic garden drive are protected through a password (the KU Leuven Enterprise Box via the KU Leuven Login). The accessibility to the data can be controlled in both storage services.

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

There won't be additional cost for the storage and backup of the data. KU Leuven Enterprise Box is centrally financed and the costs of Meise Botanic Garden drive are covered by Meise Botanic Garden.

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...).

1. The physical samples (herbaria specimens, silica dried leaf samples, seeds) will be stored at the Herbarium of Meise Botanic Garden for unlimited amount of time.
2. DNA sequences will be deposited on GenBank.
3. Other datasets (as well as codes and scripts) used for analyses and publications will be retained for at least 10 years after the end of the project.

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

1. The physical samples (herbaria specimens, silica dried leaf samples, seeds) will be stored at the Herbarium of Meise Botanic Garden for unlimited amount of time.
2. DNA sequences will be deposited on GenBank.
3. Datasets will be stored at the KU Leuven Enterprise Box and the Meise Botanic Garden Drive.
4. Scripts and codes will be deposited on GitLab
5. Additionally, LIRIAS, ResearchGate, PANGAEA and Zenodo will be used.

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

No additional costs are expected:

- LIRIAS, ResearchGate, PANGAEA and Zenodo are freely to use.
- deposits on GitLab and GenBank have no additional cost.
- Storage on the KU Leuven Enterprise Box and the Meise Botanic Garden Drive has no additional costs
- The storage of the Physical samples in the Herbarium of Meise Botanic Garden has no additional costs.

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

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

Deposited data will be available to anyone, on certain conditions e.g. depending on
their purpose. Creative Commons Licences (CC BY) will be attached to the deposited
data, enabling researchers to access, mine and reproduce our data.
For genomic data, sharing depends on the permission of UNIKIS.

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.

- Yes, Other

All coffee samples (Herbaria specimens, silica dried leaf samples, seeds) are subject to the Nagoya protocol. In accordance with the Nagoya protocol there is an agreement singed between Meise Botanic Garden (MeiseBG) and University of Kisangani (UNIKIS). MeiseBG is allowed to use and analyse the collected material, but UNIKIS remains the legitimate owner. Sharing this material with a third party must always be approved by UNIKIS.

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

1. Relevant datasets will be uploaded in a standardized format in Zenodo.
2. Codes will be made available on GitLab.

When will the data be made available?

-Datasets used for publications will be made available immediately after the publication of the results. Additional datasets will be made available within three years after the end of the project.
-The sharing of genomic data depends on the permission of UNIKIS.

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

As previously described, the Creative Commons Licences (CC BY) will be used.

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

Not yet available.

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

There are no expected costs for the data sharing (GitLab & Zenodo are free of charge).

6. Responsibilities

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

An-Sofie Tas (KU Leuven), Olivier Honnay (KU Leuven), Filip Vandelook (Meise Botanic Garden), Piet Stoffelen (Meise Botanic Garden)

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

An-Sofie Tas (KU Leuven), Olivier Honnay (KU Leuven), Filip Vandelook (Meise Botanic Garden), Piet Stoffelen (Meise Botanic Garden)

Who will manage data preservation and sharing?

An-Sofie Tas (KU Leuven), Olivier Honnay (KU Leuven), Filip Vandelook (Meise Botanic Garden), Piet Stoffelen (Meise Botanic Garden)

Who will update and implement this DMP?

An-Sofie Tas (KU Leuven), Olivier Honnay (KU Leuven), Filip Vandelook (Meise Botanic Garden), Piet Stoffelen (Meise Botanic Garden)