
On a spicy journey: In search of climate resilient vanilla production systems

A Data Management Plan created using DMPOnline.be

Creators: Charlotte Watteyn  <https://orcid.org/0000-0003-4145-3228>, Bart Muys

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Charlotte Watteyn  <https://orcid.org/0000-0003-4145-3228>, Bart Muys

Data Manager: Charlotte Watteyn  <https://orcid.org/0000-0003-4145-3228>, Bart Muys

Project Administrator: Charlotte Watteyn  <https://orcid.org/0000-0003-4145-3228>, Bart Muys

Grant number / URL: PDMT1/23/009

ID: 203917

Start date: 01-10-2023

End date: 30-09-2024

Project abstract:

Both staple food and cash crops provide livelihoods to millions of smallholder farmers in the tropics, with cash crops offering great opportunities for rural development. Vanilla is a high value tropical cash crop and a globally very important spice used in a broad spectrum of food products. Future vanilla bean supply is, however, at risk due to the combined effect of a genetically eroded crop species with an over-intensified farming system. The negative effects associated with low genetic diversity are expected to exacerbate under climate change. As a consequence, there is an urgent need to enhance vanilla production resilience through diversification at both crop and system level. Despite the promising role of the wild relatives of the crop species *Vanilla planifolia*, an in-depth evaluation to integrate these so-called crop wild relatives (CWR) into breeding programs with improved plant performance under climate change, is currently lacking. Our project therefore seeks to guide the development of a climate-smart vanilla production strategy that involves novel *Vanilla* genetics. We will apply an innovative, interdisciplinary approach that combines state-of-the-art techniques within the field of species distribution modelling and populations genetics, plant phenotyping, and participatory stakeholder assessments. The obtained insights will ultimately lead to the formulation of an action plan in close collaboration with key stakeholders along the chain of custody.

Last modified: 03-01-2024

On a spicy journey: In search of climate resilient vanilla production systems

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		<i>Indicate: N(ew data) or E(xisting data)</i>	<i>Indicate: D(igital) or P(hysical)</i>	Indicate: Audiovisual Images Sound Numerical Textual Model Software Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Vanilla DNA samples	DNA samples extracted from the leaves of Vanilla plants naturally growing in Mexico, Costa Rica and Colombia.	N	P	Other (physical DNA samples)	NA		
Vanilla sequencing data	DNA sequencing and SNP data of Vanilla samples, generated via DArTseq technologies.	N	D	Other (sequencing data)	FASTQ		
DArTseq processing and analysis script (pipeline)	Script created for the processing and analysis of the DNA sequencing data derived via DArTseq technologies.	N	D	Other (programming languages and code)	Text-based format (Phyton)		
Vanilla occurrence data	Longitude and latitude of Vanilla populations naturally occurring across the Neotropics.	N and E	D	Other (statistical data, images, geospatial data, spreadsheets)	R, TIFF, GeoTIFF, csv		
Environmental data for species distribution modelling	Climate, soil and topographical variables for species distribution modelling.	E	D	Other (geospatial data)	GeoTIFF, Esri shapefiles, ASCII GRID		
Species distribution modelling script	Script created for Vanilla species distribution modelling.	N	D	Other (programming languages and code)	Text-based format (R)		

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:

Vanilla occurrence data: GBIF (<https://www.gbif.org/>), Tropicos (<https://www.tropicos.org/home>), Literature (e.g.: DOI:10.15517/lank.v0i0.12065; DOI:10.15517/lank.v20i3.45203; DOI:10.1002/ecs2.3056).
Environmental data: Climate (<https://www.worldclim.org/>), Soil (<https://soilgrids.org/>), Topography (<https://data.usgs.gov/datacatalog/data/USGS:EROS5e83a205d69371bb>).

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- 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 or 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

Vanilla leaf and DNA samples: Both Material transfer agreements and collaboration agreements (including collection permits) have been developed with collaborating institutions in Mexico, Costa Rica and Colombia to ensure the successful collection of Vanilla leaf and DNA samples. This implies the involvement of these institutions in data processing and scientific publications, and the shared ownership of the collected material. The Vanilla leaf and DNA samples may only be used by the collaborating institutions.

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

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).

README.txt files will be created to accompany the different datasets listed in section 1 when data will be stored in the KU Leuven data repository. The README.txt files will include information on data collection procedures (e.g. Vanilla leaf sampling and DNA extraction), data processing and analysis (e.g. processing and analysis of the FASTq files generated via DArTseq technologies).

Will a metadata standard be used to make it easier to find and reuse the data?

If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

DataCite.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Personal network drive (I-drive)
- Sharepoint online

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

Personal back-ups: C-drive, external hard drives, Google Drive.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

Physical data: Vanilla DNA samples will be stored at collaborating institutions Meise Botanical Garden and University of Costa Rica - Lankester Botanical Garden. The samples will be stored in their laboratories with badge access.

Digital data: SharePoint Online is a protected and secure platform that supports online collaboration within a group (e.g. joint editing of documents and the exchange of information and ideas);

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

We do not expect any costs for data storage and backup during the research project.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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)?

- KU Leuven RDR

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

We do not expect any costs for data preservation during the expected retention period.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?

Please explain per dataset or data type which data will be made available.

- Yes, as open data

Vanilla sequencing data: FASTQ files.

DARtseq processing and analysis script (pipeline): Script created for the processing and analysis of the DNA sequencing data derived via DARtseq technologies.

Vanilla occurrence data: Longitude and latitude of Vanilla populations naturally occurring across the Neotropics.

Environmental data for species distribution modelling: Climate, soil and topographical variables for species distribution modelling.

Species distribution modelling script: Script created for Vanilla species distribution modelling.

Note: The Vanilla DNA samples (physical data) will be stored at the collaborating botanical gardens (Lankester and Meise Botanical Garden) and will only be available for reuse after the project by the collaborating institutions. (see Material transfer agreements)

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

The Vanilla DNA samples will have restricted access (restricted to collaborating institutions via Material transfer agreements).

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 per dataset or data type where appropriate.

- Yes, other

The Vanilla DNA samples will have restricted access (restricted to collaborating institutions via Material transfer agreements).

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

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.

- CC-BY 4.0 (data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

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

We do not expect any costs related to data sharing.

Responsibilities

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

Charlotte Watteyn

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

Charlotte Watteyn

Who will manage data preservation and sharing?

Bart Muys

Who will update and implement this DMP?

Charlotte Watteyn