

INSTRUCTIONS TO SUBMIT A DATASET

Type of datasets accepted

We accept a dataset if it contains genotypic data for at least one species. However, we highly recommend that it also includes one or more phenotypes.

Details needed to add a dataset

Please have the following details ready to ensure a smooth dataset submission process:

- DOI of the location where the data is hosted*
- Direct link to access the dataset*
- Link to the publication related to the dataset or the associated research*
- Species to which the data belongs
- Number of samples and markers in the dataset*
- A simple code snippet to read the data after downloading. Please refer to the examples provided in the web application to write clear and readable code. The code is intended to help researchers easily access the dataset and relevant files from the hosting websites. An example is shown below:

```
# Get geno
geno <- read.csv('location/filename.csv')
dim(geno)

# Get pheno
pheno <- read.table('location/filename.tab')
head(pheno)
```

Some suggestions to help you write clear and readable code:

- ⇒ The dataset may come in various formats such as VCF, BED, TXT, CSV, TAB, etc. Please use the appropriate libraries and syntax to read the data correctly.
- ⇒ Include helpful comments in the code to guide researchers, such as explanations of the steps, file structures, or any preprocessing that might be required.

STEPS TO ADD THE DATASET

- 1) Please enter all required fields (*) and any additional details into the web application as prompted.
 - If a dataset matching the provided details already exists in our database, its information will be displayed.
 - If the dataset includes new information—such as a different species, a different number of markers, etc.—please continue filling in the remaining fields accordingly.
- 2) Make sure to validate the fields by clicking the appropriate buttons.
- 3) Once all the fields are completed, click the **Preview** button to review the dataset information. The preview will appear similar to the example shown below:

Input	Value
Unique Name	00138_BlueberryInsightsPolyploidBlackberry
Data DOI	https://doi.org/10.5061/dryad.kd4jq6h
Article Publication link	
Title of the Article	Insights Into the Genetic Basis of Blueberry Fruit-Related Traits Using Diploid and Polyploid Models in a GWAS Context
Abstract/Description	
Species Scientific Name	
Species Common Name	Blackberry
Data Sharing Link	https://datadryad.org/dataset/doi:10.5061/dryad.kd4jq6h
Authorization for Accessing Data	false
Phenotypic Data	false
Genetic Map	false
Pedigree Information	false
Number of samples	3
Number of markers	100
Data Downloading Instructions	
Article Tags	Genetic Basis; Blueberry Traits Diploid; Polyploid Models; GWAS Context NULL

Proceed

- If the preview table does not appear, it may be due to one of the following reasons:
 - ⇒ Required fields are left empty
 - ⇒ The Data DOI, Data sharing link, or Publication link has not been validated
 - ⇒ The dataset already exists in our database
 - ⇒ If all the information seems correct but the preview still doesn't appear, please contact us—we'll be happy to look into it.
- 4) In the preview, verify that all fields are correct and fill in any additional details that may be helpful. You can edit certain fields by double-clicking on them. Note that some fields cannot be edited, as they were validated in the previous steps.

- 5) If everything looks correct, click the **Proceed** button.
- 6) Download the dataset folder by clicking the **Download Dataset Folder** button.
- 7) Unzip the downloaded content and verify its contents. The folder should include the following items:
 - **read_data_code.R**: Contains the code to read the dataset. Please review this file if a code snippet was submitted.
 - **Output folder**: This folder contains the following three files:
 - ⇒ **meta_data.md**: A markdown file with a description of the related publication.
 - ⇒ **meta_data.json**: A JSON file containing metadata details about the dataset.
 - ⇒ **citation.bib**: A BibTeX file with the citation for the dataset.
- ⚠ *We recommend not editing any of these files.*
- 8) Follow the instructions provided on the web application, or refer to the steps below, to submit the dataset to our repository.
 - Go to [GPDatasets GitHub Repository](#)
 - Sign in with your GitHub credentials
 - Drag the **unzipped** dataset folder into the DATASETS folder in GitHub.
 - Open Source Control from the left panel, or press CTRL+SHIFT+G to access it quickly.
 - Copy and paste the commit message provided by the web application, which may look similar to the example below:


```
Added 00XXX_UniqueName
```
 - Press CMD + ENTER to continue, then Press ENTER 3 times to successfully create a pull request.
- 9) After creating the pull request, return to the web application and click the **Submit** button at the bottom of the page to verify the pull request was created successfully.
 - If an error message appears after clicking *Submit*, it means the pull request was not created. In that case, please carefully follow the steps above to resubmit the pull request.
- 10) A member of our team will review the pull request and approve it once all the information has been verified.

If there was any problem in submitting the dataset, please email us at gustavoc@msu.edu (or) neelamha@msu.edu with the following details.

- Description of the problem
- Some basic dataset information

We sincerely thank you for contributing your dataset.