





















# The James Hutton Institute Information & Computational Sciences



User Guide Version 3.4.0 September, 2017

Developers:

Paul Shaw Sebastian Raubach Iain Milne Gordon Stephen David Marshall Germinate 3 Contents

# Contents

1	Intr	roduction	1	
2	Ove	erview	2	
	2.1	Authentication	$\overline{2}$	
	2.2	Page layout	2	
3	Feat	tures	4	
	3.1	User interface language	4	
	3.2	Menu	4	
	3.3	Data filtering	4	
	3.4	Search	5	
	3.5	Groups	5	
		3.5.1 Creating a group	6	
		3.5.2 Marked item lists	8	
	3.6	Charts	8	
	3.7	Help	8	
4	Dat	a Types	10	
	4.1	Passport Data	10	
	4.2	Pedigree Data	11	
	4.3	Genotypic Data	11	
		4.3.1 Allele Frequency Data	11	
		4.3.2 Genotypic Maps	11	
		4.3.3 Genetic Markers	11	
	4.4	Phenotypic Trials Data	11	
	4.5	Location Data	12	
	4.6	Climate Data	12	
	4.7	Chemical Compound Data	12	
	4.8	Attribute Data	12	

Germinate 3 Introduction

# 1 Introduction

Welcome to the introduction to Germinate.

Germinate 3 Overview

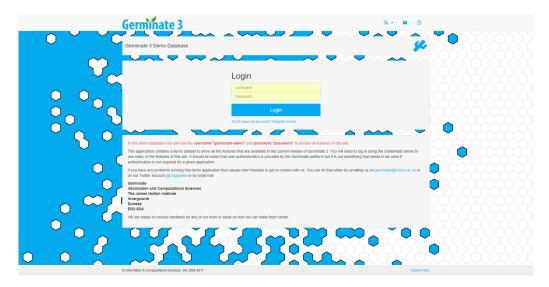


Figure 1: Depending on the configuration of Germinate, you may be asked to log in.

# 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.

To modify your user account, you can log in to Germinate Gatekeeper, which is Germinate's user authentication portal. It can be accessed from the help popup on the login page.

## 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 (cf. 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, a "contact us" button and, finally, a help button that can be clicked to get more information about the current page (cf. Section 3.7). An overview of the number of database items for certain types is shown in Figure 2D. Recent news about both the Germinate interface and the contained data are available in the news section shown in Figure 2E. Finally, a section about other projects that are related to the project you are currently looking at are available in

Germinate 3 Overview

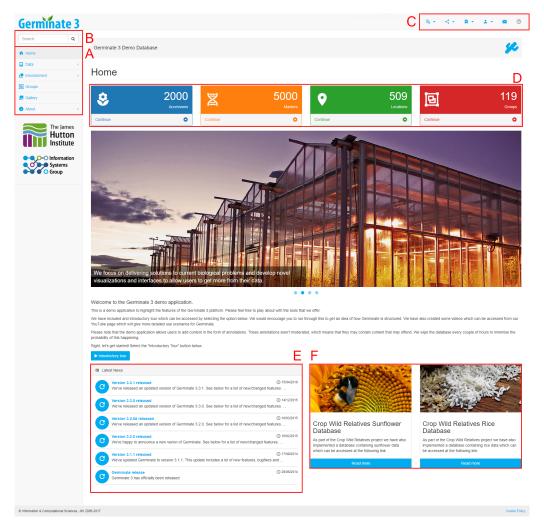


Figure 2: The home page of Germinate is the first page you will see. (A) The main menu of Germinate used to navigate the page. (B) The search box used for free-text searches of the database. (C) Language selector, social media buttons, marked item lists, user menu and the help button. (D) An overview of number of data objects that are stored in Germinate. (E) Latest news about this instance of Germinate and the contained data. (F) Other projects that have a relation to the current project.

Figure 2F.

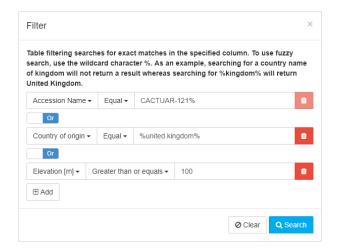


Figure 3: The table filter interface allows you to define boolean combinations of restrictions for the columns. These support wildcards as well as a variety of different comparison types.

#### 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 filters.

The filter interface can be seen in Figure 3. Each row represents a restriction. It consists of a table column, a comparison operator and, depending on this operator, either one or two values. Multiple rows are combined with boolean combinations. 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. Once this filter is applied, the table will only show items that match the given overall filter query.

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

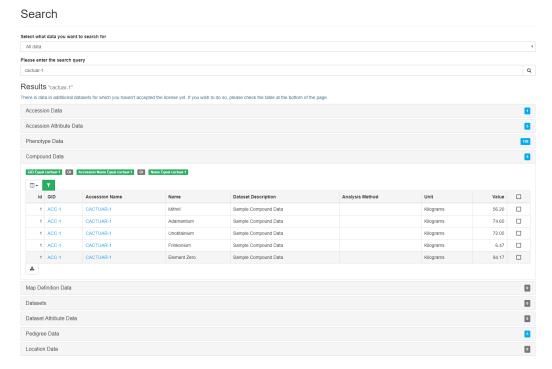


Figure 4: The search results page for the search term "cactuar-1". Results are grouped into categories with the number of matching database items highlighted. The result tables itself can be filtered further to restrict the number of results. In this example the "Compound Data" section is expanded and shows the five results that match the search query.

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 4 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 filter (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).

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 5 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

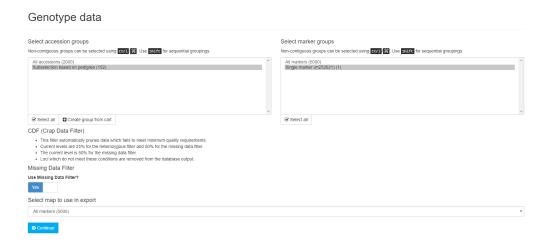


Figure 5: Example of the group sub-selection feature: You can select both accession and marker groups during the genotypic data export process to subset the dataset.

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 7 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 7A). 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.

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 7C.

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 6 shows what the group member search interface looks like. The query is specified in Figure 6A and the result shown in Figure 6B.

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 7B.

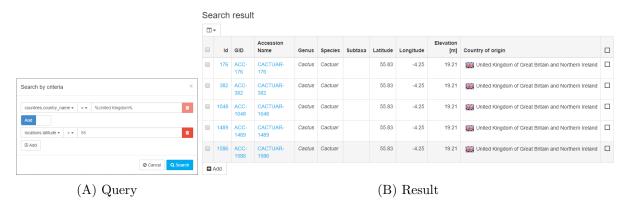


Figure 6: The group member search interface. (A) shows the boolean query interface where you can define the attributes an item needs to fulfil. In this case, the accession has to be collected in the UK at a latitude larger than 55. (B) The table showing the matching items. There are five accessions that match the query.

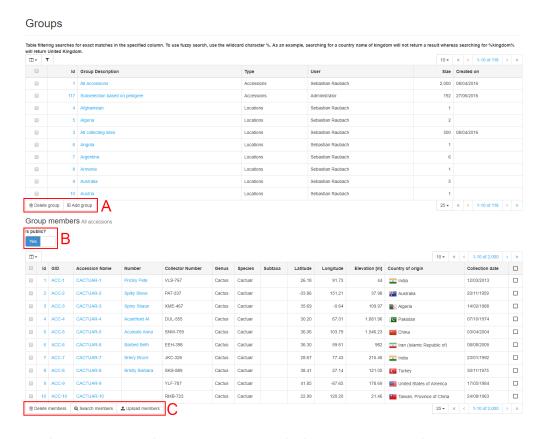


Figure 7: The groups page has many ways in which you can manipulate existing groups and create new groups. The top table shows all the groups that are visible to the current user. The bottom table shows the members of this group, i.e. the database items that are part of it; depending on the type of the 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.

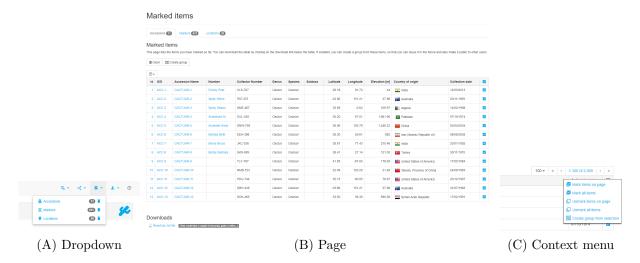


Figure 8: The marked item lists keep track of items of interest. (A) The dropdown menu shows a quick overview of how many items of each type are currently marked. The bin icon lets you easily empty the marked item list of a certain type. (B) The actual marked item lists page shows three tabs, one for 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 8A or go directly to the marked item lists page that is shown in Figure 8B.

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 8C).

#### 3.6 Charts

Most of the charts we use in Germinate are custom-built for this interface and therefore closely integrate with the way Germinate operates. All charts will offer some kind of interactivity. This can range from simple tooltips to the ability to select data points with a lasso. All charts can be downloaded as a Portable Network Graphics (PNG) or a Scalable Vector Graphics (SVG) file. Additionally, the underlying data for each chart can be downloaded in a tab-delimited data file. If the selection of individual data points is enabled, the selected items can be added to or removed from the marked item list.

#### 3.7 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.

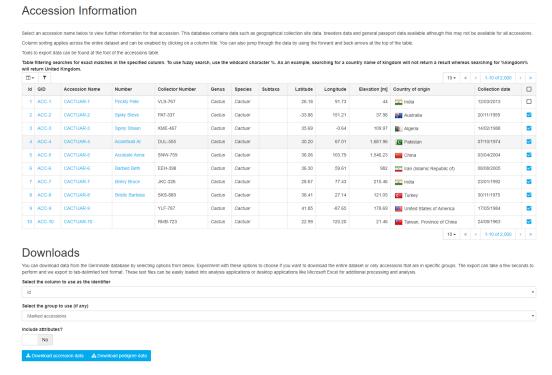


Figure 9: The accession overview page shows all the accessions that are part of this Germinate instance in a table. This table supports filtering. Accessions can be added to the marked items list by selecting the checkbox in the last column of the table. The download section below the table offers various ways to download data.

# 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

The core data of Germinate evolves around accessions and their passport data. Passport data can be described as the sum of all meta-data that relates to the accession.

Figure 9 shows the accession overview page of Germinate. The main feature is the table that shows all accessions held in this instance of Germinate in a central location. The table supports column sorting and filtering as described in Section 3.3 as well as adding accessions of interest to the marked items list (cf. Section 3.5.2). The table itself shows the most valuable information about each of the accessions, which is a subset of the Multi-Crop Passport Descriptors (MCPD) [1].

The MCPD 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.

The download section located below the table can be used to download the whole set of passport data including accession attributes (cf. Section 4.8) for either the whole dataset, a group or a custom marked item list of accessions. Additionally, the pedigree information for the same constellation of accessions can be downloaded.

Selecting an accession in the table redirects you to the passport data page. This page shows all the information related to a single accession. This includes the MCPD information, holding

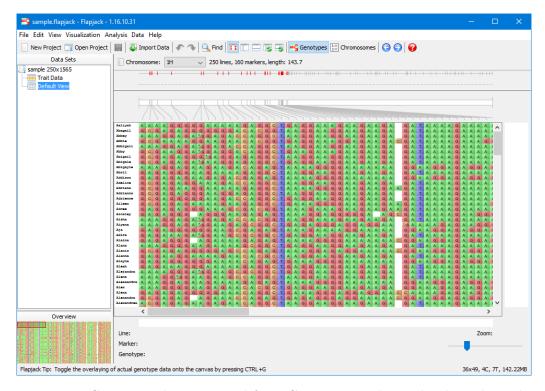


Figure 10: Genotypic data exported from Germinate and visualized in Flapjack.

institute, synonyms, pedigree data, collecting site location, images, groups, datasets, additional attributes and finally there is a section where users can add annotations to this accession (if this feature is enabled available).

#### 4.2 Pedigree Data

## 4.3 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 5. 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 4.3.2), e.g. physical vs. genetic marker positions.

The data is exported into a tab-delimited text format as well as Flapjack [2] format. Figure 10 shows an example of data exported from Germinate visualized in Flapjack.

#### 4.3.1 Allele Frequency Data

Paul

# 4.3.2 Genotypic Maps

## 4.3.3 Genetic Markers

## 4.4 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.

#### 4.5 Location Data

Whether you are interested in geographic locations at which accessions have been collected or where trials have been conducted, location data is one of the central data types in Germinate and we put a lot of thought into how we display this information and how we can make it available.

Location information can be used to filter down accessions by country, region, site name, latitude, longitude or even elevation. The geographic search page allows you to search for accessions or locations based on a point of interest query (Figure 11A) or a custom polygon search (Figure 11B). The results are shown below the map grouped into accessions and locations.

The locations page displays the location data in different ways. The locations are clustered in Figure 12A, visualized utilizing a heat map in Figure 12B and structured in a treemap in Figure 12C. All of these visualizations are interactive, so feel free to zoom in and out and select items.

- 4.6 Climate Data
- 4.7 Chemical Compound Data
- 4.8 Attribute Data

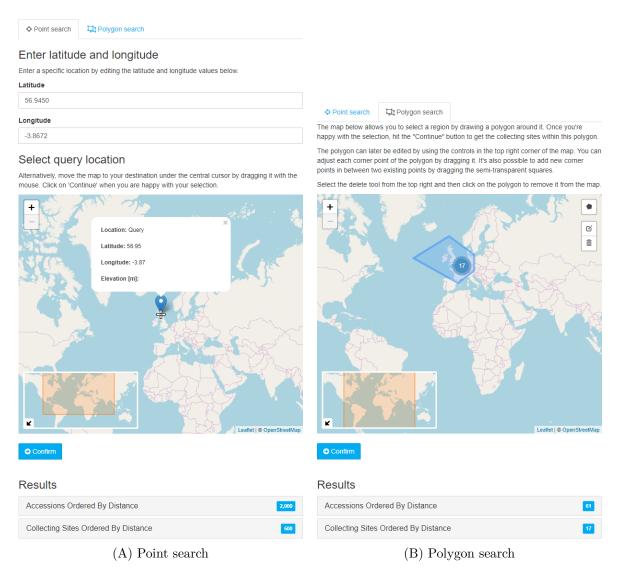


Figure 11: The location search interface. The results are shown at the bottom of each page. (A) The point search takes a user-specified query location on the map and returns all accessions and collecting sites ordered by their distance to the query. (B) With the polygon query you can define an arbitrary polygon on the map. Germinate will only return items that are located within the specified polygon.

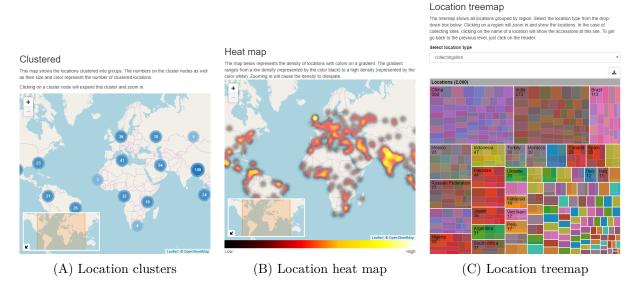


Figure 12: Different visualizations of location data on the locations page within Germinate. (A) The locations are clustered based on their latitude and longitude. Zooming in will show finer clusters. (B) The heatmap shows the distribution of locations. Yellow and white areas have a high density. (C) The locations are structured in a treemap using the country as the top level. The size of each rectangle represents the number of accessions that have been collected at this site. Selecting any of the countries zooms in to reveal the individual collecting sites.

Germinate 3 References

# 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.