
Plan Overview

A Data Management Plan created using DMPonline

Title: Resolution of resistance genes within melon QTL regions by optical mapping

Creator: Dick de Ridder

Affiliation: Other

Funder: ZonMw (Netherlands)

Template: Data management ZonMw-template 2019

ORCID iD: 0000-0002-4944-4310

Project abstract:

Plant genomes are highly complex, variable and difficult to assemble. Most assemblies in plants are performed to provide information on complex genomic regions associated with important traits, such as resistance to pests, pathogens and stress. Detecting those regions and their propagation is of high importance for plant breeding and can speed up development of new varieties. Genotyping has insufficient resolution for these complex genomic regions, but the problem may be solved by optical mapping. In this project, we will use dual-color-labelling of key genomic features in melon plants to resolve resistance genes within specific trait-related genomic (QTL) regions.

ID: 62015

Last modified: 20-08-2020

Grant number / URL: 435005027 (40-43500-98-5038)

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Resolution of resistance genes within melon QTL regions by optical mapping

1. General features of the project and data collection

1.1 Project leader contact details

Prof. dr. ir. D. de de Ridder (Projectleader and secretary)
T: +31317484074 | E: dick.deridder@wur.nl
Wageningen University
Departement of Plant Sciences
Bioinformatics
Postbus 1
6700 AP WAGENINGEN

1.2 I have composed my DMP with the assistance of a data stewardship (or management) expert. List his or her name, function, organisation/department, phone number and email address.

- The expert is connected to my department or institution (please explain his/hr expertise related to data stewardship)

Dr. Sven Warris
T: +621877849 | E: sven.warris@wur.nl
Wageningen Research
Applied Bioinformatics
Postbus 1
6700 AP WAGENINGEN
Appointed data steward and bioinformatics researcher

1.3 In collecting data for my project, I will do the following:

- Use existing data (please specify)
- Generate new data

Existing data will comprise of the melon genome, related plants and their genome annotations.

1.4 In my research, I will use:

- Exclusively quantitative data

1.5 I will be reusing or combining existing data, and I have the owner's permission for that.

- Yes, I have permission to use the data

Resources are publicly available.

1.6 In collecting new data, I will be collaborating with other parties.

- Yes, the new data will be (partly) provided by a project partner or supplier
- Yes, I will collect the new data in conjunction with other researchers or research groups

1.7 I am a member of a consortium of 2 or more partners. Clear arrangements have been made regarding data management and intellectual property. (also consider the possible effect of changes within the consortium on issues of data management and intellectual property)

- Yes, clear arrangements have been made regarding data management and intellectual property through a consortium agreement

1.8 I can give an estimate of the size of the data collection; specifically, the number of participants or subjects ("n=") in the collection and its size in GB/TB

- Yes (please specify)

Around 100GB (2 Bionano runs and subsequent data analyses of about 10 plant samples).

1.9 The following end products I will make available for further research and verification (please elaborate briefly)

- Software
- Data documentation
- Documentation of the research process, including documentation of all participants
- (Several versions of) processed data
- Raw data

Metadata for all plants used will be collected, documented and shared. The raw Bionano data will be made available through a public repository, linking to these metadata. The resulting genome mapping data will be described in a document and also be made publicly available.

OptiTools is already OSS.

1.10 During the project, I will have access to sufficient storage capacity and sites, and a backup of my data will be available. (please elaborate briefly)

- Yes, I will make use of my institution's standard facilities for storage and backup of my data

Both local infrastructures at the guest and hotel have large storage facilities, including redundant data storage. Data will also be delivered on external HDD.

2. Legislation (including privacy)

2.1 I will be doing research involving human subjects, and I am aware of and compliant with laws and regulations concerning privacy sensitive data.

- No, I will not be doing research involving human subjects; proceed to section 3 (Making data findable)

2.2 I will be doing research involving human subjects, and I have (a form of) informed consent from the participants for collecting their data.

Question not answered.

2.3 I will be doing research involving human subjects, and I will protect my data against misuse.

Question not answered.

2.4 I will stick to the privacy regulations of my organisation

Question not answered.

3. Making data findable

3.1 The data collection of my project will be findable for subsequent research. E.g., on a catalogue, a web portal, or through the search engine of the repository (note: this is key item 3, which you should report to ZonMw at the end of your project).

- Yes, it can be found through the search engine of the archive or repository in which it is stored (please specify)
- Yes, it can be found through an online (metadata) catalogue or web portal (please specify)

Data will be deposited on EBI ENA or NCBI SRA, depending on the data type. Large Bionano data will be made available through DANS EASY. All three databases create DOIs and require metadata.

3.2 I will use a metadata scheme for the description of my data collection (note: this is key item 7, which you should report to ZonMw at the end of your project).

- Yes, I will use a metadata scheme specific for my field of research (please specify)

We will use the metadata scheme used by either NCBI/SRA or EBI/ENA, whichever we will choose for submission:

<https://www.ncbi.nlm.nih.gov/sra/docs/submitmeta/>

<https://ena-docs.readthedocs.io/en/latest/submit/analyses/bionano-maps.html>

3.3 I will be using a persistent identifier as a permanent link to my data collection (note: this is key item 1, which you should report to ZonMw at the end of your project).

- Yes, I will be using the DOI code

4. Making data accessible

4.1 Once the project has ended, my data will be accessible for further research and verification.

- Yes, immediately

4.2 Once the project has ended, my data collection will be publicly accessible, without any restrictions (open access).

- Yes, proceed to section 5 (Making data interoperable)

4.3 I have a set of terms of use available to me, which I will use to define the requirements of access to my data collection once the project has ended (please provide a link or persistent identifier; also note that this is a key item 4, which you should report to ZonMw at the conclusion of your project).

Question not answered.

4.4 In the terms of use restricting access to my data, I have included at least the following:

Question not answered.

5. Making data interoperable

5.1 I will select a data format, which will allow other researchers and their computers (machine actionable) to read my data collection (note: this is key item 5, which you should report to ZonMw at the end of your project).

- Yes (please specify)

All supplied data files, such as Bionano data and genome assembly and annotations, will be made available in by industry accepted data format. These include BNX, fasta and GFF3 files as well as a tabular format for the metadata.

5.2 I will select a terminology for recording my data (e.g., code, classification, ontology) that allows my dataset to be linked or integrated with other datasets (note: this is key item 6, which you should report to ZonMw at the conclusion of your project).

- Yes, metadata standard (please specify)

Plant (phenotype) data will be classified according to the Plant Trait Ontology (<https://www.ebi.ac.uk/ols/ontologies/to>)

5.3 I will be doing research involving human subjects, and I have taken into account the reuse of data and the potential combination with other data sets when taking privacy protection measurements.

- No (please explain)

No human data are being used or generated.

6. Making data reusable

6.1 I will ensure that the data and their documentation will be of sufficient quality to allow other researchers to interpret and reuse them (in a replication package).

- I will document the research process (please explain)
- I will perform quality checks on the data to ensure that they are complete, correct and consistent (please explain)
- I will document the software used in the course of the project (please specify)

The data measurement process will be documented in the WUR E-Labjournal. Software documentation will include a description of the approach, regular well-described commits and version controlled. Further computer analysis will be detailed in a report/publication.

The Bionano data will go through a QC at the hotel.

6.2 I have a number of selection criteria, which will allow me to determine which part of the data should be preserved once the project has ended. (see also question 1.9 and 6.1)

- Yes

Intermediate data, eg. data which can be derived from the available raw data following the described approaches and which are not the end result, will be deleted.

6.3 Once the project has ended and the data have been selected, I can make an estimate of the size of the data collection (in GB/TB) to be preserved for long-term storage or archival.

- Yes (please specify)

It will be around 50 GB to 100 GB.

6.4 I will select an archive or repository for (certified) long-term archiving of my data collection once the project has ended. (note: this is a key item, which you should report to ZonMw at the conclusion of your project)

- Yes, and this archive has a data seal of approval (please specify the archive)

Main data will be stored in DANS EASY. However, the project (meta)data, genome data and annotations will be made available through EBI/NCBI. These are the industry standards and well-funded, but not certified.

6.5 Once the project has ended, I will ensure that all data, software codes and research materials, published or unpublished, are managed and securely stored. Please specify the period of storage.

- Yes, in accordance with VNSU guidelines (please specify the number of years)

Data and software will not be removed from public repositories.

6.6 Data management costs during the project and preparations for archival can be included in the project budget. These costs are:

Storage and compute costs are included in the costs of the project hours and sample delivery. DANS and EBI/NCBI are free of charge.

6.7 The costs of archiving the data set once the project has ended are covered.

- Yes (please elaborate)

DANS and EBI/NCBI are free of charge.