Genetic characterization of the S-locus in pear (Pyrus communis): exploiting haplotype diversity to characterize the functional role of SFBBs in the S-RNase-based GSI system

A Data Management Plan created using DMPonline.be

Creators: Hanne Claessen, First Name Surname

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: First Name Surname

Data Manager: Hanne Claessen

Grant number / **URL:** 1230623N

ID: 196780

Start date: 01-10-2022

End date: 30-09-2026

Project abstract:

Pear, like many other Rosaceae species, exhibits an S-RNase-dependent gametophytic self-incompatibility system (GSI) that inhibits fertilization by self- or cross-incompatible pollen. The GSI system in Pyrus species is regulated by a polymorphic S-locus which contains a stylar expressed S-RNase gene and multiple S-locus F-box (SFBB) genes that are expressed in the pollen. However, despite extensive characterization of the female S-RNase factor, little is yet known about the identity and functioning of the SFBB male S-determinants, nor about their specific interaction with the various S-RNases. The aim of proposed research project is therefore to identify all SFBB genes in the Pyrus S-locus, and to take advantage of natural genetic variation in S-haplotypes to unravel the complex interaction between SFBBs and their S-RNase targets. Following initial identification of putative SFBB genes based upon long-read nanopore sequencing of the genome of the two Pyrus communis cultivars “Bartlett” and “Conference”, and assembly of the S-locus haplotypes, we will validate and functionally characterize predicted S-RNAse-SFBB interactions using a combination of dedicated in vitro and in vivo assays (Y2H, CRISPR gene editing, etc.). Results of this study will greatly contribute to our understanding of the genetic determination and molecular functioning of the GSI system in pear, with major relevance for future plant reproductive research as well as horticultural applications.

Last modified: 03-02-2023

Genetic characterization of the S-locus in pear (Pyrus communis): exploiting haplotype diversity to characterize the functional role of SFBBs in the S-RNase-based GSI system

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	
Whole genome sequencing	Reads from whole genome sequencing of pear varieties using Nanopore technology	Use existing data	digital	observational	FASTQ, FASTA	<1TB	
SFBB sequences	Sequencing results of SFBB gene PCR products	Generate new data	Digital	Observational	FASTA	<100GB	
Protein modeling	Results of protein structure modeling and interaction modeling with S-RNase	Generate new data	Digital	Simulation data	mmCIF, PDB	<1TB	
Y2H screen	Descriptive dataset of Y2H screen results	Generate new data	Digital	Experimental	.csv	<1GB	
Pollen tube imaging	Microscope images of pollen tube growth through the style for the determination of compatibility	Generate new data	Digital	Experimental	.JPG, .TIFF	<100GB	
Pollen tube count dataset	Dataset with pollen tube counts using semi-in vivo technique to determine compatiblity after a cross	Generate new data	Digital	Experimental	.csv	<1GB	
Optimized protocols	Optimized protocols	Generate new data	Digital	Observational	.doc	<100MB	

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:

Before the start of the funding term of this project, the lab had already performed Nanopore sequencing of the pear varieties Conference and Sweet Sensation with this project in mind. Those sequencing data will be used as input for this project.

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.

- 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).

For each experiment, a clear description of the hypothesis, goal and methods will be noted in the lab notebooks. For each dataset or other experimental result, a seperate README.txt file will be maintained containing minimally the information listed in the MIBBI metadata standard that is applicable for each experiment.

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

Results of each experiment and each dataset will be accompagnied by a README.txt file providing minimally the information listed in the MIBBI metadata standard that is applicable.

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

Where will the data be stored?

Data is stored on the researcher's laptop (locally), backed up on the associated KU Leuven Onedrive, backed up on external hard drives and saved on the shared KU Leuven drivers of the lab.

How will the data be backed up?

Data is automatically backed up on the KU Leuven Onedrive and manually backed up every week on the KU Leuven driver and manually backed up every month on 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

There is currently sufficient storage for the project. If necessary, storage space in onedrive, KU Leuven driver and external hard drives is expanded.

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

The KU Leuven Onedrive is password protected using KU Leuven authenticator. The KU Leuven drivers are only accessible by Lab members and access to specific folders can be limited as needed. Data stored on external hard drives is encrypted using VeraCrypt software.

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

Costs for data storage and back up are covered by project bench fee and lab funds.

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...).

We intend to retain all generated data for at least five years after the end of the project and for 10 years total according to KU Leuven RDM policy.

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

All generated data will be stored in electronic (ELNs) and traditional paper lab notebooks. In the ELN system, all data associated with lab approaches, protocols, raw data and processed results, literature, etc. are stored in the cloud (KU Leuven onedrive) and in a central database, which is administered by the KU Leuven ICTS group ‘Facilities for Research’. KU Leuven provides sufficient storage and

back-up capacity during and after the project; i.e. by centrally organizing long term data storage and data archiving.

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

Costs will be covered by the project bench fee or lab funds

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
- Yes, in a restricted access repository (after approval, institutional access only, …)

How data will be stored differs for each dataset and will depend on the further development of the project. Data will be made available after the project, however the exact details regarding this availability still need to be decided.

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

Access is restricted until data is published.

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.

This will be decided for each dataset at the time when it is made available.

When will the data be made available?

The will be made available after publication of the results.

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

Data usage licenses still need to be decided based on the further development of the project.

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?

Costs will be covered by the project bench fee or lab funds.

6. Responsibilities

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

Hanne Claessen (researcher who generates the data)

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

Hanne Claessen (researcher who generates the data)

Who will manage data preservation and sharing?

Prof. Nico De Storme (Promotor)

Who will update and implement this DMP?

Hanne Claessen (researcher who generates the data)