



The James
Hutton
Institute

The James Hutton Institute

Information & Computational Sciences

Germinate 3

User Guide

Version 3.4.0

September, 2017

Developers:

Paul Shaw
Sebastian Raubach
Iain Milne
Gordon Stephen
David Marshall

Contents

1	Introduction	1
2	Overview	2
2.1	Authentication	2
2.2	Page layout	2
3	Features	4
3.1	User interface language	4
3.2	Menu	4
3.3	Data filtering	4
3.4	Search	4
3.5	Groups	4
3.5.1	Creating a group	5
3.5.2	Marked item lists	8
3.6	Help	8
4	Data Types	9
4.1	Passport Data	9
4.1.1	Multi-Crop Passport Descriptors	9
4.2	Genotypic Data	9
4.2.1	Allele Frequency Data	9
4.2.2	Genotypic Maps	10
4.2.3	Genetic Markers	10
4.3	Phenotypic Trials Data	10
4.4	Climate Data	10
4.5	Chemical Compound Data	10

1 Introduction

TEST

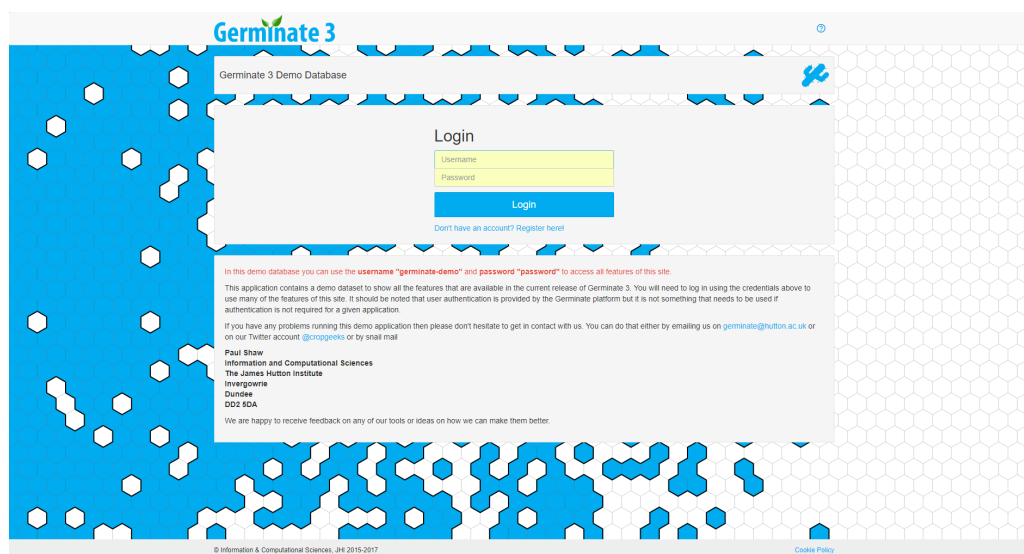


Figure 1: Depending on the configuration of Germinate, you may be asked to log in.

2 Overview

In this section, we will explain the overall structure of Germinate along with an overview of the various data types that Germinate supports.

2.1 Authentication

Germinate can be used with or without user authentication. If the administrator of Germinate decided to enable authentication, you will be asked to log in using a username and password. Figure 1 shows the login page of Germinate. If you already have a user account, simply enter the username and password into the provided text boxes.

If you do not have a user account, click on the link below the login button to create an account. You may get asked to agree to a license agreement before being able to create an account.

2.2 Page layout

Figure 2 shows the main layout of the Germinate web interface. The main menu of Germinate is positioned to the left (Figure 2A). It is used to navigate between pages. Submenus can be expanded by simply clicking on the items showing the caret symbol. The menu is explained in more detail in Section 3.2. The overall search feature, shown in Figure 2B, allows you to run a full-text search across the whole database. The results will be grouped into topical categories. See Section 3.4 for more details. The interface has a banner along the top containing the Germinate logo and a few dropdown menu items in the top right corner (c.f. Figure 2C). These items include the language selector which will be covered in Section 3.1, social media buttons, the marked item lists covered in Section 3.5.2, a user menu with specific functions based on your type of account and, finally, a help button that can be clicked to get more information about the current page (c.f. Section 3.6). Recent news about both the Germinate interface and the contained data are available in the news section shown in Figure 2.

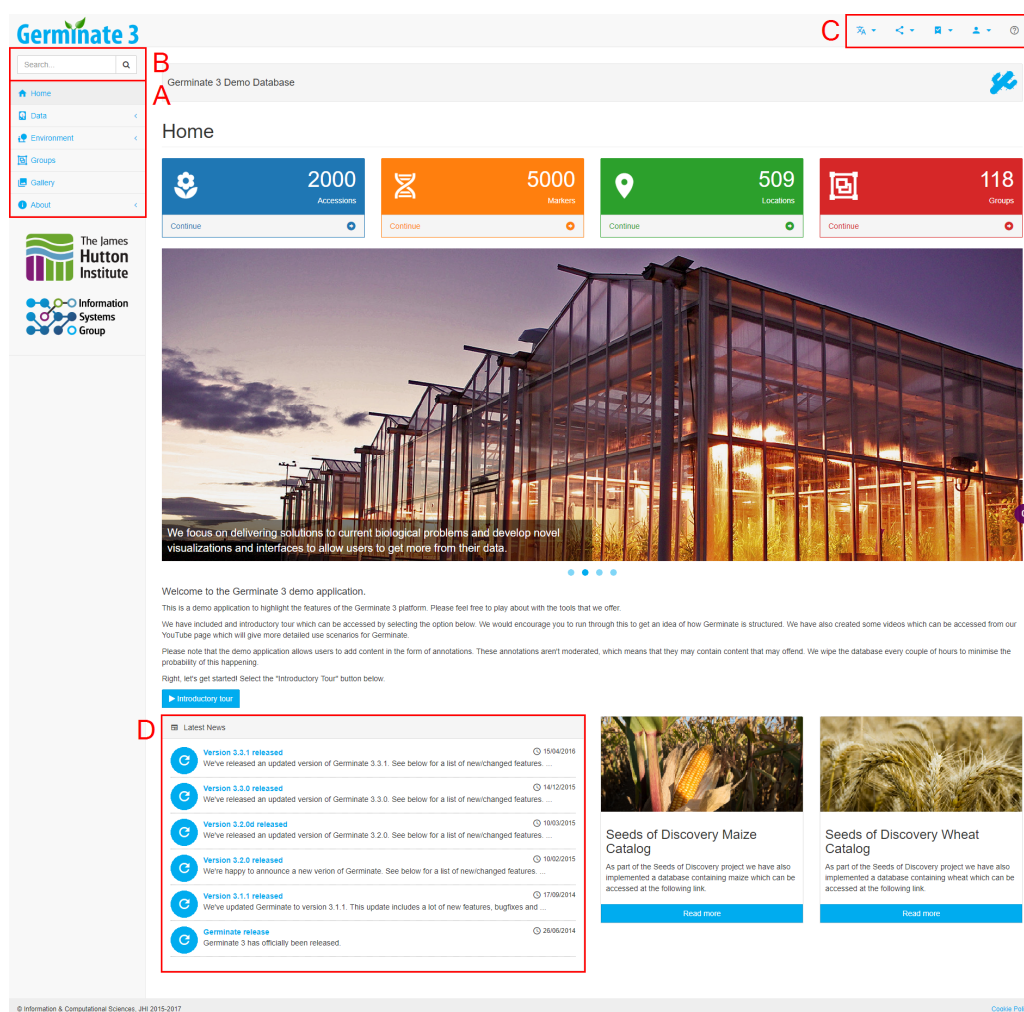


Figure 2: The home page of Germinate is the first page you will see. (A) The main menu of Germinate used to navigate the page. (B) The search box used for free-text searches of the database. (C) Language selector, social media buttons, marked item lists, user menu and the help button. (D) Latest news about this instance of Germinate and the contained data.

3 Features

We are now going to highlight some of the most important features of Germinate. All features shown in this section are applicable across various data types and pages.

3.1 User interface language

Germinate fully supports internationalization. This means that the interface can be translated into any number of languages. By default, Germinate is distributed in English. Depending on the configuration of Germinate, other languages may be available. You can switch between them by selecting a language from the first dropdown box shown in Figure 2C.

3.2 Menu

Germinate's main menu (see Figure 2A) is positioned along the left of the web interface on desktops, although it may collapse into an expandable menu along the top on smaller devices. Clicking on a submenu will expand the corresponding category whereas selecting a menu item will navigate to that page.

Please note, that depending on the configuration of the Germinate instance you are using, the menu might have a different structure.

3.3 Data filtering

A lot of tables you encounter in Germinate support filtering based on columns. These tables will have a filter icon in the top-left corner which, when clicked on, opens the table column header filters. These are text boxes that, depending on the data type of the column, allow you to specify a single value/date or an interval of values/dates. Once this filter is applied, the table will only show items that match the given filter query. By default the individual filters are combined with a boolean "And", meaning that all of them have to be true at the same time. You can switch this to a boolean "Or" to get results where at least one of the individual filters is fulfilled.

This is a particularly useful feature if you want to find items that, e.g., are from a specific country, or that have a phenotype value within a certain range. The resulting items can then be marked or put into a group.

3.4 Search

Germinate supports full-text search across various data types and columns within this type. You can search by typing a search query into the text box shown in Figure 2B. Upon search, Germinate shows the search results page. This page shows all matching database objects grouped into categories, each category representing a different data type. Figure 3 shows an example of the search results that Germinate provides. The number of matching items is shown for each section on the right. Upon expanding of a section, a result table will show the matching database items. You can then either download the data, mark specific items (cf. Section 3.5.2) or filter down further by adjusting the search criteria in the table header (cf. Section 3.3).

3.5 Groups

In Germinate we define the concept of a group to be an arbitrary grouping of database items of a certain type. Germinate supports groups of *accessions*, *markers* and *locations*. These groups can be pre-created by an administrator or user-defined, which means that you can create your own groups (assuming user authentication is enabled).

Search

Select what data you want to search for
All data

Please enter the search query
cactuar-1

Results "cactuar-1"

Accession Data 1

Accession Attribute Data 5

Phenotype Data 102

Compound Data 5

Table filtering searches for exact matches in the specified column. To use fuzzy search, use the wildcard character %. As an example, searching for a country name of kingdom will not return a result whereas searching for %kingdom% will return United Kingdom.

Id From To

Id	ACC-1	Accession Name	Name	Dataset Description	Analysis Method	Unit	Value
1	ACC-1	CACTUAR-1	Mithril	Sample Compound Data		Kilograms	56.20
1	ACC-1	CACTUAR-1	Adamantium	Sample Compound Data		Kilograms	74.60
1	ACC-1	CACTUAR-1	Unobtainium	Sample Compound Data		Kilograms	72.05
1	ACC-1	CACTUAR-1	Finikonium	Sample Compound Data		Kilograms	6.47
1	ACC-1	CACTUAR-1	Element Zero	Sample Compound Data		Kilograms	94.17

Map Definition Data 4

Datasets 4

Dataset Attribute Data 4

Pedigree Data 4

Location Data 4

Figure 3: The search results page for the search term "cactuar-1". Results are grouped into categories with the number of matching database items highlighted. The result tables itself can be filtered further to restrict the number of results. In this example the "Compound Data" section is expanded and shows the five results that match the search query.

The purpose of these groups becomes clear once you start exporting data. All types of data can either be exported for the whole dataset or the data can be subset into smaller chunks by selecting a single or a selection of groups. The exported data will then contain information about the selected groups only.

An example of this is shown in Figure 4 where we selected a group of accessions (192 out of 2000 accessions) and a group containing a single marker (out of 5000). The resulting data file will consequently contain at most 192 rows and at most a single column of data (it may contain less, if the selected database items are not actually part of the selected dataset).

3.5.1 Creating a group

This section is only applicable if the Germinate instance you are using has user authentication enabled.

In addition to using the predefined groups, you can create new groups of your own. There are multiple ways in which you can create a new group and add items to it. One option is to go to the *Groups* page of Germinate. This page shows you all the existing groups in a table and upon selection, shows you its group members. Figure 6 shows you an example of what the groups page can look like. In this example, Germinate contains 118 different groups that the current user can see. New groups can be added and existing ones deleted by pressing the buttons below the groups table (Figure 6A). Deleting a group requires you to select the checkbox in the corresponding table row as well as to have sufficient permissions to do so. When creating a new group you will be asked to select the group type and to decide on a name for the group. When you do so, the group will be associated with your user account.

Genotype data

Select accession groups

Non-contiguous groups can be selected using **Ctrl** + **Click**. Use **Shift** for sequential groupings.

All accessions (2000)
Subselection based on pedigree (192)

☒ Select all ☐ Create group from cart

CDF (Crap Data Filter)

- This filter automatically prunes data which fails to meet minimum quality requirements.
- Current levels are 25% for the heterozygous filter and 50% for the missing data filter.
- The current level is 50% for the missing data filter.
- Loci which do not meet these conditions are removed from the database output.

Missing Data Filter

Use Missing Data Filter?

☒ Yes

Select map to use in export

All markers (5000)

Select marker groups

Non-contiguous groups can be selected using **Ctrl** + **Click**. Use **Shift** for sequential groupings.

All markers (5000)
Single marker (m252621) (1)

☒ Select all

Figure 4: Example of the group sub-selection feature: You can select both accession and marker groups during the genotypic data export process to subset the dataset.

Search by criteria

countries.country_name >= %United Kingdom%

locations.latitude >= 55

Search result

	Id	GID	Accession Name	Genus	Species	Subtaxa	Latitude	Longitude	Elevation [m]	Country of origin	
<input checked="" type="checkbox"/>	176	ACC-176	CACTUAR-176	Cactus	Cactuar		55.83	-4.25	19.21	United Kingdom of Great Britain and Northern Ireland	<input type="checkbox"/>
<input checked="" type="checkbox"/>	382	ACC-382	CACTUAR-382	Cactus	Cactuar		55.83	-4.25	19.21	United Kingdom of Great Britain and Northern Ireland	<input type="checkbox"/>
<input checked="" type="checkbox"/>	1048	ACC-1048	CACTUAR-1048	Cactus	Cactuar		55.83	-4.25	19.21	United Kingdom of Great Britain and Northern Ireland	<input type="checkbox"/>
<input checked="" type="checkbox"/>	1489	ACC-1489	CACTUAR-1489	Cactus	Cactuar		55.83	-4.25	19.21	United Kingdom of Great Britain and Northern Ireland	<input type="checkbox"/>
<input checked="" type="checkbox"/>	1586	ACC-1586	CACTUAR-1586	Cactus	Cactuar		55.83	-4.25	19.21	United Kingdom of Great Britain and Northern Ireland	<input type="checkbox"/>

(A) Query
(B) Result

Figure 5: The group member search interface. (A) shows the boolean query interface where you can define the attributes an item needs to fulfil. In this case, the accession has to be collected in the UK at a latitude larger than 55. (B) The table showing the matching items. There are five accessions that match the query.

Once this is done, the group will be created and Germinate will automatically select it and show the group members table (empty at this point) below the groups table. You can now manipulate the group itself by adding and removing members using the buttons below the table as shown in Figure 6C.

Adding members to an existing group can be achieved in two ways. You can upload a list of those items from a text file or your clipboard and Germinate will look these items up based on their identifier. Once found they will be added to the group. The other option is to use a boolean search feature that is similar to the way the table filtering works (cf. Section 3.3). You can choose fields from the database tables and specify values that the items in questions should equal, smaller or larger to. Figure 5 shows what the group member search interface looks like. The query is specified in Figure 5A and the result shown in Figure 5B.

Groups can be made public so that other users have the option to use them as well. If you decide to make your group public, toggle the switch button shown in Figure 6B.

Groups

Table filtering searches for exact matches in the specified column. To use fuzzy search, use the wildcard character %. As an example, searching for a country name of kingdom will not return a result whereas searching for '%kingdom%' will return United Kingdom.

	id	Group Description	Type	User	Size	Created on
<input type="checkbox"/>	1	All accessions	Accessions	Sebastian Raubach	2,000	08/04/2016
<input type="checkbox"/>	117	Subselection based on pedigree	Accessions	Administrator	192	27/06/2016
<input type="checkbox"/>	4	Afghanistan	Locations	Sebastian Raubach	1	
<input type="checkbox"/>	5	Algeria	Locations	Sebastian Raubach	2	
<input type="checkbox"/>	3	All collecting sites	Locations	Sebastian Raubach	500	08/04/2016
<input type="checkbox"/>	6	Angola	Locations	Sebastian Raubach	1	
<input type="checkbox"/>	7	Argentina	Locations	Sebastian Raubach	6	
<input type="checkbox"/>	8	Armenia	Locations	Sebastian Raubach	1	
<input type="checkbox"/>	9	Australia	Locations	Sebastian Raubach	5	
<input type="checkbox"/>	10	Austria	Locations	Sebastian Raubach	1	

☐ Delete group ☐ Add group **A**

Group members All accessions

Is public? ☒ Yes ☐ No **B**

	id	GID	Accession Name	Number	Collector Number	Genus	Species	Subtaxa	Latitude	Longitude	Elevation [m]	Country of origin	Collection date	
<input type="checkbox"/>	1	ACC-1	CACTUAR-1	Prickly Pete	VLS-767	Cactus	Cactuar		26.18	91.73	44	India	12/03/2013	<input type="checkbox"/>
<input type="checkbox"/>	2	ACC-2	CACTUAR-2	Spiky Steve	PAT-337	Cactus	Cactuar		-33.86	151.21	37.98	Australia	20/11/1959	<input type="checkbox"/>
<input type="checkbox"/>	3	ACC-3	CACTUAR-3	Spiny Shaun	XME-467	Cactus	Cactuar		35.69	-0.64	109.97	Algeria	14/02/1988	<input type="checkbox"/>
<input type="checkbox"/>	4	ACC-4	CACTUAR-4	Acanthoid Al	DUL-555	Cactus	Cactuar		30.20	67.01	1,681.96	Pakistan	07/10/1974	<input type="checkbox"/>
<input type="checkbox"/>	5	ACC-5	CACTUAR-5	Aculeate Anna	SNW-759	Cactus	Cactuar		36.06	103.79	1,546.23	China	03/04/2004	<input type="checkbox"/>
<input type="checkbox"/>	6	ACC-6	CACTUAR-6	Barbed Beth	EEH-398	Cactus	Cactuar		36.30	59.61	982	Iran (Islamic Republic of)	08/08/2005	<input type="checkbox"/>
<input type="checkbox"/>	7	ACC-7	CACTUAR-7	Briery Bruce	JKC-326	Cactus	Cactuar		28.67	77.43	215.46	India	23/01/1992	<input type="checkbox"/>
<input type="checkbox"/>	8	ACC-8	CACTUAR-8	Bristly Barbara	SKS-889	Cactus	Cactuar		38.41	27.14	121.05	Turkey	30/11/1975	<input type="checkbox"/>
<input type="checkbox"/>	9	ACC-9	CACTUAR-9		YLF-787	Cactus	Cactuar		41.85	-87.65	178.69	United States of America	17/05/1984	<input type="checkbox"/>
<input type="checkbox"/>	10	ACC-10	CACTUAR-10		RMB-723	Cactus	Cactuar		22.99	120.20	21.46	Taiwan, Province of China	24/08/1963	<input type="checkbox"/>

☐ Delete members Search members **C**

Figure 6: The groups page has many ways in which you can manipulate existing groups and create new groups. The top table shows all the groups that are visible to the current user. The bottom table shows the members of this group, i.e. the database items that are part of it; depending on the type of the group, this can be either *accessions*, *markers* or *locations*. (A) Groups can be added and deleted by clicking on the buttons located just below the groups table. Deleting groups requires the checkbox of the corresponding table row to be selected. (B) The group visibility can be changed by toggling this switch. A public group is visible to every user whereas an invisible group is only visible to the owner. (C) Group members can be added and removed by clicking on the buttons below the group member table.

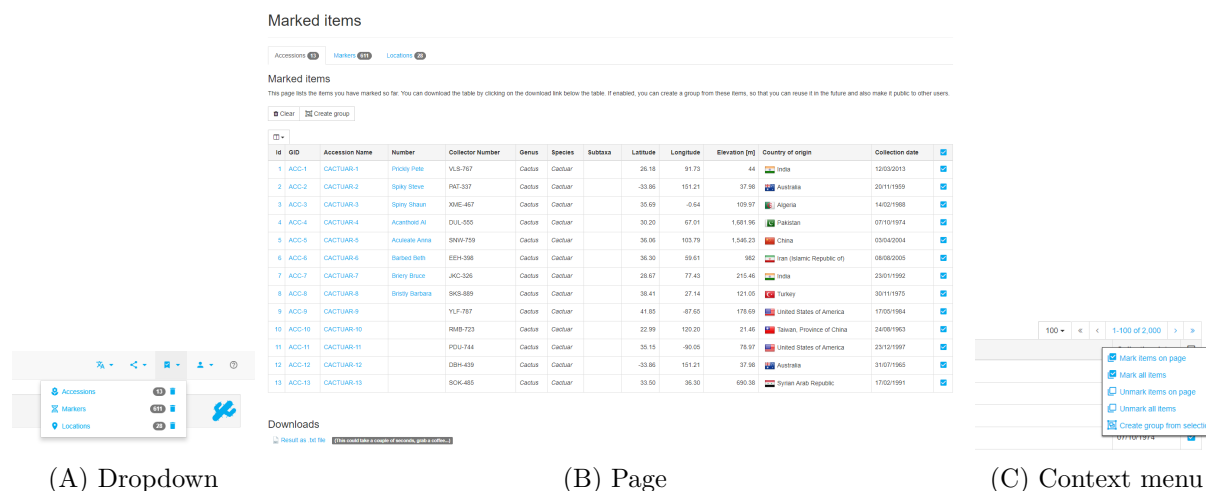


Figure 7: The marked item lists keep track of items of interest. (A) The dropdown menu shows a quick overview of how many items of each type are currently marked. The bin icon lets you easily empty the marked item list of a certain type. (B) The actual marked item lists page shows three tabs, one for each item type. The table under each tab shows all marked items of this type. You can either download this information or create a group from the list (if this is enabled). (C) Many tables have a checkbox column that you can use to mark/unmark items. You can also create a group from this context menu (if this is enabled).

3.5.2 Marked item lists

Another useful feature of Germinate is the concept of *marked item lists*. A marked item is either an accession, a marker or a location that is of interest to the user. While you are browsing the page, a lot of the tables will have a checkbox column as the last column which you can use to mark certain items. Germinate will keep track of these items for you.

To see how many items you currently have marked, you can click on the menu item as shown in Figure 7A or go directly to the marked item lists page that is shown in Figure 7B.

Once you have marked all the items that you are interested in, you can create a group of these items and use them to export data against them. To create a group, you can either go to the marked item lists page or by clicking on the header of the checkbox column and selecting "Create group from selection" (see Figure 7C).

3.6 Help

Whenever you are stuck and are not sure what to do or how to use the web interface, you can click on the last item in Figure 2C. This will open a popup with information about the page you are looking at. Sometimes this information may provide links to external resources or external tools, other times it will just explain certain parts of the page in more details.

Please note that help information may not be available for all pages. In those cases, the menu item will be disabled. We are trying to make the user interface as intuitive as possible and are working towards adding help information to all the pages. If at any point you have questions about Germinate in general or specific features of the web interface, please do not hesitate to contact us on: germinate@hutton.ac.uk.

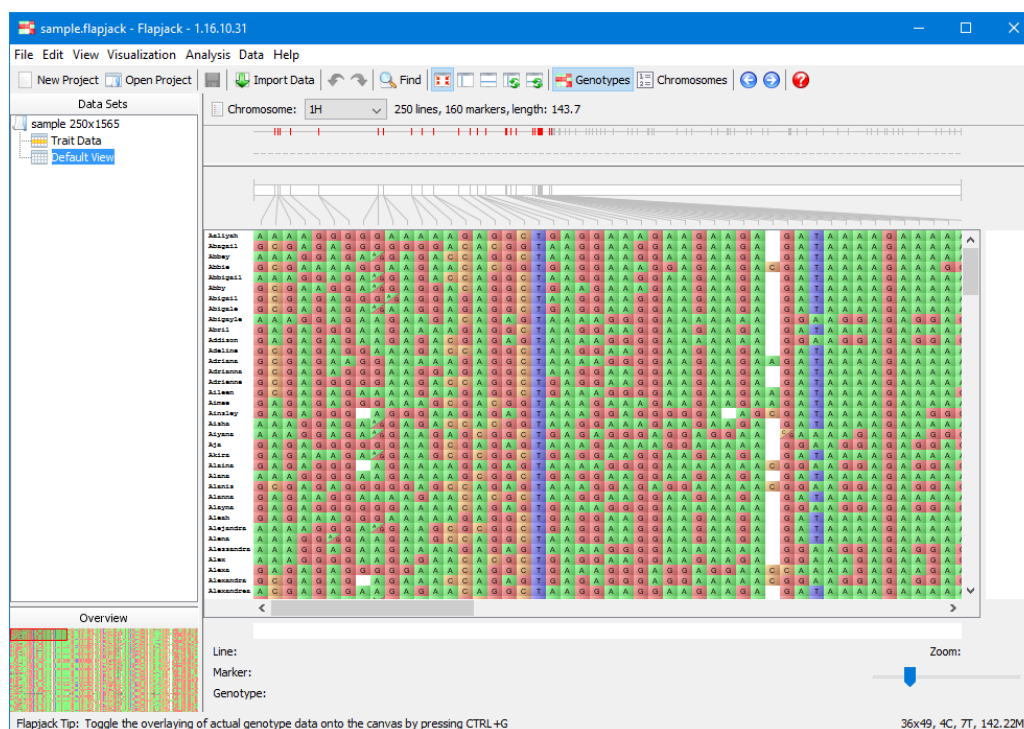


Figure 8: Genotypic data exported from Germinate and visualized in Flapjack.

4 Data Types

The following section will describe each data type that Germinate can handle in more detail. We will describe both the web interface that is used to display this data as well as what the export formats for each of the types are.

4.1 Passport Data

4.1.1 Multi-Crop Passport Descriptors

The Multi-Crop Passport Descriptors (MCPD) [1] is a widely used international standard to facilitate germplasm passport information exchange defined by the FAO. Germinate is fully MCPD V.2.1 compatible. The MCPD standard is used by many genebanks and genetic resources tools and utilities.

4.2 Genotypic Data

When we talk about genotypic data in the context of Germinate, we are referring to Single Nucleotide Polymorphic (SNP) or Simple sequence repeat (SSR) data. The data export process is shown in Figure 4. After selecting the dataset you want to export, you can decide which accessions and markers should be included in the output. Data can be exported against different maps (cf. Section 4.2.2), e.g. physical vs. genetic marker positions.

The data is exported into a tab-delimited text format as well as Flapjack [2] format. Figure 8 shows an example of data exported from Germinate visualized in Flapjack.

4.2.1 Allele Frequency Data

4.2.2 Genotypic Maps

4.2.3 Genetic Markers

4.3 Phenotypic Trials Data

Phenotypic data is a big part of Germinate. We put a lot of effort into developing meaningful visualizations as well as functionality and interoperability with our other software tools.

After selecting a dataset (or multiple datasets), you will have the choice between different visualizations and the data download. The first tab shows an overview over the data within the selected datasets, whereas the second tab lets you plot two phenotypes against each other in a scatter plot. This is particularly useful to see if there is any correlation between them. Hovering over data points shows the values per dimension as well as the accession that is responsible for this data point. Clicking on this data point will take you to the passport page for this accession. You can draw a shape around data points of interest by clicking and dragging the mouse across the chart. This will highlight the data points within the shape. You can then either right-click or use the icon in the top right of the chart to add/remove these items to/from the marked item list.

4.4 Climate Data

4.5 Chemical Compound Data

References

- [1] A. Alercia, S. Diulgheroff, and M. Mackay. FAO/Bioversity Multi-Crop Passport Descriptors V.2.1. <https://www.bioversityinternational.org/e-library/publications/detail/faobioversity-multi-crop-passport-descriptors-v21-mcpd-v21/>, 2015. Last accessed on 2017-09-15.
- [2] I. Milne, P. Shaw, G. Stephen, M. Bayer, L. Cardle, W. T. B. Thomas, A. J. Flavell, and D. Marshall. Flapjack—graphical genotype visualization. *Bioinformatics*, 26(24):3133–3134, 2010.