
Characterizing natural occurrence and temperature-induced formation of 2n gametes in Musa spp. to facilitate triploid breeding in banana.

A Data Management Plan created using DMPonline.be

Creator: Astrid Severyns

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Astrid Severyns

Project Administrator: First Name Surname

Grant number / URL: 1SF3323N

ID: 193461

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Despite being the most important fruit crop worldwide, banana (*Musa* spp.) cultivation currently faces enormous threats by several biotic and abiotic stresses, as overall agronomic biodiversity is

low. Conventional breeding of improved triploid varieties faces among long generation times and parthenocarpy, many bottlenecks linked to interploidy hybridization such as; reduced male fertility,

genomic incompatibilities, and low seed yield and viability.

A more direct route for generating triploid banana genotypes, is the exploitation of ectopic formation

of diploid (2n) gametes from 2x parents. This would strongly increase breeding efficiency as selection

can be carried out at the 2x level, avoiding inefficient interploidy crosses and complex inheritance

patterns. Additionally the variable genetic make-up (FDR or SDR) of 2n gametes allows to tailor the

genetic configuration of polyploid progeny.

The proposed study will result in the identification of germplasm lines that generate 2n pollen, as well in temperature treatments that stimulate their production, and will provide information on the mechanistic basis of 2n pollen formation and their genetic make-up, and hence may pave the way for

utilising 2n gametes in conventional banana breeding. Proposed PhD research thus has great potential for fortifying a new view on banana breeding, thereby saving time, money and labour

required to create new varieties of the world's largest fruit crop.

Last modified: 16-01-2023

Characterizing natural occurrence and temperature-induced formation of 2n gametes in Musa spp. to facilitate triploid breeding in banana.

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"> • Generate new data • Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Digital • Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ... • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA 	
Dried Musa anthers	<p>Musa anthers were collected biweekly from plants in the KU Leuven greenhouse to monitor pollen development over flowering.</p> <p>A collection of dried anthers from 71 different Musa genotypes (collected in Tanzania) in order to perform a GWAS pollen size.</p>	Generate new data	Physical				20 dm ³ about one drawer in a 4°C freezer

Fixated Musa flowers	Male flowers comprising meiotic stages to mature flowers in anthesis, including several genotypes, collected in Tanzania, Botanic Garden of Meise and Heverlee greenhouses. Female flowers, at the moment of anthesis, and at several timepoints during seed development. Collected in Tanzania	Generate new data	Physical				4 m³ in a cold room
Pollen germination data	On 30 timepoints, pollen germination was measured for several genotypes, values were noted down in paper lab journal and afterwards stored in excel files	Generate new data	Digital	Observational	.xlsx	<1MB	
Microscopy images	Pictures are made of pollen grains in a high throughput fashion on a fluorescence microscope Images of nuclei in microspores Images of prepared meiotic chromosome spreads Images of pollen tube growth in banana stigmas	Generate new data	Digital	Experimental	.tif, .jpg	<1TB	

Pollen size and viability data	Using cellprofiler, a pipeline was set up to automatically identify pollen grains in fluorescence microscopy images, and measure their size and FDA intensity. Data is stored in csv files	Generate new data	Digital	Observational	.csv	<100MB	
Climate data greenhouse	csv file containing radiation, temperature and relative humidity measurements for full ground compartment	Generate new data	Digital	Observational	.csv	<1MB	
Pollen development combined files	csv file containing combined greenhouse, pollen germination, pollen viability and pollen size data, + html files containing statistical analysis	Generate new data	Digital	Compiled data	.csv, .html	<1MB	
Flow cytometry data	single pollen nuclei extraction Ploidy measurements on developing seeds	Generate new data	Digital	Experimental	.pzfx	<100MB	
SNP markers Musa	set of 5,544 robust SNP markers as generated within the study of Sardos et al. 2016	Reuse existing data	Digital	Experimental	.FASTQ	<1MB	
Genomic data	Certain known 2n gamete producing genes will be genotyped	Generate new data	Digital	Experimental	.FASTA, .VCF	<1MB	
qPCR data	The expression of certain known 2n gamete producing genes will be assessed via qPCR analysis	Generate new data	Digital	Experimental	.xlsx	<1MB	

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:

SNP markers Musa: [SNP dataset in musa acuminata \(banana\) panel - Harvard Dataverse](#)

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.

- Yes

The induction of 2n gamete formation in Musa spp. can be interesting to speed up the breeding of new triploid banana varieties, hence it could be commercially exploited. We use two approaches: external treatments to induce 2n pollen production, discovery of genes that induce 2n gamete production in banana. In the first case, a combination of the 'pollen size and viability data', 'microscopy images', and 'flow cytometry data' might generate potential for commercial valorization. In the latter case, a combination of 'SNP markers Musa', 'genomic data' and 'qPCR data'.

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

I collaborate with Meise Botanic Garden and IITA Arusha, however there are no data dissemination restrictions in these partnerships.

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).

For the protocols I have developed and used, I keep a separate word document, stored on KU Leuven servers that describes in detail how I used my samples to obtain data.

Every performed experiment, is noted down in a paper lab journal.

I will create a document, listing all datasets I have generated, including instructions on how to access them.

The dried Musa anthers and Fixated Musa flowers are all labeled, and I keep an excel file with details on the material in question, which I update every time new samples are added to the collection. In this same excel file, I keep track of what analysis have been carried out on the samples.

All pollen germination data, pollen size and viability data and climate greenhouse data are stored in hierarchical folders and organized by genotype. The same tags are used throughout the entire sample collection and data processing. All data processing scripts are

annotated. I will also publish a paper on this subject with a section dedicated to the management and availability of the data, as well as a project report.

All microscopy images are structured in folders by type of microscopy image, and genotype. this way they can be linked to the aforementioned excel sheet that gives an overview of all samples. To provide structure they are also assembled in PowerPoint, afterwards a report, and eventually a paper.

Flow cytometry data will be stored per project in a hierarchical manner, and will also be linked by date to the paper lab journal. On this data I will regularly generate reports.

GWAS analysis using SNP markers, genomic data and qPCR data will also be stored in separate folders with a clear structuring, annotations in the data, and regular reports.

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

For all image sets, metadata is generated by the used equipment.

In case I would publish data in a public database, I would use the international standard for metadata preferred by that database.

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

Where will the data be stored?

Digital data: The data will be stored in personal and shared network drives provided by the KU Leuven.

Physical data: 4°C cold room and fridges PGCI lab Nico De Storme

How will the data be backed up?

The KU Leuven network drives are regularly backed up.

An 'archive' drive for large volume storage, provided by the KU Leuven, will be used for backing up the data.

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

Personal network drive: 50 GB

Shared network drive: 1 TB

Archive drive: 1 TB

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

Access to KU Leuven drives is password protected. The shared drive contains only members of our lab. Each researcher receives a separate folder on the shared drive, which can only be modified by that researcher.

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

Shared drive: 503,66 euro per 1 TB

Archive drive: 100 euro per 1 TB

The amount of data generated within the project falls within the limits of the 3 drives that are financed by the PGCI lab budget. The lab uses internal funding for this.

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...).

There are no reasons to not store all data, so all data will be stored.

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

On the PGCI archive drive (large volume storage provided by KU Leuven)

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

500 euro for 5 years, this will be covered by internal lab budget.

Preserving smaller datasets in repositories is usually free of charge.

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

All datasets generated in this project are eligible for placing in an open access repository, as it does not contain any personal data. However, depending on the commercialization potential of certain data, we might postpone the moment on which we place the data in an Open Access repository.

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

We might wait with making data openly available until we have published papers in this data. Afterwards we make our data available in an Open Access repository.

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

There is a potential for commercial valorization, however as the treatments we will test, have already been used on other plant species, we are not going to receive IP rights

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

Most likely, data will be made available via the CIAT dataverse, the International Centre for Tropical Agriculture (CIAT) is part of the non-profit CGIAR consortium, and uses non-profit Harvard Dataverse.

When will the data be made available?

Upon publication of research results, latest by 31/10/2026.

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

Creative Commons Attribution 4.0 International (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

These are not available yet

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

As the datasets made available would be rather small, there are no extra costs foreseen for data sharing.

6. Responsibilities

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

The researcher, Astrid Severyns

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

The researcher, Astrid Severyns

Who will manage data preservation and sharing?

the supervisor, Nico De Storme

Who will update and implement this DMP?

The researcher, Astrid Severyns