
Molecular marker assisted selection for improved seed protein content in dry bean

A Data Management Plan created using DMP Assistant

Creator: K Peter Pauls

Principal Investigator: K Peter Pauls

Data Manager: K Peter Pauls

Affiliation: University of Guelph

Template: University of Guelph Agri-Food Funding Template

ORCID ID: 0000-0003-0636-9151

Project abstract:

Access to high quality protein in human diets will become increasingly challenging in a world experiencing steady population expansion and climate change. Both challenges can be met by the production of grain crops, like dry beans, that are rich sources of plant-based protein and have the capacity to fix atmospheric nitrogen. The proposed work will build the tools to accelerate the selection of superior dry bean varieties with enhanced levels of seed protein. This will include an assessment of the range of seed protein levels that are in the current elite, high yielding varieties in the dominant market classes, an extended survey of dry germplasm from various sources for high protein genotypes and the identification of single nucleotide polymorphism (SNP) markers associated with high protein levels that can be used to rapidly introgress that trait into superior bean varieties for commercial production.

Identifier: 7063

Last modified: 01-03-2022

Grant number / URL: UG-T1-2021-101066

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

Molecular marker assisted selection for improved seed protein content in dry bean

Data collection

Provide an overview of the data that will be generated, collected or acquired to support this project. If data will be acquired from a third party, specify the source.

Bean variety identification information, seed protein content and genomic DNA sequence information. The data will be used to identify molecular markers related to protein content. Data is not sensitive in nature.

What method(s) of data collection will be employed?

Data will be derived from analysis of seeds and DNA collected from leaves of different bean varieties.

What types of data will be included?

Numeric data.

What software or digital formats will be used to collect, manage and analyze the data?

MS Excel.

Provide an indication of the scope of the data?

Approximately 250 bean varieties will be analysed for seed protein and genomic DNA variants.

Data storage

Estimate the size of data storage that will be required.

Approximately 1 GB of data will be generated.

Where will your data be stored during the collection, collation and analysis phases of the project?

All data is stored on a password protected laptop and a second copy is stored in a drive on a server in the Plant Agriculture Department.

What backup strategy will be employed?

The data will be backed up on a remote hard drive that is backed up when the laptop is attached to the desk station in the departmental office. The system is installed and maintained by desktop management services CCS.

How will your data files be organized? What file naming conventions will you use? A brief overview or example would be adequate.

Folders will be created for the data in each phase: raw, cleaned, collated and final. Files within each folder will be named with a combination of Project name_Creator surname and initials_Date of creation YYYYMMDD. Example: NavyProtein_SmithT_20210319.xlsx.

What metadata will be developed for your data? Will there be supplemental documentation prepared to assist with the interpretation and analysis of your data?

Each column in the spreadsheet will have a simple text heading [e.g. "bean variety" or "bean market class" or "seed protein content (%)"]. Supplementary text documents to describe methods used to obtain coding or numeric information will be added to assist understanding of the data.

Data archiving and preservation

Will you deposit your data in the UG data repository or an external data repository? If you are opting to not archive your data in a repository, where will your data be housed after completion of your project?

The data will be archived in the UG data repository for long-term preservation.

Discuss any data transformations that will be needed so your data is preserved in appropriate, non-proprietary formats.

The data will be exported from Excel and preserved as plain text CSV files.

If some of your data will not be preserved, how long will you retain it? Will the non-preserved data be destroyed?

The raw data will be retained for three years and then destroyed.

Sharing and reuse

Will the data that you archive in a data repository be made available for sharing and reuse by other researchers?

The data will be freely and openly shared through the UG data repository.

Explain which version of your data or subset of your data will be shared.

A version of the final data will be shared.

When will your data be available for discovery by other researchers? Will you impose an embargo on publication of your data? If so, please provide details on the duration of the embargo.

The data will be available to others after it has been published.

Will you limit who can access your data? If so, who will that be and why are you limiting the data's reuse?

Final, data will be openly available. The raw data may be provided to researchers who submit a request to the PI.

Are there specific license terms you will assign to users of your data?

The data will be licensed with a CC Attribution-ShareAlike license.

Restrictions/limitations

Are there limitations or constraints on how you manage your data resulting from legal, ethical or intellectual property concerns?

There are no limitations or constraints on how the data is managed.

Would your data need to be anonymized or de-identified before being shared with others?

The data will not need to be de-identified before sharing.

Confidential information

What information do you want to include in your DMP that should not be publicly shared?

None