

# User Manual of Breedbase

Breedbase team

2025-05-21



# Contents

<b>1 Basic Website Usage</b>	<b>11</b>
1.1 Creating a User Account . . . . .	11
1.1.1 Verifying first that you do not already have an account	11
1.1.2 Creating a user account . . . . .	11
1.2 Managing your Account . . . . .	12
1.2.1 Login . . . . .	12
1.2.2 Editing Account Settings . . . . .	13
1.2.3 Changing Your Account Status: From “User” to “Submitter” . . . . .	14
1.2.4 Submitting Feedback on an SGN Database . . . . .	14
1.3 Menu Layout . . . . .	15
1.3.1 Menu Options . . . . .	15
1.4 Working with Lists . . . . .	17
1.4.1 Creating lists . . . . .	17
1.4.2 Viewing and editing lists . . . . .	21
1.5 User Permissions . . . . .	22
<b>2 Searching the Database</b>	<b>23</b>
2.1 The Search Wizard . . . . .	24
2.1.1 How the Search Wizard Works . . . . .	24
2.1.2 How to use retrieved data . . . . .	26
2.1.3 Updating the Wizard . . . . .	29
2.2 Accessions and Plot Search . . . . .	30
2.3 Trials Search . . . . .	32
2.4 Trait Search . . . . .	33
2.5 Ontology Browser . . . . .	35
2.6 Search Seedlots . . . . .	36

<b>3 Managing User Roles</b>	<b>39</b>
3.1 What are User Roles? . . . . .	39
3.2 The Manage User Roles page . . . . .	40
<b>4 Managing Breeding Programs</b>	<b>41</b>
<b>5 Managing Locations</b>	<b>43</b>
<b>6 Managing Accessions</b>	<b>45</b>
6.1 Add Accessions Using A List . . . . .	45
6.2 Uploading Accessions and Accession's Info From A File . . . . .	48
6.3 Email alert for accession upload . . . . .	51
6.4 Add Parentage (Pedigree) Information to Accessions . . . . .	52
6.5 Working with grafts . . . . .	53
6.6 Bulk renaming of accessions . . . . .	54
<b>7 Managing Seed Lots</b>	<b>57</b>
7.1 Add New Seedlot(s) . . . . .	58
7.2 Seedlot Transactions . . . . .	59
7.3 Seed Inventory . . . . .	60
7.4 Find Seedlots For a List of Accessions . . . . .	61
7.5 Create a seedlot for an Accession or Cross . . . . .	63
7.6 Add quality data to a seedlot . . . . .	64
7.7 Seedlot Maintenance Events . . . . .	64
7.7.1 Setup . . . . .	64
7.7.2 Adding Events . . . . .	65
7.7.3 Displaying Events . . . . .	70
7.7.4 Downloading Events . . . . .	72
7.8 Deleting Seedlots . . . . .	72
<b>8 Managing Populations</b>	<b>73</b>
<b>9 Managing Crosses</b>	<b>75</b>
9.1 Crossing Experiment . . . . .	75
9.1.1 Add New Crossing Experiment . . . . .	76
9.2 Cross . . . . .	77
9.2.1 Add New Crosses . . . . .	77
9.2.2 Update Crosses by Uploading . . . . .	81
9.3 Cross Wishlist . . . . .	83

<b>CONTENTS</b>	<b>5</b>
-----------------	----------

9.3.1 Create a Cross Wishlist . . . . .	83
9.4 Crossing Experiment Detail Page . . . . .	89
9.5 Cross Detail Page . . . . .	91
<b>10 Managing Field Trials</b>	<b>95</b>
10.1 Trial Detail Page . . . . .	96
10.2 Adding Trials . . . . .	98
10.2.1 Prerequisites . . . . .	98
10.2.2 Adding a trial by using “Add Trial” form . . . . .	99
10.2.3 Adding a trial from an uploaded file . . . . .	106
10.2.4 Multi-location trials . . . . .	112
10.2.5 Email alert for multiple trial design upload . . . . .	113
10.2.6 Viewing Plot Layout and Trait HeatMap . . . . .	113
10.2.7 Adding additional information in the “Trial Detail” page	120
10.2.8 Downloading the Trial Layout from the “Trial Detail” page . . . . .	129
10.2.9 Adding Plant Entries To Your Trial . . . . .	131
10.2.10 Adding Tissue Sample Entries To Your Trial . . . . .	133
10.2.11 Uploading GPS Coordinates For Plots . . . . .	136
10.2.12 Repetitive Measurements Section . . . . .	137
10.2.13 Uploading Additional Files To Trial . . . . .	138
10.3 Updating Trial Data . . . . .	139
10.4 Deleting Trial Data . . . . .	141
<b>11 Managing Genotyping Plates</b>	<b>143</b>
11.1 Adding a New Genotyping Plate . . . . .	145
11.2 Genotyping Plate Detail Page . . . . .	147
<b>12 Using Field Book App</b>	<b>149</b>
12.1 A typical workflow . . . . .	150
12.2 Creating Field Layout Files for the Field Book App . . . . .	150
12.2.1 Creating “Field Layout Files” by using “Field Book Tools” page. . . . .	151
12.2.2 Creating “Field Layout Files” by using “Trial Detail” page. . . . .	153
12.3 Creating Trait Files for the Field Book App . . . . .	155
12.3.1 Creating a Trait List . . . . .	155
12.3.2 Creating a Trait File . . . . .	155

12.4 Transferring Files from Your Computer to Android Tablet . . . . .	157
12.4.1 Files on your computer . . . . .	157
12.4.2 Files on your Android tablet . . . . .	158
12.5 Setting up “Field Book App” for data collection . . . . .	160
12.6 Exporting Files from Field Book App . . . . .	169
12.7 Uploading Phenotype Files to an SGN database . . . . .	172
<b>13 Managing Phenotypic Data</b>	<b>175</b>
13.1 Uploading Fieldbook Phenotypes . . . . .	175
13.1.1 Export Field Book Database File . . . . .	175
13.1.2 Upload Field Book Database File . . . . .	176
13.2 Uploading Spreadsheet Phenotypes . . . . .	176
13.2.1 Generating Spreadsheet File . . . . .	177
13.2.2 Uploading Spreadsheet File . . . . .	179
<b>14 Managing Barcodes</b>	<b>181</b>
<b>15 Using the Label Designer</b>	<b>185</b>
15.0.1 First Select a Datasource . . . . .	185
15.0.2 Set Page and Label Size . . . . .	186
15.0.3 Design Your Label . . . . .	187
15.0.4 Adjust Formatting, Save, and Download . . . . .	188
<b>16 Managing Downloads</b>	<b>191</b>
<b>17 Managing ODK Data Collection</b>	<b>193</b>
17.1 ONA Crossing Information . . . . .	194
17.1.1 Managing ONA Crossing Information . . . . .	194
17.1.2 Reviewing Plant Status . . . . .	195
17.1.3 Graphical Summary For Performed Crosses . . . . .	195
17.1.4 Summary Information For Performed Crosses . . . . .	196
<b>18 Managing Tissue Samples</b>	<b>197</b>
18.1 Tissue samples from field trials . . . . .	197
18.2 Genotyping Plate Tissue Samples (96 or 384 well plates) . . . . .	200
<b>19 Managing Observation Variables</b>	<b>203</b>
19.1 Managing Observation Variables with Traits, Methods, and Scales . . . . .	203

<b>20 Managing Image Data</b>	<b>209</b>
20.1 Image-Phenotyping Dashboard . . . . .	209
20.2 Image Input . . . . .	209
20.3 Standard Process . . . . .	214
20.4 Ground Control Points . . . . .	222
<b>21 Managing VCF Data</b>	<b>223</b>
21.1 Uploading VCF Data . . . . .	223
21.2 Searching and Downloading VCF Data . . . . .	226
21.3 Searching Protocols . . . . .	228
21.4 Detail Pages and Deletion . . . . .	230
<b>22 Managing Spectral Data</b>	<b>233</b>
22.1 Upload Spectral Data . . . . .	234
22.2 Evaluate and Remove Outliers . . . . .	235
22.3 Plot Spectra . . . . .	235
22.4 Aggregate Spectra . . . . .	236
22.5 References . . . . .	237
<b>23 Managing Sequence Metadata</b>	<b>239</b>
23.1 What is Sequence Metadata? . . . . .	240
23.2 Loading Sequence Metadata . . . . .	240
23.3 Searching Sequence Metadata . . . . .	241
23.3.1 Basic Search . . . . .	241
23.3.2 Advanced Search . . . . .	243
23.4 Marker Integration . . . . .	243
23.5 Sequence Metadata API . . . . .	243
<b>24 Managing Outliers in Dataset</b>	<b>245</b>
24.1 What is Outliers Functionality in Dataset ? . . . . .	245
24.2 Accessing Trait Visualization . . . . .	246
24.3 Interpreting Visual Elements . . . . .	246
24.4 Choosing Cut-Off Values . . . . .	247
24.5 Setting Deviation Multiplier . . . . .	247
24.6 Utilizing Graph Controls . . . . .	248
<b>25 Data Analysis Tools</b>	<b>249</b>
25.1 Selection Index . . . . .	249

25.2 Genomic Selection . . . . .	253
25.2.1 Building a Model - Method 1: . . . . .	254
25.2.2 Building a Model - Method 2 . . . . .	265
25.2.3 Building a Model - Method 3 . . . . .	273
25.3 Genome Browsing . . . . .	274
25.3.1 Browsing Genotype data by Accession . . . . .	274
25.3.2 Browsing Genotype data by Trial . . . . .	275
25.4 Principal Component Analysis (PCA) . . . . .	276
25.5 ANOVA . . . . .	278
25.6 Clustering (K-Means, Hierarchical) . . . . .	278
25.7 Genetic Gain . . . . .	281
25.8 Kinship and Inbreeding Coefficients . . . . .	282
25.9 Creating Crossing Groups . . . . .	283
25.10 Search Wizard Genomic Relationship Matrix (GRM) Download	286
25.11 Search Wizard Genome Wide Association Study (GWAS) . .	287
25.12 Spectral Analysis . . . . .	291
25.12.1 Dataset selection . . . . .	292
25.12.2 Cross-validation . . . . .	293
25.12.3 Preprocessing . . . . .	294
25.12.4 Algorithms . . . . .	295
25.12.5 Output: common model summary statistics . . . . .	296
25.12.6 Export model for later use . . . . .	297
25.12.7 Predict phenotypes from an exported model (routine use) . . . . .	297
25.12.8 FAQ . . . . .	298
25.13 General Mixed Model Tool . . . . .	299
25.14 Genomic Prediction of Cross Performance (GPCP) . . . . .	301
25.15 Tool Compatibility . . . . .	302

# Introduction

Welcome to the Breedbase manual!

This manual is intended for database users.

If you are a developer looking for software implementation details, please visit the developer wiki instead: <https://github.com/solgenomics/sgn/wiki>



# **Chapter 1**

## **Basic Website Usage**

### **1.1 Creating a User Account**

#### **1.1.1 Verifying first that you do not already have an account**

Before creating an account, please verify first that you don't already have an account. You can use "Search" menu to check if you already registered as a user.

In the "Search" menu, selecting the "People" tab and search your name. If nothing is found, proceed with the instructions below. Otherwise, clicking the "Login" button. If you have forgotten your password, you can retrieve it by clicking the "Forgot your password?" link on the login page.

#### **1.1.2 Creating a user account**

On the right of the toolbar, clicking on "Login." It will take you to the login page. On the login page, clicking on the link "sign up for an account." It will take you to the page below:



The screenshot shows a web page titled "CASSAVABASE". At the top, there is a navigation bar with links for "Search", "Manage", "Analyze", "Maps", and "About". To the right of the navigation bar is a search icon and a "Login" button, which is highlighted with a red circle. Below the navigation bar, the title "Create New Account" is displayed in green. A "Notice" section contains instructions: "Before creating a new account, please check if you already have an account using the directory search.", "A link will be emailed to you. Please click on it to activate the account.", and "All fields are required.". The main form consists of several input fields: "First Name", "Last Name", "Organization", "Username" (with a note "Username must be at least 7 characters long."), "Password" (with a note "Password must be at least 7 characters long and different from your username."), "Confirm Password", and "Email Address" (with a note "An email will be sent to this address requiring you to confirm its receipt to activate your account."). At the bottom of the form are two buttons: "Reset" and "Create Account".

Filling in all of the information, then clicking “Create Account.”

After you submit the information, an email will be sent to the provided email address. Checking your email and clicking on the link to activate your account.

## 1.2 Managing your Account

### 1.2.1 Login

To login, clicking the “Login” link in the toolbar on any page and enter your username and password.

If you have forgotten your password, you can retrieve it by clicking the “Forgot your password?” link on the login page.



### 1.2.2 Editing Account Settings

Account settings can be edited by clicking on the “my profile” link displayed as your user name, on the right of the toolbar. You must login, in order to access and change account settings.



You can add personal information to your account using the “View or update personal information” link.

To change your password, username, or your contact email, clicking on “Update account information” link. You must provide your old password before you can make any changes.

The screenshot shows a user profile page for 'Gregor Mendel' on the CASSAVABASE website. At the top, there's a navigation bar with links for Search, Manage, Analyze, Maps, and About. On the right, there are search, user profile ('Gregor\_Mendel'), lists, and other account-related buttons. Below the header, it says 'Welcome Gregor Mendel' and 'Not Gregor Mendel? [log out]'. The main content area is divided into sections: 'General Tools' (with a red box around 'View or update personal (contact and research) information', 'Update account information', and 'Post to SGN forum'), 'QTL data submission' (with a link to 'Upload and analyse your QTL data'), 'SGNS submitted analysis jobs' (noting 'You have no submitted jobs.'), 'Loci with Editor Privileges' (noting 'None.'), 'Annotated Loci' (with a link to '[View annotated loci by date]'), and 'User Status' (noting 'Your current user status is submitter. You have the maximum user privileges on SGN. Please contact SGN if you would like to change your user status.').

### 1.2.3 Changing Your Account Status: From “User” to “Submitter”

After you create an account, your account has a “user” status. This account has limited privileges.

Accounts with “user” status are able to:

- Change personal information
- Post comments on pages
- Post to the forum

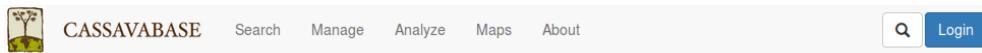
To upgrade your account status to “submitter,” contact the database curators using the “contact” link provided at the footer of each page. Submitter accounts can add data, such as new plots, accessions, phenotype data and images.

### 1.2.4 Submitting Feedback on an SGN Database

We appreciate your feedback! Feel free to submit any questions or suggestions by using the “Feedback” link provided at the footer of each page.

## 1.3 Menu Layout

SGN Database websites have a toolbar on the top of each page with a number of menus for convenient access of major functions. The menus, as pictured below, are “search,” “manage,” “analyze,” and “maps.” The toolbar also provides a quick search, a “log in” button, and a “new user” button.



### 1.3.1 Menu Options

#### Search

In the Search menu, the options are:

Tab	Description
Wizard	Search different accessions and plots by location, year, trial, and trait data. Can also be used to create lists of different types.
Accession and plots	Search accessions and plots using a variety of criteria
Trials	Search trials by name, description, breeding program, year, location, and trial type.
Markers	Search different markers
Images	Search images contained in the SGN database
People	Search database users

#### Manage

In the Manage menu, the options are:

Tab	Description
Breeding Programs	View, add and delete breeding programs
Locations	View, add and delete locations
Accessions	Manage and search different accessions
Seedlots	Manage and search different seedlots

Tab	Description
Crosses	Create new crosses in the database
Field Trials	Manage field trials. Create trials using different field layouts.
Genotyping Plates	Manage genotyping plates. Create 96 or 384 well plates.
Phenotyping	Upload phenotyping files from the Tablet Field Book application
Field Book App	Manage the field book app data (download files to tablet)
Barcodes	Refers to the old barcode system, mainly historical
Download	Download information in the database based on lists

## Analyze

**Clicking on the “Analyze” link will give a full menu of all analysis functions**

In the Analyze menu, the options are:

Tab	Description
<b>Breeder Tools</b>	
Breeder Home	Access breeding functionalities. Lists important and helpful links.
Barcode Tools	Manage, create, and download barcodes. Also access barcode tools.
Genomic Selection	Can search for traits, start building a GS model, and predict values based on genotypes
<b>Sequence Analysis</b>	
BLAST	Sequence homology search
<b>Other</b>	
Ontology Browser	Browse all recorded ontologies

## 1.4 Working with Lists

Lists are collections of identifiers that are stored in the database. Lists can be composed of accessions, plots, traits, locations, and trials. Lists are attached to the individual user's account, and can only be created and seen by the user while logged in. SGN databases make heavy use of lists in a number of tools on the website. For example, trials are created using lists of accessions.

### 1.4.1 Creating lists

Lists can be generated in various ways:

One way to create a list is by clicking on the “Lists” link located on the toolbar.



To create a new list, enter the name of your new list and then clicking on the “New List” button. The name of the list can be anything, but should be unique and should be something to help you easily identify.

List Name	Count	Type	View	Delete	Download	Share	Group
acc88	1	accessions	[View]	[X]	[Download]	[Share]	[Group]
acc_vk_1	1	accessions	[View]	[X]	[Download]	[Share]	[Group]
accessions_for_solsq_tests	374	accessions	[View]	[X]	[Download]	[Share]	[Group]
accessions_for_trial2	307	genotyping_trials	[View]	[X]	[Download]	[Share]	[Group]
desynonymize_test_list	6	accessions	[View]	[X]	[Download]	[Share]	[Group]
geno_trial	1	genotyping_trials	[View]	[X]	[Download]	[Share]	[Group]
janedoe_1_private	2	null	[View]	[X]	[Download]	[Share]	[Group]
janedoe_1_public	2	null	[View]	[X]	[Download]	[Share]	[Group]
m1	1	accessions	[View]	[X]	[Download]	[Share]	[Group]
m2	1	accessions	[View]	[X]	[Download]	[Share]	[Group]

Showing 1 to 10 of 20 entries

Previous [1] 2 Next

View Public Lists Close

You can find the list that you entered on the “Your Lists” page. To add items to your list, click on the “View” icon to open “List Contents” page.

Your Lists

Create New List. Type New List Name Here New List

Show 10 entries Search:

List Name	Count	Type	View	Delete	Download	Share	Group
acc88	1	accessions		x			
acc_vk_1	1	accessions		x			
accessions_for_sols_tests	374	accessions		x			
accessions_for_trial2	307	genotyping_trials		x			
desynonymize_test_list	6	accessions		x			
geno_trial	1	genotyping_trials		x			
janedoe_1_private	2	null		x			
janedoe_1_public	2	null		x			
m1	1	accessions		x			
m2	1	accessions		x			

Showing 1 to 10 of 20 entries Previous 1 2 Next

View Public Lists Close

On the “List Contents” page, enter items that you want to add to the list, then click on “Add” button.

List Contents

List ID: 26

List name: MyNewList Update

Type: (none) Validate

Add New Items: Add

001D  
001B

Sort Ascending Sort Descending

Search: No data available in table

Showing 0 to 0 of 0 entries

Close

The page will be updated and will display your items in a table at the bottom of the page. It is possible to sort the list if you need.

### List Contents

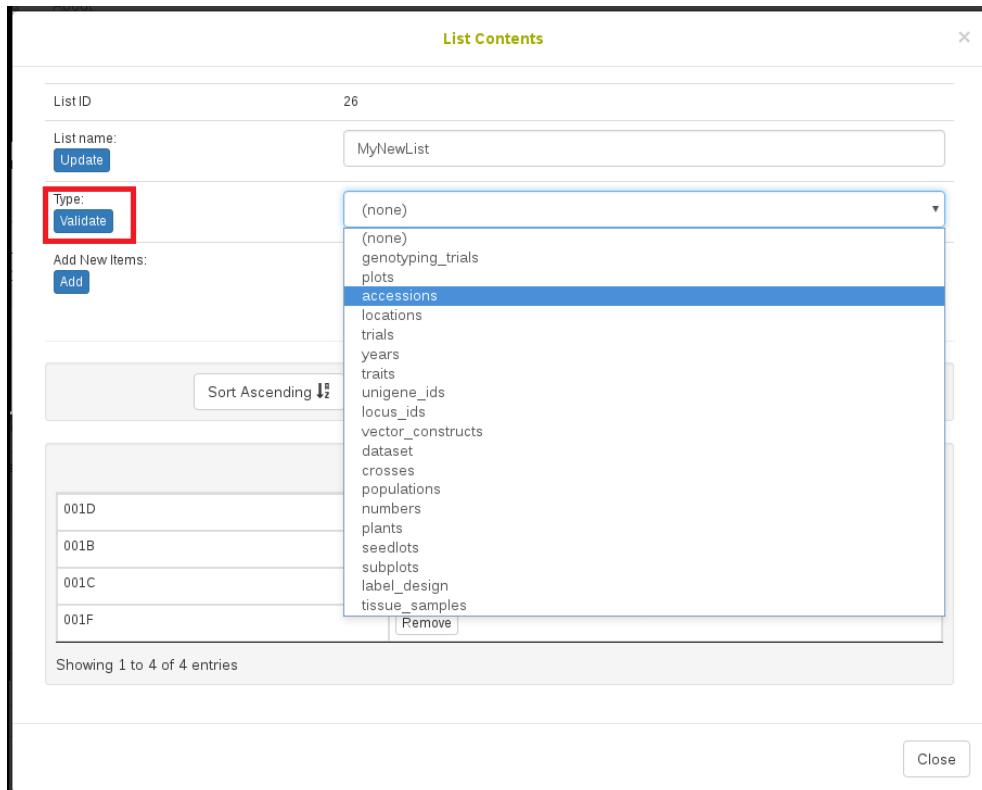
List ID	26
List name:	<input type="text" value="MyNewList"/>
Type:	<input type="button" value="Update"/> <input type="button" value="Validate"/>
Add New Items:	<input type="button" value="Add"/>  <div style="border: 1px solid #ccc; padding: 5px; width: 100%;">Add Item(s) To List. Separate items using a new line to add many items at once.</div>

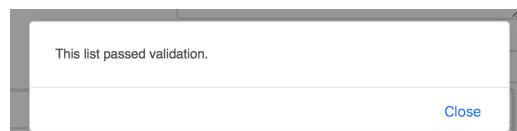
Sort Ascending	Sort Descending
Search: <input type="text"/>	
001D	<input type="button" value="Remove"/>
001B	<input type="button" value="Remove"/>
001C	<input type="button" value="Remove"/>
001F	<input type="button" value="Remove"/>

Showing 1 to 4 of 4 entries

Select the type of items in your list. To verify that the items that you added to your list are already stored in the database and that you selected a correct type for the items, click on the “Validate” button.



If those items are already in the database, a message will indicate that “This list passed validation”



Note that a list cannot contain duplicate elements. If a duplicate item is entered, the list manager will inform the user that the element is already in the list and will not add it again.

Another easy way to create a list is to use [2.1](#), which can be accessed from the Search menu.

### 1.4.2 Viewing and editing lists

Lists can be viewed and edited using the “Lists” link on the toolbar. Clicking on the link will open a window that displays all of your lists, as well as an option to create new lists.

List Name	Count	Type	Actions
IITA_WKSHP_D2	20	accessions	
IITAwksp16_accessions_list	24	accessions	
new_accession_list	6	accessions	

This page shows all lists that have been created, including those created by using the Search Wizard. You can view and edit your lists by using “Actions” buttons.

1. Clicking on the “view” icon will open a new window called “List Contents” that allows you to change the list name, the type of the list, add new items, or delete existing items.
2. Clicking on the “delete” icon will delete your list. **Caution: this action cannot be undone.**
3. Clicking on the “download” icon will download the contents of your list to your computer.
4. Clicking on the “make public” icon will make your list available for other users to view and use your list.

List Name	Count	Type	Actions
IITA_WKSHP_D2	20	accessions	
IITAwksp16_accessions_list	24	accessions	
new_accession_list	6	accessions	

## 1.5 User Permissions

Breedbase accounts are assigned one or more of four different roles to determine the level of access they have within the database. The possible roles are **User**, **Submitter**, **Sequencer**, and **Curator**. Each role grants specific permissions, and careful management of them helps prevent data from being altered or deleted in error.



Accounts are also assigned Breeding Program role(s) to grant access to the specific breeding program(s) they work with.

- The **User** role gives an account permission to view and download data throughout the database.
- The **Submitter** role gives an account permission to design field experiments and to upload and edit data using the tools in the “Manage” section. In order to submit and manage breeding data within a given breeding program, a submitter also must have a matching Breeding Program role.
- The **Sequencer** role gives an account permission to design genotyping experiments and submit plates to a genotyping service.
- The **Curator** role gives an account permission to do all of the above, as well as to delete data within the database. The Curator role also enables the addition or deletion of roles for all database accounts in the ‘Manage User Roles’ tool.

# Chapter 2

## Searching the Database

You can search for information on the database by using the following search options: Wizard, which uses combined criteria specified by users; Accessions and Plots; Trials; Markers; Images; People; FAQ.



## 2.1 The Search Wizard

The screenshot shows the 'Search Wizard' interface. At the top right are three buttons: 'Don't see your data?' (with a link to 'Refresh Lists'), 'Update Wizard', and a magnifying glass icon. Below this are four identical search panels, each consisting of a dropdown labeled 'Select Column Type', a search input field, a 'Select All' button (with '0/0' next to it), and a 'Clear' button. The panels are arranged in a 2x2 grid. At the bottom left is a section titled 'Load/Create Datasets using' with dropdown menus for 'Match' and 'Columns'. It includes a 'Load Dataset' dropdown, a 'Load' button, and a 'Create New Dataset' dropdown with a 'Create' button. To the right are three sections: 'Related Genotype Data', 'Related Trial Metadata', and 'Related Trial Phenotypes'.

### 2.1.1 How the Search Wizard Works

The search wizard presents a number of select boxes, which are initially empty. You start searching by picking a category of data from the dropdown above the left-most select box.

Once a category has been picked, the database will retrieve all the options within this category and display them within the first select box. You then select one or more options from the first select box, which activates the second dropdown.

You can then select a category from the second dropdown, and repeat this same search process through all four dropdowns and select boxes.

The screenshot shows the 'Search Wizard' interface with four main sections: Locations, Years, Accessions, and Select Column Type. The Locations section has a search bar and a list of locations including Granada, Meta, Colombia; Guanambi (BA)-IF Bajano; Hombolo; IBARAPA; Igbariam; and Ibadan. The Years section has a search bar and a list of years from 2009 to 2015, with 2011 and 2012 selected. The Accessions section has a search bar and a list of accession numbers including 20\_20, 462, 50395, 58308, and 79-106. The Select Column Type section is empty.

- In the example above, the “locations” category was chosen in the first dropdown. The first select box then displayed all the possible locations in the database. The option Ibadan was selected.
- This activated the second dropdown. The category “years” was chosen in the second dropdown. The second select box then displayed all the years that are linked in the database to the location Ibadan. From that list, the options 2011 and 2012 were selected.
- This activated the third dropdown. A final category, “accessions”, was chosen in the third dropdown. The third select box was then populated with the 3847 accessions in the database that are linked with the location Ibadan in the years 2011 or 2012.

In addition to the basic search operations demonstrated above, users can take advantage of two more features:

### Load Selection from List

Load Selection from List:

- Instead of picking a category in the first dropdown, users can instead populate the first selectbox from a list by scrolling down in the first dropdown to the “Load Selection from List” subheading and selecting a list. This is useful for starting queries with a list of plots, as this category is not among the options in the first dropdown.

### ANY/MIN/ALL Toggle



- By default, the search wizard combines options within a category using an OR query. In the example above, in the third panel the wizard retrieved accessions associated with the location ‘Ibadan’ in **ANY** of the years “2011 **OR** 2012”
- If the user clicked the toggle below the second select box to change it to **ALL** before choosing accessions in the third dropdown, the wizard would instead retrieve accessions associated with the location ‘Ibadan’ in the years “2011 **AND** 2012”. This will be a smaller set of accessions, because any accessions used only in 2011, or only in 2012 will be excluded.
- A more advanced search could use the **MIN** toggle option. This allows the user to make a query in between an ANY or ALL query, where a minimum number of matches from the selected column will be used as a filter for the next column. The minimum can be provided as either a percentage (%) or an actual count of items (#). In the example above, if the years 2011, 2012, and 2013 were selected in the second column, the user could enter ‘2’ in as the minimum and select ‘#’ as the minimum match type. This would select accessions in the third column that were used in 2 or more of the selected years.

Match	ANY	<b>MIN</b>	ALL
>=	2	%	#

### 2.1.2 How to use retrieved data

#### Getting more Info

Any option in the wizard select boxes (except for years) can be clicked to open a page with more details. The new page is opened in a new tab.

#### Saving to a list

You can store the highlighted items in any selected box to lists. This is done using the inputs and buttons directly below the select box. **Don’t forget, you must be logged in to work with lists!**



- To **add items to an existing list**, first pick an existing list using the “Add to List...” dropdown on the left. Then click the “Add” button. A popup window will confirm the action, and display the number of items added to your existing list.
- To **store items to a new list**, first type a new list name in the “Create New List...” text input on the left. Then click on the “Create” button. A popup window will confirm the action, and display the number of items added to your new list.

## Downloading Data

You can download trial metadata, phenotypes and genotypes associated with the highlighted items in the wizard select boxes. This is done using the buttons in the download section at the bottom of the page. **Don't forget, you must be logged in to download data!**



**Metadata** Trial metadata can be downloaded by selecting a subset of trials from the database or based on your search categories. To download, click on “Related Trial Metadata”, a dialog will appear. Select download format and click the “Metadata” button to complete your download.

The screenshot shows a modal window titled "Related Trial Metadata". It contains a text input field with "3 trials" and a dropdown menu set to "CSV". A blue button labeled "Metadata" is at the bottom.

**Phenotypes** The phenotypes download is quite flexible, and can download a subset of all the trial data in the database based on whichever categories and options you currently have selected. Simply click on the “Related Trial Phenotypes” link, review the options, changing or adding any additional parameters you like, then click ‘Download Phenotypes’.

The screenshot shows a modal window titled "Related Trial Phenotypes". It includes a message "Too few trials", download format options ("Default", "CSV", "All"), and several checkboxes: "Include timestamps", "Include accession entry numbers", "Suppress user defined phenotype outliers", and "Repetitive measurements: Averaged value". It also features date filters for "From" (1960-01-01) and "To" (2024-12-15), a checkbox for "include items without a date", and trait name filtering ("Trait Name Contains: e.g. plant height"). A blue button labeled "Download Phenotypes" is at the bottom.

**Genotypes** The genotype download is more stringent. It requires a minimum of one accession and one genotyping protocol to be selected in the wizard select boxes. The text box in the download section of the page will help track what has been selected. Once clicked, the “Download Genotypes” button will download a genotype file for the selected accessions.

### Saving the wizard selections

As discussed above, the selections of the individual select boxes in the wizard can be saved separately to a list. The lists can be used as inputs in other tools on the site. However, sometimes creating a selection is quite time consuming and restoring the selections from four different lists would be cumbersome too. Therefore, the selections can be saved together in a dataset, and named for later retrieval. This is done in the section “Load/Create Datasets” that is below the first two wizard select boxes. To select an existing dataset, one uses the “Load Dataset” dropdown. A particular dataset can be chosen, and the “Load” button can be clicked to retrieve and display the dataset in the wizard. To create a new dataset using items that are selected in the wizard, one can enter the name of the new dataset in the “Create New Dataset” text box. Once the dataset has been given a name, clicking the “Create” button will save the new dataset.



### 2.1.3 Updating the Wizard

The search wizard uses a copy of the database, or a cache, to return results quickly. If data appears to be missing, it usually means that the cache needs to be updated. Users with submitter privileges or above can do this using the ‘Update Wizard’ button. One can also use the ‘Refresh Lists’ button to update the available lists.



This will take just a few seconds in small databases, but may take a few hours to complete in larger databases.

## 2.2 Accessions and Plot Search

Accessions and their related materials (cross, plant, plot, population, tissue\_sample, training population) can be searched by using “Search Accessions and Plots” page. On this page, “accession” is the default stock type; however, you can change stock type by selecting an option from the dropdown list. From this page you can construct detailed queries for stock types. For example, by using the “Usage” section, the “Properties” section, and the “Phenotypes” section you could search for accessions which were diploids used in a specific year and location and were also phenotyped for height. You can also search for accessions based on genetic properties, such as the location of an introgression on a specific chromosome.

The screenshot shows the 'Search Accessions and Plots' interface. At the top, there is a search bar with a dropdown menu set to 'Uniquename'. Below the search bar are three sections: 'Properties', 'Usage', and 'Phenotypes'. A 'Search' button is located at the bottom of this section. The main area is titled 'Search Results' and contains a table with columns: Stock Name, Stock Type, Organism, Synonyms, Owners, and Organization. The table lists several entries, including 'BLANK', 'IITA-TMS-IBA011412', 'IITA-TMS-IBA30572', 'IITA-TMS-IBA880002', 'IITA-TMS-IBA880081', 'new\_acc\_ppp001', 'new\_acc\_ppp002', 'new\_acc\_ppp003', 'new\_test\_crossP001', and 'new\_test\_crossP002'. The table has a 'Show 10' dropdown and a 'Previous'/'Next' navigation bar at the bottom.

Stock Name	Stock Type	Organism	Synonyms	Owners	Organization
BLANK	accession	<i>Manihot esculenta</i>		John Doe	
IITA-TMS-IBA011412	accession	<i>Manihot esculenta</i>		John Doe	
IITA-TMS-IBA30572	accession	<i>Manihot esculenta</i>		John Doe	
IITA-TMS-IBA880002	accession	<i>Manihot esculenta</i>		John Doe	
IITA-TMS-IBA880081	accession	<i>Manihot esculenta</i>		John Doe	bti
new_acc_ppp001	accession	<i>Manihot esculenta</i>	synp001	Jane Doe	
new_acc_ppp002	accession	<i>Manihot esculenta</i>		Jane Doe	bti
new_acc_ppp003	accession	<i>Manihot esculenta</i>		Jane Doe	
new_test_crossP001	accession	<i>Solanum lycopersicum</i>		John Doe	
new_test_crossP002	accession	<i>Solanum lycopersicum</i>		John Doe	

It is possible to query over any of the available properties, such as “ploidy\_level”, “country of origin”, “introgression\_chromosome”, etc.

**Search Accessions and Plots**

Search

Uniquename

Stock Name or Description:  Type search here...

Properties

Stock Type:  Organism:

Stock Owner:  Organization:

Search By Another Property:

accession number:

country of origin:

introgen\_start\_position\_bp:

Usage

Phenotypes

In the search result table it is possible to select any of the available properties to view.

Search Results

View Another Property:

Show 10 entries

Stock Name	Stock Type	Organism	Synonyms	Owners	organization	ploidy_level
BLANK	accession					
IITA-TMS-IBA011412	accession	Manihot esculenta		John Doe		
IITA-TMS-IBA30572	accession	Manihot esculenta		John Doe		
IITA-TMS-IBA980002	accession	Manihot esculenta		John Doe		
IITA-TMS-IBA980581	accession	Manihot esculenta		John Doe	bti	
new_acc_ppp001	accession	Manihot esculenta	symp0001	Jane Doe		2
new_acc_ppp002	accession	Manihot esculenta		Jane Doe	bti	
new_acc_ppp003	accession	Manihot esculenta		Jane Doe		3
new_test_crossP001	accession	Solanum lycopersicum		John Doe		
new_test_crossP002	accession	Solanum lycopersicum		John Doe		

Showing 1 to 10 of 482 entries

Previous      ...  Next

At the bottom of the accession search there is a phenotype graphical filtering tool. Here you can filter down accessions based on combinations of trait performance. The filtered down accessions are then able to be saved to a list.



For information on adding Accessions please see the Managing Accessions help. For information on how field trial plots, plants, tissue samples, and subplots are added to the database, please see the Managing Field Trials help.

## 2.3 Trials Search

Trials on the database can be searched based on trial name, description, breeding program, year, location, trial type, design, planting date, and harvest date.

Trial Search						
Show 10 entries		Search: <input type="text"/>				
Trial name	Description	Breeding program	Folder	Year	Location	Trial type
CASS_6Genotypes_Sampling_2015	Copy of trial with postcomposed phenotypes from cassbase.	test		2017	test_location	Preliminary Yield Trial
Kasese solgs trial	This trial was loaded into the fixture to test solgs.	test		2014	test_location	Clonal Evaluation
new_test_cross	new_test_cross	test				
selection_population	selection_population			2015		
test_genotyping_project	test_genotyping_project			2015		
test_population2	test_population2			2015		
test_t	test tets	test		2016	test_location	
test_trial	test trial	test		2014	test_location	
trial2 NaCRRI	another trial for solGS	test		2014	test_location	

Showing 1 to 9 of 9 entries      Previous  Next

**Copy Results to a List**      Copy the trial names currently showing in the search results table to a new or existing list

New list...

## 2.4 Trait Search

On the Trait Search page (menu item **Search > Traits**), traits in the database can be searched by ID, name, or description. Optionally, a starting list of traits can be selected to filter down results.

Trait Search		
Subset Traits:		Select A Subset
Show 10 entries		
Trait ID	Trait Name	Definition
<input type="checkbox"/> CO_334.0000008	sprouting proportion	Proportion of stakes germinated scored one month after planting.
<input type="checkbox"/> CO_334.0000009	initial vigor assessment 1-7	Visual assessment of plant vigor during establishment scored one month after planting.
<input type="checkbox"/> CO_334.0000010	plant stands harvested counting	A count of the number of plant stands at harvest.
<input type="checkbox"/> CO_334.0000011	root number counting	A count of the total number of storage roots harvested per plot.
<input type="checkbox"/> CO_334.0000012	f.root.weight	Total fresh weight of storage roots harvested per plot measured in kilogram (kg).
<input type="checkbox"/> CO_334.0000013	fresh root yield	Fresh weight of harvested roots expressed in tons per hectare (tha) per plant.
<input type="checkbox"/> CO_334.0000014	dry yield	Dry weight of harvested roots derived by multiplying fresh storage root yield by dry matter content expressed in tons per hectares (tha).
<input type="checkbox"/> CO_334.0000015	harvest index variable	Proportion of fresh root weight in total biomass.
<input type="checkbox"/> CO_334.0000016	fresh shoot weight measurement in kg	Total fresh weight of harvested foliage and stems in kilograms per plot.
<input type="checkbox"/> CO_334.0000017	top yield	Total fresh weight of harvested foliage and stems expressed in tons per hectare (tha).

Showing 1 to 10 of 245 entries

[Select All](#) [Deselect All](#)

Previous 1 2 3 4 5 ... 25 Next

Selecting traits in the results of the search allows one to add the selected results to a trait list, or create a new trait list from the select results.

Show 10 entries			Search:
Trait ID	Trait Name	Definition	
<input checked="" type="checkbox"/> CO_334:0000008	sprouting proportion	Proportion of stakes germinated scored one month after planting.	
<input type="checkbox"/> CO_334:0000009	initial vigor assessment 1-7	Visual assessment of plant vigor during establishment scored one month after planting.	
<input type="checkbox"/> CO_334:0000010	plant stands harvested counting	A count of the number of plant stands at harvest.	
<input type="checkbox"/> CO_334:0000011	root number counting	A count of the total number of storage roots harvested per plot.	
<input checked="" type="checkbox"/> CO_334:0000012	fresh root weight	Total fresh weight of storage roots harvested per plot measured in kilogram (kg).	
<input checked="" type="checkbox"/> CO_334:0000013	fresh root yield	Fresh weight of harvested roots expressed in tons per hectares (t/ha) per plant.	
<input type="checkbox"/> CO_334:0000014	dry yield	Dry weight of harvested roots derived by multiplying fresh storage root yield by dry matter content expressed in tons per hectares (t/ha).	
<input type="checkbox"/> CO_334:0000015	harvest index variable	Proportion of fresh root weight in total biomass.	
<input checked="" type="checkbox"/> CO_334:0000016	fresh shoot weight measurement in kg	Total fresh weight of harvested foliage and stems in kilograms per plot.	
<input type="checkbox"/> CO_334:0000017	top yield	Total fresh weight of harvested foliage and stems expressed in tons per hectare (t/ha).	

Showing 1 to 10 of 245 entries **4 rows selected**

Previous 1 2 3 4 5 ... 25 Next

[Select All](#) [Deselect All](#)

[Copy Selected Results to a List](#) Copy the trait names currently selected in the search results table to a new or existing list

4 trait(s) selected.

New list... [add to new list](#)

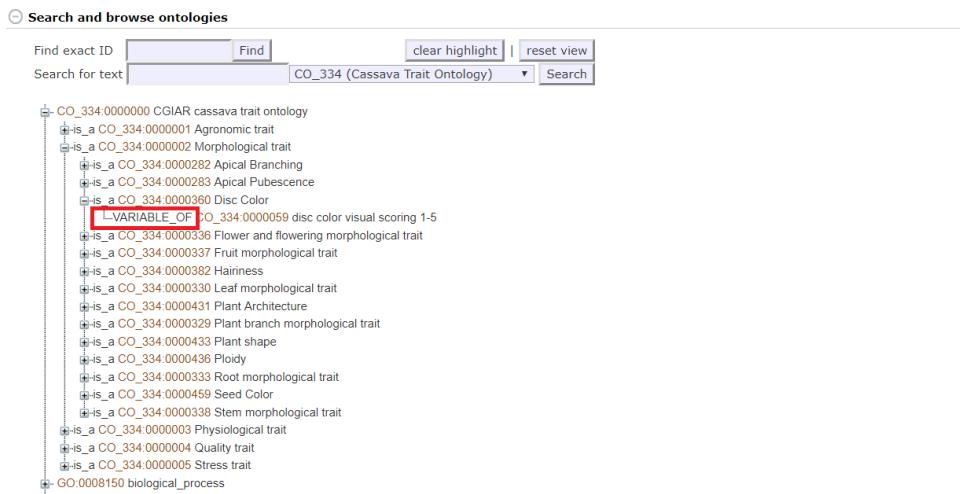
[add to list](#)

## 2.5 Ontology Browser

A more advanced tool for searching for Traits is the ontology browser, available by clicking on Analyze and Ontology Browser. From here you can search ontologies and see the various classifications of terms in a tree display.



The terms which appear in the Trait Search in 2.4 are only variable terms. The ontology browser shows these variables as different from their grouping terms by indicating VARIABLE\_OF like in the following screenshot.



## 2.6 Search Seedlots

Seedlots are different from Accessions in that they represent the physical seed being evaluated in an experiment. Seedlots have things like physical storage locations and seed quantities, which accessions do not. To search for available seedlots you go to Manage and then click Seed Lots. By clicking Search Seedlots, you can specify query information. The results from your search will be in the table below the search form.

**Available Seedlots**

**Search Seedlots**

Seedlot Name:	<input type="text"/>
Breeding Program:	<input type="text"/>
Contents (Accession):	<input type="text"/>
Location:	<input type="text"/>
Minimum Count:	<input type="text"/>

**Search**

**About Seedlots**

**What are seedlots?**

- Seedlots represent physical seed in packets.
- This can come from different breeding accessions.
- Seedlots can have a specific location, box, weight(g), and count.
- Seedlots can belong to breeding programs and organizations.

**How do I inventory my seed?**

- 1) Make sure your seedlots are in the database. Use "Add New Seedlot" to add a single seedlot or "Upload New Seedlots" to add many.
- 2) Make sure your seedlots are barcoded. You can print these barcodes from the database.
- 3) Use the "Seed Inventory" Android Application to scan seedlot barcodes and record weight. Then use "Upload Inventory" to upload this info into database. If you prefer you can create your own CSV file and upload that, if you do not want to use the Seed Inventory Application.
- **For more info about the "Seed Inventory" Android Application go to [Seed Inventory](#).**
- It is also possible to manually enter a transaction by going to the seedlot detail page and clicking "Add New Transaction".

**Seedlots**

[Add New Seedlot] [Upload New Seedlots] [Upload Inventory]

Show 10 <input type="button" value="▼ entries"/>	Seedlot Name	Breeding Program	Contents	Seedlot Location	Count	Weight(g)	Owners	Delete
	new_test_crossP001_001	test	new_test_crossP001 (accession)	NA	1		X	
	new_test_crossP002_001	test	new_test_crossP002 (accession)	NA	1		X	
	new_test_crossP003_001	test	new_test_crossP003 (accession)	NA	1		X	
	new_test_crossP004_001	test	new_test_crossP004 (accession)	NA	1		X	
	new_test_crossP005_001	test	new_test_crossP005 (accession)	NA	1	-7	X	
	new_test_crossP006_001	test	new_test_crossP006 (accession)	NA	1		X	
	test_accession1_001	test	test_accession1 (accession)	NA	-1	-72	X	
	test_accession5_001	test	test_accession5 (accession)	NA	1		X	

Showing 1 to 10 of 515 entries      Previous      ...  Next

<input type="text" value="seedlots"/>	<input type="button" value="add to new list"/>
<input type="text"/>	<input type="button" value="▼ add to list"/>



# Chapter 3

## Managing User Roles



The screenshot shows a web browser window titled "BreedBase" with the URL "breedbase.org/breeders/manage\_roles/". The page is titled "Manage User Roles" and displays a table of users and their assigned roles. The table has two columns: "User" and "Roles". The "User" column lists three entries: "Fred Sanger", "Jane Doe", and "John Doe". The "Roles" column contains colored buttons representing different roles: red for "curator", yellow for "submitter", blue for "ITTA", and light blue for "test". There are also small "X" and "+" icons next to each role button. At the bottom of the table, it says "Showing 1 to 3 of 3 entries". The page footer includes the "BREEDBASE" logo and the text "BREEDBASE is located at the Boyce Thompson Institute.", followed by the BTI logo.

User	Roles
Fred Sanger	X curator X submitter X ITTA +
Jane Doe	I test +
John Doe	I ITTA +

### 3.1 What are User Roles?

Every user account in Breedbase has one or more associated “roles” that determine the authorizations (what the user is allowed to do) in the database.

There are three fundamental roles, “curator”, “submitter”, and “user”, which determine basic read/write levels. The “curator” status can read and write everything in the database. The “submitter” status can add information and edit or delete previously submitted information. The “user” type can only read data. Additional roles represent the breeding programs, and are sometimes used to fine-tune write and edit capabilities, as it necessary for multiple users in a breeding program to edit each other’s data.

## 3.2 The Manage User Roles page

In the “Manage” menu, select the item “User Roles”. This will show the current users in the database with their associated roles. If you are logged in as a curator, the table will show system roles as well as breeding program roles; if you are logged in as a submitter or user, it will show breeding program membership.

If logged in as a “curator”, the roles can be added or deleted.

- To delete a role, click on the X in the role name. A confirm dialog will be displayed to prevent accidental deletion.
- To add a role, click on the plus sign next to the roles. A dialog will pop up with a list of roles. Select the desired role and click “Submit”.
- The new role should be displayed next to the user immediately.
- Role deletions and additions will be effective immediately.

It is recommended that few users be given the “curator” privileges to avoid confusion over data ownership and accidental data overwriting and deletion.

@ref(managing\_user\_roles)

# Chapter 4

## Managing Breeding Programs

New breeding programs can be added by using “Add New Program” button on the “Manage Breeding Programs” page.

The screenshot shows a web-based application interface for managing cassava breeding programs. At the top, there is a navigation bar with links for Search, Manage, Analyze, Maps, and About. On the right side of the header are search, user, and list management buttons. The main content area is titled "Manage Breeding Programs". Below the title is a table listing various breeding programs, each with a name and a brief description. The table includes rows for IITA, NRCRI, NaCRRRI, ARI Tanzania, NaCRRRI Germplasm Collection, CIAT, CARI, Rayong, KU, CSIR, SCP, and GN. At the bottom of the table is a blue button labeled "Add New Program", which is highlighted with a red rectangular box. The entire screenshot is framed by a thin gray border.

Name	Info
IITA	IITA cassava breeding program, Ibadan, Nigeria
NRCRI	NRCRI cassava breeding program, Umudike, Nigeria
NaCRRRI	NaCRRRI cassava breeding program, Namulonge, Uganda
ARI Tanzania	ARI Tanzania Cassava Program
NaCRRRI Germplasm Collection	NaCRRRI Landraces Germplasm collection
CIAT	CIAT cassava breeding, Cali, Colombia
CARI	Central Agricultural Research Institute, Suakoko, Liberia
Rayong	Rayong Field Crop Research Center, Thailand
KU	Kasetsart University cassava breeding program, Thailand
CSIR	Crops Research Institute, Ghana
SCP	New Cassava Varieties and Clean Seed to Combat CBSD and CMD
GN	BREEDING FOR VALUE

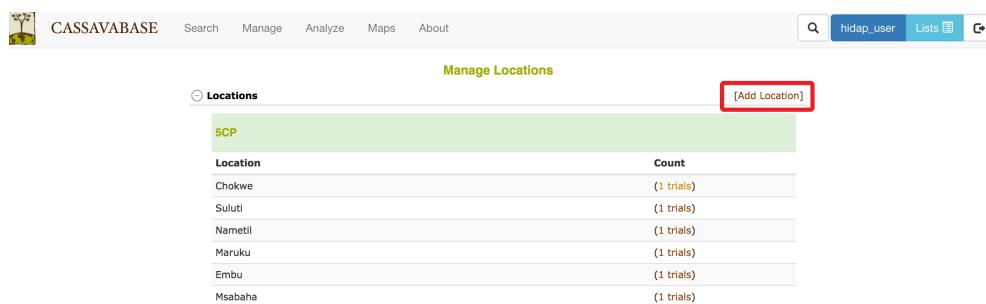
Clicking on the “Add New Program” button will generate a blank form for you to fill out the name and description of the breeding program that you want to add. After completing the form, click on “Add Breeding Program” button to finish the process.

The screenshot shows the CASSAVABASE web application. At the top, there is a navigation bar with links for Search, Manage, Analyze, Maps, and About. On the right side of the header, there is a search bar, a user account link ('hidap\_user'), a 'Lists' button, and a refresh icon. The main content area displays a list of breeding programs under the heading 'NaCRII Germplasm Collection'. This list includes entries for CIAT, CARI, Rayong, KU, CSIR, 5CP, and GN, each with a brief description. A modal dialog box titled 'Add New Breeding Program' is overlaid on the page. It contains fields for 'Name:' and 'Description:', both currently empty. Below these fields are two buttons: 'Close' and a blue 'Add Breeding Program' button. At the bottom of the main content area, there is a blue 'Add New Program' button.

# Chapter 5

## Managing Locations

Field locations can be managed using the “**Manage Locations**” page. On this page, locations in the database are organized based on their breeding programs. Each location has a link to trials conducted in that location. To add a new location, click on the “Add Location” button that links to the “Add New Location” form.



The screenshot shows the CASSAVABASE interface with the title "Manage Locations". At the top, there is a navigation bar with links for Search, Manage, Analyze, Maps, and About. On the right side of the header, there are buttons for "hidap user", "Lists", and a refresh icon. Below the header, the page title "Manage Locations" is displayed above a table. The table has a header row with columns for "Location" and "Count". The data rows list locations such as Chokwe, Suluti, Nametil, Maruku, Embu, and Msabaha, each with a "(1 trials)" entry. At the bottom right of the table, there is a red-bordered button labeled "[Add Location]".

Location	Count
Chokwe	(1 trials)
Suluti	(1 trials)
Nametil	(1 trials)
Maruku	(1 trials)
Embu	(1 trials)
Msabaha	(1 trials)

On the “Add New Location” form, fill out the location name that you want to add. Latitude, longitude, and altitude are optional. Submit the new location by clicking on the “Add Location” button at the bottom right of the form.



# Chapter 6

## Managing Accessions

The “Manage Accession” page provides links for adding new accessions. You can choose to add accessions into the database by either using a List you have created or by uploading XLS or XLSX file. Both options will be detailed below. To begin click on the “Add Accessions or Upload Accession Info” link.

The screenshot shows the "Manage Accessions" page with three main sections:

- Accessions:** Shows a total of 137066 accessions. Includes a "Search Accessions" input field and buttons for "Add Accessions Or Upload Accession Info" (highlighted with a red box) and "Upload Pedigree File".
- Find Trials in Common:** Allows selecting an accession list from a dropdown menu and clicking "Find Trials". A note says "Use a list of accessions to search for trials that contain them all".
- Populations:** Includes a "[Create Population]" button.

This will open a dialog allowing you to select either “Using Lists” or “Uploading a File”.

### 6.1 Add Accessions Using A List

First we will show how to add accessions “Using Lists”.

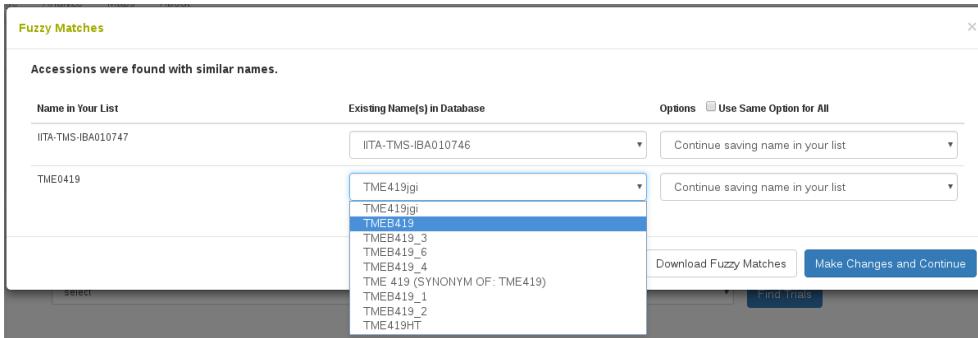


Here you select an accession list which you have previously made. If you need to create or edit your list you can do so now by clicking “Manage Lists”. Once you have selected your list you can click “Continue”.

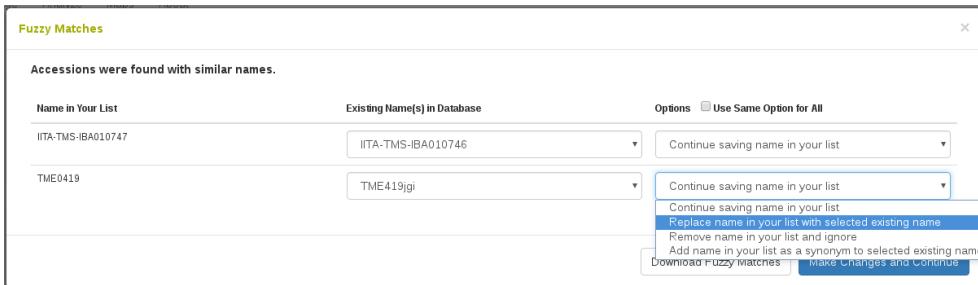
The first dialog which can appear will show the accessions which already exist in the database.

Found Accessions	
<b>The following accessions already exist and cannot be added:</b>	
Total number already in the database(7)	
Show <b>10</b> entries	Search: <input type="text"/>
Search Name	Found in Database
IITA-TMS-IBA010746	IITA-TMS-IBA010746
IITA-TMS-IBA010758	IITA-TMS-IBA010758
IITA-TMS-IBA010760	IITA-TMS-IBA010760
IITA-TMS-IBA010779	IITA-TMS-IBA010779
IITA-TMS-IBA010797	IITA-TMS-IBA010797
IITA-TMS-IBA010816	IITA-TMS-IBA010816
IITA-TMS-IBA010819	IITA-TMS-IBA010819
Showing 1 to 7 of 7 entries	
Previous <b>1</b> Next	
<b>Continue</b>	

Click “Continue”. The next dialog which can appear will show accessions which have very similar matches to the accession names you are adding. In the example below, there are two accession names that are very similar to accession names already in the database. ‘TME0419’ is very similar to ‘TME419’, and actually is probably a mistake that should not be added to the database.



To avoid situations in adding a mistaken duplicate accession, the database gives you options for moving forward with these very similar looking accession names. You can either “continue saving the name in your list”, “replace name in your list with selected existing name”, “remove name in your list and ignore”, or “add name in your list as a synonym to selected existing name”.



Clicking “Download Fuzzy Matches” will return a tabular result of the “fuzzy” accession name results shown. Click “Make changes and continue” to move on.

The final dialog shows the accessions that will be added. Here you need to assign the species of these accessions. You can optionally group the accessions into a population and/or add an organization for the accessions.

The dialog box has a title bar 'Accessions to Be Added'. It contains three input fields: 'Species name for added accessions' (Manihot esculenta), 'Population name for added accessions (optional)', and 'Organization name for added accessions (optional)'. Below these fields is a section titled 'The following accessions are new and will be added to the database:' with a note 'Total number to be added(2)'. Underneath, it lists 'ITA-TMS-IBA010747 completely\_new\_accession'. At the bottom right are 'Close' and 'Add Accessions' buttons.

Once you click “Add Accessions”, the new accessions will be created in the database and you will see the following confirmation dialog, which includes links to the newly created accessions.



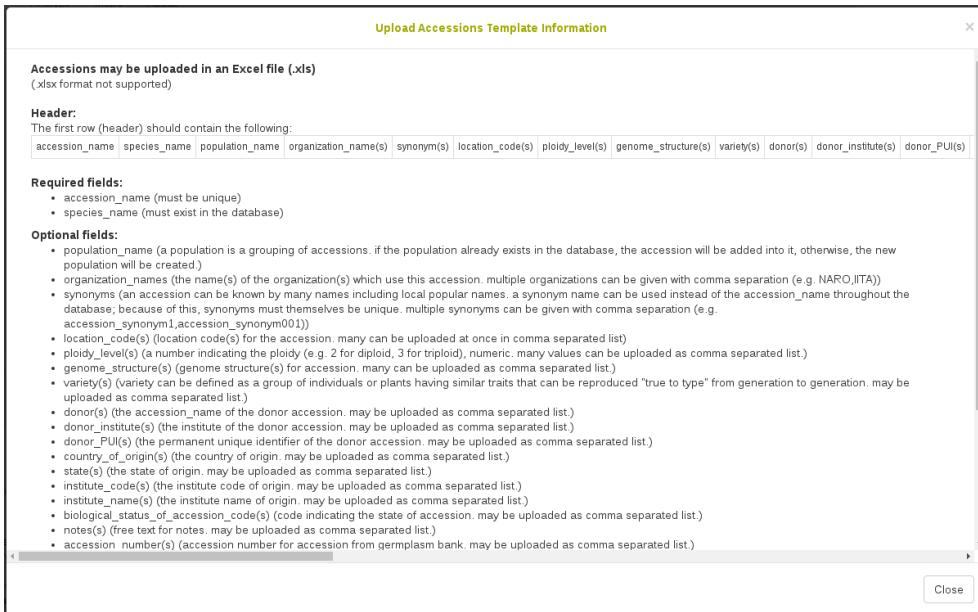
## 6.2 Uploading Accessions and Accession's Info From A File

The process to upload accessions is very similar to using a list, but enables you to add a variety of properties, such as synonyms, to the accessions in bulk.

## 6.2. UPLOADING ACCESSIONS AND ACCESSION'S INFO FROM A FILE49



Clicking on “Spreadsheet format” will show the following dialog. Here it shows that the file must be XLS or XLSX format and can contain a number of header columns as attributes. It is important that you use exactly the same header column names as listed here. In columns that indicate that many attribute values can be passed at once using (s), such as synonym(s), you can pass a comma separated list of values, such as ‘synonym1,synonym2’.



Once you have selected your XLS or XLSX file for upload, click “Continue”.

The following process is the same way as with lists:

The first dialog which can appear will show accession names which are already in the database.

Click “Continue” and the next dialog that can appear will show “fuzzy” matches for the accession names you are trying to upload. Here you can choose to prevent adding accession names which look very similar to each other as wrongly duplicated accessions.

Click “Continue” and the final dialog that will appear will show the information to be added into the database. Here it is divided into accession names that are new and accession names that already exist in the database; however, for the accession names that already exist it will show additional attributes that originated from your file that will be added to these accessions.

uniquename	properties
new_test_accession01	state:Oyo germplasmName:new_test_accession01 ploidyLevel:2 countryOfOriginCode:Nigeria species:Manihot esculenta defaultDisplayName:new_test_accession01 organizationName:test_organization populationName:test_population locationCode:ITH synonyms:new_test_accession_synonym1,new_test_accession_synonym2,new_test_accession_synonym3
new_test_accession02	organizationName:test_organization populationName:test_population defaultDisplayname:new_test_accession02 synonyms:germplasmName:new_test_accession02 state:Oyo countryOfOriginCode:Nigeria species:Manihot esculenta
new_test_accession03	germplasmName:new_test_accession03 state:Oyo countryOfOriginCode:Nigeria species:Manihot esculenta defaultDisplayName:new_test_accession03 organizationName:test_organization populationName:test_population synonyms:new_test_accession3_synonym1
new_test_accession04	synonyms: populationName:test_population organizationName:test_organization defaultDisplayname:new_test_accession04 species:Manihot esculenta countyOfOriginCode:Nigeria state:Oyo germplasmName:new_test_accession04

uniquename	properties
ITA-TMS-IBA010746	stock_id:4867 synonyms:ITA-TMS-IBA010746_synonym1,ITA-TMS-IBA010746_synonym2 species:Manihot esculenta defaultDisplayName:ITA-TMS-IBA010746 germplasmName:ITA-TMS-IBA010746 organizationName:null populationName:test_population

Once you click “Add Accessions”, the new accessions and information will be created in the database and you will see the following confirmation dialog, which includes links to the created and updated accessions.



## 6.3 Email alert for accession upload

When uploading accessions from a file, you have the option to receive email notifications about the status and results of your upload by clicking the “Email Alert” checkbox. By default, the system will use the email address associated with your account, but you have the option to enter a different email address if you prefer. After submitting, the upload process runs in the background, allowing you to continue using the interface without interruptions. Once the process completes, you will receive an email with the upload results, including any warnings or errors that may have occurred during the upload.

A screenshot of a web-based form titled "Add Accessions". The form has two main tabs at the top: "Using Lists" (which is currently selected) and "Uploading a File".  
  
Below the tabs, there is a section for "File format information" with links to "Spreadsheet format".  
  
The main configuration area includes:

- "Upload File:" with a "Browse..." button and the file path "demo\_accessions.xlsx".
- "Email Alert:" with a checked checkbox labeled "On" and an input field "Email: noreply@breedbase.com".
- "Use Fuzzy Search:" with a checked checkbox and a note: "Note: Use the fuzzy search to match similar names to prevent uploading of duplicate accessions. Fuzzy searching is much slower than regular search. Only a curator can disable the fuzzy search."
- "Append Synonyms:" with a checked checkbox and a note: "When checked, add synonyms of existing accession entries to the synonyms already stored in the database. When not checked, remove any existing synonyms of existing accession entries and store only the synonyms in the upload file."

  
At the bottom of the form, there is a yellow callout box with the following text:

Accessions may be uploaded using any of the supported file types: MS Excel (.xls or .xlsx), comma-separated file (.csv), tab-delimited file (.txt or .tsv), or semicolon-separated file (.ssv).  
Optional columns may be left out, if not used in your data.

  
At the very bottom are "Close" and "Continue" buttons.

## 6.4 Add Parentage (Pedigree) Information to Accessions

Pedigree data can be uploaded from your computer by clicking on “Upload Pedigree File”

The screenshot shows the 'Manage Accessions' interface. At the top right, there is a red box highlighting the '[Add] [Upload Pedigree File]' button. The page displays a total of 127130 accessions. There are sections for 'Find Trials in Common' and 'Populations'. A note at the bottom right says 'Maximize to see population info'.

**IMPORTANT!** Please use only tab-delimited text file format (.xls or .xlsx formats are NOT supported).

You can find detailed information on how to prepare pedigree file by clicking on “File format information”

The currently supported format has four tab separated columns:

progeny name female parent accession male parent accession type

Type can be biparental, self, backcross, sib, polycross, reselected, or open. In the case of the open type, the male parent accession field can remain blank. For all other types, both columns should be filled, even if they contain the same information as another column (such as self).

The screenshot shows a modal dialog titled 'Upload Pedigrees'. It contains a note about file format (Spreadsheet Format) and a file upload input field labeled 'Choose File' with the message 'no file selected'. There are 'Close' and 'Submit File' buttons at the bottom.

**Template Information**

Pedigrees may be uploaded in tab-delimited text file format  
(.xls or .xlsx formats are NOT supported)

**Header:**  
The first row (header) should contain the following:

progeny name	female parent accession	male parent accession	type
--------------	-------------------------	-----------------------	------

**Required fields:**

- progeny name (must exist in the database and can be accession uniquename or accession synonym)
- female parent accession (must exist in the database)
- type (biparental, open, self)

**Optional fields**

- male parent accession (can be accession uniquename or accession synonym or population name).

**Notes**

- Always specify the type of the cross (biparental, open, or self).
- If the type is open and no potential parents are known, leave the male parent field empty

**Close**

## 6.5 Working with grafts

Grafts are plants that are composed of a rootstock and a scion, which are genetically different and fused together, usually at the stem level.

To work with grafts, the grafts interface needs to be activated by adding a configuration parameter in the sgn\_local.conf file. The parameter is show\_grafting\_interface. It should be set to 1 in sgn\_local.conf, the default is 0 (in sgn.conf).

Grafts to be created need to be specified using an Excel file (xlsx format) with two columns. The first column should have the header “scion accession” and should list accession names that will be scions. The second column should have the header “rootstock accession” and should list accession names that will be rootstocks.

In the database, the graft accessions will be created as single accessions. The

graft accession will have two relationships, one to the scion accession (`scion_of` relationship) and one to the rootstock (`rootstock_of` relationship). These relationships are displayed on the pedigree viewer. The graft accession name is created from the scion accession name and the rootstock accession name, separated by the graft separator character. By default, the graft separator character is the plus sign ‘+’. The graft separator character can be changed in the `sgn_local.conf` file, using the parameter `graft_separator_string`. The graft separator string should not occur in any other accession names that are not grafts.

When the grafting interface is activated, a new button will be shown on the manage accessions page, called “Upload Grafts”.

Clicking the button brings up the upload grafts dialog.

Select the Excel file containing the grafting information. The system will validate the file, for example, check whether the accessions are in the database, and if the headers are correct.

The validation result will be presented, and if problems are found, they will be listed. In addition, if there are problems, the Upload button will be grayed out and upload will not be possible. Conversely, if there are no problems, the Upload button will be activated and can be clicked to upload the data.

If the upload completes, a completion message is displayed with a summary what was uploaded.

Grafted accessions can be used like any other accession, for example, they can be used on field layouts. If you create a list of graft accessions, use the list type ‘accessions’.

Note that you shouldn’t create new grafts based on other grafts. The scion accession and the rootstock accession have to be different, otherwise they will not be created.

## 6.6 Bulk renaming of accessions

Accessions can be renamed in bulk using the rename accessions feature. To rename accessions, prepare a tab delimited file with two columns: the first column should have the header “old name” and contain the accession names

that need to be changed. The second column should have the header “new name” and contain the names that the accessions in column 1 should be renamed to.

The accession renaming feature is available from the Manage->Accessions page. Click on the “Rename Accessions” button. The first step is the upload of the file with a verification step. The verification step checks whether all the accession names in column 1 exist in the database, and whether all the accession names given in column 2 do NOT exist in the database. Only if both conditions are met, will the “rename” button become active, otherwise an error message is displayed listing the offending accession names.

Optionally, the old name can be automatically added as a synonym to the renamed accession, using the checkbox on the submit form. This option is clicked by default. Unclick the checkbox to NOT save any old names as synonyms.

Note that accession renaming should not be undertaken lightly. This feature is intended for special use cases, such as where accessions are created in a nursery with a name that is different from the accession name in the downstream breeding program.

It can also be used to rename accessions in bulk that have spelling mistakes and other issues. Please note however, that the tool does not make any attempt to change the names of associated elements, such as plots, that may have been constructed using accession names.

Because of the many implications of accession renaming, the feature is limited to accounts with the curator role.



# Chapter 7

## Managing Seed Lots

Seedlots are different from Accessions in that they represent the physical seed being evaluated in an experiment. Seedlots have things like physical storage locations and seed quantities, which accessions do not. The seed in seedlots can be from crosses or can be named accessions. Seedlots from crosses would represent seed harvested. Click Manage and then Seed Lots to begin.

Available Seedlots

About Seedlots

What are seedlots?

Seedlots represent physical seed in packets.

- This can be from crosses or named accessions.
- Seedlots can have a specific location, box, weight(g), and count.
- Seedlots can belong to breeding programs and organizations.

How do I inventory my seed?

1) Make sure your seedlots are in the database. Use "Add New Seedlot" to add a single seedlot or "Upload New Seedlots" to add many.  
2) Make sure your seedlots are barcoded. You can print these barcodes from the database.  
3) Use the "Seed Inventory" Android Application to scan seedlot barcodes and record weight. Then use "Upload Inventory" to upload this info into database. If you prefer you can create your own CSV file and upload that, if you do not want to use the Seed Inventory Application.  
For more info about the "Seed Inventory" Android Application go to [Seed Inventory](#).  
It is also possible to manually enter a transaction by going to the seedlot detail page and clicking "Add New Transaction".

Add New Seedlot | Upload New Seedlots | Upload Inventory

Seedlots

Search Seedlots

Show 10 entries

Seedlot Name	Breeding Program	Contents	Seedlot Location	Count	Weight (g)	Owners	Delete
new_test_crossP001_001	test	new_test_crossP001 (accession)	NA	1		X	
new_test_crossP002_001	test	new_test_crossP002 (accession)	NA	1		X	
new_test_crossP003_001	test	new_test_crossP003 (accession)	NA	1		X	
new_test_crossP004_001	test	new_test_crossP004 (accession)	NA	1		X	
new_test_crossP005_001	test	new_test_crossP005 (accession)	NA	1	-7	X	
new_test_crossP006_001	test	new_test_crossP006 (accession)	NA	1		X	
test_accession1_001	test	test_accession1 (accession)	NA	-1	-72	X	
test_accession5_001	test	test_accession5 (accession)	NA	1		X	

Showing 1 to 10 of 515 entries

Previous 1 2 3 4 5 ... 52 Next

seedlots

## 7.1 Add New Seedlot(s)

To add a single new seedlot, click on “Add Seedlot”. This will bring up the following dialog where you enter information about where the seedlot exists, what accession or cross is contained in it, and how many seeds there are. A seedlot must contain either an accession or a cross, and not both. A seedlot must have a weight in grams or a seed count or both of these.

**Create New Seedlot**

<b>Name:</b>	Required
<b>Breeding Program:</b>	test
<b>Location:</b>	Required
<b>Box Name:</b>	Optional
<b>Contents:</b>	
<b>Accession name:</b>	One Content Required
OR	
<b>Cross name:</b>	One Content Required
<b>Amount [number of seeds]:</b>	Amount OR Weight(g) Required
<b>Weight (g):</b>	Amount OR Weight(g) Required
<b>Organization:</b>	Optional
<b>Timestamp:</b>	Wed Mar 14 10:44:34 2018
<b>Description:</b>	Optional

**OK**

In the case where you have many seedlots to add to the database, you can upload an excel XLS or XLSX file instead. Click “Upload Seedlots” to see the following dialog.

**Upload Seedlots**

**File format information**  
Spreadsheet format

**Breeding Program:** NelsonLab

**Location:** Required

**Population Name:** Optional

**Organization Name:** Optional

**Upload File (.xls):** Choose File No file chosen

**Submit**

## 7.2 Seedlot Transactions

Seedlots are capable of tracking where seeds came from, such as from crosses, and to where seeds go, such as to plots in the field. If you navigate to a seedlot detail page you will see the following.

**Seedlot test\_accession2\_001**

**Details**

Breeding Program	test
Seedlot Name	test_accession2_001
Organization	my org
Location Code	NA
Box Name	box2
Contents	test_acquisition2 (acquisition)
Current count	1
Current weight(g)	34

**Transactions**

Show 10 entries

Transaction Id	Transaction Date	From	To	Transaction Num Seeds	Transaction Weight(g)	Operator	Description	Options
40088	Mon Sep 18 11:44:00 2017	test_acquisition2 (acquisition)	test_acquisition2_001 (seedlot)	+1	NA	nmares	Auto generated seedlot from accession. DbPatch 00085	[Edit]
41456	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	-34	some user	Seed inventory CSV upload.	[Edit]
41459	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	-34	some user	Seed inventory CSV upload.	[Edit]
41460	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	-34	some user	Seed inventory CSV upload.	[Edit]
41464	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	-34	some user	Seed inventory CSV upload.	[Edit]
41466	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	+170	some user	Seed inventory CSV upload.	[Edit]
41470	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	NA	some user	Seed inventory CSV upload.	[Edit]
41473	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	+0	some user	Seed inventory CSV upload.	[Edit]
41477	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	+0	some user	Seed inventory CSV upload.	[Edit]
41478	2018-04-01-02-04-32	test_acquisition2_001 (seedlot)	test_acquisition2_001 (seedlot)	NA	+0	some user	Seed inventory CSV upload.	[Edit]

Showing 1 to 10 of 10 entries

On this page you see and can edit information regarding a single seedlot,

such as its name and location. You will also see a table indicating all the transactions that a seedlot has been involved in, such as if it was planted in a plot in the field. Transactions to field plots are created when adding or uploading a new trial or from a trial's detail page. Clicking on "Add New Transaction" let you add a transaction from between this seedlot and another seedlot. This kind of transaction is useful for representing if you have distributed seed to different locations.

Add New Seedlot Transaction

**Transaction Type:** Added to this Seedlot (test\_accession2\_001)

**Taken From Existing Seedlot:** test\_accession4\_001  
Only showing seedlots with matching content

**Transaction Amount (number seeds):** Amount OR Weight Required

**Transaction Weight (g):** Amount OR Weight Required

**Timestamp:** Wed Mar 14 10:54:18 2018

**Description:** Required

OK

### 7.3 Seed Inventory

To inventory your seed: 1) Make sure your seedlots are in the database. Use "Add New Seedlot" to add a single seedlot or "Upload New Seedlots" to add many. 2) Make sure your seedlots are barcoded. You can print these barcodes from the database. 3) Use the "Inventory" Android Application to scan seedlot barcodes and record weight. Then use "Upload Inventory" to upload this info into database. If you prefer you can create your own CSV file and upload that, if you do not want to use the Inventory Application. For more info about the "Inventory" Android Application go to Inventory.

Clicking the "Upload Inventory" button will bring the following dialog:

**Upload Seedlot Inventory**

**How do I inventory my seed?**

- 1) Make sure your seedlots are in the database. Use "Add New Seedlot" to add a single seedlot or "Upload New Seedlots" to add many.
- 2) Make sure your seedlots are barcoded. You can print these barcodes from the database.
- 3) Use the "Seed Inventory" Android Application to scan seedlot barcodes and record weight. Then use "Upload Inventory" to upload this info into database. If you prefer you can create your own CSV file and upload that, if you do not want to use the Seed Inventory Application.
- For more info about the "Seed Inventory" Android Application go to [Seed Inventory](#).**
- It is also possible to manually enter a transaction by going to the seedlot detail page and clicking "Add New Transaction".

i
File format information  
Spreadsheet format

Upload File (.csv):

No file chosen

The CSV file that should contain your inventory should meet these Template requirements. The Seed Inventory Android Application exports this exact file.

**Upload Template Information**

**Seedlots may be uploaded in a CSV file (.csv)**  
(Excel .xls and .xlsx format not supported)

**Header:**  
The first row (header) should contain the following:

box_id	seed_id	inventory_date	inventory_person	weight_gram
--------	---------	----------------	------------------	-------------

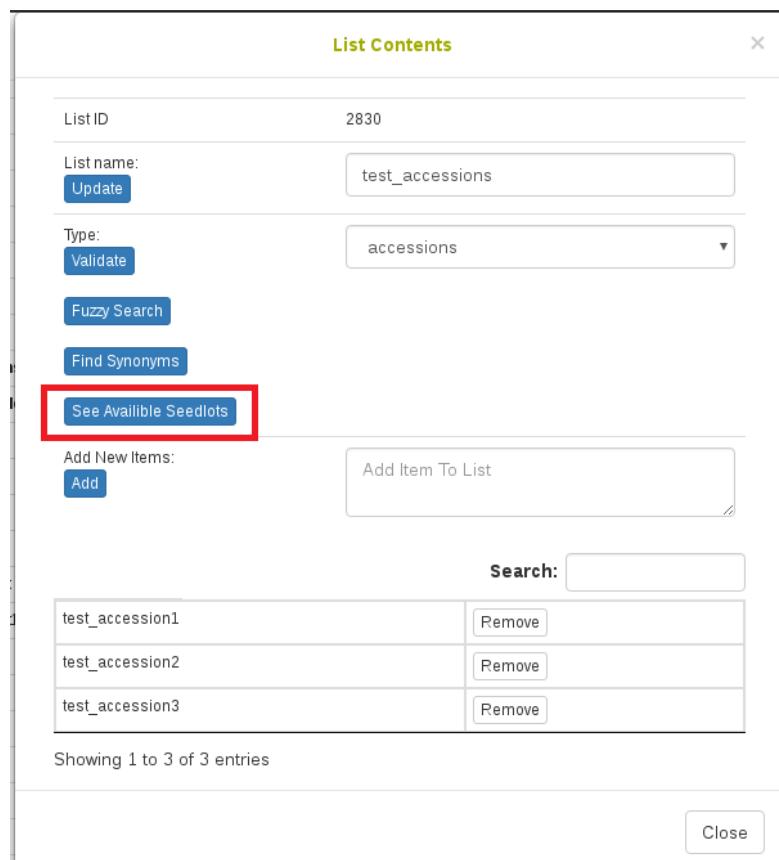
**Required fields:**

- box\_id (the name of the box that the seedlot is in. also called box\_name.)
- seed\_id (unique identifier for the seedlot. must exist in the database. also called seedlot\_name)
- inventory\_date (a timestamp for when the seedlot was inventoried)
- inventory\_person (the name of the person doing the inventory. can be any name. also called operator\_name)
- weight\_gram (the weight in grams of the seedlot)

## 7.4 Find Seedlots For a List of Accessions

A convenient tool for searching available seedlots for a list of accessions is available in the list tool. First open up your list of accessions. For help

opening a list of accessions please see the List section help. There is a button called “See Available Seedlots”.



Once you click this, you will see the following table in a dialog. From here you can create a list of seedlots using the checkboxes and the input at the bottom.

Available Seedlots					
Accessions	Seedlots				
	Breeding Program	SeedlotName	Contents	SeedlotLocation	Count
test_accession3	IITA	test_accession3_001	test_accession3	NA	-71
	IITA	seedtest004	test_accession3	x1 location	99
	IITA	seedtest004	test_accession3	x2	48
	IITA	seedtest004	test_accession3	x2	99
test_accession1	NPCRI	UG120243_0015	test_accession1	NA2	1
test_accession2	IITA	test_accession2_001	test_accession2	NA	-138
	IITA	seedtest003	test_accession2	x1 location	0
	IITA	seedtest003	test_accession2	x2	0
	IITA	seedtest003	test_accession2	x2	0
	IITA	seednx10	test_accession2	x2	0
	IITA	seednx11	test_accession2	x2	26

Create a New List from Selected Seedlots:

## 7.5 Create a seedlot for an Accession or Cross

Complementary to what we saw above for creating seedlots from the “Manage Seedlots” page, it is possible to create a new seedlot from an accession’s detail page or from the cross detail page. On the accession detail page, this is visible in the “Related Stocks” section as seen below. The cross detail page has an identical section. Notice the link for creating a new seedlot, which streamlines adding the seedlot.

Related stocks  
 Related stocks in trials  
 Seedlots of this Accession [\[Create New Seedlot\]](#)

Show 10 entries	Search:			
SeedlotName	Breeding Program	Contents	SeedlotLocation	Count
002B_001	IITA	002B (accession)	NA	-17
002B_test1_001	IITA	002B (accession)	Abuja	90

Showing 1 to 2 of 2 entries Previous 1 Next

Copy Seedlots to a List Copy the seedlot names showing in table to a new or existing list

Progenies  
 Groups /members  
 Related stocks for tissue sample

## 7.6 Add quality data to a seedlot

Quality information can be added to a seedlot in the quality field. This is also available as a column in the file upload format. It is recommended to use a controlled vocabulary, defined by the user, for the quality field. For example, good quality seed should be labelled “ok”, whereas other quality descriptors could be “moldy”, “insect damage”, or “low sprouting”, etc.

## 7.7 Seedlot Maintenance Events

For some crops, such as sugar kelp, a “seedlot” requires routine maintenance for the successful long-term storage of the seedlot. (For example, a Seedlot Maintenance Event for sugar kelp would be the routine change of the water that gametophytes are kept in). Breedbase can now store a record of these Seedlot Maintenance Events associated directly with existing Seedlots. Maintenance Events can be uploaded using a simple Excel template or recorded directly on the website.

### 7.7.1 Setup

Each Breedbase instance needs to be configured to support the storage of Seedlot Maintenance Events since each crop will have their own distinct set of maintenance events for their seedlots. To check if your Breedbase instance supports this feature, go to the Manage menu and select the Seed Lots page. Make sure you are logged in and look for the **Seedlot Maintenance** button near the top, next to the **Create Seedlot(s)** and **Upload Inventory** buttons. If you don’t see this button, contact the developer(s) supporting your Breedbase instance and ask if they can setup this feature.

**Available Seedlots**

The screenshot shows a web interface for managing seed lots. At the top, there are two tabs: 'About Seedlots' and 'Seedlots'. Below these are three buttons: 'Create Seedlot(s)', 'Upload Inventory', and 'Seedlot Maintenance'. The 'Seedlot Maintenance' button is circled in orange. A large search form titled 'Search Seedlots' is centered. It contains fields for 'Seedlot Name', 'Breeding Program', 'Contents (Accession UniqueName)', 'Contents (Cross Unique ID)', 'Location', 'Minimum Count', 'Minimum Weight (g)', and 'Quality'. Each field has an 'exact match' checkbox next to it. A 'Search' button is at the bottom of the form. Below the form, a note says 'Enter search terms and click the "Search" button to view results.'

*The location of the Seedlot Maintenance button on the Manage > Seed Lots page*

### 7.7.2 Adding Events

Seedlot Maintenance Events can be added using two methods: 1) Uploading an Excel template or 2) Recording events directly on the website

#### Uploading Events with Excel Template

To bulk-upload a file of Seedlot Maintenance Events, first create an Excel (.xls or .xlsx) file with the following headers:

- **seedlot** - the name of the Seedlot to associate the event with (must exactly match an existing Seedlot in the database)
- **type** - the name of the Seedlot Maintenance Event type (these vary between Breedbase instances, a list of supported event types is displayed)

on the upload page)

- **value** - the value of the Seedlot Maintenance Event (these may be different for each event type and vary between Breedbase instances, a list of supported event values is displayed on the upload page)
- **notes** - optional, additional notes/comments about the event
- **operator** - the username of the Breedbase user that recorded the event
- **timestamp** - the date/time the event was recorded, in ‘YYYY-MM-DD HH:MM:SS’ format

Once you have an Excel file with the events filled out, follow these steps to upload the events to the database:

1. Make sure you are logged in to your Breedbase instance
2. Go to the Manage > Seed Lots page
3. Select the **Seedlot Maintenance** button
4. Select the **Upload Maintenance** button
5. Choose your Excel (.xls or .xlsx) file to upload
6. Select the **Upload** button

The dialog box has a header "Upload Seedlot Maintenance Events" and a close button "X". It contains the following sections:

- Select an Excel (.xls) file with the Seedlot Maintenance Events to upload**
- Requirements:**
  - The Maintenance Events are associated with Seedlots, so the name of the Seedlot in the file must match an existing Seedlot in the database. If a Seedlot is not yet in the database, go to the [Manage Seedlots](#) page to create it first.
  - The name of the Maintenance Event must be a valid event type. Valid event types include:

Event Type Name	Event Type Values
Water Change	Successful, Unsuccessful
Blended	Successful, Unsuccessful
Container Scraped	Successful, Unsuccessful
Light Intensity	<10, 20, 30-45, 50-75
Light Color	red, white
Container Size	1 L, 500 mL, 250 mL, 125 mL, vial
Form	Backup vial, Flask, Backup vial sibling
Biomass	high, medium, low
Health	healthy, not healthy
Color	1, 2, 3, 4, 5, 6, 7
Stickiness	yes, no
Clumping	yes, no
Contaminants	green, bacteria, other, all, green and bacteria, green and other, bacteria and other, none
Additional Notes	Any Value

**i File format information**  
Spreadsheet Format

Upload File:  No file chosen

*The Seedlot Maintenance upload dialog, showing the supported event types and values (for sugar kelp)*

## Recording Events on Website

To add individual Seedlot Maintenance Events to the database in real time, as they're being recorded, use the **Record Maintenance** page. Follow these steps to record Seedlot Maintenance Events:

1. Make sure you are logged in to your Breedbase instance

2. Go to the Manage > Seed Lots page
3. Select the **Seedlot Maintenance** button
4. Select the **Record Maintenance** button
5. Enter the **Seedlot Name** or scan a barcode that has the Seedlot Name encoded. Once entered, the box at the top of the page will display basic information about the Seedlot as well its recently recorded events.
6. Select or Enter the values of individual events
7. Optionally, notes button next to each event to add additional notes/comments about that specific event
8. Make sure the operator/username and timestamp are correct
9. Select the **Submit** button to add the recorded events to the database.  
NOTE: any events that remain selected as “Not Recorded” will not be submitted to the database.

**Record Seedlot Maintenance**

**Seedlot**

Name:	TEST_SEEDLOT_1-LOTA		Barcode
Contents:	SA18-CB-S1-FG1 (accession)		
Location:	WHOI		
Box:	Shelf 1 / Tray 1		
Recent Events:	Event	Value	Notes
	Water Change	Successful	additional notes
	Container Size	1 L	2021-07-22 13:04:24

**Maintenance Events**

**Actions**

- Water Change
 

<input checked="" type="button"/> Not Recorded	<input type="button"/> Successful	<input type="button"/> Unsuccessful	
--	-----------------------------------	-------------------------------------	--
- Blended
 

<input checked="" type="button"/> Not Recorded	<input type="button"/> Successful	<input type="button"/> Unsuccessful	
--	-----------------------------------	-------------------------------------	--
- Container Scraped
 

<input checked="" type="button"/> Not Recorded	<input type="button"/> Successful	<input type="button"/> Unsuccessful	
--	-----------------------------------	-------------------------------------	--

**Observations**

- Light Intensity
 

<input checked="" type="button"/> Not Recorded	<input type="button"/> <10	<input type="button"/> 20	<input type="button"/> 30-45	<input type="button"/> 50-75	
--	----------------------------	---------------------------	------------------------------	------------------------------	--
- Light Color
 

<input checked="" type="button"/> Not Recorded	<input type="button"/> red	<input type="button"/> white	
--	----------------------------	------------------------------	--
- Container Size
 

<input checked="" type="button"/> Not Recorded	<input type="button"/> 1 L	<input type="button"/> 500 mL	<input type="button"/> 250 mL	<input type="button"/> 125 mL	<input type="button"/> vial	
--	----------------------------	-------------------------------	-------------------------------	-------------------------------	-----------------------------	--

**Additional Notes**  
Any additional notes, usually concerning culture termination or partial use for an experiment

**Username/Timestamp**

Operator:	dwaring87
Timestamp:	2021-08-18 14:29:49

*The Seedlot Maintenance record page, as configured for sugar kelp*

### 7.7.3 Displaying Events

Recently recorded Seedlot Maintenance Events are displayed in a table from the main Seedlot Maintenance page, as well as the detail page for individual Seedlots.

Seedlot Maintenance									
<a href="#">About Seedlot Maintenance</a>		<a href="#">Seedlot Maintenance Tools</a>							
<a href="#">Seedlot Maintenance Events</a>		<a href="#">Record Maintenance</a> <a href="#">Upload Maintenance</a>							
<a href="#">Filter Events</a>		Filter maintenance events based on date, type, and/or value							
<a href="#">Excel</a> <a href="#">CSV</a>		Search: <input type="text"/>							
Seedlot	Event ID	Event Date	Event Type	Value	Notes	Operator	Options		
TEST_SEEDLOT_1-LOTA	381860	Thu Jul 22 13:04:24 2021	Water Change	Successful	additional notes	dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381866	Thu Jul 22 13:04:24 2021	Color	3		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381865	Thu Jul 22 13:04:24 2021	Biomass	medium		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381864	Thu Jul 22 13:04:24 2021	Container Size	1 L		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381863	Thu Jul 22 13:04:24 2021	Light Color	red		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381862	Thu Jul 22 13:04:24 2021	Light Intensity	<10		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381861	Thu Jul 22 13:04:24 2021	Blended	Unsuccessful		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381819	Fri Jul 9 13:22:24 2021	Clumping	yes		dwaring87	[Remove]		
TEST_SEEDLOT_1-LOTA	381818	Fri Jul 9 13:21:46 2021	Blended	Successful		dwaring87	[Remove]		
TEST_SEEDLOT_2	381816	Fri Jul 9 13:19:08 2021	located_in	Successful		dwaring87	[Remove]		
TEST_SEEDLOT_2	381807	Fri Jul 9 13:18:05 2021	Light Intensity	50-75		dwaring87	[Remove]		
TEST_SEEDLOT_2	381804	Fri Jul 9 13:18:05 2021	Water Change	Successful		dwaring87	[Remove]		
TEST_SEEDLOT_2	381805	Fri Jul 9 13:18:05 2021	Blended	Successful		dwaring87	[Remove]		
TEST_SEEDLOT_2	381806	Fri Jul 9 13:18:05 2021	Container Scraped	Successful		dwaring87	[Remove]		
TEST_SEEDLOT_2	381808	Fri Jul 9 13:18:05 2021	Light Color	red		dwaring87	[Remove]		
TEST_SEEDLOT_2	381809	Fri Jul 9 13:18:05 2021	Container Size	1 L		dwaring87	[Remove]		

*Unfiltered table of recent Seedlot Maintenance events*

The events displayed in these tables are sorted by timestamp, with the most recently recorded events displayed first. The displayed events can be filtered using any number of supported filter criteria, such as:

- seedlot names (as entered on the page or using an existing seedlot list)
- dates (on, on or before, before, on or after, and/or after the entered dates)
- event types
- event type values
- operator/username

Select the properties of the filter(s) you want to apply, then select the **Add** button next to the button to add the filter to the list of applied filters. Once

you're done adding filters, select the **Filter** button to search the database for the filtered events.

**Seedlot Maintenance**

- [About Seedlot Maintenance](#)
- [Seedlot Maintenance Tools](#)
- [Seedlot Maintenance Events](#)

[Record Maintenance](#) [Upload Maintenance](#)

**Filter Events** Filter maintenance events based on date, type, and/or value

Add one or more filters to apply to the table of displayed maintenance events. To add a filter, enter the properties for a filter type and click the add button to add the filter to the list. Once you're done adding filters, click the Filter button to display the results.

**Seedlot(s):** Enter the name(s) of the Seedlot(s) - one per line [Add](#)

TEST\_SEEDLOT\_2

**OR**

**Select a List of Seedlots** [Add](#)

**Date:** mm/dd/yyyy [Add](#)

**Type:** Health [Add](#) healthy  
not healthy

**Operator:** dwaring87 [Add](#)

**Applied Filters:**

Property	Comparison	Value	Remove
name	includes	TEST_SEEDLOT_2	<a href="#">X</a>
Water Change	includes	Any Value	<a href="#">X</a>
Health	includes	Any Value	<a href="#">X</a>

**Filter**

[Excel](#) [CSV](#)

Search:

Seedlot	Event ID	Event Date	Event Type	Value	Notes	Operator	Options
TEST_SEEDLOT_2	381804	Fri Jul 9 13:18:05 2021	Water Change	Successful		dwaring87	[Remove]
TEST_SEEDLOT_2	381811	Fri Jul 9 13:18:05 2021	Health	healthy		dwaring87	[Remove]
TEST_SEEDLOT_2	381789	Fri Jul 9 13:15:11 2021	Water Change	Successful		dwaring87	[Remove]

EVENTS: 1 - 3 / 3  
PAGE: 1 / 1

[« Prev](#) [Next »](#)

*A filtered table of Seedlot Maintenance events*

The filtered events can be downloaded directly from the table using the **Excel** or **CSV** buttons at the top of the table. Or Seedlot Maintenance Events can be bulk-downloaded (this includes all events for a Seedlot) using a list of Seedlots from the main downloads page (see below).

#### 7.7.4 Downloading Events

To bulk-download all events for a specific subset of Seedlots:

1. Create a list containing the Seedlots you are interested in.
2. Go to the **Download Using Lists** page (Manage > Download)
3. Find the **Download Seedlot Maintenance Events** section
4. Select your list of Seedlots
5. Select the **Download** button to generate the download file

The downloaded file will follow the same format as the upload template and will contain all recorded Seedlot Maintenance Events for each Seedlot in the list.

### 7.8 Deleting Seedlots

Seedlots can be deleted on the Manage Seedlots page (/breeders/seedlots) by search the seedlot and then clicking the X to delete one seedlot at a time. To delete a seedlot, the logged in user needs the required delete privileges on the seedlot. The seedlot also should not have any transactions associated with it (except for the initial transaction).

To delete seedlots in bulk, generate a list of type seedlot, for example, using the wizard. Open the section “Delete seedlots using a list” on the Manage Seedlots page and select the list. Seedlot deletion using a list is only available to user with curator status.

# Chapter 8

## Managing Populations

Populations are modeled as groups of accessions. This grouping can be useful in downstream analyses. To manage these populations go to Manage Accessions and scroll tp the bottom.

The screenshot shows a web-based application for managing accessions. At the top, there's a header titled "Manage Accessions". Below it, there are three main sections:

- Accessions:** Shows a total of 137103 accessions. It includes a "Search Accessions" input field and links to "Add Accessions Or Upload Accession Info" and "Upload Pedigree File".
- Find Trials in Common:** Allows selecting an accession list from a dropdown menu and clicking "Find Trials". A note says "Use a list of accessions to search for trials that contain them all".
- Populations:** This section is highlighted with a red border. It has a "Create Population" link.

To add a new population click “Create Population”. The following dialog will appear where you choose a list of accessions and give a name to the new population. Please note it is also possible to create a population when you are uploading new accessions into the database.

**Create A Population**

<b>Population Name:</b>	<input type="text"/>
<b>Choose a List of Accessions to Add:</b>	<input type="text" value="119acc"/>
<input type="button" value="Close"/> <input type="button" value="Submit"/>	

Click on the plus (+) button next to Populations to see all the available populations. Click on a population name to see the accessions in the population.

<input checked="" type="radio"/> Populations		[Create Population]
<a href="#">new_test_population</a>	[Go To Population Page]	[Add Accessions To Population] [Delete Population]
<a href="#">NARITA</a>	[Go To Population Page]	[Add Accessions To Population] [Delete Population]

From here you can delete accessions from a population as well as add new accessions to the population.

new_test_population		[Go To Population Page]	[Add Accessions To Population] [Delete Population]
<b>Show 10 ▾ entries</b>			
Accession Name	Description	Synonyms	Remove From Population
037B			X
037D			X
037F			X
038F			X
039B			X
039D			X
039F			X
040B			X
040D			X
041B			X
Showing 1 to 10 of 119 entries			
<a href="#">new_test_population</a>		<input type="button" value="add to new list"/>	
<a href="#">119acc</a>		<input type="button" value="add to list"/>	

# Chapter 9

## Managing Crosses

Information for crosses can be managed using the “Crosses” option in the Manage menu.



### 9.1 Crossing Experiment

Different crosses in the same trial/nursery/project are grouped via “**crossing experiment**”. Crossing experiments are organized based on their breeding programs. To find a crossing experiment, you can either type the crossing experiment name in the “Search” box, or look for the crossing experiment directly in its breeding program by clicking on the “+” icon. In each breeding program, crossing experiments can be placed directly in the breeding program, or organized in folders. The “**Folders**” section allows you to place crossing experiments in folders, move a crossing experiment in a folder to another folder, or rearrange your folders within a breeding program.

**Manage Crosses**

[\[Add Crossing Trial\]](#) [\[Add Cross\]](#) [\[Upload Crosses\]](#) [\[Add Cross Wishlist\]](#)

Information	Breeding Programs – Folders – Crossing Trials	Refresh
<b>Search</b> <input type="text" value="Search"/>  <i>Double click crossing trial (⊖) or folder (📁) to view detail page.</i> <b>Breeding programs (⊕)</b>	Breeding Programs – Folders – Crossing Trials <ul style="list-style-type: none"> <li>+ IITA</li> <li>+ NRCRI</li> <li>+ NaCRRRI</li> </ul>	<a href="#">[Refresh]</a>
<b>Folders</b> <input type="button" value="Create new folder"/>  <input type="button" value="Move crossing trial to folder"/>  <input type="button" value="Move folder"/>		

**Manage Crosses**

[\[Add Crossing Trial\]](#) [\[Add Cross\]](#) [\[Upload Crosses\]](#) [\[Add Cross Wishlist\]](#)

Information	Breeding Programs – Folders – Crossing Trials	Refresh
<b>Search</b> <input type="text" value="Search"/>  <i>Double click crossing trial (⊖) or folder (📁) to view detail page.</i> <b>Breeding programs (⊕)</b>	Breeding Programs – Folders – Crossing Trials <ul style="list-style-type: none"> <li>+ IITA           <ul style="list-style-type: none"> <li>- 2017               <ul style="list-style-type: none"> <li>⊖ IITA_crossingtrial_2017</li> <li>⊖ yam_crossingtrial_2017</li> </ul> </li> </ul> </li> <li>+ NRCRI</li> <li>+ NaCRRRI           <ul style="list-style-type: none"> <li>⊖ crossing_nursery_2018</li> </ul> </li> </ul>	<a href="#">[Refresh]</a>

### 9.1.1 Add New Crossing Experiment

To add a new crossing experiment, click on “Add Crossing Experiment” link.

**Manage Crosses**

[\[Add Crossing Trial\]](#) [\[Add Cross\]](#) [\[Upload Crosses\]](#) [\[Add Cross Wishlist\]](#)

Information	Breeding Programs – Folders – Crossing Trials	Refresh
<b>Search</b> <input type="text" value="Search"/>	Breeding Programs – Folders – Crossing Trials <ul style="list-style-type: none"> <li>+ IITA</li> <li>+ NRCRI</li> <li>+ NaCRRRI</li> </ul>	<a href="#">[Refresh]</a>

Required Information:

- **“Crossing Experiment Name”**: enter a name for the crossing experiment. The crossing experiment name must not already exist in the database.

- “**Breeding program**”: select a breeding program that is available in the database. New breeding programs can be added on the “Breeding program” page, accessible from the “Manage” menu. *Breeding Program Page*
- “**Location**”: select a location for the crossing experiment. New locations can be entered on the “**Locations**” page, accessible from the “**Manage**” menu. *Location Page*
- “**Year**”: select a year.
- “**Description**”: enter a description for the crossing experiment.

After filling in the information, click “**Submit**” to generate the crossing experiment.

Add New Crossing Trial ×

---

<b>Crossing Trial Name:</b>	IITA_crossing_trial_2017
<b>Breeding Program:</b>	IITA
<b>Location:</b>	Ibadan
<b>Year:</b>	2017
<b>Description:</b>	To improve disease resistance

---

Close
Submit

## 9.2 Cross

### 9.2.1 Add New Crosses

Add a cross by using the “Add New Cross” dialog

To add a single new cross, click on “Add Cross” link.

The screenshot shows the 'Manage Crosses' interface. At the top, there are buttons for '[Add Crossing Trial]', '[Add Cross]' (which is highlighted with a red box), '[Upload Crosses]', and '[Add Cross Wishlist]'. Below this is a search bar and a 'Double click' instruction. On the left, there's a 'Breeding Programs - Folders - Crossing Trials' sidebar with a tree structure: IITA (with 2017 and NRCRI), NaCRRI (with crossing\_nursery\_2018). The main area displays a table with columns for 'Information' and 'Breeding Programs - Folders - Crossing Trials'.

Enter cross information in the popup dialog.

The screenshot shows the 'Add New Cross' dialog. It includes a help section for 'Cross type information' and 'Descriptions of cross types'. The 'Required' section contains fields for 'Crossing Trial' (IITA\_crossingtrial\_2017), 'Location' (Ibadan), 'Cross Name' (UG120001xUG120002), 'Cross Type' (biparental), 'Female Parent' (UG120001), and 'Male Parent' (UG120002, highlighted with a blue border). The 'Optional' section includes fields for 'Field Trial' (empty), 'Female Plot' (Choose trial first), and 'Male Plot' (Choose trial first). A 'Search Plots' button is also present.

Required Information:

- “**Crossing experiment**”: select a crossing experiment available in the database.
- “**Location**”: select a location available in the database.

- “**Cross name**”: enter a name for the cross. The cross name must not already exist in the database.
- “**Cross type**”: the options for cross types are: biparental, self, open pollinated, bulk, bulk selfed, bulk and open pollinated, double haploid, polycross, reciprocal and multicross.

**Create New Crosses**

<b>Crossing Trial:</b>	Select Crossing Trial
<b>Location:</b>	Cornell Biotech
<b>Cross Name:</b>	
<b>Cross Type:</b>	Select a cross type biparental self open pollinated bulk bulk selfed bulk and open pollinated doubled haploid polycross reciprocal multicross
<b>Optional:</b>	
<b>Field Trial:</b>	
<b>Female Plot:</b>	
<b>Male Plot:</b>	

- The “**Female Parent**” and “**Male Parent**” field are auto-complete fields for accessions that are already in the database. The parents specified will be entered in the pedigree of the new accessions generated by this cross.

Optional Information:

- “**Female Plot and/or Male Plot**”: In addition to the accession names, specific plots used in the cross can also be added to the database. To retrieve plot names associated with each female/male accession, enter your trial name, then click “**Search Plots**”. Plot names of each parental accession in that field trial will be shown in the drop-down list, you can then select the plot used in the cross.

**Optional:**

<b>Field Trial:</b>	Kasese solgs trial	Search Plots
<b>Female Plot:</b>	KASESE_TP2013_842	
<b>Male Plot:</b>	KASESE_TP2013_1591	

Additional crossing experimental information such as pollination date, number of flowers, number of fruits, number of seeds can be specified during adding new cross. Alternatively, this information can be updated or edited directly on the “**Cross Details**” page.

If you know the number of accessions that are generated from the cross, they can be instantiated immediately in the database by clicking the “**Add accessions for progeny**” checkbox and specifying the number.

Specify Pollination Date:

Specify Number of Flowers:

Specify Number of Fruits:

Specify Number of Seeds:

Add New Accessions for Progeny:

**Submit**

Click “Submit” to generate the cross.

### Upload New Crosses

To upload new crosses from an Excel file (.xls or .xlsx), click on “Upload Crosses” link.



Select a crossing experiment and a location available in the database from drop-down lists and choose a file that you want to upload, then click “**Upload File**”.

**Upload Crosses**

 File Format Information  
Spreadsheet Format

**Crossing Trial:** IITA\_crossingtrial\_2017

**Location:** Ibadan

**Upload File:** Choose File No file chosen

**Close** **Upload File**

Please check spreadsheet format carefully. The file must be an Excel file (.xls or .xlsx).

**Template Information**

Crosses may be uploaded in an Excel file (.xls)  
(.xlsx format not supported)

**Header:**  
The first row (header) must contain the following:

cross_name	cross_type	female_parent	male_parent	Tag Number	Pollination Date	Number of Flowers	Number of Fruits	Fruit Harvest Date	Number of Seeds	Seed Harvest Date	Number of Seeds Sown	Number of Seeds Germinated
------------	------------	---------------	-------------	------------	------------------	-------------------	------------------	--------------------	-----------------	-------------------	----------------------	----------------------------

**Required columns:**

- cross\_name (must not conflict with an existing cross name)
- cross\_type (must be one of the following: biparental, self, open, bulk, bulk\_self, bulk\_open, or doubled\_haploid)
- female\_parent (accession names must exist in the database)
- male\_parent (required in the header, but value may be left blank for most cross types. Must be specified for biparental and bulk crosses.  
When specified, accession names must exist in the database)

**Optional columns (dates must be in the format YYYY/MM/DD):**

- Tag Number
- Pollination Date
- Number of Flowers
- Number of Fruits
- Fruit Harvest Date
- Number of Seeds
- Seed Harvest Date
- Number of Seeds Sown
- Number of Seeds Germinated

**Close**

## 9.2.2 Update Crosses by Uploading

To upload progenies and/or experimental info of crosses already in the database, go to “**Manage-Upload**” page.

In the “Crosses” section, there are links for uploading progenies and experimental info.

Crosses			
Plan	Add	Manage	Search
<a href="#">Create Cross Wishlist</a>	<a href="#">Upload Many New Crosses</a> <a href="#">Add A Cross</a> <b><a href="#">Upload Progenies of Existing Crosses</a></b> <a href="#">Upload Experimental Info of Existing Crosses</a>	<a href="#">Go To Manage Crosses Page</a>	<a href="#">Go To Search Crosses</a>

Please check spreadsheet format in each link carefully. The file must be an Excel file (.xls or .xlsx).

**Template Information** ×

**Progenies of existing crosses may be uploaded in an Excel file (.xls)**  
(.xlsx format not supported)

**Header:**  
The first row (header) must contain the following:

cross_name	progeny_name
------------	--------------

**Required columns:**  
**-cross\_name** (must exist in the database)  
**-progeny\_name** (must not already exist in the database, must have only one progeny for each row, you can add many progenies by adding more rows)

Close

**Template Information** ×

**Experimental Info of existing crosses may be uploaded in an Excel file (.xls)**  
(.xlsx format not supported)

**Header:**  
The first row (header) must contain the following:

cross_name	At least one column of experimental info listed below
------------	---

**Required columns:**  
**-cross\_name** (must exist in the database, must not have duplicate cross name in the upload file)  
**-At least one of the following columns:** (all of the experimental info of a cross must be in a single row)  
 Tag Number  
 Pollination Date  
 Number of Bags  
 Number of Flowers  
 Number of Fruits  
 Number of Seeds

Note: crossing experimental information is customized based on the need for

each crop. As a result, column headers for experimental info in your database may be different from the information shown in this manual.

## 9.3 Cross Wishlist

An Android ODK application is being developed to record cross information on a mobile device in the field. To link this mobile application with the database, the Cross Wishlist can be used to create a plan for which crosses to perform.

This tool is available on the Manage Cross page. It is currently only available on certain databases, so when you click this link you may see an alert mentioning that the cross wishlist is not available on your database.

The screenshot shows a web-based application titled 'Manage Crosses'. At the top right, there is a toolbar with three buttons: '[Create Cross Wishlist]', '[Create New Crosses]', and '[Upload Crosses File]'. The main area displays a hierarchical tree structure under the heading 'BREEDING PROGRAMS -- FOLDERS -- CROSSES'. The tree includes nodes for IITA, NRCRI, NaUHR, ARI Tanzania, QAT, CARI, Rayong, KU, CSIR, SCP, BTI, SRI Kibaha, and CNPA\_PRT. On the left side of the interface, there is a sidebar with sections for 'Information' (containing a 'Search' input field), 'Folders' (with buttons for 'Create new folder', 'Move cross to folder', and 'Move folder'), and 'Breeding programs' (with a link to 'BREEDING PROGRAMS').

### 9.3.1 Create a Cross Wishlist

#### Step 1. Select the accessions to be crossed in your trial

There are two interfaces for this step, either “Not Using Lists” or “Using Lists”. Depending on if you already have a list of female and male accessions to use, you can decide on which interface to use. The end result of using either interface is the same.

The screenshot shows a modal window titled "Create Cross Wishlist". At the top, there are two tabs: "Using Lists" (orange) and "Not Using Lists" (blue, selected). Below the tabs, there are two input fields: "Trial Name(s)" and "Female Accession Name(s)". The "Trial Name(s)" field contains the placeholder "Please select a trial". The "Female Accession Name(s)" field contains the placeholder "First Select A Trial". Below these fields is a dropdown menu with the heading "Available Cross Wishlist(s) for ODK Use". At the bottom right of the modal are three buttons: "Next", "Close", and "Available Cross Wishlist(s) for ODK Use".

We will start by showing “Not Using Lists”. First select the trial in which the crosses are to be performed. This will populate a select box with all the accessions used in that trial. From here, one or many accessions can be selected as the female accession.

The screenshot shows the same modal window as the previous one, but with the "Female Accession Name(s)" dropdown expanded. The dropdown list contains the following items:

- IITA-TMS-IBA000203
- IITA-TMS-IBA000210
- IITA-TMS-IBA000211
- IITA-TMS-IBA000214
- IITA-TMS-IBA000222
- IITA-TMS-IBA000252
- IITA-TMS-IBA30572

Once the female accessions are selected, a table is populated. Each row in this table begins with the female accession that was selected, followed by a select box with all the accessions used in the trial. From here, one or many accessions can be selected as the male to use in the cross.

**Create Cross Wishlist**

Using Lists      Not Using Lists

Trial Name(s):	05uyt20InterB
Female Accession Name(s):	ITTA-TMS-IBA000203 <b>ITA-TMS-IBA000210</b> ITTA-TMS-IBA000211 ITTA-TMS-IBA000214 ITTA-TMS-IBA000222 ITTA-TMS-IBA000252 ITTA-TMS-IBA30572 ITTA-TMS-IBA000224
Female Parent	ITTA-TMS-IBA000210  ITTA-TMS-IBA000214  ITTA-TMS-IBA000222
Select Male Parent	ITTA-TMS-IBA000203 ITTA-TMS-IBA000210 ITTA-TMS-IBA000211 ITTA-TMS-IBA000214 ITTA-TMS-IBA000222 ITTA-TMS-IBA000252 ITTA-TMS-IBA30572 ITTA-TMS-IBA000224
Priority ( 1 : highest, 10 : lowest )	Select Male(s)  Select Male(s)  Select Male(s)

Available Cross Wishlist(s) for ODK Use      Next      Close

Once the male accessions are selected to cross with each female accession, a table indicating priorities appears. Priority is meant to indicate an order in which to attempt the cross; first the highest priority male will be considered, but if this cross is not possible then subsequent males will be considered. An equal priority can be given and this will not indicate a specific order to follow.

**Create Cross Wishlist**

Using Lists      Not Using Lists

Trial Name(s): 05uyt20interIB

Female Accession Name(s):

- ITA-TMS-IBA000203
- ITA-TMS-IBA000210
- ITA-TMS-IBA000211
- ITA-TMS-IBA000214
- ITA-TMS-IBA000222
- ITA-TMS-IBA000252
- ITA-TMS-IBA30572
- ITA-TMS-IBA30574

Female Parent	Select Male Parent	Priority { 1 : highest, 10 : lowest }								
ITA-TMS-IBA000210	<ul style="list-style-type: none"> <li>ITA-TMS-IBA000203</li> <li>ITA-TMS-IBA000210</li> <li>ITA-TMS-IBA000211</li> <li>ITA-TMS-IBA000214</li> <li>ITA-TMS-IBA000222</li> <li>ITA-TMS-IBA000252</li> <li>ITA-TMS-IBA30572</li> <li>ITA-TMS-IBA30574</li> </ul>	<table border="1"> <thead> <tr> <th>Male Parent</th> <th>Priority</th> </tr> </thead> <tbody> <tr> <td>ITA-TMS-IBA000203</td> <td>1</td> </tr> <tr> <td>ITA-TMS-IBA000211</td> <td>3</td> </tr> <tr> <td>ITA-TMS-IBA000252</td> <td>2</td> </tr> </tbody> </table>	Male Parent	Priority	ITA-TMS-IBA000203	1	ITA-TMS-IBA000211	3	ITA-TMS-IBA000252	2
Male Parent	Priority									
ITA-TMS-IBA000203	1									
ITA-TMS-IBA000211	3									
ITA-TMS-IBA000252	2									
ITA-TMS-IBA000214	<ul style="list-style-type: none"> <li>ITA-TMS-IBA000203</li> <li>ITA-TMS-IBA000210</li> <li>ITA-TMS-IBA000211</li> <li>ITA-TMS-IBA000214</li> <li>ITA-TMS-IBA000222</li> <li>ITA-TMS-IBA000252</li> <li>ITA-TMS-IBA30572</li> <li>ITA-TMS-IBA30574</li> </ul>	<table border="1"> <thead> <tr> <th>Male Parent</th> <th>Priority</th> </tr> </thead> <tbody> <tr> <td>ITA-TMS-IBA000203</td> <td>1</td> </tr> <tr> <td>ITA-TMS-IBA000210</td> <td>1</td> </tr> </tbody> </table>	Male Parent	Priority	ITA-TMS-IBA000203	1	ITA-TMS-IBA000210	1		
Male Parent	Priority									
ITA-TMS-IBA000203	1									
ITA-TMS-IBA000210	1									

Available Cross Wishlist(s) for ODK Use      Next      Close

Alternatively, we could have used the “Using List” interface instead. Here we select the trial in which the crosses will be performed and we provide a list of accessions to consider for the females and the males to be crossed.

Create Cross Wishlist

Using Lists      Not Using Lists

Trial Name(s):	05uyt20interB
Female Accession List:	acc_test
Male Accession List:	acc_test

Set Cross Priorities: 1 is highest and 10 is lowest  
 Female Accessions Are in First Column and Male Accessions Are in Header

Female Accessions	IITA-TMS-IBA010746	IITA-TMS-IBA010758	IITA-TMS-IBA010760	IITA-TMS-IBA010779	IITA-TMS-IBA010797	IITA-TMS-IBA010816	IITA-TMS-IBA010819
IITA-TMS-IBA010746							
IITA-TMS-IBA010758							
IITA-TMS-IBA010760							
IITA-TMS-IBA010779							
IITA-TMS-IBA010797							

Available Cross Wishlist(s) for ODK Use      Next      Close

### Step 2. Select the female plots to be considered in the crosses

After selecting your lists, the table below is populated. The first column has all the female accessions specified and the header row has all the male accessions specified. The males to consider crossing with each female are indicated with priority.

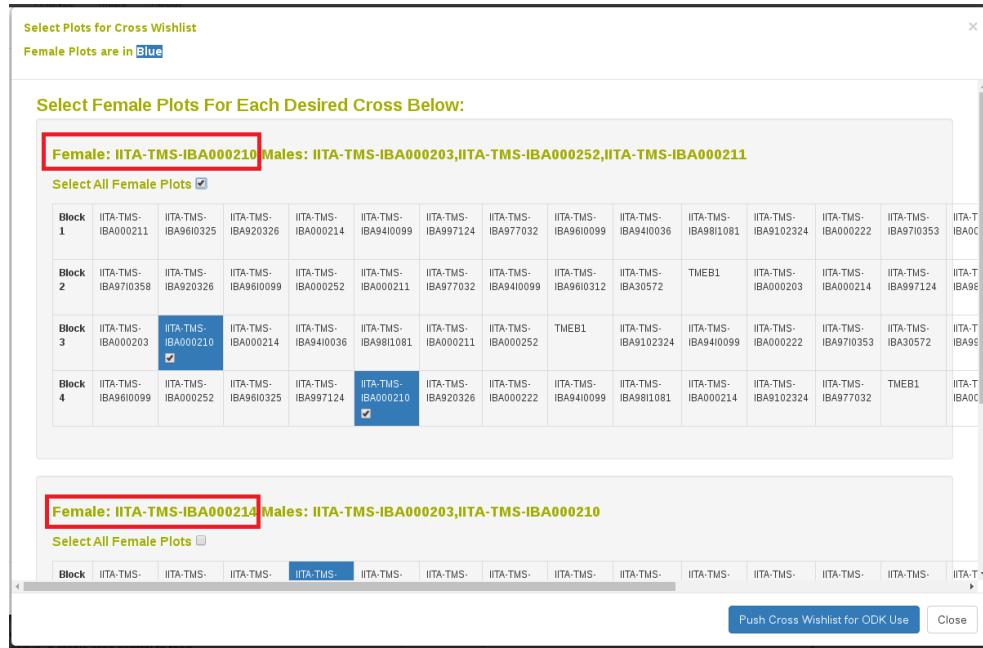
**Create Cross Wishlist**

Female Accessions	ITA-TMS-IBA010746	ITA-TMS-IBA010758	ITA-TMS-IBA010760	ITA-TMS-IBA010779	ITA-TMS-IBA010797	ITA-TMS-IBA010816	ITA-TMS-IBA010819
ITA-TMS-IBA010746							
ITA-TMS-IBA010758		3		1			2
ITA-TMS-IBA010760							
ITA-TMS-IBA010779				1	1	1	
ITA-TMS-IBA010797							
ITA-TMS-IBA010816							
ITA-TMS-IBA010819							

**Set Cross Priorities: 1 is highest and 10 is lowest**  
**Female Accessions Are in First Column and Male Accessions Are in Header**

Available Cross Wishlist(s) for ODK Use    Next    Close

After female and male accessions are selected to cross, either by the “Not Using List” or “Using List” interface, click Next. The next dialog will allow selection of specific female plots to use for the cross. Sections for each female accession selected will appear with the field layout displayed. Selecting all plots in which the female is present indicates that the cross should be performed on all plots where that female accession is present.



### Step 3. Transfer the cross wishlist to your mobile crossing application

Clicking “Push Cross Wishlst for ODK Use” will send the cross wishlist plan to the ONA server for use by the mobile ODK application. Crosses can then be performed and recorded in the field using the mobile application. Afterwards, the crosses are sent back to our database and stored.

## 9.4 Crossing Experiment Detail Page

Information for crosses in the same crossing experiment is compiled in the crossing experiment detail page.

**Details for IITA\_crossingtrial\_2017**

**⌚ Crossing Trial details**

Crossing Trial Name	IITA_crossingtrial_2017
Breeding Program	IITA
Location	Ibadan
Year	2018
Trial Type	crossing_trial
Planting Date	[No Planting Date]
Harvest Date	[No Harvest Date]
Description	To improve disease resistance

**Folder** [New Folder] | [Change]

2017

---

**⌚ Crosses in this trial**

Show **10** entries Search:

Cross Name	Female Parent	Male Parent	Cross Type	Female Plot	Male Plot
UG120030xUG120031	UG120030	UG120031	biparental	KASESE_TP2013_1627	KASESE_TP2013_909
UG120030xUG120032	UG120030	UG120032	biparental		
UG120030xUG120033	UG120030	UG120033	biparental		

Showing 1 to 3 of 3 entries Previous **1** Next

**Crossing Experimental Info**

Show 10 entries	Search:					
Cross Name	Tag Number	Pollination Date	Number of Bags	Number of Flowers	Number of Fruits	Number of Seeds
UG120030xUG120031	1627	2017/02/21	4	30	25	40
UG120030xUG120032	367	2018/02/02	5	48	30	50
UG120030xUG120033		2018/02/02		40		

Showing 1 to 3 of 3 entries

Previous 1 Next

**Progeny Info**

Show 10 entries	Search:
Cross Name	Number of Progenies
UG120030xUG120031	10
UG120030xUG120032	0
UG120030xUG120033	0

Showing 1 to 3 of 3 entries

Previous 1 Next

Each cross name, female parent, male parent, female plot and male plot has a link to its own detail page, which contains information specific to each one. Note: crossing experimental information is customized based on the need for each crop. As a result, the details of the information in your database may be different from the information shown in this manual.

## 9.5 Cross Detail Page

Information of each cross can also be viewed in its detail page.

Detail for cross 'UG120030xUG120031'

Cross information	
<a href="#">[Edit]</a>	
Organism	Solanum lycopersicum
Stock type	cross
Stock name	UG120030xUG120031
Uniquename	UG120030xUG120031
Description	

Parents											
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;">Cross Type</th> <th style="width: 15%;">Female Accession</th> <th style="width: 15%;">Male Accession</th> <th style="width: 15%;">Female Plot</th> <th style="width: 15%;">Male Plot</th> </tr> </thead> <tbody> <tr> <td style="padding: 2px;">biparental</td> <td style="padding: 2px;">UG120030</td> <td style="padding: 2px;">UG120031</td> <td style="padding: 2px;">KASESE_TP2013_1627</td> <td style="padding: 2px;">KASESE_TP2013_909</td> </tr> </tbody> </table>	Cross Type	Female Accession	Male Accession	Female Plot	Male Plot	biparental	UG120030	UG120031	KASESE_TP2013_1627	KASESE_TP2013_909	<a href="#">[Edit]</a>
Cross Type	Female Accession	Male Accession	Female Plot	Male Plot							
biparental	UG120030	UG120031	KASESE_TP2013_1627	KASESE_TP2013_909							

Crossing Experimental Info													
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;">Tag Number</th> <th style="width: 15%;">Pollination Date</th> <th style="width: 15%;">Number of Bags</th> <th style="width: 15%;">Number of Flowers</th> <th style="width: 15%;">Number of Fruits</th> <th style="width: 15%;">Number of Seeds</th> </tr> </thead> <tbody> <tr> <td style="padding: 2px;">1627</td> <td style="padding: 2px;">2017/02/21</td> <td style="padding: 2px;">4</td> <td style="padding: 2px;">30</td> <td style="padding: 2px;">25</td> <td style="padding: 2px;">40</td> </tr> </tbody> </table>	Tag Number	Pollination Date	Number of Bags	Number of Flowers	Number of Fruits	Number of Seeds	1627	2017/02/21	4	30	25	40	<a href="#">[Edit]</a>
Tag Number	Pollination Date	Number of Bags	Number of Flowers	Number of Fruits	Number of Seeds								
1627	2017/02/21	4	30	25	40								

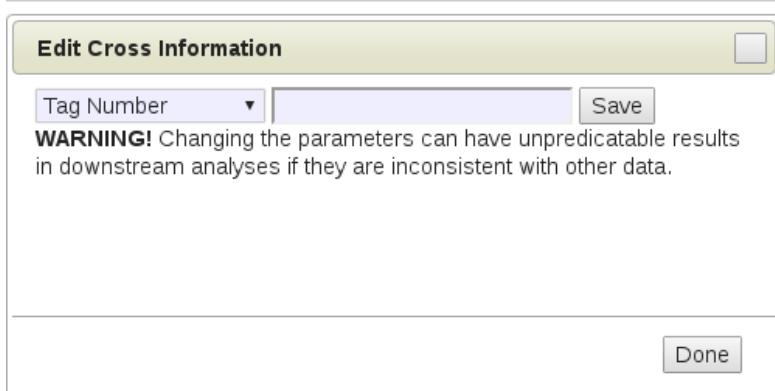


The screenshot shows a table titled "Seedlots of this Cross". The columns are "Seedlot Name", "Breeding Program", "Contents", "Seedlot Location", and "Count". There is one entry: UG120030xUG120031\_2018, IITA, UG120030xUG120031 (cross), Ibadan, 50. Below the table, it says "Showing 1 to 1 of 1 entries". At the top right, there is a link "[Create New Seedlot]".


The screenshot shows a list of progeny items under the heading "Progeny". The items are: UG120030xUG120031P001, UG120030xUG120031P002, UG120030xUG120031P003, UG120030xUG120031P004, UG120030xUG120031P005, UG120030xUG120031P006, UG120030xUG120031P007, UG120030xUG120031P008, UG120030xUG120031P009, UG120030xUG120031P010. Below the list are buttons for "Select All", "Items: 10", "New list...", "add to new list", "accessions\_for\_soilgs\_tests", and "add to list".

This page allows you to update or edit crossing experimental information and add progenies related to that cross. Note: crossing experimental information is customized based on the need for each crop. As a result, the details of the information in your database may be different from the information shown in this manual.



The screenshot shows a dialog box titled "Edit Cross Information". It has a "Tag Number" dropdown menu and a "Save" button. A warning message reads: "WARNING! Changing the parameters can have unpredictable results in downstream analyses if they are inconsistent with other data." At the bottom right is a "Done" button.



# Chapter 10

## Managing Field Trials

To view trial details on the database, click on the “Field Trials” link under the “manage” menu on the toolbar.



Clicking on the “Field Trials” link will bring you to the “Manage Trials” page. On this page, trials are organized according to their breeding programs. To access trial details, click on the + icon next to your breeding program.

A screenshot of the 'Manage Trials' page. The top navigation bar includes 'Search', 'Manage', 'Analyze', 'Maps', 'About', and a user profile for 'Gregor\_Mendel'. Below the header, there is a section titled 'Manage Trials' with buttons for '[Upload Trial]', '[Add Trial]', and '[Add Multi-location Trial]'. The main content area is divided into two columns: 'Information' on the left and 'Breeding Programs -- Folders -- Trials' on the right. The 'Information' column contains a search bar and a button for 'Download Trial Phenotypes'. The 'Breeding Programs' column shows a tree-like hierarchy of breeding programs: IITA, NRCRI, NaCRI, ARI Tanzania, NaCRI Germplasm Collection, CIAT, CARI, and Rayong. Each program has a plus sign icon next to its name, indicating it can be expanded to view trials.

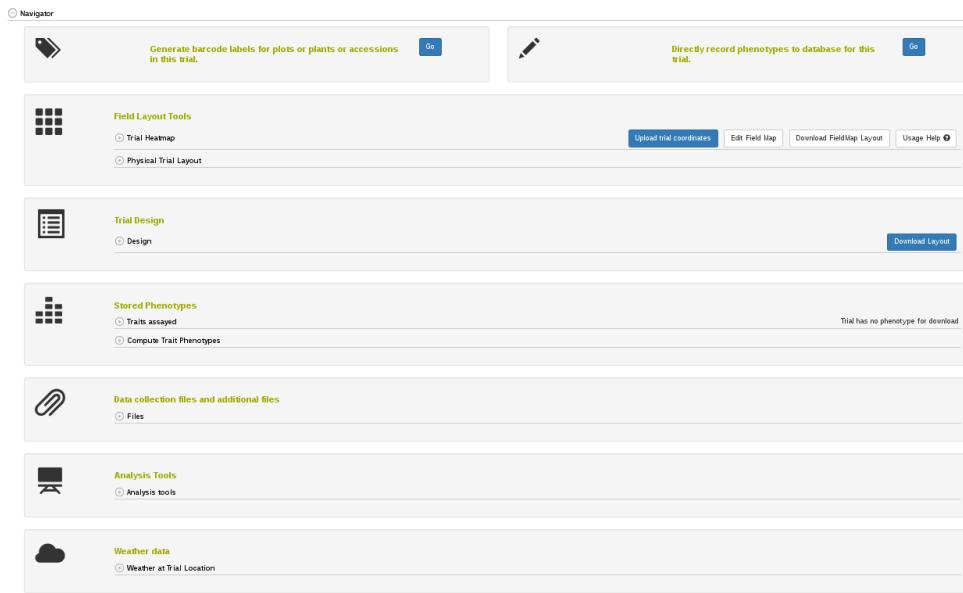
Trials can be placed directly in their breeding program. Alternatively, they can be organized by using folders within each breeding program. Clicking on trial name will take you directly to the trial details page.



## 10.1 Trial Detail Page

Trial detail page displays important information about individual trial including breeding program, location, year, description of the trial, design, and any files associated with that trial.

The “Navigator” section on the trial detail page allows easy access to all aspects of your trial. This section contains subsections for printing labels for your plots or plants, recording phenotypes, viewing your trial layout or design, viewing phenotypes for this trial, or conducting analyses.



The “transplanting date” field feature will only be shown if it has a value. To add a transplanting date after creating a trial, change the show\_transplanting\_date parameter from 0 to 1 in the SGN config file. As a result, you will be able to add a date under the transplanting date field by clicking the “Edit Trial Details” on the trial detail page.

The screenshot shows a software interface for managing field trials. The main title is 'Edit Trial Details'. The form includes the following fields:

- Trial Name:** 199934HBEPR\_cara
- Breeding Program:** demo
- Location:** test\_location
- Year:** 1999
- Trial Type:** Preliminary Yield Trial
- Planting Date:** Clear, 06/04/1999
- Transplanting Date:** Clear, Jul 2024 (calendar shows 11th selected)
- Harvest Date:** Clear
- Description:** EPR
- Field Size (ha):** 8
- Plot Width (m):** 5
- Plot Length (m):** 5
- Trial Will Be Genotyped:** No
- Trial Will Be Crossed:** No

A note at the bottom left indicates 'Indicates pending change'. At the bottom right are 'Cancel' and 'Save Changes' buttons.

## 10.2 Adding Trials

Only users with the account status of “submitter” may create trials. To learn how to change your account status from “user” to “submitter” visit the [1.2](#) page.

### 10.2.1 Prerequisites

- To add a trial, all of your accessions should already exist in the database before you begin to design a trial. If you have accessions that are not in the database, see the instructions for adding accessions .
- Breeding program and location for your trial should also exist in the database. If you need to add breeding program and/or location to the database, see instructions for adding breeding program and location in the “Managing Breeding Programs” and “Managing locations” respectively.

On the “Manage Trials” page, there are three alternative methods for you to add new trials: by using “Add Trial” form, “Upload Trial” form, or “Add Multi-location Trial” form.

The screenshot shows the 'Manage Trials' interface. On the left, there's a sidebar with 'Information' and a search bar. On the right, a tree view shows a folder named 'test' containing three sub-folders: '2018TrialUpload01', '2018t1', and '2018t2'. At the top right, there are three buttons: 'Upload Existing Trial' (disabled), 'Design New Trial' (highlighted in red), and 'Design New Multi-location Trial'.

### 10.2.2 Adding a trial by using “Add Trial” form

#### Step 1. Begin the “Design new trial” workflow

Click on “Design New Trial” to begin.

This screenshot is identical to the one above, showing the 'Manage Trials' page with the 'Design New Trial' button highlighted in red.

The first step in this workflow is an introduction that looks like:

The screenshot shows the 'Design New Trial' workflow. At the top, a progress bar indicates six steps: 'Intro' (selected), 'Trial Information', 'Design Information', 'Field Map Information', 'Custom Plot Naming', and 'Review Designed Trial'. Below the progress bar, a text box contains the following information:

**This workflow will guide you through designing a new trial in the database**

A field trial represents plots in the field where each plot has a globally unique `plot_name`, a sequential `plot_number` that is unique in the trial (e.g. 101, 102, 103 for three separate plots), and an accession representing the genotype being tested in that plot. Each plot can belong to different blocks and reps depending on the experimental design you are using (e.g. complete block design vs augmented). Each plot can have a `row_number` and `col_number` indicating the relative position of the plot in the field.

To design a trial you need to provide a globally unique trial name. The `plot_names` will be generated based on the trial name you provide (e.g. if the trial name is 2018MyTrial, `plot_names` will be generated like 2018MyTrial\_101, 2018MyTrial\_102, etc.).

You also need to provide a list of accessions to use. Based on the design you have picked, the accessions will be randomized over the blocks or replicates in the trial.

Depending on the design you have picked, you will need to provide different design parameters (e.g. for complete block you will need to provide number of blocks, while for alpha lattice you will need to provide block size and number of replicates).

At the bottom of the text box is a blue 'Go to Next Step' button, and at the bottom right is a 'Close' button.

Here it gives information about what is required for a trial, including that to create a new trial, you need to create a list of the accessions that you would like to use in the trial. Lists can be viewed, created, and modified with the

“lists” tool at the upper right of the screen. For more information on lists, click [here](#).

### Step 2. Enter “Trial Information”

On this screen you need to enter basic information about the trial, such as breeding program and location(s). You must also select a design type, such as Complete Block Design. The design is important because it influences how your genotypes are distributed and randomized over the trial. You must first click validate before proceeding to the next step.

The screenshot shows the 'Design New Trial' application. The current step is 'Trial Information' (step 2). The interface includes a top navigation bar with numbered steps 1 through 6. Step 2 is highlighted with a green circle and a green underline. Below the navigation is a form titled 'Enter basic information about the trial'. The form fields are as follows:

- Trial Name:** 2018MyTrial1
- Breeding Program:** test
- Location:** Cornell Biotech
- Trial Type:** phenotyping\_trial
- Year:** 2018
- Description:** Phenotyping trial in 2018
- Design Type:** Complete Block

A note below the 'Design Type' dropdown reads: "generates Randomized Complete Block Design, using the methods of random number generation in R. Creates plot entities in the database." At the bottom of the form are two buttons: 'First validate the form' and 'Continue to Next Step'. A 'Close' button is located in the bottom right corner of the window.

### Step 3. Enter “Design Information”

On this screen you need to specify a list of accessions to use in the experiment. This list must be a valid list of accessions. You must also specify all required design information, such as number of blocks in this case.

The screenshot shows a software window titled "Design New Trial". At the top, there is a horizontal navigation bar with six steps: "Intro" (step 1), "Trial Information" (step 2), "Design Information" (step 3, highlighted in green), "Field Map Information" (step 4), "Custom Plot Naming" (step 5), and "Review Designed Trial" (step 6). Below the navigation bar, the main area is titled "Design your trial layout" and contains the following fields:

- List of accessions to include (required):** A dropdown menu showing "test\_stocks".
- List of checks to include. Checks list should be separate from accessions list. (optional):** An empty dropdown menu.
- Number of blocks (required):** An empty input field.

At the bottom of the form is a blue "Continue to Next Step" button and a "Close" button in the bottom right corner.

#### Step 4. Enter “Field Map Information” (Optional)

On this screen you can specify how the row and column numbers will be generated for the plots in the trial. The row and column number represent a relative position of the plot in the field. If you are not exactly sure of how you will plant the plots in the field or you have an irregular (non-rectangular) layout, you can skip this step for now. This information can be added on the Trial Detail Page once the trial is saved in the database in order to reflect exactly how the plots were planted in the field.

Specify the number of rows and columns for the entire field

If you do not know exactly in which rows and columns you will end up planting the plots, do not provide this and go to the next step.

If you will plant your plots in an irregular (non-rectangular) layout, do not provide this and go to the next step.

You can upload the exact row and column information for your plots (in any layout shape) on the Trial Detail Page after you have created the trial in the database and actually planted the experiment.

**Field map display:**

**Number of rows (optional):**

**Plot layout format:**

**Continue to Next Step**

### Step 5. Custom Plot Naming (Optional)

On this screen it is possible to change the format in which plot names will be generated for your trial. It is recommended to skip this step and just use the format generated by the database by default.

If you want to change the way in which plot names will be generated by the database

It is recommended to skip this step and move on to the Next Step

**Custom plot naming/numbering:**

**Plot prefix:**

**Plot start number:**

**Plot number increment:**

**Continue to Next Step**

### Step 6. Review Designed Trial

On this screen you can review the trial that the database has generated. You will see a graphical representation of the trial. The numbers on the

squares represent the plot\_number of each plot and on mouse hover you can see further information about the plot.



You will also see a table representation of all the plots and their information. If you want to redo the randomization, you can click the “Redo Randomization” button.

**Design New Trial**

Plot Name	Accession Name	Check Name	Plot Number	Row number	Col number	Block Number	Block Row Number	Block Col Number	Rep Number	Seedlot Name	Num Seeds Per Plot
2018MyTrial1_rep1_test_accession1_1001	test_accession1		1001	1	1	1			1		
2018MyTrial1_rep1_test_accession5_1002	test_accession5		1002	1	2	1			1		
2018MyTrial1_rep1_test_accession4_1003	test_accession4		1003	1	3	1			1		
2018MyTrial1_rep1_test_accession3_1004	test_accession3		1004	1	4	1			1		
2018MyTrial1_rep1_test_accession2_1005	test_accession2		1005	1	5	1			1		
2018MyTrial1_rep2_test_accession4_2001	test_accession4		2001	2	5	2			2		
2018MyTrial1_rep2_test_accession2_2002	test_accession2		2002	2	4	2			2		
2018MyTrial1_rep2_test_accession5_2003	test_accession5		2003	2	3	2			2		
2018MyTrial1_rep2_test_accession1_2004	test_accession1		2004	2	2	2			2		
2018MyTrial1_rep2_test_accession3_2005	test_accession3		2005	2	1	2			2		
2018MyTrial1_rep3_test_accession3_3001	test_accession3		3001	3	1	3			3		
2018MyTrial1_rep3_test_accession1_3002	test_accession1		3002	3	2	3			3		
2018MyTrial1_rep3_test_accession2_3003	test_accession2		3003	3	3	3			3		
2018MyTrial1_rep3_test_accession5_3004	test_accession5		3004	3	4	3			3		
2018MyTrial1_rep3_test_accession4_3005	test_accession4		3005	3	5	3			3		

**Redo Randomization**

**Close**

At the bottom there is a brief summary of the trial followed by two buttons.

**Design New Trial**

Trial Is Valid  
The following trial will be added

**Design type**  
Randomized Complete Block Design

**Number of locations**  
1

**Number of accessions**  
5

**Number of blocks**  
3

**Number of accessions per block**  
Block 1: 5 accessions  
Block 2: 5 accessions  
Block 3: 5 accessions

**Number of reps**  
3

**Treatments:**

**Add Field Management Factor(s) to Design** **Confirm (Saves Trial In Database)**

**Close**

### Step 7. Add Field Management Factors to your design (Optional)

You can add Field Management Factors by clicking “Add Field Management Factor(s) to Design”. Clicking this opens a dialog to name your factor. You can name this to account for fertilizer or watering regime or inoculation or anything else. This is optional and can be added from the trial detail page afterwards.

Add Field Management Factor Name: Fertilizer\_Nov\_9\_2017\_20%

Applied To: Plots

Continue Close

Click “Continue” and a dialog will appear where you can specify plots for which the factor was applied. There is a select all button also.

plot_name	accession	plot_number	block_number	rep_number	is_a_control	row_number	col_number	Fertilizer_Nov_9_2017_20% [Select all]
2017_tutorial_trial_1001	002D	1001	1	1	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1002	0000test2	1002	1	1	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1003	003D	1003	1	1	undefined	undefined	undefined	<input checked="" type="checkbox"/>
2017_tutorial_trial_1004	004D	1004	1	1	undefined	undefined	undefined	<input checked="" type="checkbox"/>
2017_tutorial_trial_1005	002B	1005	1	1	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1006	002B	1006	1	2	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1007	002D	1007	1	2	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1008	001B	1008	1	1	1	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1009	0000test1	1009	1	1	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1010	002B	1010	1	3	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1011	003D	1011	1	2	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1012	004D	1012	1	2	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1013	0000test1	1013	1	2	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1014	001F	1014	1	1	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1015	001F	1015	1	2	undefined	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1016	001D	1016	1	1	1	undefined	undefined	<input type="checkbox"/>
2017_tutorial_trial_1017	001D	1017	1	2	1	undefined	undefined	<input type="checkbox"/>

Apply Field Management Factors to Plants (if available):  Continue Close

### Step 8. Saving new trial in the database

Once you are done reviewing the trial you can click “Confirm” to save the generated trial into the database. Once the trial has saved you will see the

final completion screen:



### 10.2.3 Adding a trial from an uploaded file

If you already have trial design layout in a spreadsheet, you can add your trial into the database by using “Upload Trial” form. To access “Upload Trial” form, click on “Upload Existing Trial(s)” button on the “Manage Trials” page.

When you click “Upload Existing Trial(s)” you will see the following workflow. Notice that there are 5 numbered sections to the workflow.

**Step 1:**

The first step is to understand what the format of the trial upload is. It is important to understand that the field layout represents plots in the experiment. Each plot has a globally unique plot\_name, a sequential plot\_number that is unique in the trial (but not globally unique. e.g. 101, 102, 103 for three separate plots), an accession\_name representing what genotype is planted in that plot, and a block\_number representing design replication. Each plot can be thought of as having a row\_number and a column\_number representing the relative position of the plot in a grid (e.g. the top left plot is row 1 column 1 following by row 1 column 2). Each plot can be planted with an amount of seed from a seedlot, where the seedlot\_name represents the specific seed packet that was used, and num\_seed\_per\_plot and weight\_gram\_seed\_per\_plot represent amount that were transferred from the seedlot\_name to the plot\_name. Treatments (management factors) can be applied onto plots using additional column names in your file, where a 1 represents if the factor was applied to the plot and an empty cell means it was not applied.



This information and more can be found by clicking “Information about file format”, which shows the following:



### Minimum File requirements

- All accession names in the file must exist in the database. See adding accessions for more information.
- The uploaded file should be XLS or XLSX file format (NOT CSV).
- The first row (header) must contain the column names: plot\_name accession\_name plot\_number block\_number is\_a\_control rep\_number range\_number row\_number col\_number seedlot\_name num\_seed\_per\_plot weight\_gram\_seed\_per\_plot

Minimal Example:

plot_acces	plot	is_a_c	rep_numb	range_numb	row_numb	col_numb	seedlot_n	num_se	weight_g	plot_se
2018plot1	ac10sion11	1								
2018plot2	ac20sion22									
2018plot3	ac10sion12									
2018plot4	ac20sion21	1								

### File validation

- In case of errors in the uploaded file such as missing or invalid data, a window will appear listing the specific errors in the file that must be corrected before a successful upload.

### Uploading a trial with Field Management Factors

- You can upload a trial with field management factor(s) by adding additional column(s). The column header will be the factor e.g. fertilizer, watering regime, inoculation, etc. and the values in these columns will be either 1 or empty, indicating that the factor was applied to the plot or not.

#### Step 2:

Once you feel that your experiment field layout is in the right format, click on to the Next Step. You will see the following form which must be filled in completely:

Upload Existing Trial

File Formatting      Enter trial information      Fix missing accessions problem      Fix missing seedlots problem      Try submitting trial again

Enter information about the experiment and upload your trial layout

File format information  
Spreadsheet format

Trial Name: 2018TrialUpload01

Breeding Program: test

Location: Cornell Biotech

Trial Type: phenotyping\_trial

Year: 2018

Description: Testing of upload

Design Type: Complete Block

Upload File: Choose File | wk17trialupload

First validate the form      Upload Trial

Close

The trial name must be globally unique in the database. Please try to follow standard naming conventions for your group.

First you need to validate the form, and then you can click “Upload Trial”.

### Step 3:

In the case where you have uploaded an experiment using accession\_names that are not already present in the database, you will be taken to this screen. If the accession\_names in your file are all already in the database, this step will be skipped. The reason it is necessary for your accessions to be in the database before you can add a trial using them is that a single accession can be used among many trials and therefore must exist as a separate entity in the database; because of this it is also very important to be careful about adding wrongly duplicated accession\_names into the database. From this screen it is possible to make a new list with the missing accession\_names and then click “Add Accessions to the database” to immediately resolve the issue. Once all your accessions are in the database, click to move to the Next Step.



### Step 4:

In the case where you have uploaded an experiment using seedlot\_names that are not already present in the database, you will be taken to this screen. If the seedlots in your file are all already in the database, this step will be skipped. The reason it is necessary for your seedlots to be in the database before you can add a trial using them is that a given seedlot can be used among many trials and therefore must exist as a separate entity in the database. From this screen it is possible to add the missing seedlots; you can either upload an XLS or XLSX file to add many at once or you can add them one by one. Once all your seedlots are in the database, click to move to the Next Step.



### Step 5:

If there are any other errors with your file, such as if the plot\_names are not globally unique in the database or your plot\_numbers are not unique in your trial or row\_number is not an integer or any other error, you will see the errors listed in the red box. It is up to you to correct these errors in your file. Simply open up the file you selected earlier in Excel and correct the issues and then save the file. Then you can click "Submit Trial" and it will resubmit it for you. You can continue to edit your file here and submit

as many times as you need until it is accepted.



### Completion screen

Whether you were lucky enough to submit your trial successfully on Step 2 or if you tried many times on Step 5, once your trial has been saved in the database you will see the following screen:



#### 10.2.4 Multi-location trials

To add multi-location trials, simply select the multiple locations while using the ‘Add Trial’ form.

This will create a separate trial for each selected location, but they will share the same design and will be grouped in a single folder.

By default each trial design will have a fresh randomization, but if desired you may check the “Use same randomization for all locations” option.

### 10.2.5 Email alert for multiple trial design upload

When uploading multiple trials from a file, you have the option to receive email notifications by clicking the “Email Alert” checkbox. By default, the system will use the email address associated with your account, but you have the option to enter a different email address if you prefer. After submitting, the upload process runs in the background, allowing you to continue using the interface without interruptions. Once the process completes, you will receive an email with the upload results.



### 10.2.6 Viewing Plot Layout and Trait HeatMap

#### 10.2.6.1 Viewing plot layout

In the “Field Layout Tools and Phenotype Heatmap” section of a Trial Detail page, the trial physical layout is displayed by default. The relative position of the plots will be displayed based on the row and column positions given to the plots during the trial creation or upload steps. The plots are color-coded based on the plot’s rep and block numbers and whether or not it is used as a check. Hover the mouse over the plot to see details about a specific plot.

Select:  Display Trials in Same Field:

Border Plots and Filler Plots  Even Block Numbers (e.g. 2,4,...)  Odd Block Numbers (e.g. 1,3,...)  Checks  
 Odd Rep Numbers (e.g. 1,3,...)  Even Rep Numbers (e.g. 2,4,...)  Overlapping Plots (2+ plots at some position)  
 Plot Has Image

Download Plot Order    Include Borders

Plot Layout   Invert Rows

Top Border  Left Border  Right Border  Bottom Border

3	301	302	303	304	305	306	307	308	309	310
2	201	202	203	204	205	206	207	208	209	210
1	101	102	103	104	105	106	107	108	109	110
	1	2	3	4	5	6	7	8	9	10

If there is more than one trial grown in the same physical field, the trial layouts of all of the trials can be shown together if the trials share these properties:

Each trial has the same year

Each trial has the same location

The location type of the trials' location is set to Field

The row and column positions of all of the plots (across the related trials) don't overlap. For example, trial #1 starts at row 1 and trial #2 starts at row 10.

When these conditions are met and you check the "Select Trials in Same Field" checkbox, the plots from all of the related trials will be displayed on the same field layout. The plots will be color-coded by trial. The planting order and harvest order downloads will include the plots from all of the displayed trials in the order in which the plots occur in the field.



### 10.2.6.2 Viewing plot layout for multiple trials

#### Tracking plot images on fieldMap

Plot images can be seen on fieldMap if a plot is associated to any image.



To view plot image(s), click on a plot, a dialog will appear.



On the appeared dialog, click on View plot images. To see more images if a plot has more than 2 images, click on See more images... Medium size of an image can be viewed by clicking on an image.



### Viewing assayed trait heatmap

Phenotype heatmap can be viewed by selecting a specific assayed trait from the selectbox drop-down. Mousing over the plots, highlights the plot in green and also displays the plot's field information including the selected trait's phenotype value.



### Suppressing Plot Phenotype

Clicking on a plot on the heatmap would display a dialog that has a button for suppressing a plot phenotype value for a given trait. A suppressed plot value can be excluded during trial analysis and phenotype download.



### 10.2.7 Adding additional information in the “Trial Detail” page

After you added a new trial to the database, you can edit trial details or add more information for that trial through the “Trial Detail” page.

#### Uploading Physical Trial Layout

You can upload physical trial layout by clicking on the “Upload trial coordinates” button on the “Trial Detail” page.



Please check file format carefully. You can find file format information by clicking on the “Spreadsheet format” on the “Upload trial coordinates” window.



Spreadsheet format:



### Physical Trial Layout File requirements

- All plot names in the file must exist in the database.
- The uploaded file should be tab delimited (txt).
- The first row (header) must contain the column names

Example:

plot_name	row_number	col_number
plot1	1	1
plot2	1	2
plot3	1	3

Select the trial layout coordinates file that you want to upload for this trial, then click “OK” button to upload the file.



The following message is displayed after the coordinates are uploaded.



The field layout can be viewed by clicking on the “Trial Heatmap Section” to see a drop-down of the field map.



### Downloading Field Map Spreadsheet

Field map spreadsheet can be downloaded if the trial has field coordinate (row and column numbers) uploaded for it plots. To download, click on the Download FieldMap Layout link on the Trial Heatmap section.



A dialog will appear, click on the submit button to download.



Click to view downloaded spreadsheet.

N16													
	A	B	C	D	E	F	G	H	I	J	K	L	
1	Columns	1	2	3	4	5	6	7	8	9	10	11	
2	Rows	1	TDri0000066	TDri0000079	TDri0000339	TDri9602024	TDri9618948	TDri0000332	TDri0000308	TDri0000358	TDri0000344	TDri000021	TDri99.8
3		2	TDri9618948	TDri0000361	TDri0000308	TDri0000339	TDri000021	TDri000079	TDri9602024	TDri0000332	TDri99.8	TDri000189	TDri000021
4													
5													
6													
7													

## Editing Physical Trial Layout

"Usage Help" link contains information on how to edit physical trial layout.



#### Background:

Field map is a tool that enable users to view the physical layout of plots in a trial. Maps can be generated on the fly while adding or uploading a trial, if that option is enabled or rows and column numbers provided in the trial upload files respectively. Field map coordinates can also be uploaded independently after trials have been added or uploaded. It's a very intuitive, flexible and user friendly tool for manipulation/making changes to field trial layouts before phenotypes are uploaded.

#### Editing Options:

##### Replace Plot Accession

A plot accession can be replaced by an accession within or outside of the trial. To do this, **click on the plot and provide the name of the new accession** (must already exist in the database).

##### Replace Trial Accession

An accession used in a trial can be replaced by a new accession or another accession from the trial. When this replace option is used, it replaces every instances (plots and plants) of that accession in the trial. To do this, **click on the Edit Field Map link** by the top right of the physical trial layout section; **click on Replace Accession button**; **select accession** to replace from the trial and **provide a new accession** (must already exist in the database); **click on Replace Trial Accession button** to complete the operation.

##### Substitute Plot Accessions

This feature allows you to switch plot accessions between any two plots. To switch the accessions of two plots, **click on the Edit Field Map link**; **click on Substitute Accession button**; **select the plots to switch their accessions**; **click on Substitute Plot Accession** to switch the accession in those plots.

#### Features:

##### Mouse Over

Displays plot field information.

##### Double Click

Double clicking on a plot, opens the stock page for that plot.

##### Download Map

Field Map can be downloaded as image using the download button below the map.

##### Delete Map

Field Map can be deleted if the user have the right privilege.

#### Note:

- You have to be a **curator** or a **submitter and associated to the breeding program** of the trial to use the features of this tool.
- **Input boxes** used within the field map tool will automatically (**autocomplete**) give accession name options from the database when you start typing the accession name.
- **Changes can only be made to the physical layout when phenotypes are not yet upload for the trial.**

[Close](#)

There are three different options for editing trial layout:

- Replacing plot accession by clicking on the plot in the layout.
- Replacing trial accession by using “Edit Field Map” link.

- Substituting plot accessions by using “Edit Field Map” link.

When you move a cursor over a plot on the trial layout, information for that plot appears.

Field Layout Columns						
Field Layout Rows	1	2	3	4	5	6
12	11	10	9	8	7	
13	rep_number: 2 block_number: 2 row_number: 3 col_number: 5 accession_name: NR070391					18
24	23	22	21	20	19	
25	26	27	28	29	30	
36	35	34	33	32	31	

To edit a specific plot, clicking on that plot. Entering new accession on the “Replace Plot Accession” form, then clicking on “Replace Plot Accession” button.



To replace an accession (in every plot/plant of that accession), clicking on “Edit Field Map” button.



On the “Edit Field Map” window, clicking on “Replace Accession” button.



Selecting any accession that you want to replace and entering your new accession, then clicking “Replace Trial Accession” button.



You can switch plot accessions between any two plots by clicking on “Substitute Accession” button.



On the “Substitute Plot Accession” form, selecting the two plots that you want to switch, then clicking on the “Substitute Plot Accession” button.



### 10.2.8 Downloading the Trial Layout from the “Trial Detail” page

Click on “Download Layout” on the Trial Detail page.

The screenshot shows the 'Physical Trial Layout' section of a software application. On the left, there's a sidebar with sections like 'Physical Trial Layout' (selected), 'Design', 'Attributes', 'Accessions', 'Seedlots', 'Controls', 'Plots', 'Plant Entities', and 'Treatments'. The main area displays trial design details: 'Value' (e.g., 'RCBD', '4', '4'), 'Upload Coordinates' (link), 'Edit Field Map' (link), 'Usage Help' (link), and a prominent red-bordered 'Download Layout' button. To the right, there's a 'Completion' sidebar with sections for 'Trial Info' (Trial Name, Breeding Program, Location, Trial ID, Trial Type, Planning Date, Harvest Date, Description, In Folder) and 'Field Design' (Phenotyping).

The trial layout includes all information regarding the observation units in the experiment. The observation units can be plots, plants, or subplots. The trial layout can include trial design information such as the block\_number and rep\_number. It can also include physical map information such as the row\_number and col\_number, if that information is available for the trial. The trial layout also includes information regarding treatments that have been applied in the field. Optionally, the layout can give information regarding accession's global performance for a list of traits.

This screenshot shows a 'Download Trial Layout' dialog box. At the top, it says 'Download Trial Layout for 05uyt20interUB'. It has fields for 'Trial' (05uyt20interUB), 'Format' (CSV), 'Treatment' (Include All Treatments In Download), and 'Data Level' (Plots). Below these are two sections: 'Included Columns:' and 'Not Included Columns:'. The 'Included Columns:' list includes plot\_name, block\_number, plot\_number, rep\_number, row\_number, col\_number, accession\_name, and is\_a\_control. The 'Not Included Columns:' list includes synonyms, trial\_name, location\_name, year, pedigree, tier, seedlot\_name, seed\_transaction\_operator, and num\_seed\_per\_plot. At the bottom, there's a note about average performance of accessions and a dropdown menu labeled 'select'. The dialog has 'Close' and 'Submit' buttons at the bottom right.

### 10.2.9 Adding Plant Entries To Your Trial

After you added a new trial to the database you can choose to add plant entries to your trial. Adding plant entries enables plant level phenotyping. It is generally better to enter data at the plant level into the database because it is always possible to calculate plot level phenotypes from the individual plant data.

Plant entries can be added to your trial in two ways: 1) Automatically generated by the database. The only input required is the number of plants per plot. 2) Uploaded in an XLS or XLSX file. This allows you to specifically name your plant entries.

These two options are available in the “Plant Entries” section on the Trial Detail Page, as shown in the screen shot below.

The screenshot shows a web-based application interface for managing trial designs. At the top left is a 'Design' button with a gear icon. At the top right is a 'Download Layout' button. Below these are sections for 'Attribute' and 'Value'. The 'Attribute' section includes fields for 'Number of Blocks' (RCBD), 'Number of Replicates' (2), 'Plot Length', 'Plot Width', and 'Plants Per Plot'. Below this is a sidebar with radio buttons for 'Accessions', 'Seedlots', 'Controls', 'Plots', and 'Plant Entries'. The 'Plant Entries' section is highlighted with a red box. Inside this box are two buttons: 'Add plant entries' and 'Upload plant entries'. To the right of the 'Upload plant entries' button is a link '[Upload Seedlots Planted In Trial]'. At the bottom of the sidebar is another radio button for 'Field Management Factors' with a link '[Add Management Factor]' to its right.

#### Automatically Generate Plant Entries

Clicking on “Add plant entries” opens the following dialog box. The only input required is the number of plants per plot. This will create plant entries that are named as a concatenation of the plot\_name and the plant’s index number e.g. plot\_name\_plant\_1



## Upload Plant Entries

Alternatively, you can choose to upload an XLS or XLSX file that contains the names of the plant entries. Clicking on “Upload plant entries” opens the following dialog box.



Clicking on “Spreadsheet format” will give you information about the XLS or XLSX file to upload. Clicking this will open the following dialog box.



This shows you that the files requires the header to contain “plot\_name” and “plant\_name”. The plot\_name must exist in the database already and the plant\_name must be unique in the database.

Along with the file, you must specify “number of plants per plot”. This is intended to be the total number of plants that were plants. If the file you upload shows three plants in one plot and four plants in another plot, that is fine.

### 10.2.10 Adding Tissue Sample Entries To Your Trial

Some trials require tissue samples to be collected from plants in a field trial. The database will generate these tissue sample identifiers for you and will maintain all relationships with the plant, plot, accession, etc. To begin, go to the Design section of a trial’s detail page and open the “tissue sample entries” section. Please note that tissue samples are directly related to plants, therefore your trial requires plants before you can add tissue samples.

Design [Download Layout]

Attribute	Value
Design	CRD
Number of Blocks	2
Number of Replicates	2
Plot Length	
Plot Width	
Plants Per Plot	

Accessions  
 Seedlots  
 Controls  
 Plots  
 Plant Entries  
 Tissue Sample Entries [Upload Seedlots Planted In Trial]  
 Field Management Factors [Upload GPS Coordinates] [Add Management Factor]

Add tissue sample entries

When you click on “Add tissue sample entries” you will see a dialog where you specify the number of tissue samples you require per plant. Once you have specified how many tissues samples, you can give specific words to distinguish samples, such as “root” or “stem”, as seen below.

Add tissue samples to plotbox

**WARNING:** This trial does not have plant entries. Tissue samples are added for each plant entry, so you must add plant entries first. You can do so on the "Plant Entries" section of the trial detail page.

Number of tissue samples per plant:

**Tissue Name 1:** leaf

**Tissue Name 2:** root

**Tissue Name 3:** examples: leaf or root or stem

Inherits Management Factor(s) From Plots:

Once you have added tissue sample entries they will appear in the design section of the trial as seen below.

Each tissue sample has a detail page where you can add information about the sample, such as if it is in transit or in storage somewhere.

Organism	Manihot esculenta
Stock type	tissue_sample
Stock name	spt9903_1003_subplot_2_plant_4_leaf1
Unique name	spt9903_1003_subplot_2_plant_4_leaf1
Description	

SGN stock 787404 (spt9903\_1003\_subplot\_2\_plant\_4\_leaf1)

The related stocks section near the bottom of this detail page displays the relationships between all stocks, including tissue samples.

Related stocks

- Related stocks in trials
- Seedlots of this Accession
- Progenies
- Groups /members
- Related stocks for tissue sample

[Create New Seedlot]

Name	Type
II6_1015_plant_1_leaf1	tissue_sample
II6_1015_plant_1_root2	tissue_sample
II6_1015_plant_1_stem3	tissue_sample
II6_1015_plant_2_leaf1	tissue_sample
II6_1015_plant_2_root2	tissue_sample
II6_1015_plant_2_stem3	tissue_sample
II6_1015_plant_3_leaf1	tissue_sample
II6_1015_plant_3_root2	tissue_sample
II6_1015_plant_3_stem3	tissue_sample
II6_1018_plant_1_leaf1	tissue_sample

Show 10 entries Search: ▾

Showing 1 to 10 of 234 entries Previous 1 2 3 4 5 ... 24 Next

Copy Stocks to a List Copy the stock names showing in table to a new or existing list

### 10.2.11 Uploading GPS Coordinates For Plots

You can upload GPS coordinates for the plots in your trial. There is a link on the Trial Detail Page as shown below.

Design

[Download Layout]

Attribute	Value
Design	CRD
Number of Blocks	1
Number of Replicates	3
Plot Length	
Plot Width	
Plants Per Plot	3

Accessions

Seedlots

Controls

Plots

Plant Entries

Field Management Factors

[Upload Seedlots Planted in Trial]

[Upload GPS Coordinates]

[Add Management Factor]

Clicking on this link will bring up the following dialog.



Here you can upload an XLS or XLSX file. To see information on the format of the file that should be uploaded, click on “Spreadsheet format”. This will bring up the following dialog.



This dialog tells you that the file must be XLS or XLSX and must contain: plot\_name WGS84\_bottom\_left\_x WGS84\_bottom\_left\_y WGS84\_bottom\_right\_x WGS84\_bottom\_right\_y WGS84\_top\_right\_x WGS84\_top\_right\_y WGS84\_top\_left\_x WGS84\_top\_left\_y The GPS coordinates should be WGS84 format and specify a four-pointed polygon around the plot.

### 10.2.12 Repetitive Measurements Section

If a trial includes repetitive traits or time-series values, you can effectively view and analyze these values through the Repetitive Measurements Section. Start by selecting the desired trait from the trait drop-down menu. Next,

define the date range by either using the date-range picker or an interactive slider, which allows you to dynamically adjust the period you wish to examine. Once the date range is set, determine how to handle the repetitive measurements by choosing from various options such as First Value, Last Value, Averaged Value, Sum Values, or All Values. Choosing the “All Values” option enables an additional feature that visualizes the trend of the values over time, helping you identify patterns and trends within the data.

The screenshot shows a web-based application titled "Repetitive Measurements". At the top, there is a search bar labeled "View a plot by repetitive measurements level". Below it, a form is displayed with the following fields:

- Select Trait: cassava mosaic disease incidence 12-month evaluation
- Start Date: 2000-01-01
- End Date: 2021-02-01
- Choose DateRange slider: (A horizontal slider is shown)
- Observation Level: Plots
- Repetitive Measurements Level: All values
- Submit button

Below the form, there is a table with the following data:

Observation Unit Name	Accession Name	Repetitive Values	Repetitive line graph
199934HBEPR_cara_rep1_UG120002_1	UG120002	4, 7, 1	
199934HBEPR_cara_rep1_UG120004_2	UG120004	8, 2, 5	
199934HBEPR_cara_rep1_UG120006_3	UG120006	3, 6, 9	

At the bottom of the table, it says "Showing 1 to 3 of 3 entries". There are also "Previous" and "Next" buttons.

### 10.2.13 Uploading Additional Files To Trial

It may be of interest to you to upload additional documents, images, or recordings to your trial. To do this, scroll down to the “Uploaded Additional File” section on the trial detail page. From here you can view and download any of these additional files.

<input checked="" type="radio"/> <a href="#">Uploaded Additional Files</a> <span style="float: right;"><a href="#">[Upload Additional Files]</a></span>			
<input type="button" value="Show 10 ▾ entries"/> <span style="float: right;">Search: <input type="text"/></span>			
Filename	Date Uploaded	Uploaded By	Options
2018-01-17_15:30:45_2016_mchare_pollination_block	2018-01-17 15:30:49.967178+00	nmorales	<a href="#">Download</a>
2018-01-17_18:03:42_2016_mchare_pollination_block	2018-01-17 18:03:47.092829+00	nmorales	<a href="#">Download</a>
2018-01-17_18:12:36_Screenshot from 2017-04-28 12:35:05.png	2018-01-17 18:12:40.924951+00	nmorales	<a href="#">Download</a>
2018-01-17_18:14:26_Screenshot from 2017-04-28 12:35:05.png	2018-01-17 18:14:30.73281+00	nmorales	<a href="#">Download</a>
2018-01-17_18:15:38_Screenshot from 2017-04-28 12:35:05.png	2018-01-17 18:15:42.328389+00	nmorales	<a href="#">Download</a>
2018-01-17_18:17:25_Screenshot from 2017-04-28 12:35:05.png	2018-01-17 18:17:29.467101+00	nmorales	<a href="#">Download</a>

Showing 1 to 6 of 6 entries      Previous      1      Next

To upload an additional file, click on the “Upload Additional Files” link. A dialog will appear where you simply select your desired file. For information, you can click “Upload information” to see the following message.



## 10.3 Updating Trial Data

To update the trial-level metadata (such as the planting date, design type, description, etc) of one or more existing trials, click the “Update Existing Trial(s)” button from the Manage > Field Trials page. This upload can also be used to rename trials or move trials to a different breeding program. In order to update a trial, you must be a curator or a submitter (that is associated with the breeding program of the trials).



Here you can upload a file that contains the new metadata for the existing

trials in the database. The first column is labeled ‘trial\_name’ and includes the name of the existing trial. Additional columns can be included for the metadata you want to update. Any columns not included in the file or values left blank will leave the existing metadata unchanged. The columns that can be included are:

- new\_trial\_name: A new name for the trial, must not already exist in the database
- breeding\_program: The name of breeding program that managed the trial, must exist in the database.
- location: The name or abbreviation of the location where the trial was held, must exist in the database.
- year: The year the trial was held.
- transplanting\_date: The transplanting\_date of the trial was conducted. Date in YYYY-MM-DD format or ‘remove’ to remove the date
- planting\_date: Date of Planting in YYYY-MM-DD format or ‘remove’ to remove the date
- harvest\_date: Date of Harvest in YYYY-MM-DD format or ‘remove’ to remove the date
- design\_type: The shorthand for the design type, must exist in the database. Possible values include CRD: Completely Randomized Design, RCBD: Randomized Complete Block Design, RRC: Resolvable Row-Column, DRRC: Doubly-Resolvable Row-Column, ARC: Augmented Row-Column, Alpha: Alpha Lattice Design, Lattice: Lattice Design, Augmented: Augmented Design, MAD: Modified Augmented Design, greenhouse: undesigned Nursery/Greenhouse, splitplot: Split Plot, p-rep: Partially Replicated, Westcott: Westcott Design
- description: Additional text with any other relevant information about the trial.
- trial\_type: The name of the trial type, must exist in the database. Possible values include Seedling Nursery, phenotyping\_trial, Advanced Yield Trial, Preliminary Yield Trial, Uniform Yield Trial, Variety Release Trial, Clonal Evaluation, genetic\_gain\_trial, storage\_trial, heterosis\_trial, health\_status\_trial, grafting\_trial, Screen House, Seed Multiplication, crossing\_block\_trial, Specialty Trial
- plot\_width: plot width in meters
- plot\_length: plot length in meters

- field\_size: field size in hectares

## 10.4 Deleting Trial Data

To delete a trial data, click on the “Delete trial data” section. There are links to delete traits, layout and trial entry data.

The screenshot shows the YamBase software interface. At the top, there is a navigation bar with links for 'YamBase', 'Search', 'Manage', 'Analyze', 'Maps', 'About', and user-specific options like 'ogbalex', 'Lists', 'Calendar', and a profile icon. Below the navigation bar, a search bar displays the placeholder 'No data available in table'. Underneath the search bar, it says 'Showing 0 to 0 of 0 entries' and includes 'Previous' and 'Next' buttons. At the bottom of the interface, there is a section titled 'Data Agreement' with a link '[Add/edit data agreement]'. A red box highlights the 'Delete trial data' button, which is accompanied by the note 'Deletion cannot be undone'. Other buttons in this section include 'Delete trait data', 'Delete layout data', and 'Delete trial entry'.

To delete assayed trait data, click on “Delete trait data” link. On the appeared dialog, confirm deletion by clicking on the “Select Traits For Deletion” button, then select one or more traits to delete from the trial.



To delete trial layout data, click on the “Delete layout data” link. Confirm deletion on the appeared dialog.

To Delete trial entry, click on “Delete trial entry” link. Confirm deletion on the appeared dialog.

# Chapter 11

## Managing Genotyping Plates

Genotyping Plates represent the content of a genotyping plate sent to a genotyping facility (e.g. samples in specific wells). To streamline this process, it is possible to upload this information or let the database create a plate for you. Once the genotyping plate is saved in the database it is then possible to export the information directly to genotyping facilities that are BrAPI compliant. The genotyping facility can then provide status information to us via BrAPI.

To begin go to Manage->Genotyping Plates.

**Manage Genotyping Trials**

[About Genotyping Trials](#)

**What are genotyping trials?**

- Genotyping trials represent 96 or 384 well plates.
- Each well in the plate has a unique source sample ID.
- The "contents" of each well can be either a tissue sample, plant name, plot name, or accession name. This "source" name must be in the database (e.g. as tissue samples or plants or plots from a trial, or just as accession names). Ideally you will have the barcodes from the field with you.
- Use the "Coordinator" Android Application to scan your "source" barcodes and record the position of the tissue sample in the 96 or 384 well plate. If you prefer you can create your own XLS file and upload that, if you do not want to use the Coordinator Application. Alternatively you can let the database generate the genotyping trial for you, and then produce the plate in that layout.
- For more information on the "Coordinator" Android Application go to [Coordinator](#).
- Click "Add Genotyping Trial" and fill in the form completely.
- To ease shipping materials to the genotyping facility, we can generate the required templates for you after the data is in the database.

**Genotyping Trials**

[Add Genotyping Trial]

Information	Breeding Programs – Folders – Genotyping Trials
<b>Search</b> <input type="text" value="Search"/> <p>Double click genotyping trial ( ) or folder ( ) to view detail page.</p> <p>Breeding programs ( )</p> <b>Folders</b> <a href="#">Create new folder</a> <a href="#">Move genotyping trial(s) to folder</a> <a href="#">Move folder</a>	<a href="#">Refresh</a> <ul style="list-style-type: none"> <li>test           <ul style="list-style-type: none"> <li>18DIA0001</li> <li>testing_folder_g</li> <li>test_coord2</li> </ul> </li> </ul>

Here the genotyping plates are divided by Breeding Program. These sections can be expanded by clicking on one.

- SCP
- ARI Tanzania
- BTI
- CARI
- CIAT
- CNRA\_PRT
- CSIR
- IITA
- KU
- **NRCRI**
  - NRCRI\_POLYCROS2
  - NRCRI\_POLYCROS10
  - NRCRI\_GS4
  - NRCRI\_GMS
  - NRCRI\_GS3
  - NRCRI\_GS7
  - NRCRI\_GS5
  - NRCRI\_POLYCROS1
  - NRCRI\_GM1
  - NRCRI\_GM2
  - NRCRI\_PP1
  - NRCRI\_GS6
  - NRCRI\_GM4
  - NRCRI\_POLYCROS3
  - NRCRI\_POLYCROS7
  - NRCRI\_POLYCROS9
  - NRCRI\_POLYCROS8
  - NRCRI\_POLYCROSS
  - NRCRI\_GS9
  - NRCRI\_GS8
  - NRCRI\_POLYCROS6
  - NRCRI\_GS2
  - NRCRI\_GMS
  - NRCRI\_POLYCROS4
  - NRCRI\_GS1
- NaCRRI
- Other
- Rayong
- SRI Kibaha

## 11.1 Adding a New Genotyping Plate

To begin, click on “Add Genotyping Plate”. Notice that this form is split into three sections: “Plate Information”, “Well Information”, and “Confirm”. The first section is for defining information about the genotyping plate, such as a Plate identifier, plate format (96 well), etc. The second section is for defining the samples in the wells, such as sample names, sample concentrations, well position, etc. The final section is for Submitting the info.

All fields in the Plate Information section are required.

The screenshot shows a modal dialog box titled "Add Genotyping Trial". At the top, there are three tabs: "Plate Information" (selected), "Well Information", and "Confirm". The "Plate Information" tab contains the following fields:

- Genotyping Project Name:** Should match Vendor Project (text input: e.g. NextGenCassava)
- Genotyping Plate ID:** (text input: e.g. 18DNA00001)
- Plate Format:** (dropdown menu: 96 Well)
- Sample Type:** (dropdown menu: DNA)
- Breeding Program:** (dropdown menu: IITA)
- Location:** (dropdown menu)
- Year:** (dropdown menu: 2017)
- Description:** (text area)
- Genotyping Facility:** (dropdown menu: None)

At the bottom right of the dialog box is a "Close" button.

In the Well Information section you can choose to either 1) Upload an XLS or XLSX spreadsheet with your sample layout or 2) let the database create the sample layout.

**Add Genotyping Trial**

**Plate Information**   **Well Information**   **Confirm**

**1. Do you already have a plate layout created?**

**File format information**: Spreadsheet format

Select Plate Layout XLS File:  No file chosen

**2. Or do you want us to generate a plate layout for you?**

- Select a list for the source material going into each well. Your list should be a one to one pairing to each well e.g. if you want to fill 95 wells you should supply a list of 95 elements.
- Note: From the most desirable to least desirable list type you can choose: tissue samples, plants, plots, or accessions

**Source Observation Unit List:** 119acc

**Blank Well:** (Cornell IGD requires a specific well to be blank.) e.g. A01

**Well Concentration (ng/u):** (If you used the same conc for all wells)

**Well Volume (uL):** (If you used the same vol for all wells)

**TISSUE:** (If used the same tissue for all wells)

**Extraction:** (If used the same extraction for all wells)

**Person:** (If same person prepared all wells.)

If you choose to upload an XLS or XLSX spreadsheet, the Spreadsheet Template info requires the following:

**Upload Template Information**

This is for uploading a pre-existing genotyping plate layout.  
File must be Excel file (.xls)  
(.xlsx format not supported)

**Header:**  
The first row (header) must contain the following:  

date	sample_id	well_A01	row	column	source_observation_unit_name	dna_person	notes	tissue_type	extraction	concentration	volume	is_blank
------	-----------	----------	-----	--------	------------------------------	------------	-------	-------------	------------	---------------	--------	----------

**Required fields:**

- date (should be YYYY/MM/DD)
- sample\_id (the unique identifier for the sample in the well)
- well\_A01 (the position of the sample in the plate)
- row (the row position of the sample in the plate e.g. A)
- column (the column position of the sample in the plate e.g. 10)
- source\_observation\_unit\_name (must exist in the database, the identifier of the origin material, in order of most desirable identifier to least desirable identifier that can be used here: tissue sample name, plant name, plot name, accession name)

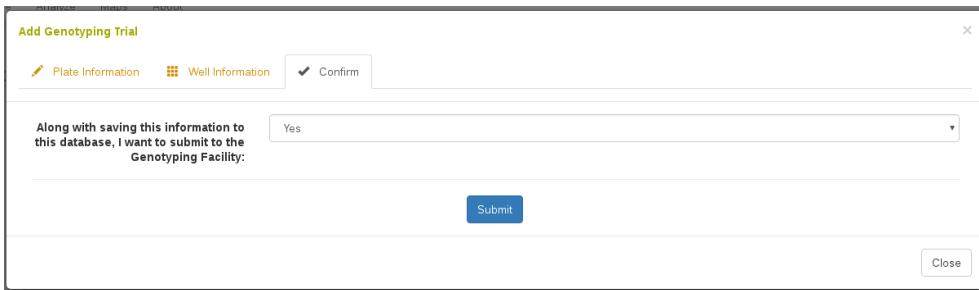
**Optional fields:**

- dna\_person (the name of the person who prepared the well)
- notes (any additional notes on the well)
- tissue\_type (free-text for what type of tissue is present in the well)
- extraction (free-text for the extraction method e.g. CTAB)
- concentration (concentration in ng/u)
- volume (volume in uL)
- is\_blank (indicates if well is blank, write 1 if blank, otherwise leave empty.)

In either case, the sample identifier is generally a concatenation of Plate name and well position, e.g. MyGenotypingTrial1\_A01. In either case, you need to provide a “source\_observation\_unit\_name” for each sample. This

can be a tissue sample name, a plant name, a plot name, or an accession name; however, in any case, the identifier must already exist in the database. This allows us to link the sample in the well to specific field trial plots, or, plants, or tissue\_samples. If you only know which accession is in the well, you can use the accession name.

In the final Confirm section you can decide whether to submit this information to the genotyping facility you selected. This requires that the genotyping facility is BrAPI compliant to work.



## 11.2 Genotyping Plate Detail Page

If you open a specific genotyping plate, it will take you to the detail page. Here you can see the Accessions used in the plate (if you created the trial and the source\_observation\_unit\_names you used were plots, this will still work because we know the accession of the plot or plant or tissue sample).

**Genotyping trial genou31**

Breeding Program	IITA (IITA cassava breeding program, Ibadan, Nigeria)
Trial Type	Genotyping Trial
Plate Format	96
Plate Sample Type	DNA
Genotyping Facility	lgd
Submitted to Genotyping Facility	yes
Genotyping Facility Status	

Live Status From Genotyping Facility	
<a href="#">Download PDF</a>	


  
SGN trial 3391 (genou31)

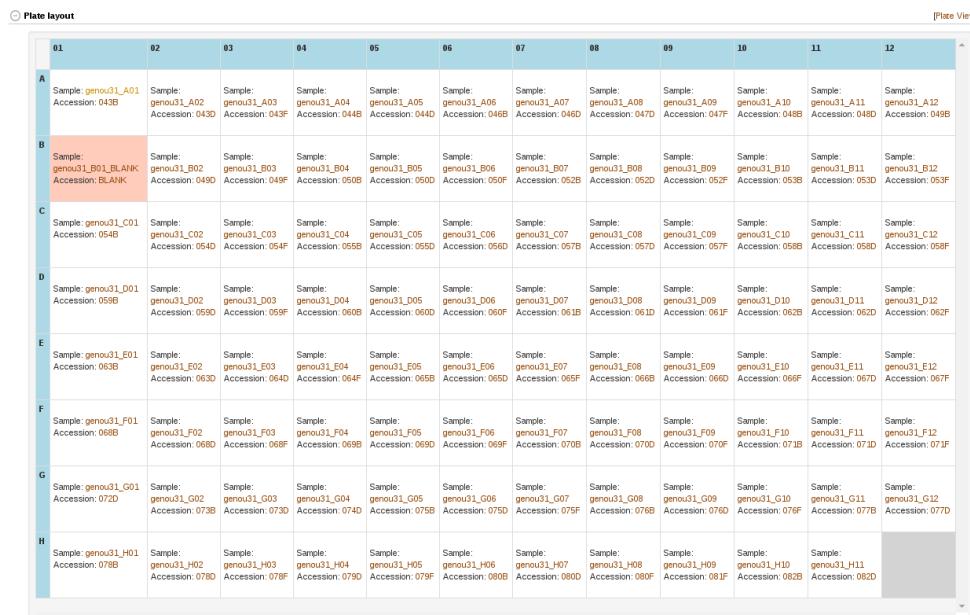
Design      [Download layout \[xls\] \[cs\]](#)

Accessions

Tissue Sources

Tissue Samples

Further down you can see a graphical representation of your plate with well positions. This can be 96 well or 384 well depending on your plate format.



# Chapter 12

## Using Field Book App

### Field Book

Field Book is a one of the apps developed as part of the One Handheld Per Breeder initiative. Our ultimate goal is to create a program that can be used by a wide range of researchers to take notes on field research plots. While there have been several versions of similar software or attempts to create a digital note-taking tool released in the past, most of these were either simple spreadsheet applications that didn't allow for fast and flexible data entry or were created for proprietary and expensive hardware, inaccessible for most research groups.

By working with an open-source software and relatively inexpensive hardware (\$200-\$300), we have created a platform that will not only allow researchers the ability to replace hard-copy field books (alleviating the possibility of transcription errors), but allow current technology to be used in environments where cost and inflexibility have been limiting factors. We have worked with an independent developer to create a flexible, open-source, note-taking program. Input data can be numeric, percentage, categorical, date, and Boolean. These data types can be used for a variety of traits.

Field Book is written in Java and can be used on any 7 or 10 inch Android tablet. The file can be installed by moving the .apk file to the tablet memory, allowing installations from non-Market apps, and opening the .apk file. Examples of import files are available within the app. Please backup any data before upgrading from earlier versions of Field Book.

[App](#) | [Manual](#) | [Source](#)

### Developers

Jesse Poland, *Project Lead, USDA-ARS / Kansas State University*

Technology Projects, *Project Programmer*

Trevor Rife, *Project Designer and Programmer, Kansas State University*

### Translators

Irazema Fuentes-Bueno, *Spanish, USDA-ARS*

Sandra Dunckel, *German, Kansas State University*



SGN databases support the Android Field Book App for collecting phenotypic data in the field with tablet computers. The app is available here:

<https://play.google.com/store/apps/details?id=com.fieldbook.tracker>

- The app can also be downloaded directly from the Google Play store.

There is no charge for the app.

- Field Book App requires two files for collecting data: Field layout file and trait file.
- SGN databases can generate the field layout file and trait file, which can be downloaded onto your computer, then transferred to an Android tablet device.

## 12.1 A typical workflow

1. Creating a *field layout file* based on the design of field trial
2. Creating a *trait file* from the list of traits
3. Downloading the field layout file and trait file from the database to your computer
4. Downloading the field layout file and trait file to the tablet (where the Field Book App is installed)
5. Collecting phenotypes
6. Exporting phenotypes from Field Book App to your computer
7. *Uploading the exported phenotype file* from your computer to the database

## 12.2 Creating Field Layout Files for the Field Book App

There are two alternative methods for creating “Field Layout Files”.

1. Using “Field Book Tools” page
2. Using “Trial Detail” page.

## 12.2. CREATING FIELD LAYOUT FILES FOR THE FIELD BOOK APP151

### 12.2.1 Creating “Field Layout Files” by using “Field Book Tools” page.

To access “Field Book Tools” page, clicking on “Field Book App” in the “Manage” menu.



On the “Field Book Tools” page, clicking on “New”



On the “Download Fieldbook” window, selecting trial name and data level (plots or plants), then clicking on “Submit” button. A treatment can be selected, which allows you to record phenotypes based on treatment application. A list of traits can be selected, which provides a summary of an accession’s global performance for those traits in the Fieldbook.



If the field book layout file was successfully created, a pop-up window will indicate that the field book layout file was saved successfully. Clicking on the file name will immediately download the file onto your computer. The file is also available to download on the “Field Book Tools” page, if you need to re-download it.



To download field layout file to your computer, clicking on “Download File”, the file can then be transferred to your tablet. If you no longer want to keep the field layout file, clicking on “Delete Layout File”.

## 12.2. CREATING FIELD LAYOUT FILES FOR THE FIELD BOOK APP153

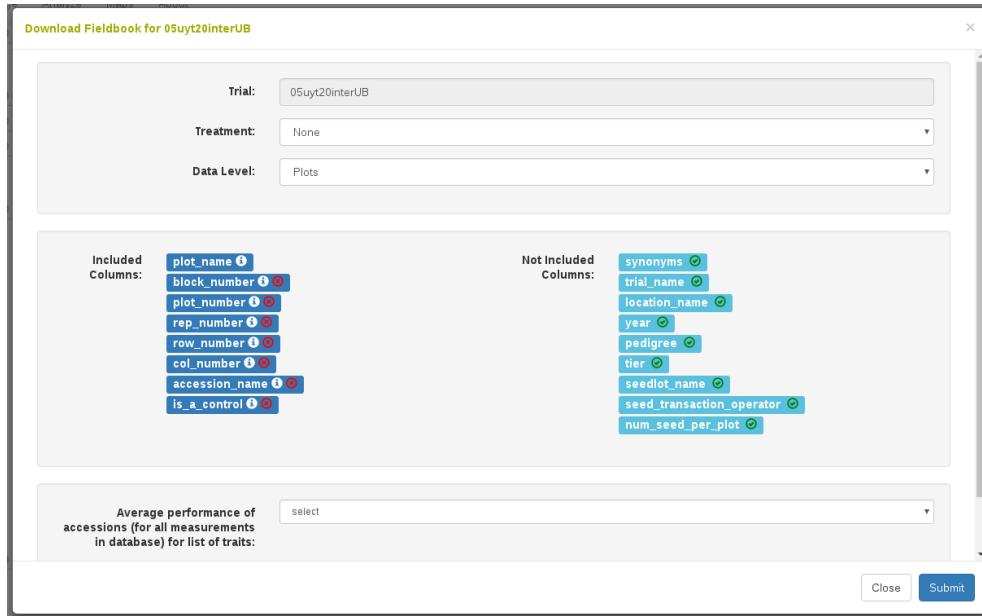
The screenshot shows the CASSAVABASE Field Book Tools interface. At the top, there is a navigation bar with links for Search, Manage, Analyze, Maps, About, and a user profile for Gregor\_Mendel. Below the navigation bar is a search bar and a button labeled 'Lists' with a grid icon. The main content area is titled 'Field Book Tools'. A sidebar on the left contains a link to 'Field Book Software'. The main content area has a heading 'Field Layout Files' with a 'New' button. Below this, there is a timestamp '2016-10-12\_20:08:09\_05uyt20lnterUB.xls' followed by two red-bordered buttons: 'Download File' and 'Delete Layout File'. Underneath this, there are sections for 'Trait Files', 'Uploaded Phenotype Files' (with a 'None' status and a '[Upload]' button), and 'Removed Phenotype Files' (with a 'None' status).

### 12.2.2 Creating “Field Layout Files” by using “Trial Detail” page.

To create “Field Layout Files”, go to the “Trial Detail” page of the trial that you want to create the file. On the “Trial Detail” page, scrolling down to the bottom of the page to find “Android Field Book Layout” in the “Files” section, then clicking on the “Create Field Book” link.

The screenshot shows the 'Files' section of a Trial Detail page. It includes a 'Data Collection Files' section with 'Phenotyping Spreadsheets', 'Android Field Book Layout' (which has a red box around the '[Create Field Book]' button), and 'Data Collector Spreadsheet'. There is also a 'Create Spreadsheet' button next to 'Phenotyping Spreadsheets'. Below this is an 'Uploaded Phenotyping Files' section with 'Phenotyping Spreadsheets', 'Android Field Book Exported', and 'Data Collector Spreadsheet', each with an '[Upload]' button.

Clicking on the “Create Field Book” link will open a new window showing the name of the trial that you selected, as well as data level (plots or plants). A treatment can be selected, which allows you to record phenotypes based on treatment application. A list of traits can be selected, which provides a summary of an accession’s global performance for those traits in the Fieldbook. To proceed, clicking on “Submit” button.



If the field book layout file was successfully created, a pop-up window will indicate that the field book layout file was saved successfully. Clicking on the file name will immediately download the file onto your computer. The file is also available to download on the “Field Book Tools” page, if you need to re-download it.



To download field layout file to your computer, clicking on “Download File”, the file can then be transferred to your tablet. If you no longer want to keep the field layout file, clicking on “Delete Layout File”.

The screenshot shows the CassavaBase Field Book Tools interface. At the top, there's a navigation bar with links for Search, Manage, Analyze, Maps, and About. On the right, there's a search bar, a user profile for 'Gregor\_Mendel', and buttons for 'Lists' and 'Field Book Tools'. Below the navigation, a box titled 'Field Book Tools' contains an information icon and text about the app. A section for 'Field Layout Files' shows a single entry: '2016-10-12\_20:08:09\_05uyt20InterUB.xls'. Buttons for 'Download File' and 'Delete Layout File' are shown in red boxes. Below this, there are sections for 'Trait Files', 'Uploaded Phenotype Files' (with a 'New' button), and 'Removed Phenotype Files'.

## 12.3 Creating Trait Files for the Field Book App

Steps to Create a Trait File:

### 12.3.1 Creating a Trait List

After you logged in, lists can be created and managed using the Search Wizard or the “Lists” link. For more information on how to create lists, click [here](#).

The screenshot shows the main CassavaBase homepage. At the top, there's a navigation bar with links for Search, Manage, Analyze, Maps, and About. On the right, there's a search bar, a user profile for 'Gregor\_Mendel', and buttons for 'Lists' and 'Field Book Tools'. A sidebar on the left has a 'Wizard' menu with options like 'Accessions and Plots', 'Trials', 'Markers', 'Images', 'People', and 'FAQ'. A large image in the center shows a person working in a cassava field. A callout box points to the 'Field Book Tools' button with text: 'Search accessions and trials', 'Make crossings', 'Fieldbook App & uploading', and 'Cassava Trait Ontology'.

### 12.3.2 Creating a Trait File

After you have your trait list, clicking on the “Field Book App” link found under the “Manage” menu tab. This will take you to the “Field Book Tools” page.



To create a new trait file, finding the heading “Trait Files”, then clicking on the “New” link.



Clicking on the “New” link will open a dialogue box titled “Create Trait File”. Please enter your “Trait file name” and select “List of traits to include” from drop-down list that you previously created. You can only use traits included in the list. Check the box titled “Include Notes Trait” if you would also like to record and upload general plot notes in the field. Click “OK” to submit.



If your trait file was successfully created, a new window will indicate that the trait file was saved, then clicking on “Close”.

## 12.4. TRANSFERRING FILES FROM YOUR COMPUTER TO ANDROID TABLET 157



After the trait file was saved, you will see your file listed in the “Field Book Tools” page. Clicking on “Download” link to download the trait file to your computer.



After downloading the trait file to your computer, the file can be transferred to an Android Tablet. You need the Android Field Book App to open the file. The Android Field Book App can be downloaded at: <http://www.wheatgenetics.org/bioinformatics/22-android-field-book>

## 12.4 Transferring Files from Your Computer to Android Tablet

### 12.4.1 Files on your computer

After downloading, Field Layout files and Trait files can be found in the “Downloads” folder of your computer. Field Layout files on your computer will have a prefix “fieldbook\_layout\_” added to the beginning of the file name. For example: “**2014-01-28\_19:14:34\_Trial Demo\_location 6767.xls**” on the the database website will be saved as “**field\_book\_layout\_2014-01-28\_19:14:34\_Trial Demo\_location**

6767.xls” on your computer.



The files can be transferred to Android tablet by copying the files into the tablet’s Internal Storage File.

#### 12.4.2 Files on your Android tablet

To transfer Field Layout file and Trait file to your Android tablet, connecting an Android tablet to your computer, then clicking on tablet icon on your computer. Clicking on the tablet icon will open a window showing an “Internal Storage” file.



After you installed the Android Field Book App, all files for the app are stored in the “fieldBook” folder within the “Internal storage” folder.

## 12.4. TRANSFERRING FILES FROM YOUR COMPUTER TO ANDROID TABLET159



Within the “fieldBook” folder, there are five sub-folders:

- field\_export
- field\_import
- plot\_data
- resources
- trait

Field Layout files must be copied into the “field\_import” folder.



Trait files must be copied into the “trait” folder.



You can either drag and drop, or copy the Field Layout file and the Trait file from your computer to the folders in your Android tablet.

## 12.5 Setting up “Field Book App” for data collection

After you transferred the Field Layout file and Trait file from your computer to Android tablet, you still need to set up “Field Book App” on your tablet for data collection.

To set up the Field Book App:

1. To open the Field Book App in the Android Tablet, clicking on the Field Book App icon, which is a green rectangle.

## 12.5. SETTING UP “FIELD BOOK APP” FOR DATA COLLECTION161



2. To import Field Layout files, clicking on the “Fields” section of the main menu of the Field Book App.



Clicking on the “Fields” tab will open a new dialogue that will let you select the file that you want to import.

## 12.5. SETTING UP “FIELD BOOK APP” FOR DATA COLLECTION163



Choosing a Field File will generate a new dialogue that will ask you to choose between an Excel or CSV format. Since the data from the database is in Excel format, choose the Excel option.



After submitting the file format, a final dialogue box will appear. Please provide information about the file that you want to import. Please ensure that “plot\_name” is set as the unique identifier. To finalize the process, clicking “OK” button.

## 12.5. SETTING UP “FIELD BOOK APP” FOR DATA COLLECTION165



3. To import Trait Files, clicking on the “Traits” tab on the main menu of the Field Book App.



Then, clicking on the three dots symbol found on the upper right corner of the Field Book screen. This will open a drop down menu with the choices “Import” and “Export”. Clicking on “Import”

## 12.5. SETTING UP “FIELD BOOK APP” FOR DATA COLLECTION167



Clicking on “import” will open a new dialogue that displays a list of trait files that you can select to import to the Field Book App.



The trait file is now imported into the Field Book App. The traits page will show all trait files and available traits.



## 12.6 Exporting Files from Field Book App

Data that were collected on the Field Book App can be exported back to your tablet folder, which can then be transferred to your computer.

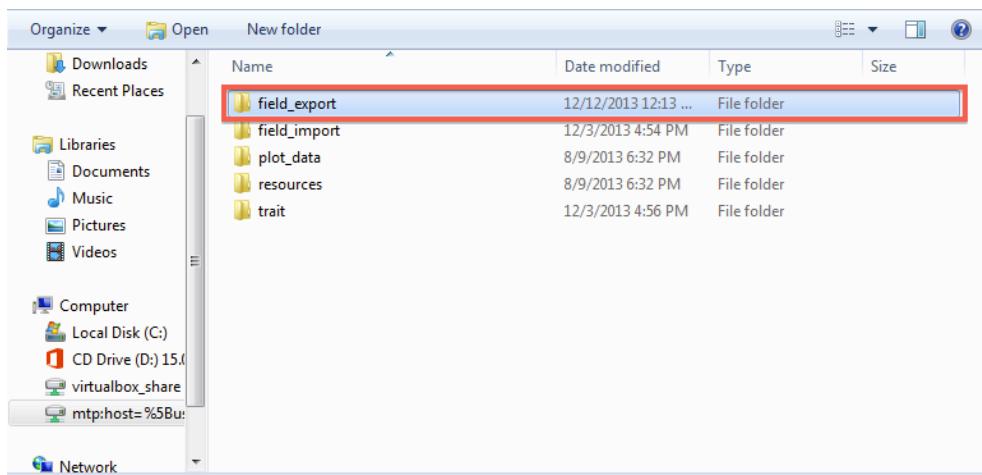
To export files containing data from the Field Book App to your tablet, clicking on the “Export” link on the main menu page of the Field Book App.



Clicking on the “Export” link will open a new dialogue window. To ensure that data are exported in a correct format for the database, checking the “Database Format” box, then clicking on “OK” button.



The exported file can then be found in the “field\_export” sub-folder within the “fieldBook” folder on your tablet. Once you connect your tablet to your computer, you can directly transfer the file to your computer.





## 12.7 Uploading Phenotype Files to an SGN database

To upload phenotype files to the database, clicking on “Field Book App” in the “Manage” menu.



On the “Field Book Tools” page, clicking on “Upload” link in the “Uploaded Phenotype Files” section.



Clicking on the “Upload” link will open a new dialogue asking you to choose

## 12.7. UPLOADING PHENOTYPE FILES TO AN SGN DATABASE 173

a file that you want to upload to the database website. Please ensure that “plot\_name” is the first column of the file to be uploaded. To make sure that the file has the correct format for uploading, click on the “Verify” button. After the file format has been verified, click on the “Store” button.



The list of uploaded phenotype files can be found on the Field Book Tools page

A screenshot of the 'Trait Files' page. It features sections for 'Trait Files' (listing several .trt files with download links) and 'Uploaded Phenotype Files' (listing a single file: '2014-01-09\_08:23:09\_2014.01.09\_fieldbook\_layout\_2014-01-09\_04-35-14\_Trial Demo\_location 6767.xls\_database.csv' with an 'Upload' link). A red arrow points to the 'Uploaded Phenotype Files' section. Below these are sections for 'Removed Phenotype Files' (empty) and 'None'.

The uploaded files will also be seen in the corresponding “Trial Detail” page.

Trial detail for Trial Demo\_location6767

Breeding program	change
Demo (Demonstration Breeding Program) <b>X</b>	
Year(s)	
6767	
Location(s)	
Demo_location	
Description	
Demonstration in SanDiego	
<b>Design</b>	
Design: RCB	
Number of blocks: 4	
Number of replicates: 1	
<b>Accessions</b>	
<b>Plots</b>	
<b>Traits assayed</b>	
storage root size (3 assays)	
leaf lobe number (3 assays)	



# Chapter 13

## Managing Phenotypic Data

To facilitate uploading process for phenotypic data, “Manage Phenotypic Data” page provides two options for uploading: Field Book Phenotype file in database format and phenotype file in Excel (.xls or .xlsx) file format. To access “Manage Phenotypic Data” page, clicking on “Phenotyping” in the “Manage” menu.



### 13.1 Uploading Fieldbook Phenotypes

#### 13.1.1 Export Field Book Database File

The database upload of Field Book phenotype data relies on the “Database” format from the Field Book. Please make sure to export the “Database” format from the Field Book if you intend to upload the data using the Field Book Upload we describe below. If you prefer to use the “Table” format that the Field Book exports, you can modify this format to work with the Speadsheet Upload we describe below.

### 13.1.2 Upload Field Book Database File

To upload a Field Book Phenotype file in a database format, click the “Upload Fieldbook” link

The screenshot shows the CASSAVABASE interface with the title "CASSAVABASE" at the top. Below it is a navigation bar with links for Search, Manage, Analyze, Maps, and About. A search bar and user profile "Gregor\_Mendel" are also present. The main content area is titled "Manage Phenotypic Data". It has sections for "Phenotype Search", "Uploaded Files" (which includes a link "[Upload Fieldbook]" highlighted with a red box), and "Removed Files".

The “Upload Fieldbook” link on this page and “Upload” link on the “Field Book Tools” page open the same dialogue. Please follow instructions for uploading phenotypic files on the 12 page.



## 13.2 Uploading Spreadsheet Phenotypes

To upload a phenotype file in an Excel (.xls or .xlsx) file format, click the “Upload Spreadsheet” link.

The screenshot shows the CASSAVABASE interface with the title "CASSAVABASE" at the top. Below it is a navigation bar with links for Search, Manage, Analyze, Maps, and About. A search bar and user profile "Gregor\_Mendel" are also present. The main content area is titled "Manage Phenotypic Data". It has sections for "Phenotype Search", "Uploaded Files" (which includes a link "[Upload Spreadsheet]" highlighted with a red box), and "Removed Files".

Please specify “Data Level” (Plots or Plants) and select the Excel file that you want to upload.



### 13.2.1 Generating Spreadsheet File

You can find more file format information by clicking on “Spreadsheet Format” link. Clicking on “Spreadsheet Format” will open the following dialog.



Clicking on “Create Phenotyping Spreadsheet” will bring up a dialog where you can indicate the trial(s) you are interested in and the trait list you are interested in. Clicking “Submit” will download the xlsx file onto your computer, where you can then fill in the phenotypes.



### 13.2.2 Uploading Spreadsheet File

To ensure that the file has a correct format for uploading, click on the “Verify” button. This will check the contents of the file and also perform quality checks on the values in the file. These checks include checking the trait definition for categorical values, minimum and maximum values, and data type checking. It will also check if there are already values uploaded for the given observation units and traits. If there are, there is an option to overwrite the existing values with the new values in your file. If the file is valid, only then can you click “Store” to store the information in the database.

Upload Phenotype Spreadsheet

**i** File format information  
Spreadsheet Format

**Timestamps**  Included:

**Data Level:** Plots

**Phenotype Spreadsheet:** Choose File No file chosen

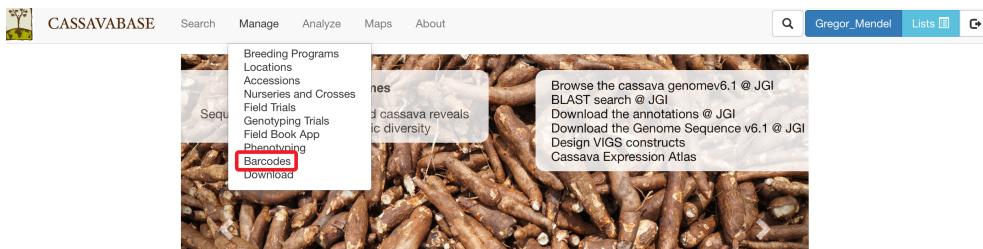
**Close** **Verify** **Store**



# Chapter 14

## Managing Barcodes

SGN databases provide tools for generating barcodes for stock identification. To access “Barcode Tools” page, clicking on “Barcodes” in the “Manage” menu.



“Barcode Tools” page provides four options for generating barcodes:

- Single barcode
- Multiple barcodes
- Plot phenotyping barcodes
- Trial barcodes

To generate single barcode, clicking on “Generate Barcode” link on the “Barcode Tools” page.



In the “Generate Barcode” section, specify the name of the barcode, size of the barcode, then clicking on “Generate Barcode”



The database will generate a barcode for your stock. The barcode can be printed for your stock identification. It also appears on its corresponding stock page.



If you have a list of stocks that you want to generate barcodes, you can use “Download Stock Barcodes” section. You have three options for entering stock names:

1. Typing in stock names, or copy and paste from other file into the box (1)
2. Choosing a list of stocks from your “Lists” (2), and transferring the list into the box (1) by clicking on “paste” button.
3. Uploading a “Tab-delimited Text File” with stock names.
4. Select an optional printing format from the available formats.

You can select printer settings that you prefer in the “Printer Settings” sec-

tion. After you enter stock names and specify printer settings, clicking on “Download Barcodes” button at the bottom of the page.

The screenshot shows the 'Barcode Tools' section of the CASSAVABASE website. At the top, there are navigation links: Search, Manage, Analyze, Maps, About, and a user profile section with 'Gregor\_Mendel', 'Lists', and a refresh icon. Below the navigation is a search bar with a magnifying glass icon and a blue 'Search' button.

**Generate Barcode**

**Download Stock Barcodes**

**1** Enter a List of Stock Names:

**2** Or Paste From a List:

**3** Or Upload Tab-delimited Text File With Stock Names:   
no file selected

Print Duplicate Labels Per Row:

Print Field Information For Plots:  Useful for Printing Field Information of Trials.

Print Parents For Nurseries:  Useful for Printing Pedigree Information for Nurseries.

**Printer Settings**

Number of Label Rows:	10	Number of Label Columns Per Page:	3
Page Format:	Letter	Add text to label, e.g. location:	<input type="text"/>
Top Margin (mm):	12	Left Margin (mm):	70
Bottom Margin (mm):	12	Right Margin (mm):	20

If you have a list of plots that you want to generate phenotyping barcodes, you can use “Download Plot Phenotyping Barcodes” section. You have three options for entering plot names:

1. Typing in plot names, or copy and paste from other file into the box (1)
2. Choosing a list of plots from your “Lists” (2), and transferring the list into the box (1) by clicking on “paste” button.
3. Uploading a “Tab-delimited Text File” with plot names.

[Download Plot Phenotyping Barcodes](#)

---

Enter a List of Stock Names:

Or Paste From a List:

Or Upload Tab-delimited Text File With Stock Names:  Choose File No file chosen

Add Text to Label, e.g. location:

If you have a list of trials that you want to generate barcodes, you can use “Download Trial Barcodes” section. You have three options for entering trial names:

1. Typing in trial names, or copy and paste from other file into the box (1)
2. Choosing a list of trial from your “Lists” (2), and transferring the list into the box (1) by clicking on “paste” button.
3. Uploading a “Tab-delimited Text File” with trial names.

[Download Trial Barcodes](#)

---

Enter a List of Trial Names:

Or Paste From a List:

Or Upload Tab-delimited Text File With Trial Names:  Choose File No file chosen

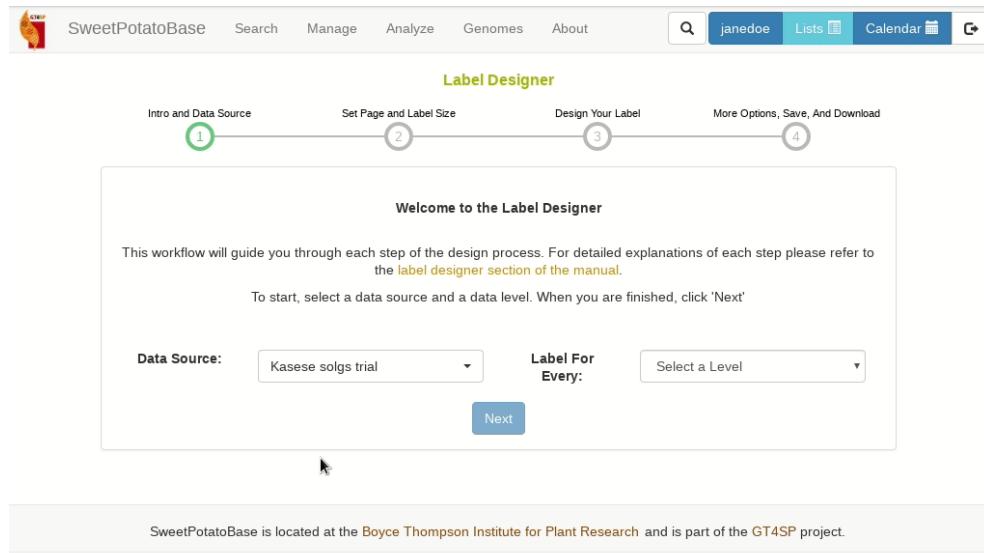
# Chapter 15

## Using the Label Designer

Breedbase provides an interactive design tool for creating custom labels. To access the Label Desginer, click on “Label Designer” in the “Manage” menu. The following sections explain your many options as you advance through each step of the design workflow.

### 15.0.1 First Select a Datasource

The first step is to select a data source. Since the label designer can generate labels for different data types, you can optionally filter the source selection by the data type you’re interested in. Then, select a field, genotyping, or crossing trial to populate your labels with the trial design information. Or select a list to populate your label with the list contents. For data sources with multiple levels of information you will also be asked to pick a level (plot, plant, etc.) before proceeding. To generate plot-level labels for more than one trial at once, select a list of trials as the source and plot as the level.



### 15.0.2 Set Page and Label Size

Now choose whether to create a new design or load a saved design. If you choose new, you will be prompted to select a page size and label size. If you do not see your page or label size as an option, then select Custom and enter your desired dimensions in pixels, or 1/72nds of an inch. If you choose saved, you will be prompted to select a saved design then will be taken directly to the design step with the saved design elements preloaded.



### 15.0.3 Design Your Label

Below is a draw area where you can begin adding elements to your label. First select a type, then field, size, and font, then click ‘Add’ You can add text to an exisiting field or create a completely custom field by clicking ‘Create Custom Field’ Once added, you can drag and drop elements, or delete them by clicking on the red box in their upper left corners. Barcodes can also be resized by dragging on the green box in their lower right corners. If you are creating labels for a trial it is highly recommended to include a barcode encoding your plot, plant, or tissue sample names. These are your unique identifiers that will need to included with any phenotypic or genotypic measurements loaded into the database. When you are satisfied with your design, click next!



#### 15.0.4 Adjust Formatting, Save, and Download

Last step! Here you can tweak your formatting and page layout, save your design, or download your labels. The additional settings dialog will allow you to fine tune the print margins and margins between labels. The units are pixels or 1/72nds of an inch. It's not recommended to change these until you've already done a test print. You can also set the # of copies per label, filter by rep, or download just the first page for test purposes. To save your design just type a unique name and hit save. This will save your design to your list manager where you can set it to public to share it with others. Finally if you are ready just hit download to generate and download your labels!





# Chapter 16

## Managing Downloads

You can download phenotype, trial meta-data, pedigree, GBS genotype and GBS genotype QC files from the database to your computer by using “Lists”. To download, clicking on “Download” in the “Manage” menu.



For each category, you can select a list of accessions from your “Lists” to download their phenotypes, pedigree, GBS genotype, GBS genotype QC. In the case of downloading trial meta-data, you would provide a list of trials, while for downloading phenotype and GBS genotype QC, you can also use a list of trials or traits.

A screenshot of a "Download Metadata" form. The form has a title "Download Metadata" and a subtitle "Select Parameters:". It contains a table with three columns: "Trials", "Options", and "Action". The "Trials" column has a dropdown menu set to "select". The "Options" column has a "Format:" dropdown menu set to "XLS". The "Action" column contains a blue "Download" button. The entire form is enclosed in a light gray border.

CASSAVABASE Search Manage Analyze Maps About Gregor\_Mendel Lists  

**Download Using Lists**

Choose a list for each parameter and click "Download".

**Download Phenotype**

Select parameter:

Accessions	Trials	Traits	Format	Timestamps	Data Level	Action
select	select	select	<input type="checkbox"/> .xls (default)	No	All	<b>Download</b>
			<input type="checkbox"/> .csv			
			<input type="checkbox"/> html			

**Download Pedigree**

Select parameter:

Accessions	Action
select	<b>Download</b>

**Download GBS Genotype**

Select parameter:

Accessions	Genotyping Protocol	Action
select	<input checked="" type="checkbox"/> GBS ApeKI Cassava genome v5 <input type="checkbox"/> GBS ApeKI Cassava genome v6 <input type="checkbox"/> protocol <input type="checkbox"/> GBS ApeKI Cassava genome v6_Oct2015	<b>Download</b>

**GBS Genotype QC**

Select parameter:

Trials	Accessions	Action
select	select	GBS ApeKI Cassava genome v5 <b>Quality Control</b>

# Chapter 17

## Managing ODK Data Collection

To access this page go to Manage and then ODK Data Collection. ODK is used for remotely collecting data on Android and IOS devices. We currently are working to support two ODK service providers, namely ONA and SMAP. We are using ONA to collect crossing information, including all lab activities following seed production. We are using SMAP for phenotypic data collection.

## 17.1 ONA Crossing Information

### 17.1.1 Managing ONA Crossing Information

**Manage ODK Data Collection**

**What is ODK?**

- ODK is an application that allows mobile data collection using user defined forms on Android or iOS devices. Data collected on the device can be instantaneously synched to the ODK server. To find out more go to the [ODK site](#). Many services have developed web interfaces to better streamline the ODK experience. These services assist in creating forms, deploying forms to your mobile application, and visualizing data uploaded back from the mobile device. Currently we are working with [SMAP](#) and [ONA](#) as two ODK services.

**What do I do from this page?**

- ONA is currently being used for collecting crossing information. This requires exporting a crossing plan from here to the OHA server. The crossing plan guides collection of cross information and this data is synched with ONA using ODK. From here, we run a script twice a day, which pulls data on OHA into our database.
- SMAP is currently being used for collecting phenotype information. The user collects phenotypes using a form they previously created. The questions in the form map directly to terms in the ontology. As they collect data on the mobile device, the data is synched to SMAP. From here, we run a script twice a day, which pulls data on SMAP into our database.

**Crossing Data: ONA ODK Application**

Select A Cross Wishlist: cross\_wishlist\_Abuja

Select An ODK Form on ONA: BTracT\_NM2018\_01\_BTracT - Nelson Mandela

**Management**

Export Cross Wishlist (Crossing Plan) to Selected Form on ONA

Import Crossing Data from Selected Form on ONA

Schedule Import For Selected Form: Once per day at midnight

Scheduled Time: Once per day at midnight

Confirm

To begin collecting data using the ONA ODK form you must first have a crossing plan in the form of a Cross Wishlist. To do this from this page, click the “Export Cross Wishlist to ONA” button. Please refer to the “Create Cross Wishlist” help section for more information. It is possible to view the current available cross wishlists by clicking the “Export Cross Wishlist to ONA” button and then clicking “Available Cross Wishlists”.

Once your cross wishlist is available, you can use your mobile ODK application to record crosses being done realtime. You can also record all laboratory activities following seed extraction up to greenhouse plantlet hardening.

As you collect data using your mobile ODK application, your responses will be synchronized with our database. The “Schedule Import for Selected Form” section gives you options to perform the import daily or more frequently. It is also possible to initiate a data import from ONA at anytime by clicking “Import Crossing Data from Selected Form on ONA”.

### 17.1.2 Reviewing Plant Status

The mobile ODK application has options to collect information about the status of plants in the field, such as if they are flowering. Images for each plant can also be recorded. The database will report this information here in a summary table that looks like the following. Notice that images are also transferred to the database.

Summary of Received Plant Status			
Show 10 entries			Search:
PlotName	Date	Status	Search:
	null	<b>Status:</b> Accession Name: null Trial Name: null User: null Status Location: Status: fallen Note: null Image: undefined	
16-Huti-white_r8c12_plot157	2018-02-12	<b>Status:</b> Accession Name: Huti-white Trial Name: 2016 mchare pollination block User: HM Status Location: in_field Status: destroyed Note: destroyed by elephants 	
16-ITC1460-Ijihulnkundu_r17c8_plot344	2018-01-18	<b>Flowering:</b> Accession Name: ITC1460-Ijihu Inkundu Plant Sex: female	
16-ITC0712-AAcvRose_r1c1_plot1	2018-01-17	<b>Flowering:</b> Accession Name: ITC0712-AAcv Rose Plant Sex: male	
16-ITC1468-Kahuti_r1c2_plot2	2018-01-17	<b>Flowering:</b> Accession Name: ITC1468-Kahuti Plant Sex: female	

### 17.1.3 Graphical Summary For Performed Crosses

There is a section to summarize activities done for each cross. In this table each row represents a single cross performed. All the activities that have been performed will be shown here, such as “first pollination” and “embryo rescue”. The scatter plot shown tracks seed numbers generated on the Y axis and date of activity on the X axis.



### 17.1.4 Summary Information For Performed Crosses

There is a secondary section to summarize what has been done across the entire Cross Wishlist. This tree structure shows all activities performed for a cross and shows how these crosses relate to the Cross Wishlist.



# Chapter 18

## Managing Tissue Samples

To access this page go to Manage and then Tissue Samples.

### 18.1 Tissue samples from field trials

A field trial contains plots planted with a specific accession. Each plot can contain many plants, which in turn can contain many tissue samples. On the manage tissue sample page we can see the field trials that contain tissue samples already. We can choose to download the tissue sample layout as seen in the below picture.

Manage Tissue Samples

Field Trial Tissue Samples

Create tissue samples for field trial

View and create tissue samples for field trials. Tissue samples come from a plant in a plot.

Field Trials With Tissue Samples

Show 10 entries

Trial name	Description	Breeding program	Folder	Year	Location	Trial type	Design	Planting Date	Harvest Date	Download
CASS_6Genotypes_Sampling_2015	Copy of trial with postcomposed phenotypes from cassbase.	test								<a href="#">Download Layout</a>
Kasese solgs trial	This trial was loaded into the fixture to test solgs.	test								<a href="#">Download Layout</a>

Showing 1 to 2 of 2 entries

Previous 1 Next



If the field trial you want to collect tissue samples from is not in the above table, you can click the button highlighted below.

The screenshot shows a web-based application titled "Manage Tissue Samples". At the top left is a leaf icon. To its right is the title "Field Trial Tissue Samples" and a sub-instruction: "View and create tissue samples for field trials. Tissue samples come from a plant in a plot." Below this is a blue button labeled "Create tissue samples for field trial". The main area is titled "Field Trials With Tissue Samples". It includes a search bar and a table with the following columns: Trial name, Description, Breeding program, Folder, Year, Location, Trial type, Design, Planting Date, Harvest Date, and Download. Two entries are listed:

Trial name	Description	Breeding program	Folder	Year	Location	Trial type	Design	Planting Date	Harvest Date	Download
CASS_8Genotypes_Sampling_2015	Copy of trial with postcomposed phenotypes from cassbase.	test								<button>Download Layout</button>
Kasese solgs trial	This trial was loaded into the fixture to test solgs.	test								<button>Download Layout</button>

At the bottom of the table area, it says "Showing 1 to 2 of 2 entries". There are "Previous" and "Next" buttons. The entire interface has a light gray background with white boxes for the tables and forms.

Once you have clicked this button, you will enter a workflow that begins with the following introduction.

The screenshot shows a modal window titled "Create Tissue Samples for a Field Trial". It features a horizontal navigation bar with four steps: "Intro" (highlighted with a green circle), "Select a field trial", "Plant Entries", and "Create Tissue Sample Entries". Below the navigation is a yellow box containing the text: "This workflow will guide you through creating tissue samples for your field trial". Underneath this, there is descriptive text: "Tissue samples are linked to a single plant, which is in turn linked to a single plot.", "Many tissue samples can be created for each plant.", "Each tissue sample needs a globally unique name.", and "Tissue samples can then be transferred into genotyping trials (96 or 384 well plates)". At the bottom of the modal is a blue button labeled "Go to Next Step" and a "Close" button in the bottom right corner.

Once you click next, you will need to select your trial.

Create Tissue Samples for a Field Trial

Intro 1 Select a field trial 2 Plant Entries 3 Create Tissue Sample Entries 4

Select a field trial

Select	Trial name	Description	Breeding program	Folder	Year	Location	Trial type	Design	Planting Date	Harvest Date	Download
<input checked="" type="checkbox"/>	CASS_6Genotypes_Sampling_2015	Copy of trial with postcomposed phenotypes from cassbase.	test		2017	test_location	Preliminary Yield Trial	RCBD			<button>Download Plot List</button>
<input type="checkbox"/>	Kasese soils trial	This trial was loaded into the fixture to test soils.	test		2014	test_location	Clonal Evaluation	Alpha			<button>Download Plot List</button>
<input type="checkbox"/>	PVA20	asd	test		2018	Cornell Biotech	Seedling Nursery	RCBD			<button>Download Plot List</button>
<input type="checkbox"/>	new_test_cross	new_test_cross	test								<button>Download Plot List</button>
<input type="checkbox"/>	selection_population	selection_population			2015						<button>Download Plot List</button>
<input type="checkbox"/>	test_genotyping_project	test_genotyping_project			2015						<button>Download Plot List</button>
<input type="checkbox"/>	test_population2	test_population2			2015						<button>Download Plot List</button>

Show 10 entries Search:

Close

Next, if your trial currently only has plot entries saved, you will be asked to enter how many plants are in each plot.

Create Tissue Samples for a Field Trial

Intro 1 Select a field trial 2 Plant Entries 3 Create Tissue Sample Entries 4

Plant entries in your field trial

Please create plant entries for this trial.

Number of plants per plot:

Inherits Management Factor(s) From Plots:

Submit

Close

Finally you will be asked how many tissue samples you want for each plant. You can specify a string to include in the tissue sample name, such as leaf or root.

Create Tissue Samples for a Field Trial

Intro 1 Select a field trial 2 Plant Entries 3 Create Tissue Sample Entries 4

Create tissue sample entries for this trial

① Number of tissue samples per plant: 3

② Tissue Name 1: leaf

③ Tissue Name 2: leaf

④ Tissue Name 3: stem

Inherits Management Factor(s) From Plots:

Submit

Close

Afterwards you should see the following success message, indicating that the tissue samples are saved.

Create Tissue Samples for a Field Trial

Intro 1 Select a field trial 2 Plant Entries 3 Create Tissue Sample Entries 4

Complete! Your field trial's tissue samples were saved.

② Tissue samples saved successfully

- You may want to go to the trial detail page for the trial now that it has plants.
- You can print barcodes for the new tissue samples.
- You can use these tissue samples as source material for a genotyping trial (96 or 384 well plate)

Close

## 18.2 Genotyping Plate Tissue Samples (96 or 384 well plates)

A genotyping plate represents a 96 or 384 well plate. You can use the Coordinate Android application to create your plate layout, or you can upload your own Excel plate layout, or you can use the database to generate a plate layout. Ideally, you will use tissue sample names originating from a field trial as the “source” for each well tissue sample, but you can also use plant names, plot names, or accession names.

From the manage tissue samples page, you can see the genotyping plates

## 18.2. GENOTYPING PLATE TISSUE SAMPLES (96 OR 384 WELL PLATES)201

saved in the database. You can also download the layouts as shown below.

The screenshot shows a web-based application for managing genotyping trials. At the top, there's a header with a grid icon and the title "Genotyping Trial Tissue Samples". Below the header, there are two buttons: "Create or upload a genotyping trial" and "Export to Genotyping Vendor". The main area is titled "Genotyping Trials" and contains a table with two entries:

Trial name	Description	Breeding program	Folder	Year	Location	Download
18DNA101	A 96 well DNA sequencing plate	test		2017	Cornell Biotech	<a href="#">Download Layout</a>
18Ngeno1	asd	test		2017	Cornell Biotech	<a href="#">Download Layout</a>

Below the table, it says "Showing 1 to 2 of 2 entries". On the right side of the table, there are "Previous", "1", and "Next" buttons. A red box highlights the "Download Layout" button for the first entry.

If you need to create a new genotyping plate, you can click the button shown below. This will guide you through a workflow for uploading or creating the new plate layout.

This screenshot is similar to the one above, but the "Create or upload a genotyping trial" button is highlighted with a red box. The rest of the interface, including the table of trials and the download buttons, remains the same.

Genotyping vendors require you to send a plate layout during submission. You can download the plate layout as shown above, or you can go to a genotyping plate detail page to download the Intertek formatted file.

In the future you will be able to directly export your genotyping plate plate layout to vendors.



# Chapter 19

## Managing Observation Variables

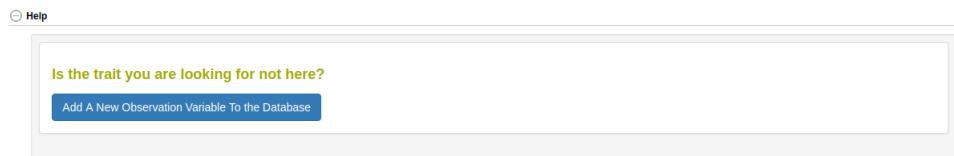
### 19.1 Managing Observation Variables with Traits, Methods, and Scales

Observation variables are the identifiers used when collecting phenotypic data. An observation variable is composed of a trait, a method, and a scale. The trait describes the attribute being measured e.g. ‘Plant Height’. The method defines the protocol in which the trait was observed e.g. ‘Using a one meter long measuring stick’. The scale defines the units or dimensions for which the measurement was taken e.g. ‘Meters’.

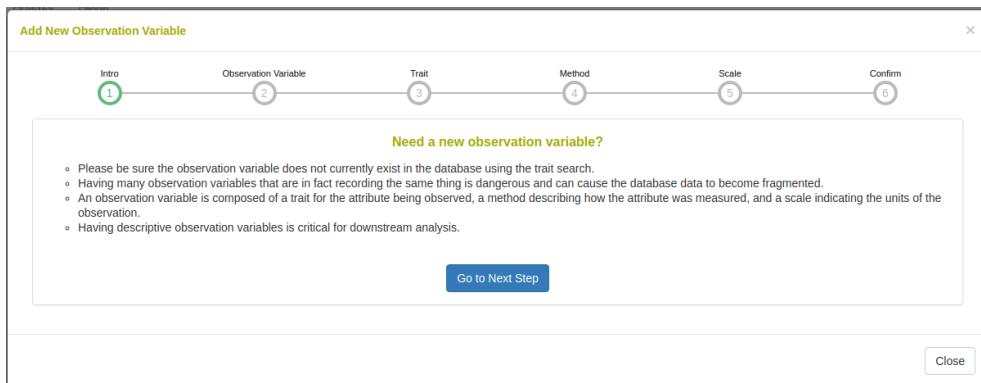
Generally, observation variables are defined in ontologies that are predefined. We often use ontologies from cropontology.org. In this case, you will not be able to define your own observation variables directly; instead, you will need to contact us and we will add the observation variable for you.

For databases where the user has greater control, we have an interface to allow addition of observation variables, along with traits, methods, and scales. To begin, go to the Search->Traits page.

If the database you are on allows you to directly add observation variables, you will see the following button at the bottom of the page.



When you click the button, the following workflow will appear. You should be logged in or else it will not allow addition of the observation variable. The workflow begins with an introduction.



On the next workflow step, you select the ontology that you want to insert the new observation variable into. You must also give a name and a definition for the new observation variable.

A screenshot of the 'Add New Observation Variable' workflow, step 2. The title bar says 'Add New Observation Variable'. The timeline shows steps 1 through 6. Step 2 is highlighted with a green circle. A note box says 'Define your observation variable' and provides a description: 'An observation variable is composed of a trait for the attribute being observed, a method describing how the attribute was measured, and a scale indicating the units of the observation.' Below are input fields:

- Observation Variable Ontology Name:** Two checkboxes:  COMP (Composed traits) and  ISOL (ISOL).
- New Observation Variable Name:** Input field: 'Plant height in meters extracted from RGB image using ImageJ'.
- New Observation Variable Definition:** Input field: 'Plant height in meters extracted from an red green blue (RGB) image using ImageJ'.

At the bottom are 'Go to Next Step' and 'Close' buttons.

On the next workflow step, you select the trait ontology to use. Once you

## 19.1. MANAGING OBSERVATION VARIABLES WITH TRAITS, METHODS, AND SCALES205

select a trait ontology, a select containing all the terms in the selected ontology will appear. You can either select a trait or if it does not exist in the select, you can create a new one by giving a name and a definition for the new trait.

Add New Observation Variable

Intro      Observation Variable      Trait      Method      Scale      Confirm

Define your trait

A trait defines the attribute being measured. It is one component of an observation variable; the others are a method and a scale.  
Here you can select a trait that already exists in an ontology in the database or you can add a new trait into an ontology.

Trait Ontology Name:

COMP (Composed traits)  
 ISOL (ISOL)

Existing Traits in Selected Ontology: None

If the trait does not exist in the ontology you selected above, you can add a new trait into the ontology here.

New Trait Name: Plant height from base of plant to highest point on branch

New Trait Definition: The height of a plant from the base of the plant to the highest possible point on a branch

Go to Next Step

Close

On the next workflow step, you select the method ontology to use. Once you select a method ontology, a select containing all the terms in the selected ontology will appear. You can either select a method or if it does not exist in the select, you can create a new one by giving a name and a definition for the new method.

Add New Observation Variable

A method defines the how it was measured. It is one component of an observation variable; the others are a trait and a scale.

Method Ontology Name:  CASSTISS (cass\_tissues)

Existing Methods in Selected Ontology: None

If the method does not exist in the ontology you selected above, you can add a new method into the ontology here.

New Method Name: ImageJ for plant height extraction from RGB image

New Method Definition: A script in [ImageJ](#) for extracting plant height from a red, green, blue (RGB) image

Go to Next Step

Close

On the next workflow step, you select the scale ontology to use. Once you select a scale ontology, a select containing all the terms in the selected ontology will appear. You can either select a scale or if it does not exist in the select, you can create a new one by giving a name and a definition for the new scale. You can also define a format, minimum, maximum, categories, and default value for the new scale.

## 19.1. MANAGING OBSERVATION VARIABLES WITH TRAITS, METHODS, AND SCALES207

Add New Observation Variable

Intro 1 Observation Variable 2 Trait 3 Method 4 Scale 5 Confirm 6

Define your scale

Scale Ontology Name:  UO (Units)

Existing Scales in Selected Ontology: m|UO:0300001

If the scale does not exist in the ontology you selected above, you can add a new scale into the ontology here.

New Scale Name: e.g. mass in kilograms

New Scale Definition: e.g. the mass in kilograms

New Scale Format: Select One

New Scale Minimum: e.g. 1

New Scale Maximum: e.g. 10000

New Scale Categories ('/' separated): e.g. 1/3/5/7

New Scale Default Value: e.g. 5

Go to Next Step

Close

On the last page of the workflow, you confirm the submission.

Add New Observation Variable

Intro 1 Observation Variable 2 Trait 3 Method 4 Scale 5 Confirm 6

Confirm you are creating this observation variable in the database

Submit

Close

Afterwards, you can use the newly created observation variable ontology term in your phenotyping.



# Chapter 20

## Managing Image Data

### 20.1 Image-Phenotyping Dashboard

1. Upload raw image-captures in a compressed file (.zip) for orthophoto-mosaic assembly or upload previously stitched orthophotomosaic raster (.PNG, .JPG) imagery.
2. Dashboard shows all field trials and uploaded imaging events in collapsible sections.
3. Follow standard processes to manually create templates for assignment of plot-polygon images to the field experiment design.
4. All imagery is shown with the spectral category within collapsible sections. Figure shows NIR imagery.
5. Apply Fourier transform filtering, thresholding, and vegetation index masking. Plot-polygon images for all image processes are shown.
6. Extract and export phenotypic values from plot-polygon images for analyses and model training.

### 20.2 Image Input

Clicking “Upload Imagery” will open the following dialog.

This workflow will guide you through uploading aerial images to the database

Your field trial must already be in the database before you can upload images for it. Please go to [Manage->Field Trials](#) if it is not.

A field trial represents plots in the field where each plot has a globally unique *plot\_name*, a sequential *plot\_number* that is unique in the trial (e.g. 101, 102, 103 for three separate plots), and an *accession\_name* representing the genotype being tested in that plot. Each plot can belong to different blocks (*block\_number*) and reps (*rep\_number*) depending on the experimental design you are using (e.g. complete block vs augmented design). Each plot can have a *row\_number* and *col\_number* indicating the relative position of the plot in the field. A field trial can represent a yield trial, a phenotyping trial, a crossing block, a greenhouse, a nursery, etc.

If you have raw aerial images that have not been stitched into an orthophotomosaic image of the whole field, your raw images should be uploaded using a zipfile (.zip). You can have several drone runs for a single field trial. For an individual drone run, once you have uploaded all photos, you can stitch an orthophotomosaic together. Afterwards, you will have options to cut the ortho image into plot polygons and extract phenotypes for those plots into the database. The maximum zipfile size is 2GB.

If you already have an orthophotomosaic image of your entire field, you can upload that image under a field trial and a drone run. Afterwards, you will have options to cut the ortho-image into plot polygons and extract phenotypes for those plots into the database. The maximum size for each image is 200MB. The preferred upload format is PNG.

**Example Data:** Micasense 5 Band Raw Images (Unstitched image-captures) (Upload zipfile for ImageBreed to stitch.)  
**Example Data:** Micasense 5 Band Panel Images (Micasense calibration panel images.) (Upload zipfile for ImageBreed to calibrate Micasense raw-captures during stitching.)  
**Example Data:** Micasense 5 Band Previously Stitched Orthophotomosaic Images (PNG Files in provided zipfile. Can upload each band separately into ImageBreed.)

[Go to Next Step](#)

[Close](#)

Raw-captures can be uploaded in a compressed (.zip) file so that they can be assembled into an orthophotomosaic. If orthophotomosaic assembly is not required, raster images (.PNG, .JPG) can be uploaded. Example data is given for raw Micasense RedEdge 5-band multispectral captures and for stitched orthophotomosaics.

Select your field trial

Field Trial: [2015\\_NYH2](#)

[Go to Next Step](#)

[Close](#)

To begin uploading images, a field trial must be selected. The field trial must already be saved in the database. For information about adding a field trial, please read the Field Trial documentation.

**Upload Drone Imagery**

Select or create new drone run

Select	Imaging Event Name	Imaging Event Type	Imaging Event Description	Imaging Event Date	Camera	Field Trial Name	Field Trial Description
<input checked="" type="checkbox"/>	2015_NYH2_07212015	Aerial Medium to High Res	Orthos from Nick Kaczmar from Pix4d	2015-July-21	micasense_5	2015_NYH2	G2F NYH2 2015

Show 10 entries Search:

Showing 1 to 1 of 1 entries Previous  Next

[Create new drone run if not present in table](#)

Imaging Event Name:

Imaging Event Type:

Camera Type:

Imaging Event Description:

Imaging Event Date:

[Go to Next Step](#)

[Close](#)

The image data is added to an imaging (drone run) event. Here you can select a previously saved imaging event or you can create a new one by defining a name, description, and date.

**Upload Drone Imagery**

**Stitched vs Unstitched and Number of Bands (Image Sets) To Upload**

- Raw images (unstitched) coming from your drone can be uploaded in a zip file. We can then stitch them together into an orthophotomosaic of the entire drone run.
- Or you can choose to upload a single image and skip any stitching

- It is possible to upload regular RGB or Black and White photos.
- For multi-spectral cameras, it is possible to upload individual spectra orthomosaicphotos.
- When uploading many separate bands of unstitched images, you will upload a single zipfile (.zip) which contains all images. In the zipfile each image is named following the template IMG\_0001\_1.tif, IMG\_0001\_2.tif, ..., IMG\_0001\_5.tif, ..., IMG\_9999\_5.tif. The final number represents the 5 bands coming from the camera, while the middle number is an index for the image capture. The middle number can be as many digits long as needed. The images should be in order in the zipfile. You will also need to upload a zipfile (.zip) containing the Micasense radiometric calibration panel images, so that ImageBreed can produce the best orthomosaic possible.

Do you require stitching an ortho image of the drone run:

Number of Spectral Bands (Image Sets) To Upload:

[Go to Next Step](#)

[Close](#)

The uploaded data can be raw image-captures or complete raster images. Here you can select whether orthophotomosaic stitching is required.

**Stitched vs Unstitched and Number of Bands (Image Sets) To Upload**

- Raw images (unstitched) coming from your drone can be uploaded in a zip file. We can then stitch them together into an orthophotomosaic of the entire drone run.
- Or you can choose to upload a single image and skip any stitching.

- It is possible to upload regular RGB or Black and White photos.
- For multi-spectral cameras, it is possible to upload individual spectra orthomosaicphotos.
- When uploading many separate bands of unstitched images, you will upload a single zipfile (.zip) which contains all images. In the zipfile each image is named following the template IMG\_0001\_1.tif, IMG\_0001\_2.tif, ..., IMG\_0001\_5.tif, ..., IMG\_9999\_5.tif. The final number represents the 5 bands coming from the camera, while the middle number is an index for the image capture. The middle number can be as many digits long as needed. The images should be in order in the zipfile. You will also need to upload a zipfile (.zip) containing the Micasense radiometric calibration panel images, so that ImageBreed can produce the best orthomosaic possible.

**Do you require stitching an ortho image of the drone run:**

Yes, I am uploading a zipfile of images to stitch

**Go to Next Step**

In the case that orthophotomosaic stitching is required, select ‘yes’. On the next step you will see the following: Upload a zipfile with the raw-captures. When uploading Micasense RedEdge raw-captures, provide images of the Micasense calibration panels in a zipfile as well.

**Select Image(s) to Upload**

Drone Images ZipFile (.zip) (2GB Maximum):  No file chosen

Micasense Radiometric Calibration Images ZipFile (.zip):  No file chosen

Working Image Scale (Megapixels):  0.6

**Submit**

In the case that orthophotomosaic assembly is not required, simple upload the raster images. Select the number of image bands that will be uploaded

e.g. for a five band multispectral camera, select 5.

In the cases that orthophotomosaic stitching is not required, select ‘no’. On the next step you will see the following:

Upload an image at each band with a unique name, description, and spectral type.

## 20.3 Standard Process

Once imagery is uploaded, it will appear on the dashboard under the field trial. Clicking the “Run Standard Process” button will begin extracting plot-polygon phenotypes from the imagery.

Manage Drone Imagery

Drone Imagery

Upload Imagery Download Image-Phenotypes Calculate Statistics

Search:

Field Trials > Imaging Events

**Field Trial: 2015\_NYH2 (2 Imaging Events)**

**2015\_NYH2\_07212015 2015-July-21**

**2015\_NYH2\_08072015 2015-August-07**

Imaging Event Name: 2015\_NYH2\_08072015  
Imaging Event Type: Aerial Medium to High Res Orthos from Nick Kaczmar from Pix4d  
Description:  
Date: 2015-August-07  
Field Trial: 2015\_NYH2

Run Standard Process For 2015\_NYH2\_08072015

Delete Imaging Event

No Plot Images Saved

Image Band(s)

Name: 2015_NYH2_08072015_Blue	Images/Actions
Description: Ortho from Nick Kaczmar from Pix4d Type: Blue (450-520nm)	<a href="#">View Images</a>
Name: 2015_NYH2_08072015_Green	<a href="#">View Images</a>
Description: Ortho from Nick Kaczmar from Pix4d Type: Green (515-600nm)	<a href="#">View Images</a>
Name: 2015_NYH2_08072015_Red	<a href="#">View Images</a>
Description: Ortho from Nick Kaczmar from Pix4d Type: Red (600-690nm)	<a href="#">View Images</a>
Name: 2015_NYH2_08072015_NIR	<a href="#">View Images</a>
Description: Ortho from Nick Kaczmar from Pix4d Type: NIR (780-3000nm)	<a href="#">View Images</a>
Name: 2015_NYH2_08072015_RedEdge	<a href="#">View Images</a>
Description: Ortho from Nick Kaczmar from Pix4d Type: Red Edge (690-750nm)	<a href="#">View Images</a>

Clicking the button will open the following dialog.

Manage Drone Imagery: Run A Standard Process

Intro 1 Drone Run Band 2 Rotate 3 Cropping 4 Thresholding 5 Plot Polygons 6 Apply 7 Indices 8 Phenotypes 9

This workflow will guide you through applying a standard process to your aerial imaging bands

Here you can take one of the drone run bands you uploaded all the way through the process to plot image saving. This will require manual steps such as image rotation, cropping, and plot polygon templating. After you have completed this process for one drone run band, you can apply it to all other drone run bands and other calculated vegetative indices.

Go to Next Step

Select a drone run band to use in this process. In the case of the Micasense 5 band multispectral camera there will be 5 bands shown here; select the NIR channel in this case because it has the highest contrast. In the case of standard color images, there will only be the RGB Color Image option here.

**Manage Drone Imagery: Run A Standard Process**

**Select a drone run band**

Please select one drone run band to take through the process. It is recommended to select a band that has high contrast, such as a NIR band.

Show 10 entries									Search: <input type="text"/>	
Select	Drone Run Band Name	Drone Run Band Description	Drone Run Band Type	Drone Run Name	Drone Run Description	Drone Run Date	Field Trial Name	Field Trial Description		
<input type="checkbox"/>	2015_NYH2_08072015_Blue	Ortho from Nick Kaczmar from Pix4d	Blue (450-520nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015		
<input type="checkbox"/>	2015_NYH2_08072015_Green	Ortho from Nick Kaczmar from Pix4d	Green (515-600nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015		
<input type="checkbox"/>	2015_NYH2_08072015_Red	Ortho from Nick Kaczmar from Pix4d	Red (600-690nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015		
<input type="checkbox"/>	2015_NYH2_08072015_NIR	Ortho from Nick Kaczmar from Pix4d	NIR (780-3000nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015		
<input type="checkbox"/>	2015_NYH2_08072015_RedEdge	Ortho from Nick Kaczmar from Pix4d	Red Edge (690-750nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015		

Showing 1 to 5 of 5 entries

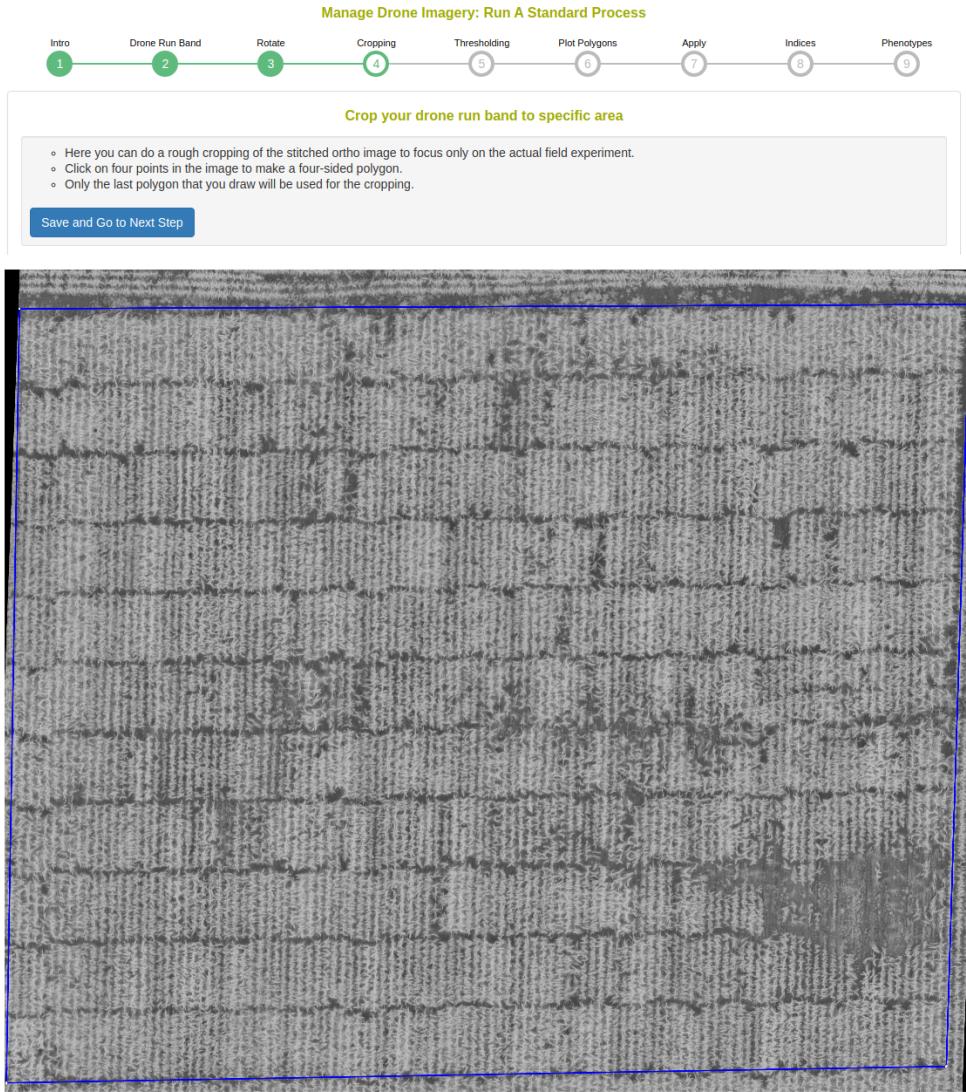
Previous 1 Next

[Go to Next Step](#)

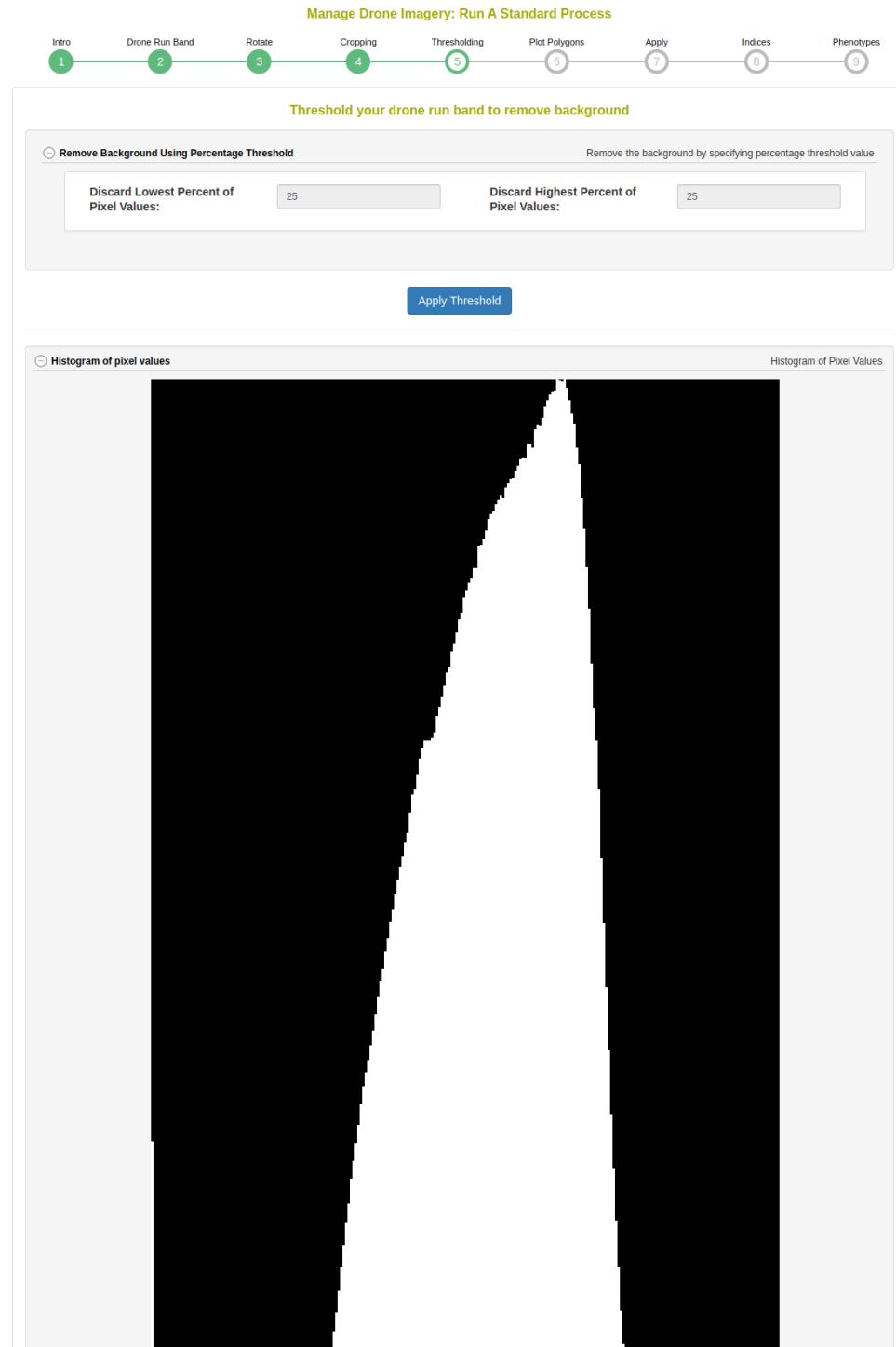
Rotate the image so that there the plots are oriented in a grid fashion. There can be a skew in the field layout, as seen in the following example.



Perform a rough cropping of the image by clicking on the four corners of the field. Cropping is important to remove any extraneous parts of the image.



This step shows a histogram of the cropped image. The standard process will magnitude threshold the top and low ends of the distribution.



In this step, the template for the plot polygons in the experimental field design are associated to the image. First, defined the number of rows and columns in the field experiment. Then click the four corners of the image, in respect to the top right, top left, bottom left, and bottom right positions. Next click on “Draw Plot Polygon Template”. Review the template and clear/repeat the process until the template matches well. It is possible to “copy/paste” templates in the case where there are large breaks in the field design. Next, scroll down to the “assign Plot Polygons to Field Trial Entities” section. Select the location of Plot Number 1 as either “top left” or “top right” and whether the field design is serpentine or zigzag. Click on “Generate Assignments” and review that the names of the plots appear correctly in the overlay on the image. Finally, click “Finish and Save Polygons to Plots” when you have confirmed the assignments.

**Manage Drone Imagery: Run A Standard Process**

Intro    1    Done Run Band    2    Rotate    3    Cropping    4    Thresholding    5    Plot Polygons    Apply    Indices    Phenotypes

Define plot polygons relative to the field layout

1    **Generate Polygon Template Tool**

Number of Rows: 10    Number of Columns: 50

Click Top Left Corner    Click Top Right Corner    Click Bottom Left Corner    Click Bottom Right Corner

Template Parameters

2    Draw Plot Polygon Template (Does not save. Apply multiple templates if needed.)

Previously Used Plot Polygon Templates

Total Image Width: 3316px. Total Image Height: 2680px.

Clear All Polygons    Clear One Polygon    Find Distance Between Points



Assign Plot Polygons to Field Trial Entities

Location of First Plot (e.g. plot number): 1<sup>3</sup>    Second Plot Follows First Plot Going: Right

Plot Number Orientation: Serrpentine<sup>4</sup>

4    5    Generate Assignments (Does Not Save)    Finish and Save Polygons To Plots

plot_name	accession_name	plot_number	block_number	is_a_control	rep_number	row_number	col_number	plot_geo_json	Polygon Assigned
2015_NYH2_plot_1	PHV63LH195	1	1	null	1	1	45		
2015_NYH2_plot_10	CG1096PHZ51	10	1	null	1	1	54		
2015_NYH2_plot_100	LH45UH82	100	4	null	1	2	45		
2015_NYH2_plot_101	PHN11_LH145_0002PHB47	101	5	null	1	3	45		
2015_NYH2_plot_102	NYH-081PHB47	102	5	null	1	3	46		
2015_NYH2_plot_103	W10005_0029PHB47	103	5	null	1	3	47		
2015_NYH2_plot_104	B109PHZ51	104	5	null	1	3	48		
2015_NYH2_plot_105	PH207_PHG47-13PHB47	105	5	null	1	3	49		
2015_NYH2_plot_106	PH207_PHG47-10PHB47	106	5	null	1	3	50		
2015_NYH2_plot_107	CG045LUH82	107	5	null	1	3	51		

Show 10 entries    Search: [ ]

Showing 1 to 10 of 500 entries    Previous    1    2    3    4    5    ...    50    Next

Next, the dialog shows you that the standard process will be repeated for all uploaded image bands.

**Apply these same steps to other drone run bands in the current drone run**

Here you can apply the same actions you did for the previous steps 1 to 6, to additional drone run bands in this drone run.  
Thresholding will be done dynamically, by removing the top and bottom 20% of pixel values.

Select	Drone Run Band Name	Drone Run Band Description	Drone Run Band Type	Drone Run Name	Drone Run Description	Drone Run Date	Field Trial Name	Field Trial Description
<input checked="" type="checkbox"/>	2015_NYH2_08072015_Blue	Ortho from Nick Kaczmar from Pix4d	Blue (450-520nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015
<input checked="" type="checkbox"/>	2015_NYH2_08072015_Green	Ortho from Nick Kaczmar from Pix4d	Green (515-600nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015
<input checked="" type="checkbox"/>	2015_NYH2_08072015_Red	Ortho from Nick Kaczmar from Pix4d	Red (600-690nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015
<input checked="" type="checkbox"/>	2015_NYH2_08072015_NIR	Ortho from Nick Kaczmar from Pix4d	NIR (780-3000nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015
<input checked="" type="checkbox"/>	2015_NYH2_08072015_RedEdge	Ortho from Nick Kaczmar from Pix4d	Red Edge (690-750nm)	2015_NYH2_08072015	Orthos from Nick Kaczmar from Pix4d	2015-August-07	2015_NYH2	G2F NYH2 2015

Showing 1 to 5 of 5 entries

Previous 1 Next

**Go to Next Step**

Next, choose which vegetation indices to apply.

**Manage Drone Imagery: Run A Standard Process**

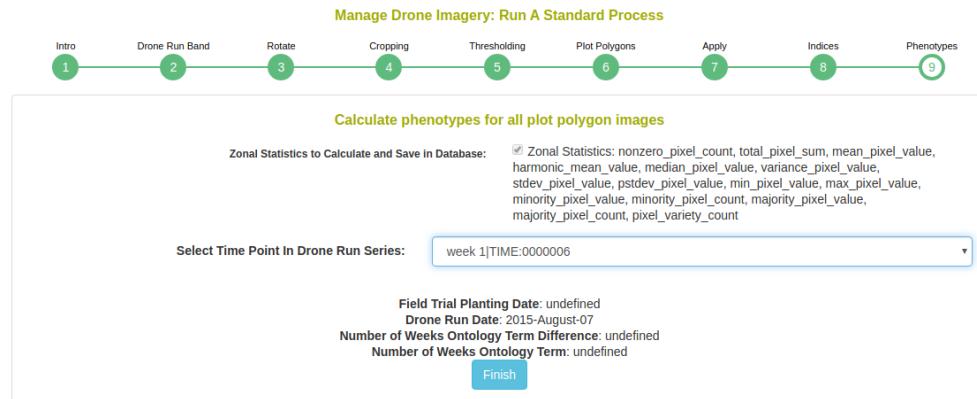
**Create and apply these same steps to vegetative indices**

Vegetative Indices To Apply:

- Triangular Greenness Index (TGI)
- Visible Atmospheric Resistant Index (VARI)
- Normalized Difference Vegetative Index (NDVI)
- Normalized Difference Red Edge Vegetative Index (NDRE)

**Go to Next Step**

Next, choose the phenotypic values to extract. You must define the time point for which the phenotype is; if the field trial has a planting date, the time point will automatically be populated as image date minus the planting date.



After completing the standard process, the job will continue in the background until it completes. You can check the status of the job from the dashboard.

## 20.4 Ground Control Points

Ground control points can be saved after an imaging event has undergone the standard process on orhomosaics. Ground control points can then be used across imaging events on the same field experiment in order to automate the entire standard process.

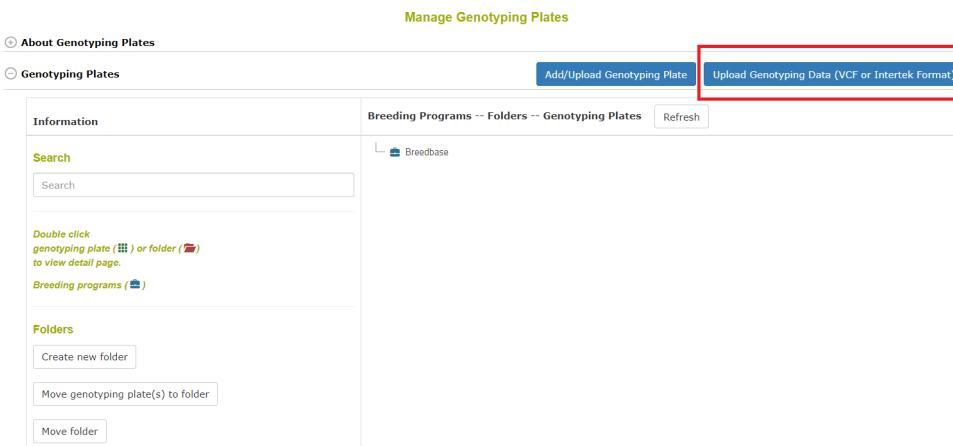
# Chapter 21

## Managing VCF Data

### 21.1 Uploading VCF Data

Genotyping data in VCF can be loaded from the web-interface. Breedbase can store any genotypic variants from a VCF, allowing for polyploids, structural variants, etc. without problems.

To begin go to Manage->Genotyping Plates and click the button seen below: Note that you do not need to have genotyping plates uploaded to upload VCF data; you may upload genotyping data to accessions or you can upload genotyping data for tissue samples in genotyping plates.

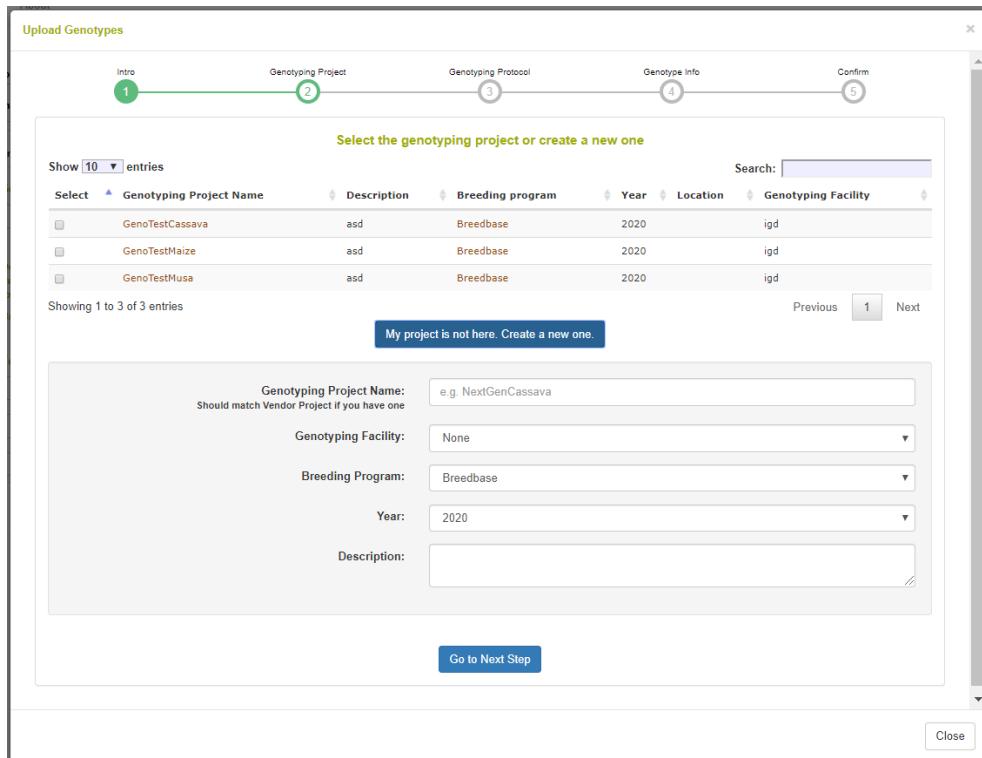


The screenshot shows the 'Manage Genotyping Plates' page. At the top, there are two tabs: 'About Genotyping Plates' and 'Genotyping Plates'. Below the tabs, there is a search bar labeled 'Search' and a note: 'Double click genotyping plate (grid) or folder (box) to view detail page.' There is also a link to 'Breeding programs (list)'. On the right side, there is a tree view showing a single node 'Breedbase'. At the bottom of the page, there are three buttons: 'Add/Upload Genotyping Plate', 'Upload Genotyping Data (VCF or Intertek Format)', and 'Refresh'. The 'Upload Genotyping Data (VCF or Intertek Format)' button is highlighted with a red box.

The workflow begins with an intro:



On the following step in the workflow, a genotyping project is defined or selected. A genotyping project is a high-level entity for grouping several genotyping events. It is defined with a name, description, name, breeding program, and genotyping facility (IGD, Intertek, etc.).



The following step is to define or select a genotyping protocol. A genotyping protocol represents the set of markers being called against a specific reference genome. A genotyping protocol is defined with a name, description, reference genome name, species name, and a location of data generation. Note in the picture that you can select whether the samples in your file are accessions or tissue samples in the database; tissue samples are for when a genotyping plate is stored in the database. There is an option to parse the sample names for appended sequencing numbers from IGD, where the sample names are like “accession:igdnumber”.

**Upload Genotypes**

Showing 1 to 3 of 3 entries

My protocol is not here. Create a new one.

Genotyping Protocol Name: e.g. GBS ApeKI Cassava genome v6 Jan2015

Genotyping Protocol Reference Genome: Mesculenta\_511\_v7.0

Species: e.g. Manihot esculenta

Description:

Choose Sample Unit:

Exported Tissue Sample Name: The sample names in your VCF are tissue\_sample\_names that already exist in genotyping plates (e.g. 96 well plates) in the database. The sample names in your VCF file can be the tissue\_sample\_name triple pipe joined to the accession\_name (e.g. tissue\_sample\_name||accession\_name) or just simply the tissue\_sample\_name corresponding to the genotyping plate well.

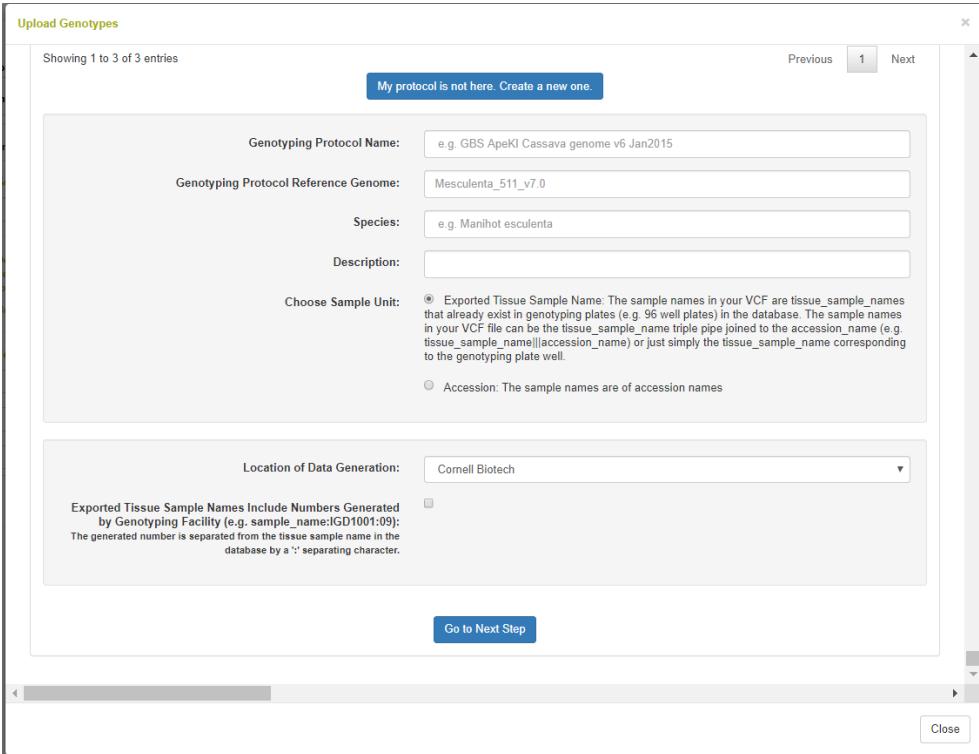
Accession: The sample names are of accession names

Location of Data Generation: Cornell Biotech

Exported Tissue Sample Names Include Numbers Generated by Genotyping Facility (e.g. sample\_name:IGD1001:09); The generated number is separated from the tissue sample name in the database by a ‘:’ separating character.

Go to Next Step

Close



The final step is to select the VCF from your computer and upload it. The web interface can be used to upload files arbitrarily large; it is a NGINX configuration to set this size.

The screenshot shows a software interface titled 'Upload Genotypes' with a progress bar at the top. The steps are numbered 1 through 5: 1. Intro, 2. Genotyping Project, 3. Genotyping Protocol, 4. Genotype Info, and 5. Confirm. Step 4 is currently active, indicated by a green circle and a green line under the number. The main area is titled 'Provide genotype information'. It contains a dropdown menu 'Select type of genotyping data:' with 'VCF' selected. Below it is a section for 'File format information' with 'VCF format' and an 'i' icon. To the right is a file selection field 'Select VCF File:' with 'Choose File | No file chosen'. At the bottom are 'Go to Next Step' and 'Close' buttons.

## 21.2 Searching and Downloading VCF Data

The Search Wizard is the primary means of querying data in the database. Go to Search->Wizard to begin.

Once genotyping protocols are stored, select Genotyping Protocols from the first dropdown menu. Then if you select one or more and select Accessions from the second dropdown menu, you will see the accessions for which genotypes were stored. As seen in the following picture, there is a section for filtering genotypes by chromosome, start position, and end position. Genotypes can be downloaded in VCF or DosageMatrix formats.

**Search Wizard**

Don't see your data? Refresh Lists | Update Wizard

**Genotyping Protocols**

**Accessions**

**Select Column Type**

**Select Column Type**

**Search**

**Search**

**Search**

**Search**

**Load/Create Datasets using Match Columns**

**Related Genotype Data**

To download related genotype data, select **1 or more Accessions** and **no more than 1 Genotyping Protocol** in the wizard. Optionally, select a Chromosome and enter a position range below. If no genotyping protocol is selected, the database default protocol will be used.

3 accessions, selected protocol

Chromosome	Start Position	End Position
All		

VCF

Related Trial Metadata

Related Trial Phenotypes

Using the “Default genotyping protocol” which is configured in a system, you can query over field phenotypic evaluations before downloading genotypes and phenotypes.



## 21.3 Searching Protocols

Genotyping protocols can be search by going to Search->Genotyping Protocols. To download genotypes accessions must be selected, though any combination of search criteria can be used to filter and select those accessions. If a genotyping protocol is not selected, then the default genotyping protocol set in the configuration will be used. Genotyping protocols can also be selected in the wizard.

The screenshot shows the 'Search Wizard' interface with four search panels:

- Traits:** Shows selected traits like 'abscisic acid content of leaf ug/g(CO\_33)' and 'amylopectin content ug/g in percentage()'.
- Accessions:** Shows selected accessions like '4N509/3IH6' and '6F629/3IH6'.
- Trials:** Shows selected trials like '2018\_NYH2'.
- Locations:** Shows selected locations like 'M3 Aurora Musgrave Research Farm' and 'P1 Aurora Musgrave Research Farm'.

Each panel has a 'Select All' button and a 'Clear' button. Below each panel are 'Match ANY ALL' buttons and 'Add' buttons for adding to a list. At the bottom are 'Create New List...' and 'Create' buttons.

The genotyping download menu on the Search Wizard presents options for filtering by chromosome, start position, and end position. Genotypes can be downloaded in VCF or Dosage Matrix formats. The genomic relationship matrix (GRM) can be downloaded for the selected accessions in a tab-delimited matrix format or in a three-column format that is useful in Asreml. Genotypes can be computed from the parents in the pedigree if those parents are genotyped by clicking on the “compute from parents” checkbox. Additionally, the GRM can be computed using genotypes of parents in the pedigree if the “compute from parents” checkbox is selected.

Related Genotype Data

To download related genotype data, select **1 or more Accessions** and **no more than 1 Genotyping Protocol** in the wizard. Optionally, select a Chromosome and enter a position range below. If no genotyping protocol is selected, the database default protocol will be used.

5 accessions, default protocol

Chromosome	Start Position	End Position	Compute From Parents
All			<input checked="" type="checkbox"/>

Dosage Matrix (tsv)

[Download Genotypes](#)

3-Column Format (tsv)

[Download Genetic Relationship Matrix \(GRM\)](#)

As is described elsewhere, the Search Wizard presents a way to filter phenotypic values by minimum and maximum values, and allow for download in CSV and Excel formats.

**Related Trial Phenotypes**

4 trials

CSV Plots

Include timestamps  Suppress user defined phenotype outliers

Trait Name Contains	Min Value	Max Value
	-∞	∞

**Phenotypes**

## 21.4 Detail Pages and Deletion

The genotyping protocol detail page will show all information about the protocol such as the reference genome used, the header information lines in the uploaded VCF file, the markers involved, and the samples genotyped.

The markers section will show all markers used and their annotations, such as position, chromosome, alternate allele, reference allele, marker format, etc.

**Markers** View information about the markers used in this protocol.

Marker Name(s):  Search

Show 10 entries

Marker Name	Chromosome	Position	Alternate	Reference	Quality	Filter	Info	Format
S0_1000880	0	1000880	T	C	.	PASS	.	GT
S0_1000890	0	1000890	.	G	.	PASS	.	GT
S0_1000912	0	1000912	.	C	.	PASS	.	GT
S0_1000916	0	1000916	.	C	.	PASS	.	GT
S0_1000922	0	1000922	.	C	.	PASS	.	GT
S0_1000924	0	1000924	G	A	.	PASS	.	GT
S0_101126	0	101126	.	G	.	PASS	.	GT
S0_1027188	0	1027188	.	T	.	PASS	.	GT
S0_1152731	0	1152731	.	C	.	PASS	.	GT

Showing 1 to 10 of 955,690 entries Previous 1 2 3 4 5 ... 95569 Next

The samples section will show all samples genotyped. Notice the Download links in the table which can be used to easily get the VCF file results for each

genotyped samples with all markers in the genotyping protocol. For getting multiple samples at once, use the Search Wizard as discussed above.



The screenshot shows a table titled "Genotype Data" with the sub-header "View and download genotyping data from this protocol." The table has columns: Protocol, Sample Name, Sample Type, Accession Name, Synonyms, Description, Number of Marker Scores, IGD Number, and Download. There are 10 entries shown, with a total of 1,577 entries. The entries are as follows:

Protocol	Sample Name	Sample Type	Accession Name	Synonyms	Description	Number of Marker Scores	IGD Number	Download
GenoProtMaize	554353-1-1-B	accession	554353-1-1-B		SNP genotypes for stock (name = 554353-1-1-B, id = 41812)	955690	100000044	<a href="#">Download</a>
GenoProtMaize	554353-1-1-B	accession	554353-1-1-B		SNP genotypes for stock (name = 554353-1-1-B, id = 41812)	955690	100000101	<a href="#">Download</a>
GenoProtMaize	554360-1-1-B	accession	554360-1-1-B		SNP genotypes for stock (name = 554360-1-1-B, id = 41813)	955690	100000106	<a href="#">Download</a>
GenoProtMaize	554363-1-1-B	accession	554363-1-1-B		SNP genotypes for stock (name = 554363-1-1-B, id = 41814)	955690	100000107	<a href="#">Download</a>
GenoProtMaize	554371-1-1-B	accession	554371-1-1-B		SNP genotypes for stock (name = 554371-1-1-B, id = 41815)	955690	100000113	<a href="#">Download</a>
GenoProtMaize	554372-1-1-B	accession	554372-1-1-B		SNP genotypes for stock (name = 554372-1-1-B, id = 41816)	955690	100000108	<a href="#">Download</a>
GenoProtMaize	554372-1-1-B	accession	554372-1-1-B		SNP genotypes for stock (name = 554372-1-1-B, id = 41816)	955690	100000460	<a href="#">Download</a>
GenoProtMaize	6F629	accession	6F629		SNP genotypes for stock (name = 6F629, id = 41817)	955690	100000797	<a href="#">Download</a>
GenoProtMaize	8M129	accession	8M129		SNP genotypes for stock (name = 8M129, id = 41818)	955690	100000153	<a href="#">Download</a>
GenoProtMaize	8M129	accession	8M129		SNP genotypes for stock (name = 8M129, id = 41818)	955690	100000450	<a href="#">Download</a>

Showing 31 to 40 of 1,577 entries

Previous 1 2 3 4 5 ... 158 Next

The genotyping protocol and all associated genotyping data can be deleted from the genotyping protocol page.



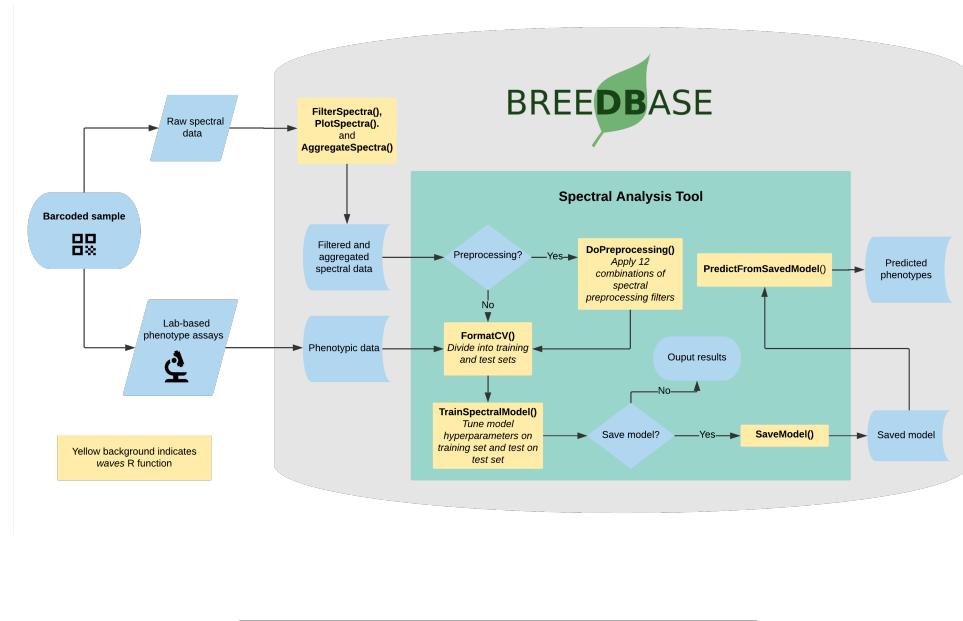
The screenshot shows a confirmation dialog with the title "Delete Genotyping Protocol and All Data". It contains a message "Delete genotyping protocol and all data from this protocol." and a large red button labeled "Delete Genotyping Data For this Protocol".



# Chapter 22

## Managing Spectral Data

Breedbase has implemented a flexible spectral data storage protocol that handles spectral data irrespective of the source spectrometer. Spectral data storage and analysis in Breedbase makes use of the R package *waves* for outlier identification, plotting, sample aggregation, and prediction model training.



## 22.1 Upload Spectral Data

Spectral data can be added as a CSV file that includes metadata in the leftmost columns followed by one column per spectral measurement to the right. Rows represent a single scan or sample, each with a unique ID that must match to a Breedbase observationUnitName. Future data transfer using BrAPI will allow for interoperability with data collection software.

To upload a spectral dataset, navigate to the ‘Manage NIRS Data’ page by selecting ‘NIRS’ in the ‘Manage’ menu and click the blue ‘Upload NIRS’ button. This will open an upload workflow. A link to the required file format and an example .csv file can be found by clicking in the light blue info box in this workflow. Another example of the file format is shown below.

- **id**: Optional identifier for each NIRS read. The id must be an integer.
- **sampling\_id**: Optional identifier for each sample. Strings are allowed.
- **sampling\_date**: Optional field. The format allowed is: YYYY-MM-DD.
- **observationunit\_name**: Required field that matches existing data in the database. It can be the plot name, subplots, plant name, or tissue sample, depending how your trial is designed.
- **device\_id**: Optional field to identify your device. Strings are allowed.
- **device\_type**: Required field. It is possible upload data for a single device type. They can be: SCiO, QST, Foss6500, BunchiN500, or LinkSquare.
- **comments**: Optional field for general comments. All other columns are required wavelengths. You can add how many columns you want upload – there is no limit.

Manage NIRS Data

NIRS

Upload and perform analyses using NIRS data

Uploaded NIRS Data

NIRS Analyses

Trained NIRS Models

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	<b>Id</b>	<b>sample_id</b>	<b>sampling_date</b>	<b>observationunit_name</b>	<b>device_id</b>	<b>device_type</b>	<b>comments</b>	<b>740</b>	<b>741</b>	<b>742</b>	<b>743</b>	<b>744</b>	<b>745</b>	<b>746</b>
2	1	7a6ac477-d291-4d07-af	2020-6-24	myTriad20_rept_acc_001	503E4BFC4E923999	SCIO		0.885707958	0.88572938	0.885590265	0.885493457	0.885493162	0.885572662	0.885628732
3	2	7a6ac477-d291-4d07-af	2020-6-24	myTriad20_rept_acc_002	503E4BFC4E923999	SCIO		0.89132994	0.8908724223	0.890824451	0.89079170	0.890797281	0.907621118	0.907593838
4	3	7a6ac477-d291-4d07-af	2020-6-24	myTriad20_rept_acc_003	503E4BFC4E923999	SCIO		0.889220207	0.889013119	0.888681812	0.888401257	0.888310362	0.888297321	0.888294052
5	4	73c648ca-5fb-4231-a1f6	2020-6-24	myTriad20_rept_acc_004	503E4BFC4E923999	SCIO		0.8900087	0.888604969	0.888191073	0.888958654	0.888933709	0.888072741	0.88824772
6	5	73c648ca-5fb-4231-a1f6	2020-6-24	myTriad20_rept_acc_005	503E4BFC4E923999	SCIO		0.89101707	0.93686202	0.93620132	0.937873867	0.937742819	0.937775128	0.9374594
7	6	73c648ca-5fb-4231-a1f6	2020-6-24	myTriad20_rept_acc_006	503E4BFC4E923999	SCIO		0.876289461	0.875981159	0.875570263	0.875289867	0.875162225	0.875171404	0.87520442
8	7	d5b55c93-4cb8-4ef6-8c1	2020-6-24	myTriad20_rept_acc_007	503E4BFC4E923999	SCIO		0.87921785	0.878925781	0.878588441	0.8783872	0.878346457	0.87842929	0.87845739
9	8	7a6ac477-d291-4d07-af	2020-6-24	myTriad20_rept_acc_008	503E4BFC4E923999	SCIO		0.890746588	0.890515672	0.89016542	0.88990562	0.889783304	0.889782826	0.889792154
10	9	7a6ac477-d291-4d07-af	2020-6-24	myTriad20_rept_acc_009	503E4BFC4E923999	SCIO		0.850444238	0.85032422	0.850094039	0.8499495	0.849942163	0.850175036	

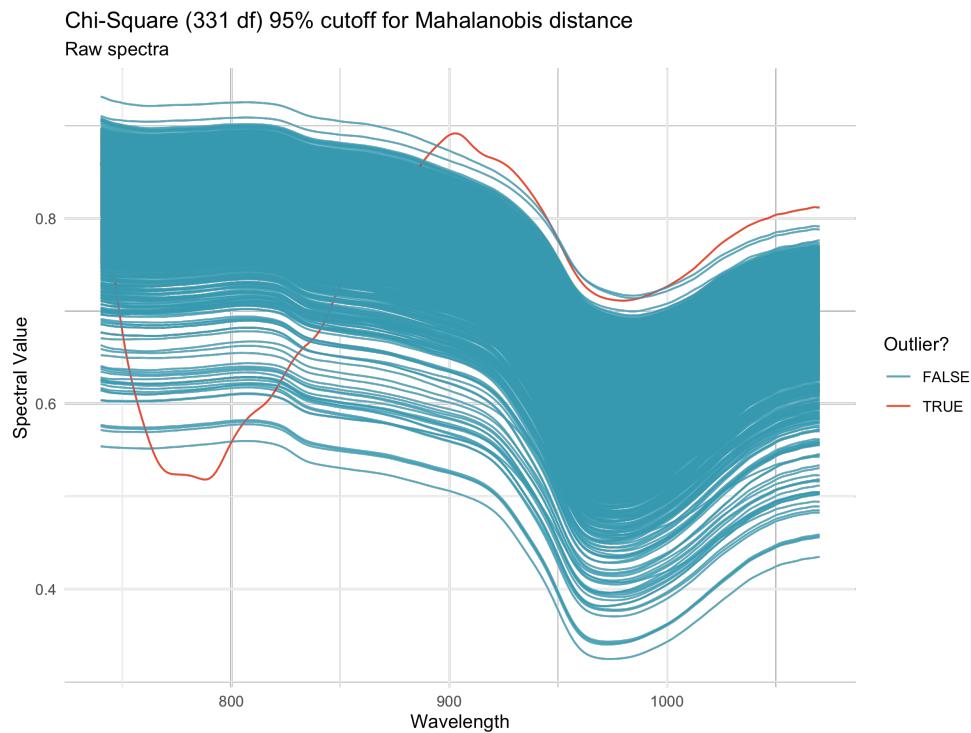
## 22.2 Evaluate and Remove Outliers

Spectral calibration models can be heavily affected by the presence of outliers, whether they come from spectrometer spectral artifacts or user errors. Mahalanobis distance (Mahalanobis, 1936) is a measure of the distance between a single observation and a larger distribution and is commonly used in the identification of outliers in a multivariate space (Des Maesschalck et al, 2000). The `FilterSpectra()` function in the R package `waves` calculates the Mahalanobis distance of each observation in a given spectral matrix using the `stats:::mahalanobis()` function. Observations are identified as outliers if the squared distance is greater than the 95th percentile of a  $\chi^2$ -distribution with  $p$  degrees of freedom, where  $p$  is the number of columns (wavelengths) in the spectral matrix (Johnson and Wichern, 2007). In Breedbase, this procedure is applied on a per-dataset basis on upload and outliers are given binary tags “Outlier.”

## 22.3 Plot Spectra

After outlier identification, a plot is generated using the `PlotSpectra()` function in `waves`. This function uses the filtered spectra and `ggplot2::ggplot()` to

create a line plot with outliers highlighted by color. A list of rows identified as outliers are shown beneath the plot. Plots are saved as .png files and linked to the original input datasets. Plot image files can be downloaded via the “Download Plot” button in the upload workflow.



## 22.4 Aggregate Spectra

To obtain a stable and reliable spectral profile, most spectrometer manufacturers recommend that multiple spectral scans are captured for each sample. While some spectrometers aggregate these scans internally, many do not, requiring the user to do so before analysis can take place. Breedbase handles these cases upon data upload following filtering steps by calling the *AggregateSpectra()* function from *waves*, saving the aggregated scans for future access through the search wizard feature. Scans are aggregated by sample mean (e.g. plot-level basis) according to the provided observationUnitName field. After aggregation, the user exits the upload workflow and the raw data file is saved in the upload archive.

## 22.5 References

- De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. (2000). The Mahalanobis distance. Chemom. Intell. Lab. Syst. 50(1): 1-18.
- Johnson, R. A. & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis (6th Edition). p 773.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. National Institute of Science of India.

**Analysis tool documentation**



# Chapter 23

## Managing Sequence Metadata

Manage Sequence Metadata

Sequence Metadata

Upload Sequence Metadata Search Sequence Metadata

Sequence Metadata Protocols View and query existing sequence metadata

GWAS Results Report of quantitative trait loci (QTLs) identified by running rrBLUP analysis on phenotype trials and genotype trials within the T3 database.

Show 10 entries Search:

Protocol Name	Description	Properties																				
Akhunov eQTL Analysis	eQTL analysis performed by the Akhunov lab.	<p>Data Type: GWAS Results Reference Genome: RefSeq_v1 Score: effect size Attributes:</p> <table border="1"><thead><tr><th>Key</th><th>Description</th></tr></thead><tbody><tr><td>effect</td><td>effect size</td></tr><tr><td>r2</td><td>coefficient of determination</td></tr><tr><td>gene</td><td>gene name</td></tr><tr><td>t</td><td>t-statistic</td></tr><tr><td>p</td><td>p-value</td></tr><tr><td>fdr</td><td>false discovery rate</td></tr><tr><td>tissue</td><td>tissue sampled, either 'seedling' or 'spike'</td></tr></tbody></table> <p>Links:</p> <table border="1"><thead><tr><th>Title</th><th>URL Template</th></tr></thead><tbody><tr><td>JBrowse - eQTL SNP</td><td><a href="https://graingenomes.org/jb/?data=/ggds/whe-lwgsc2018&amp;loc=chr{{feature}}:{{start}}..{{end}}&amp;tracks=eQTL-annot,eQTL-seedling,eQTL-spike">https://graingenomes.org/jb/?data=/ggds/whe-lwgsc2018&amp;loc=chr{{feature}}:{{start}}..{{end}}&amp;tracks=eQTL-annot,eQTL-seedling,eQTL-spike</a></td></tr></tbody></table>	Key	Description	effect	effect size	r2	coefficient of determination	gene	gene name	t	t-statistic	p	p-value	fdr	false discovery rate	tissue	tissue sampled, either 'seedling' or 'spike'	Title	URL Template	JBrowse - eQTL SNP	<a href="https://graingenomes.org/jb/?data=/ggds/whe-lwgsc2018&amp;loc=chr{{feature}}:{{start}}..{{end}}&amp;tracks=eQTL-annot,eQTL-seedling,eQTL-spike">https://graingenomes.org/jb/?data=/ggds/whe-lwgsc2018&amp;loc=chr{{feature}}:{{start}}..{{end}}&amp;tracks=eQTL-annot,eQTL-seedling,eQTL-spike</a>
Key	Description																					
effect	effect size																					
r2	coefficient of determination																					
gene	gene name																					
t	t-statistic																					
p	p-value																					
fdr	false discovery rate																					
tissue	tissue sampled, either 'seedling' or 'spike'																					
Title	URL Template																					
JBrowse - eQTL SNP	<a href="https://graingenomes.org/jb/?data=/ggds/whe-lwgsc2018&amp;loc=chr{{feature}}:{{start}}..{{end}}&amp;tracks=eQTL-annot,eQTL-seedling,eQTL-spike">https://graingenomes.org/jb/?data=/ggds/whe-lwgsc2018&amp;loc=chr{{feature}}:{{start}}..{{end}}&amp;tracks=eQTL-annot,eQTL-seedling,eQTL-spike</a>																					

## 23.1 What is Sequence Metadata?

Sequence Metadata is a feature that allows for the efficient storage and retrieval of sequence annotations for a specific region along a reference genome. The annotation data can contain a primary “score” value and any number of secondary key/value attribute data. For example, Sequence Metatadata can store MNase open chromatin scores for every 10 basepairs along the reference genome as well as genome-wide association study (GWAS) statistics, including the trait information associated with the result. This data can then be filtered by position and/or scores/attribute values and even cross-referenced with markers stored in the database.

## 23.2 Loading Sequence Metadata

Sequence Metadata can be loaded into the database using a gff3-formatted file. The following columns are used to load the data:

- **#1 / seqid:** The name of the database feature (ie chromosome) the metadata is associated with (The feature name must already exist as a feature in the database)
- **#4 / start:** The metadata’s start position
- **#5 / end:** The metadata’s end position
- **#6 / score:** (optional) The primary score attribute of the metadata
- **#9 / attributes:** (optional) Secondary key//value attributes to be saved with the score. These should be formatted using the gff3 standard (key1=value1;key2=value2). The attribute key cannot be either score, start, or end.

To upload the gff3 file:

1. Go to the **Manage > Sequence Metadata** page
2. Click the **Upload Sequence Metadata** button
3. On Step 2 of the Wizard, select the Type of data to be uploaded
  - This groups similar datasets together in the same Data Type category
4. On Step 3 of the Wizard, select an existing Protocol or create a new one
  - The Protocol is used to describe how the data was generated and define the score value and any secondary attributes. Adding the

attributes (and their descriptions) to the Protocol will allow the Sequence Metadata queries to filter the data based on the value of one or more of these attributes. Attributes not defined in the Protocol will still be stored and displayed on retrieval, but will not be able to be used in a search filter.

5. Finally, select and upload your gff3 file to the database. The database will verify the format of the file before its contents are stored.

## 23.3 Searching Sequence Metadata

To retrieve stored Sequence Metadata, go to the **Search > Sequence Metadata** page.

### 23.3.1 Basic Search

The basic Sequence Metadata search options include selecting the reference genome and species, the chromosome, and (optionally) the start and/or end position(s) along the reference genome. In addition, one or more specific protocols can be selected to limit the results.

**Search Sequence Metadata**

Filter the sequence metadata by position, sequence metadata type and/or protocol, and/or by protocol attribute value(s).

**Query Range**

Reference Genome:	RefSeq_v1 (Triticum aestivum)
Feature:	1A
Start:	1200000
	End: 1300000

---

**Protocol**

Protocol:	Gene Annotation IWGSC Assembly Variant Effect Predictor <b>GWAS Results</b> Akhunov eQTL Analysis <b>T3 Automated GWAS</b> <b>MNase</b> MNase Open Chromatin
-----------	---

---

**Advanced Search**

Filter by attribute values

**Search**

The Sequence Metadata search results are returned as a table, including the chromosome and start/stop positions of the annotation, along with the primary score value and any additional key/value attributes. The markers column will include a list of marker names of any stored markers that are found within the start/stop positions of the Sequence Metadata. The data can be downloaded as a table in an Excel or CSV file or a machine-readable (code-friendly) JSON file. If the Sequence Metadata JBrowse configuration is set, the filtered results can be displayed as a dynamic JBrowse track.

Protocol	Feature	Start	End	Score	Attributes	External Links	Markers
T3 Automated GWAS	1A	1207522	1207522	0.0454952857260828	ID: RAC875_c20883_801 Locus: TraeCS1A02G002300 Population: TCAP90K_SpringAM_panel x SW-AMPanel_2012_Saskatoon Trait: SDS sedimentation Variables: CO_321:0001138 pvalue: 0.00080725065468678 qvalue: 0.0454952857260828 zvalue: 3.350296620481	<a href="#">EnsemblPlants - Gen Summary</a> <a href="#">GrainGenes - Probe Report</a> <a href="#">JBrowse - Gene Annotations, Variants, and GWAS</a> <a href="#">Knetminer - Gene Network</a>	<b>1 marker found:</b> 1A @ 1207522 (T/C) • RAC875_c20883_801 (Infinium 90K)

### 23.3.2 Advanced Search

Any number of advanced search filters can be applied to the query. The advanced filters can limit the search results by the value of the primary score and/or any of the secondary attribute values.

The screenshot shows the 'Advanced Search' interface. At the top right is a link 'Filter by attribute values'. Below it is a note: 'Return only sequence metadata features that have attribute values that match the added comparisons. If more than one attribute filter is added, the sequence metadata feature must match all of the filters.' The interface is divided into two main sections:

- Score:** Contains fields for 'Protocol' (T3 Automated GWAS), 'Comparison' (Greater Than or Equal), and 'Value' (empty input field). A blue 'Add' button is to the right.
- Attribute:** Contains fields for 'Protocol' (T3 Automated GWAS), 'Key' (Trait), 'Comparison' (Equal), and 'Value' (empty input field). A blue 'Add' button is to the right.

Below these sections is a table titled 'Attribute Filters' with the following data:

Protocol	Attribute	Comparison	Value
T3 Automated GWAS	score	Greater Than or Equal	0.04 <span style="color:red;">X</span>
T3 Automated GWAS	Trait	Equal	grain yield <span style="color:red;">X</span>

## 23.4 Marker Integration

A table of Sequence Metadata annotations are embedded on the Marker/Variant detail page. The table will include any annotations that span the position of the marker (for data of the same reference genome and species).

## 23.5 Sequence Metadata API

A publicly accessible RESTful API (Application Programming Interface) is available to query the database for Sequence Metadata directly from your programming environment (R, python, etc) to be used in analysis. The data is returned in a JSON format. Documentation for the API can be found on the **Manage > Sequence Metadata** page



# Chapter 24

## Managing Outliers in Dataset

### 24.1 What is Outliers Functionality in Dataset ?



As in step [The Search Wizard](#) we can create a dataset.

The dataset incorporates a feature to identify outlier points, which we may choose to exclude from a specific dataset. It's important to note that these exclusions only apply at the dataset level, and no data is permanently removed from the database. Additionally, outlier categorization can be modified at any time, and these changes are visible to all other functionalities within the system.

Each dataset stores a wholly unique set of outlier points, completely independent of any other dataset in the database. Outliers are specifically designated for traits within datasets, exclusively encompassing phenotype data. If a particular dataset lacks traits as a part of wizard selection, this functionality is not available.

Each trait has its own set of defined outliers.

## 24.2 Accessing Trait Visualization

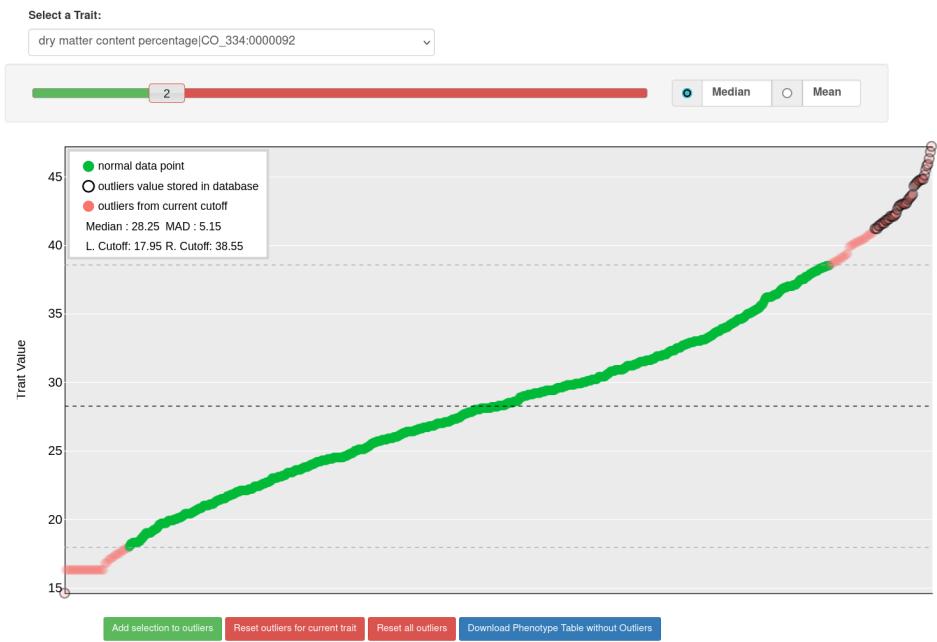
Once you've selected a specific trait, the web application provides access to a visualization of the data points associated with that trait.



## 24.3 Interpreting Visual Elements

Once you've selected a specific trait, the web application provides access to a visualization of the data points associated with that trait.

- **Green Points:** As per the legend, represent values for the selected trait that fall below the cut-off point set by the slider. (non-outliers)
- **Black Outlined Points:** These data points are outlined with black borders, indicating that they are currently designated as outliers in the database.
- **Red Points:** The red data points denote the cut-off points established by the slider for the allowable deviation value.



## 24.4 Choosing Cut-Off Values

You have two fundamental options for setting cut-off points:

- **Median with MAD:** This option involves using the median (middle value) along with the Mean Absolute Deviation (MAD) as a reference point for determining cut-off values.
- **Mean with Standard Deviation:** Alternatively, you can choose to use the mean (average) in conjunction with the Standard Deviation to set cut-off points.

## 24.5 Setting Deviation Multiplier

The slider allows you to specify the deviation multiplier from a central point, which influences the cut-off values.

## 24.6 Utilizing Graph Controls

Beneath the graph, you'll find four buttons, each serving a distinct function:

- **Add selection to outliers:** This button enables you to save the current cut-off points to the database for future reference.
- **Reset outliers for current trait:** You can use this option to reset outliers for the selected trait.
- **Reset all outliers:** This button allows you to reset outliers for the entire dataset.
- **Download Phenotype Table without outliers:** You can download the phenotype data table in a comma-separated value format file, using this feature, with outliers excluded for selected dataset.



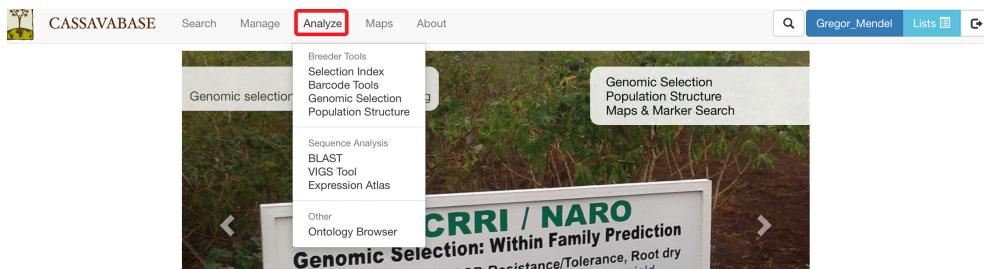
Add selection to outliers    Reset outliers for current trait    Reset all outliers    Download Phenotype Table without Outliers

These tools and functions are designed to provide you with control and insights when working with data visualization and outliers.

# Chapter 25

## Data Analysis Tools

SGN databases provides several tools for phenotype data analysis, marker-assisted selection, sequence and expression analyses, as well as ontology browser. These tools can be found in the “Analyze” menu.



### 25.1 Selection Index

To determine rankings of accessions based on more than one desirable trait, SGN databases provide a “Selection Index” tool that allows you to specify a weighting on each trait. To access the tool, clicking on “Selection Index” in the “Analyze” menu.



On the Selection Index page, selecting a trial that you want to analyze.

**Build a Selection Index**

Parameters				
<b>Trial select:</b>	<b>Trait select:</b>			
<input type="text" value="Please select a trial"/>				
<b>Traits and weights:</b>				
Trait name	Trait CO id	Trait synonym	Weight	Remove?

After you selected a trial, you can find traits that were assayed in that trial in the “Trait” box.

**Build a Selection Index**

Parameters				
<b>Trial select:</b>	<b>Trait select:</b>			
<input type="text" value="06uyt25Ncmdlk"/>	<input checked="" type="checkbox"/> Select a trait boiled tuberous root color visual 1-3 cassava bacterial blight severity 3-month evaluation cassava bacterial blight severity 9-month evaluation cassava mosaic disease incidence 1-month evaluation cassava mosaic disease incidence 3-month evaluation cassava mosaic disease severity 1-month evaluation cassava mosaic disease severity 3-month evaluation dry matter content percentage dry yield ease of peeling root cortex visual rating 1-7 fibre content estimation in percentage <b>fresh root weight</b> fresh root yield fresh shoot weight measurement in kg harvest index variable initial vigor assessment 1-7 plant stands harvested counting poundability assessment 0-4 root neck length visual rating 0-7 root number counting			
<b>Traits and weights:</b>				
Trait name	Trait CO id	Trait synonym	Weight	Remove?
<b>Additional options:</b>				
<input type="checkbox"/> Include accessions with missing phenotypes <input type="checkbox"/> Scale values to a reference accession Select a reference accession				
<b>Rankings</b>				
<a href="#">Selection Index</a> <a href="#">Raw Averages</a>				

Selecting a trait that you want to include in the analysis will open a new dialogue showing the selected trait and a box that you can assign a “Weight” of that trait. After you are done, you can continue by selecting another trait by clicking on “Add another trait” link.

**Build a Selection Index**

**Parameters**

**Trial select:** 06uyt25NcmndlK

Traits and weights:				
Trait name	Trait CO id	Trait synonym	Weight	Remove?
fresh root weight	CO:0000012	RtWt_Wgh_kg	Must be a number (+ or -, d)	X

Add another trait

**Additional options:**

- Include accessions with missing phenotypes
- Scale values to a reference accession:

SIN formula:  
SIN = 1 \* (fresh root weight)

Select a reference accession

Calculate Rankings

After you selected another trait, this page will automatically update information for you by showing all of the traits that you selected for the analysis.

**Build a Selection Index**

**Parameters**

**Trial select:** 06uyt25NcmndlK

Traits and weights:				
Trait name	Trait CO id	Trait synonym	Weight	Remove?
fresh root weight	CO:0000012	RtWt_Wgh_kg	7	X
initial vigor assessment 1-7	CO:0000009	IVig_IITAVisScg_1to7	3	X

Add another trait

**Additional options:**

- Include accessions with missing phenotypes
- Scale values to a reference accession:

SIN formula:  
SIN = 7 \* (fresh root weight) + 3 \* (initial vigor assessment 1-7)

Select a reference accession

Calculate Rankings

You also have options to choose a reference accession, choose to include accessions with missing phenotypes, scaling values to a reference accession. After you complete your setting, clicking on “Calculate Rankings”

**Build a Selection Index**

Parameters				
<b>Trial select:</b> 06uyt25Ncmdlk				
<b>Traits and weights:</b>				
Trait name	Trait CO id	Trait synonym	Weight	Remove?
fresh root weight	CO:0000012	RtWt_Wgh_kg	7	X
initial vigor assessment 1-7	CO:0000009	IV(g_IITAVisScg_1to7)	3	X
<a href="#">Add another trait</a>				
<input type="checkbox"/> Include accessions with missing phenotypes <input type="checkbox"/> Scale values to a reference accession: Select a reference accession			<b>SIN formula:</b> $SIN = 7 * (\text{fresh root weight}) + 3 * (\text{initial vigor assessment 1-7})$ <a href="#">Calculate Rankings</a>	

The Selection Index tool will generate rankings of accessions based on the information that you specified. You can copy the results to your system clipboard, convert the table data to CSV format, or print the data.

**Rankings**

Selection Index				
Raw Averages				
<a href="#">Copy</a> <a href="#">CSV</a> <a href="#">Print</a> <input type="text" value="Search:"/> <a href="#">Search</a>				
Accession	7 * (fresh root weight)	3 * (initial vigor assessment 1-7)	SIN	SIN Rank
IITA-TMS-IBA940006	156.8	21	177.80	1
IITA-TMS-IBA8200058	138.25	21	159.25	2
IITA-TMS-IBA961708	131.6	18	149.60	3
IITA-TMS-IBA990554	115.5	21	136.50	4
IITA-TMS-IBA982132	113.75	19.5	133.25	5
IITA-TMS-IBA010090	108.15	21	129.15	6
IITA-TMS-IBA9102327	108.5	18	126.50	7
IITA-TMS-IBA961432	103.25	19.5	122.75	8
IITA-TMS-IBA000028	98.7	21	119.70	9
IITA-TMS-MM961751	94.5	19.5	114.00	10

Table description: weighted\_values for trial 06uyt25Ncmdlk.

Showing 1 to 10 of 25 entries      Previous [1](#) [2](#) [3](#) Next

Clicking on “Raw Average” will display average values of the phenotypes of those ranked accessions.



Accession	fresh root weight	initial vigor assessment 1-7
IITA-TMS-IBA940006	22.40	7.00
IITA-TMS-IBA820008	19.75	7.00
IITA-TMS-IBA961708	18.80	6.00
IITA-TMS-IBA990554	16.50	7.00
IITA-TMS-IBA982132	16.25	6.50
IITA-TMS-IBA9102327	15.50	6.00
IITA-TMS-IBA010090	15.45	7.00
IITA-TMS-IBA961432	14.75	6.50
IITA-TMS-IBA000028	14.10	7.00
IITA-TMS-MM961751	13.50	6.50

Table description: raw\_avgs for trial 06uyt25NcmdIk.

Showing 1 to 10 of 25 entries

Previous 1 2 3 Next

Selection Index tool also allows you to save top ranked accessions directly to “Lists”. You can retrieve top ranked accessions by selecting a number or a percent.



Save top ranked accessions to a list:

By number:

Or percent:

## 25.2 Genomic Selection

The prediction of breeding values for a trait is a one step or two steps process, depending on what stage in your breeding cycle you are. The first step is to build a prediction model for a trait using a training population of clones with phenotype and genotype data. If you have yet to select parents for crossing for your first cycle of selection you can use the breeding values of the training population. If you are at later stages of your selection program, you need to do the second step which is applying the prediction model on your selection population. All clones in your training and selection populations must exist in the database.

To use the genomic selection tool, on [cassavabase.org](http://cassavabase.org), select “Genomic Selection” from the “analyze” pull-down menu.



The screenshot shows the CASSAVABASE homepage with a navigation bar at the top. The navigation bar includes links for home, forum, contact, help, and wiki, along with user account links for Isaak Tecle, lists, my account, and log out. Below the navigation bar is a search bar with the placeholder "Search for a trait" and a "sol search" button. A sub-navigation menu titled "solGS: start building a GS model by searching for a trait or selecting a training population" is displayed, containing three options: "Search for a trait", "Use a trial as a training population", and "Create a training population".

There are three ways to build a model for a trait.

### 25.2.1 Building a Model - Method 1:

One way to build a model is, using a trait name, to search for trials in which the trait was phenotyped and use a trial or a combination of trials to build a model for the trait. For example, if you search for “mosaic disease severity”, you will get a list of trials you can use as training populations.



The screenshot shows the CASSAVABASE homepage with a search bar containing the trait name "mosaic". Below the search bar, a list of results is shown, including "cassava mosaic disease incidence" and "cassava mosaic disease severity". At the bottom of the page, there are two additional options: "Use a trial as a training population" and "Create a training population".

You will get a list of trials (as shown below) in which the trait of your inter-

ested was phenotyped. From the list, you can use a single trial as a training population or combine several trials to form a training population for the prediction model of the trait. Let's say, you want to create a training population using individuals from trials "cassava ibadan 2001/02" and "cassava ibadan 02/03" and build a model for "cassava mosaic disease severity" using all clones from the training population.

**GS populations evaluated for cassava mosaic disease severity**

Select a training population or create a new one using one or more trials

Trial	Description	Location	Year	Tip(?)
<input checked="" type="checkbox"/> Cassava Ibadan 2002/03	Plants assayed at Ibadan in 2002/03	Ibadan	2002/03	
<input checked="" type="checkbox"/> Cassava Ibadan 2001/02	Plants assayed at Ibadan in 2001/02	Ibadan	2001/02	
<input type="checkbox"/> AYT 2011-2012	AYT 2011-2012 Trial NR09	Umudike	2011	
<input type="checkbox"/> Cassava Ibadan 2003/04	Plants assayed at Ibadan in 2003/04	Ibadan	2003/04	
<input type="checkbox"/> Cassava Ibadan 2004/05	Plants assayed at Ibadan in 2004/05	Ibadan	2004/05	
<input type="checkbox"/> Cassava Igbariam 2009	Plants assayed at Igbariam in 2009	Igbariam	2009	
<input type="checkbox"/> Cassava Ibadan 2005/06	Plants assayed at Ibadan in 2005/06	Ibadan	2005/06	
<input type="checkbox"/> Cassava Ibadan 2000/01	Plants assayed at Ibadan in 2000/01	Ibadan	2000/01	
<input type="checkbox"/> Cassava Ibadan 1999/00	Plants assayed at Ibadan in 1999/00	Ibadan	1999/00	
<input type="checkbox"/> Cassava Ibadan 2006/07	Plants assayed at Ibadan in 2006/07	Ibadan	2006/07	

**1 2 3 4 5 >**

**Done selecting**

---

Trials to combine

Trial	Description	Location	Year	
<input checked="" type="checkbox"/> Cassava Ibadan 2002/03	Plants assayed at Ibadan in 2002/03	Ibadan	2002/03	
<input checked="" type="checkbox"/> Cassava Ibadan 2001/02	Plants assayed at Ibadan in 2001/02	Ibadan	2001/02	

**Combine trials & build model**

Select the trials to combine (the same coloured), click 'done selecting', click the "combine trials and build model" button, and you will get a model and its output for the trait. On the model detail page, you can view the description of input data used in the model, output from the model and search interface for selection populations the model you can apply to predict their breeding values. The description of the input data for the model includes the number of phenotyped clones, and the number of markers, scatter and frequency distribution plots for the phenotype data, relationship between the phenotype

data and GEBVs, population structure. The model output includes model parameters, heritability of the trait , prediction accuracy, GEBVs of the individuals from the training population and marker effects.

**Genomic selection model for Cassava mosaic disease severity (CMDS) in Training population  
2907368219**

(+) **Training population summary**

Name	Training population 2907368219	No. of lines	239
Description	This training population is a combination of <b>Cassava Ibadan 2001/02</b> and <b>Cassava Ibadan 2002/03</b> .	No. of markers	97337
Owner	Peter Kulakow	Genotyping version	GBS ApeKI Cassava genome v5

(+) **Trait phenotype data**

(+) **Model accuracy**

(+) **Population structure analysis - PCA**

(+) **Model Parameters**

(+) **Genomic estimated breeding values (GEBVs) - GBLUP method**

(+) **Marker Effects**

(+) **Predict GEBVs of a selection population using the model**

Expand each section to see detailed information.

If you expand the ‘Trait phenotype data’ section, you will find plots to explore the phenotype data used in the model. You can assess the phenotype data using a scatter and histogram plots and the descriptive statistics.





A regression line between observed phenotypes and GEBVs shows the relationship between the two.



You can also explore if there is any sub-clustering in the training population using PCA.



To check the model accuracy, a 10-fold cross-validation test, expand the ‘model accuracy’ section.

**Genomic selection model for Cassava mosaic disease severity (CMDS) in Training population  
2907368219**

- [④ Training population summary](#)
- [④ Trait phenotype data](#)
- [④ Model accuracy](#)

Runs	Accuracy r)
Validation test 6	0.648
Validation test 3	0.612
Validation test 9	0.571
Validation test 4	0.556
Validation test 7	0.555
Validation test 5	0.478
Validation test 8	0.444
Validation test 2	0.422
Validation test 10	0.417
Validation test 1	0.335
Average	0.5

[ Download model accuracy report ]
- [④ Population structure analysis - PCA](#)
- [④ Model Parameters](#)
- [④ Genomic estimated breeding values \(GEBVs\) - GBLUP method](#)
- [④ Marker Effects](#)
- [④ Predict GEBVs of a selection population using the model](#)

Marker effects are also available for download. To do so, expand the ‘Marker Effects’ section and click the ‘Download all marker effects’ link and you will get a tab delimited output to save on your computer.

**Genomic selection model for Cassava mosaic disease severity (CMDS) in Training population**  
2907368219

- [Training population summary](#)
- [Trait phenotype data](#)
- [Model accuracy](#)
- [Population structure analysis - PCA](#)
- [Model Parameters](#)
- [Genomic estimated breeding values \(GEBVs\) - GBLUP method](#)
- [Marker Effects](#)

*Top 10 markers:*

Marker	Effects
S1_317473	0.00308
S9_16963656	0.00298
S14_18164196	0.00295
S19_117475272	0.00281
S19_113674348	0.00274
S8_472214	0.00256
S3_6218223	0.00256
S8_7605425	0.00253
S4_15467245	0.00249
S4_15467248	0.00249

[ [Download all marker effects](#) ]

- [Predict GEBVs of a selection population using the model](#)

The breeding values of the individuals used in the training population are displayed graphically. Mousing over each data point displays the clone and its breeding value. To examine better, you can zoom in into the plot by selecting an area on the plot. You can download them also by following the “Download all GEBVs” link.



### Estimating breeding values in a selection population

If you already have a selection population (in the database), from the same model page, you can apply the model to the selection population and estimate breeding values for all the clones in the population. You can search for a selection population of clones in the database using the search interface or you can make a custom list of clones using the *list interface*. If you click the “search for all relevant selection populations”, you will see all relevant selection populations for that model. However, this option takes long time because of the large set of populations in the database and the filtering. Therefore, the fastest way is to search for each of your selection populations by name. If you are logged in to the website you will also see a list of your

custom set of genotyped clones.

**Genomic selection model for Cassava mosaic disease severity (CMDS) in Training population  
2907368219**

- Training population summary**
- Trait phenotype data**
- Model accuracy**
- Population structure analysis - PCA**
- Model Parameters**
- Genomic estimated breeding values (GEBVs) - GBLUP method**
- Marker Effects**
- Predict GEBVs of a selection population using the model**

Go

Search for all relevant selection populations

Selection population	Description	Year	Predict GEBVs
Cassava Ibadan 2005/06	Plants assayed at Ibadan in 2005/06	2005	CMDS
Cassava Ibadan 2003/04	Plants assayed at Ibadan in 2003/04	2003	CMDS
Cassava Ibadan 2006/07	Plants assayed at Ibadan in 2006/07	2006	[Predict]

**List-based selection population**  
 Selection candidates list 2015

Go
+ Make a new list of clones

**List-based selection population**  
 Selection candidates list 2015

Predict GEBVs  
 [Predict]

To apply the model to a selection population, simply click your population name or “Predict Now” and you will get the predicted breeding values. When you see a name of (or acronym) of the trait, follow the link and you will see an interactive plot of the breeding values and a link to download the breeding values of your selection population.



### 25.2.2 Building a Model - Method 2

Another way to build a model is by selecting a trial, instead of selecting and searching for a specific trait. This approach is useful when you know a particular trial that is relevant to the environment you are targeting to breed material for. This method allows you to build models and predict genomic estimated breeding values (GEBVs) for several traits within a single trial at once. You can also calculate selection index for your clones when GEBVs are

estimated for multiple traits.

To do this select the “Genomic Selection” link found under the “analyze” menu. This will take you to the same home page as used with Method 1. However, instead of entering information to search for in “Search for a trait”, click on “Use a trait as a trial population”. This will expand a new menu that will show all available trials.

The screenshot shows the CASSAVABASE website interface. At the top, there is a navigation bar with links for home, forum, contact, help, and wiki. Below the navigation bar, there are links for search, manage, analyze, maps, and user accounts (Liana Acevedo, lists, my account, log out). A search bar with the placeholder "Search for a trait" is present. Below the search bar, a traits index is shown with links for B, C, D, E, F, H, I, L, P, R, S, and T. A large red arrow points downwards from the "analyze" menu item towards the "Use a trial as a training population" link in the expanded menu below.

home | forum | contact | help | wiki

search manage analyze maps

Liana Acevedo lists my account log out

soIGS: start building a GS model by searching for a trait or selecting a training population

Search for a trait

Traits index: [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [H](#) | [I](#) | [L](#) | [P](#) | [R](#) | [S](#) | [T](#)

Use a trial as a training population

Create a training population

↓

**solGS: start building a GS model by searching for a trait or selecting a training population**

**Select a training population or create a new one using one or more trials**

Trial	Description	Location	Year	Tip(?)
<input checked="" type="checkbox"/> Cassava Ibadan 2002/03	Plants assayed at Ibadan in 2002/03	Ibadan	2002/03	
<input type="checkbox"/> Cassava Ibadan 2001/02	Plants assayed at Ibadan in 2001/02	Ibadan	2001/02	
<input type="checkbox"/> AYT 2011-2012	AYT 2011-2012 Trial NR09	Umudike	2011	
<input type="checkbox"/> Cassava Ibadan 2003/04	Plants assayed at Ibadan in 2003/04	Ibadan	2003/04	
<input type="checkbox"/> Cassava Ibadan 2004/05	Plants assayed at Ibadan in 2004/05	Ibadan	2004/05	
<input type="checkbox"/> Cassava Igbariam 2009	Plants assayed at Igbariam in 2009	Igbariam	2009	
<input type="checkbox"/> Cassava Ibadan 2005/06	Plants assayed at Ibadan in 2005/06	Ibadan	2005/06	
<input type="checkbox"/> Cassava Ibadan 2000/01	Plants assayed at Ibadan in 2000/01	Ibadan	2000/01	
<input type="checkbox"/> Cassava Ibadan 1999/00	Plants assayed at Ibadan in 1999/00	Ibadan	1999/00	
<input type="checkbox"/> Cassava Ibadan 2006/07	Plants assayed at Ibadan in 2006/07	Ibadan	2006/07	

**1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 >**

**Select a list-based training population or create a new one**

To begin creating the model, select the existing trial that you would like to use. In this example I will be using the trial and trait data from “Cassava Ibadan 2002/03” trial. Clicking on a trial will take you to a page where you can find information such as number of markers and number of phenotypes clones.

Select one or more traits from training population "Cassava Ibadan 2002/03" to build a GS model and predict GEBVs for:

Training population summary

Name	Cassava Ibadan 2002/03	No. of lines	237
Description	Plants assayed at ibadan in 2002/03	No. of traits	20
Owner	Peter Kulakow	No. of markers	97337
		Genotyping version	GBS ApeKI Cassava genome v5

Traits

- boiled tuberous root color visual 1-3
- Cassava bacterial blight incidence
- Cassava bacterial blight severity
- Cassava green mite severity
- Cassava mosaic disease incidence
- Cassava mosaic disease severity
- dry matter content percentage
- dry yield
- top yield

**Build model**

Phenotypic correlation analysis

Trait names and their acronyms

In addition to the number of phenotype clones and number of markers, the main page for the trial selected also has information and graphs on phenotypic correlation for all of the traits. By moving your cursor over the graph you can read the different values for correlation between two traits. A key with all of the trait names of the acronyms used can be found in the tab below the graph.



Below the “Training population summary” there is a tab for “Traits”. Clicking on this tab will show all available traits for the specific trial. You can create a model by choosing one or multiple traits in the trial and clicking “Build Model”. In this example, the traits for “cassava bacterial blight severity” and “cassava mosaic disease severity” have been selected.

Select one or more traits from training population "Cassava Ibadan 2002/03" to build a GS model and predict GEBVs for:

Training population summary

Name	Cassava Ibadan 2002/03	No. of lines	237
Description	Plants assayed at Ibadan in 2002/03	No. of traits	20
Owner	Peter Kulakow	No. of markers	97337
		Genotyping version	GBS ApeKI Cassava genome v5

Traits

- boiled tuberous root color visual 1-3
- Cassava bacterial blight incidence
- Cassava bacterial blight severity
- Cassava green mite severity
- Cassava mosaic disease incidence
- Cassava mosaic disease severity
- dry matter content percentage
- dry yield
- top yield

**Build model**

Phenotypic correlation analysis

Trait names and their acronyms

Clicking on “Build Model” will take you to a new page with the models outputs for the traits. Under the “Genomic Selection Model Output” tab you can view the model output and the model accuracy. Clicking on any of the traits will take you to a page with information about the model output on that individual trait within the trial. There you can view all of the trait information that was seen in more detail in *Method 1*.

**Prediction models from Cassava Ibadan 2002/03**

Models summary

Training population	Description	Models						
Cassava Ibadan 2002/03	Plants assayed at Ibadan in 2002/03	<table border="1"> <thead> <tr> <th>Trait</th> <th>Model accuracy</th> </tr> </thead> <tbody> <tr> <td>DY</td> <td>0.46</td> </tr> <tr> <td>CMDS</td> <td>0.46</td> </tr> </tbody> </table>	Trait	Model accuracy	DY	0.46	CMDS	0.46
Trait	Model accuracy							
DY	0.46							
CMDS	0.46							

Predict GEBVs of a selection population using the models

Genetic correlation analysis

Calculate selection index

Trait names and their acronyms

You can apply the models to simultaneously predict GEBVs for respective traits in a selection population by clicking on “Predict Now” or the name of the selection population. You can also apply the models to any set of genotyped clones that you can create using the “lists” feature. For more information on lists, click [here](#). Follow the link to the trait name to view and download the predicted GEBVs for the trait in a selection population.

**Prediction models from Cassava Ibadan 2002/03**

<b>Models summary</b>			
<b>Training population</b>	<b>Description</b>	<b>Models</b>	
Cassava Ibadan 2002/03	Plants assayed at Ibadan in 2002/03	Trait	Model accuracy
		DY	0.46
		CMDS	0.46

<b>Predict GEBVs of a selection population using the models</b>			
<input type="text" value="search for a selection population"/>	<input type="button" value="Go"/>	<input type="button" value="Search for all relevant selection populations"/>	
<b>Selection population</b>	<b>Description</b>	<b>Year</b>	<b>Predict GEBVs</b>
Cassava Ibadan 2005/06	Plants assayed at Ibadan in 2005/06	2005	DY   CMDS
Cassava Ibadan 2006/07	Plants assayed at Ibadan in 2006/07	2006	[Predict]

<b>List-based selection population</b>	<b>Selection candidates list 2015</b>	<b>Go</b>	<b>+ Make a new list of clones</b>
List-based selection population			
Selection candidates list 2015			

<b>Genetic correlation analysis</b>

<b>Calculate selection index</b>

<b>Trait names and their acronyms</b>
---------------------------------------

To compare clones based on their performance on multiple traits, you can calculate selection indices using the form below. Choose from the pulldown menu the population with predicted GEBVs for the traits and assign relative weights for each trait. The relative weight of each trait must be between 0 - 1. 0 being of least weight and importance, not wanting to consider that particular trait in selecting a genotype and 1 being a trait that you give highest importance.

In this example we will be using the “Cassava Ibadan 2002/03” population and assigning values to each of the traits. Remember that there is a list of acronyms and trait names at the bottom of the page for reference. After entering whatever values you would like for each trait click on the “Calculate”

button to generate results. This will create a list of the top 10 genotypes that most closely match the criteria that you entered. The list will be displayed right below the “selection index” tab. This information can also be downloaded onto your computer by clicking on the “Download selection indices” link underneath the listed genotypes and selection indices.

**Prediction models from Cassava Ibadan 2002/03**

- Models summary**
- Predict GEBVs of a selection population using the models**
- Genetic correlation analysis**
- Calculate selection index**

Cassava Ibadan 2002/03  And assign relative weights to traits.

DY:  CMDS:

**Calculate**

**Genotype ranking based on multiple traits performance [selection index]**

*Top 10 genotypes:*

Genotypes	Selection indices
IITA-TMS-IBA974763	3.37
IITA-TMS-IBA974779	2.91
IITA-TMS-IBA930266	2.88
IITA-TMS-IBA950061	2.71
IITA-TMS-IBA940239	2.6
IITA-TMS-ZAP940153	2.49
IITA-TMS-IBA972205	2.36
IITA-TMS-IBA996016	2.11
IITA-TMS-IBA930098	1.87
IITA-TMS-ZAP930151	1.76

[ Download selection indices ]  
**Relative weights:** DY: 0.6 CMDS: 0.4  
**Name:** Cassava Ibadan 2002/03

**Correlation between selection index and GEBVs of traits**

	CMDS	DY
Index	Positive	Neutral
CMDS	Neutral	Positive
DY	Positive	Neutral

Negative  
Neutral  
Positive

### 25.2.3 Building a Model - Method 3

In addition to creating a model by searching for pre-existing traits or by preexisting trial name, models can also be created by using your own list of clones. This creates a model by using or creating a training population.

The page to use the third Method for creating a population model is the same as for the other two models. Select “Genomic Selection” from under the “analyze” menu of the main toolbar. This will take you to the Genomic Selection homepage and show you all three available methods to create a model. To see and use Method 3 scroll down and click on the tab labeled “Create a Training Population”. This will open a set of tools that will allow you to use pre-existing lists or to create a new list.



Once the “Create a Training Population” tab is opened you have the option to use a pre-existing list or create new one. To learn how to create a list, click [here](#). The “Make a new list of plots” link will take you directly to the Search Wizard that is usually used to create lists.

Please note: the only lists that can be used in Method 3 to create a model are lists of plots and trials. If the pre-existing list is not of plots or trials (for example, traits, or locations) it will not show up and cannot be used as a training population. When you create you use a list of trials, the trials data will be combined to create a training data set.

To use your custom list of plots or trials as a training population, select the list and click “Go”. This will take you to a detail page for the training population.

Select one or more traits from training population "Training population 1" to build a GS model and predict GEBVs for.

Training population summary

Name	Training population 1	No. of lines	195
Description	Uploaded on: Wed Jan 6 14:46 2016	No. of traits	26
Owner	isaaktecle	No. of markers	97337
		Genotyping version	GBS ApeKI Cassava genome v5

Traits

- Cassava anthracnose disease incidence
- Cassava anthracnose disease severity
- Cassava bacterial blight incidence
- Cassava bacterial blight severity
- Cassava green mite severity
- Cassava mosaic disease incidence
- Cassava mosaic disease severity
- dry matter content percentage
- dry yield
- top yield

**Build model**

Phenotypic correlation analysis

**Run correlation**

Trait names and their acronyms

From here on you can build models and predict breeding values as described in *Method 2*.

## 25.3 Genome Browsing

There are two ways to evaluate genotype information within the browser, from an accession detail page or a trial detail page.

### 25.3.1 Browsing Genotype data by Accession

If you are interested in browsing genotype information for a single accession, for example 'BAHKEYHEMAA', navigate to the accession detail page.

Search Results

Show 10 entries

Stock Name	Stock Type
BAHKYEHHEMAA	accession

Showing 1 to 1 of 1 entries

Near the bottom of the detail page is a collapsible section called “Accession Jbrowse”.

Genotype data

Accession JBrowse

[View the tracks for this accession in JBrowse](#)

This section will contain a link to the accession jbrowse page if the necessary genotype data is available. Clicking the link should take you to a page that looks like this, at which point you can browse the genotype data in the form of a vcf track aligned to the latest build of the genome.



### 25.3.2 Browsing Genotype data by Trial

If you are interested in browsing genotype information for the accessions within a given trial, navigate to the trial detail page.

Trial Search			
Show 10 entries			
Trial name	Description	Breeding program	Folder
12ayt30whrmK	Assessment of Varieties of Cassava for high yield, high dry matter and disease resistance using Advance Yield Trial (30 clones) in Mokwa 2012/2013 Breeding Season	IITA	12_Mokwa

Halfway down the page is a collapsible section called “Trial Jbrowse”. This section will contain a link to the trial jbrowse page if the necessary genotype data for at least two accessions planted in the trial is available.

- [Compute Trait Phenotypes](#)
- [Trial JBrowse](#)
- [View the dataset for this trial in JBrowse](#)
- [Files](#)

Clicking the link should take you to a page that looks like this, at which point you can browse the genotype data in the form of vcf tracks aligned to the latest build of the genome.



## 25.4 Principal Component Analysis (PCA)

Principal component analysis helps estimate and visualize if there is subgrouping of individuals within a dataset based on a number of variables. Currently, you can use marker data to run PCA on datasets.

You can run PCA from multiple places on the website. To do PCA on

- (1) individuals from a trial, go to the trial detail page and find the PCA tool under the “Analysis tools” section.
- (2) individuals from a training population you used in a GS modeling, do your modeling and find the PCA tool in the model output page.
- (3) individuals in a training population and selection population you applied the training model, do your modeling, apply the model on the selection population and find the PCA tool on the selection population prediction output page.
- (4) individuals in a list of accessions you created, for example using the search wizard, go to the “Analyze” menu and select the “Population Structure”, select your list of individuals and run PCA.
- (5) individuals from multiple trials, create a list of the trials using the search wizard, go to the “Analyze” menu and select the “Population Structure”, select your list of trials and run PCA.



With all the options, you will get a interactive plot of the two PCs (shown below) that explain the largest variance. Point the cursor at any data point and you will see the individual name with its corresponding PCs scores. By clicking the ‘Download all PCs’, you can also download the 10 PCs scores in the text format.

## 25.5 ANOVA

Currently, ANOVA is implemented for a single trial (single year and single location). You can do ANOVA for RCBD, CRD, Alpha and Augmented trial designs. ANOVA is done using linear mixed effects model, where the genotypes is fixed effect and the replications and blocks are random effects. Fixed effect significance level is computed using “lmer” from “lmeTest” R package.

You can do ANOVA from two places: trial detail and training population detail. In both cases, if the phenotype data was from the supported trial designs,

- Go to the ANOVA section down in the trial or training population page
- Select the trait of you want to perform ANOVA
- Click the “Run ANOVA” and wait for the result

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(> F)
genotypes	7,266.18	24.38	298	102.95	1.69	0.001

Download: [\[Anova table\]](#) | [\[Model Summary\]](#) | [\[Adjusted Means\]](#) | [\[Model Diagnostics\]](#)

## 25.6 Clustering (K-Means, Hierarchical)

The K-Means method allows you to partition a dataset into groups (K number). The hierarchical clustering, agglomerative, allows you to explore underlying similarity and visualize in a tree structure (dendrogram) the different levels of similarities (clusters) among samples. You can do clustering based on marker data, phenotype data and GEBVs. When you use phenotype data, first clone averages for each trait are calculated. Both methods use Euclidean distance as a measure of similarity. For the hierarchical clustering, the complete-linkage (farthest neighbour) method is used to link up clusters.

There are three pathways to using this tool.

(1) When you have data in the form of a list or dataset from the search wizard:

- (A) – go to the “Analyze” menu and select the clustering option
- (B) – make sure you are logged in
- (C) – Select the relevant genotyping protocol, if you are clustering using genotype data
- (D) – select your list or dataset, click “Go”
- (E) – select clustering type
- (F) – select the data type to use
- (G) – If you are running K-Means clustering, provide the number of partitions (K). If left blank it will partition the data set into optimal numbers for the dataset.
- (H) – click the “Run Cluster” and wait for the analysis to finish or queue the request and wait for an email with the analysis result.
- (I) – You can download the outputs following the download links.

(2) From the trial detail page:

- (A) – Go to the “Analysis Tools” section
- (B) – Follow steps D to G in (1)

(3) In the solGS pipeline:

- (A) – Once you are in a model output put page, you will see a section where you can do clustering in the same way as above (option 2).

K-Means clustering:



Hierarchical clustering:



[Download 34 clones : Dendrogram | Newick tree format | Analysis Report](#)

## 25.7 Genetic Gain

You can check for genetic gain by comparing the the GEBVs of a training and a selection population. You can do this in the solGS pipeline once you build a model and apply the model to predict the GEBVs of a selection population. Once at that stage, you will see a section “Check Genetic Gain”. Select a selection population to compare with the training population and click the “Check Genetic Gain” button. The genetic gain will be visualized in boxplots. You can download the boxplot(s) as well as the GEBVs data used for the plot(s).



## 25.8 Kinship and Inbreeding Coefficients

This tool allows you to estimate genetic relatedness between a pair of individuals (kinship), homozygosity across loci in an individual (inbreeding coefficient), and genetic similarity of an individual relative to the rest of the population (average kinship).

There are three pathways to using this tool.

- (1) When you have a list or dataset clones, created from the search wizard:
  - (A) – go to the “Analyze” menu and select the kinship and inbreeding
  - (B) – make sure you are logged in
  - (C) – Select the genotypic protocol for the marker data
  - (D) – select your list or dataset of clones, click “Go”
  - (E) – click the “Run Kinship” and wait for the analysis to finish, depending on the data size this may take minutes. You can choose to submit the analysis and wait for an email notice to view the results or wait for it to complete.

(F) – You can download the output following the download links.

(2) From the trial detail page:

(A) – Go to the “Analysis Tools” section

(B) – Follow steps C to G in (1)

(3) In the solGS pipeline:

(A) – Once you are in a model output put page, scroll down to the “Kinship and Inbreeding” section and run kinship.



[Download: 34 clones Kinship matrix | Average kinship | Inbreeding coefficients](#)

## 25.9 Creating Crossing Groups

If you calculate selection index based on GEBVs of multiple traits, and you want to select a certain proportion of the indexed individuals (e.g. top 10%, or bottom 10%) and then you want to partition the selected individuals into a number of groups based on their genotypes, you can use the k-means clustering method.

The procedure is:

- (1) predict GEBVs for 2 or more traits
- (2) In the models output page, calculate selection indices. Note the name of the selection index data.
- (3) Go to the clustering section,
  - select the selection index data,
  - select “K-means”,
  - select “Genotype”,
  - in the K-numbers textbox, fill in the number of groups you want to create,
  - in the selection proportion textbox, fill in the proportion of the indexed individuals you want to select, e.g. for the top 15 percent, 15. if you wish to select bottom performing, prefix the number with minus sign (e.g. -15)
  - then run cluster and wait for the result.



## 25.10 Search Wizard Genomic Relationship Matrix (GRM) Download

The genomic relationship matrix (GRM) is useful for understanding underlying structure in your population. Breedbase can compute the GRM using rrBLUP. First, select accessions in the search wizard and optionally select a genotyping protocol. If no genotyping protocol is selected, the default genotyping protocol in your system is used (as defined in sgn\_local.conf). Specify the minor allele frequency, missing marker data, and missing individuals data filters to apply. The GRM can be returned in a matrix format (.tsv) which shows all pairwise relationships between the selected accessions and is useful for visualization; alternatively, the GRM can be returned in a three-column format (.tsv) which is useful for programs like ASReml outside of Breedbase. The GRM can also be returned as a simple correlation heatmap image (.pdf). The GRM can be computed from parents of the selected accessions granted the parents were genotyped, by clicking the checkbox “compute from parents”; this is useful for programs where parental lines are genotyped and then hybrids are created and evaluated in the field.

## 25.11. SEARCH WIZARD GENOME WIDE ASSOCIATION STUDY (GWAS)287

**Search Wizard**

Don't see your data? Refresh Lists Update Wizard

**Trials**

Search

- Select All 1/7 Clear
- + CassavaTrial
- + GenoTestCassava
- + GenoTestMaize
- + GenoTestMusa
- + MaizeTrial
- MaizeInbredTrial

Match ANY ALL

Add to List... Add Create New List... Create

**Accessions**

Search

Select All 135/135 Clear

- 554353-1-1-B
- 554360-1-1-B
- 554363-1-1-B
- 554371-1-1-B
- 554372-1-1-B

Match ANY ALL

Add to List... Add Create New List... Create

**Traits**

Search

Select All 3/14 Clear

- + Plot Weight [lbs]G2F:0000011
- + Pollen DAP [days]G2F:0000013
- + Root Lodging [plants]G2F:0000015
- + Silk DAP [days]G2F:0000017
- + Other traits G2F:0000019
- Grain Moisture [percent]G2F:0000005
- Grain Yield [bu/acre]G2F:0000007
- Plant Height [cm]G2F:0000009

Match ANY ALL

Add to List... Add Create

**Select Column Type**

Search

Select All 0/0 Clear



Match ANY ALL

Add to List... Add Create

**Related Genotype Data**

Download Genotype Data      Compute From Parents  
135 accessions, default protocol

Chromosome Start Position End Position

All

Genotypes Download Format Marker Set Filter

VCF File Format Select a marker set

Download Genotypes       Download Genotypes

Marker Set Filter Manage Marker Sets

Minor Allele Frequency Marker Filter Individuals Filter

0.05 0.60 0.80

Genomic Relationship Matrix (GRM) Download Format

Matrix (.tsv)

Download GRM       Download GRM

Genome Wide Association Study (GWAS) Download Format

Manhattan + QQ Plots (.pdf)

Selected Traits Are All Repeated Measurements

No

Run GWAS       Download GWAS

**Related Trial Metadata**

## 25.11 Search Wizard Genome Wide Association Study (GWAS)

Performing a genome wide association study (GWAS) can determine genotypic markers which are significantly correlated to phenotypic traits. Breedbase can compute GWAS using rrBLUP. First, select accessions and trait(s)

in the search wizard, and optionally select a genotyping protocol. If no genotyping protocol is selected, the default genotyping protocol in your system is used (as defined in sgn\_local.conf). Several traits can be selected in the search wizard; if the traits are not to be treated as repeated measurements then select “no” in the select box and this will tell Breedbase to return GWAS results independently for the selected traits. If the selected traits are indeed all repeated measurements then select “yes” in the select box and Breedbase will return a single GWAS analysis across all the phenotypic records. Specify the minor allele frequency, missing marker data, and missing individuals data filters to apply. GWAS results can be returned in a tabular format (.tsv) where the -log10(p-values) for the selected traits are returned; alternatively, the GWAS results can be returned as Manhattan and QQ plots for the selected traits. The GWAS can be computed from parents of the selected accessions granted the parents were genotyped, by clicking the checkbox “compute from parents”; this is useful for programs where parental lines are genotyped and then hybrids are created and evaluated in the field.

The GWAS will filter the data by the input MAF and missing data filters provided. After filtering the data is imputed using an “EM” method in rrBLUP. The Kinship matrix (GRM) is computed from the imputed genotypic data and used in the GWAS model. The GWAS uses fixed effects for different field trials and replicates in the phenotypic data.

## 25.11. SEARCH WIZARD GENOME WIDE ASSOCIATION STUDY (GWAS)289

**Search Wizard**

Don't see your data? Refresh Lists Update Wizard

**Trials**

Search

- CassavaTrial
- GenoTestCassava
- GenoTestMaize
- GenoTestMusa
- MaizeTrial
- MaizeInbredTrial

Match ANY ALL

Add to List... Create New List... Create

**Accessions**

Search

- 554353-1-1-B
- 554360-1-1-B
- 554363-1-1-B
- 554371-1-1-B
- 554372-1-1-B

Match ANY ALL

Add to List... Create New List... Create

**Traits**

Search

- Plot Weight [lbs]G2F:00000011
- Pollen DAP [days]G2F:00000013
- Root Lodging [plants]G2F:00000015
- Silk DAP [days]G2F:00000017
- Other traits: G2F:00000019
- Grain Moisture [percent]G2F:00000005
- Grain Yield [bu/acre]G2F:00000007
- Plant Height [cm]G2F:00000009

Match ANY ALL

Add to List... Create New List... Create

**Select Column Type**

Search

- Select All 3/14 Clear

Match ANY ALL

Add to List... Create New List... Create

**Related Genotype Data**

Download Genotype Data      Compute From Parents  
135 accessions; default protocol

Chromosome

Start Position

End Position

Genotypes Download Format

Marker Set Filter

VCF File Format

Select a marker set

Download Genotypes

Download Genotypes

Marker Set Filter

Manage Marker Sets

Minor Allele Frequency	Marker Filter	Individuals Filter
0.05	0.60	0.80

Genomic Relationship Matrix (GRM) Download Format

Matrix (.tsv)

Download GRM

Download GRM

Genome Wide Association Study (GWAS) Download Format

Manhattan + QQ Plots (.pdf)

Selected Traits Are All Repeated Measurements

No

Run GWAS

Download GWAS

Related Trial Metadata





## 25.12 Spectral Analysis

Visible and near-infrared spectroscopy (vis-NIRS) can be related to reference phenotypes through statistical models to produce accurate phenotypic predictions for unobserved samples, increasing phenotyping throughput. This technique is commonly used for predicting traits such as total starch, protein, carotenoid, and water content in many plant breeding programs. Breedbase implements the R package [waves](#) to offer training, evaluation, storage, and use of vis-NIRS prediction models for a wide range of spectrometers and phenotypes.



### 25.12.1 Dataset selection

In order to initiate an analysis, the user must select one or more datasets using 2.1. A dataset in Breedbase can contain observationUnit-level (plot-, plant-, or sample-level) trial metadata and phenotypic data from one or more trials. After navigating to the “NIRS” webpage under the “Manage” tab in Breedbase, the user can initiate an analysis and select one of these datasets as input for model training. An optional test dataset can be selected in the second step of the workflow.



Predict Phenotypes From Spectral Model

Intro **1** Test Dataset **2** Spectral Model **3** Summary **4**

Select the dataset you are interested in predicting phenotypes for (the accessions or plots or tissues samples in the dataset need to have spectra uploaded):

Dataset: Show [2] entries Search:

Select	Dataset Name	Contents						
<input type="checkbox"/>	dataset1	<table border="1"> <thead> <tr> <th>Trials</th> <th>Accessions</th> <th>Traits</th> </tr> </thead> <tbody> <tr> <td>field_tr</td> <td>test_acces<sup>↑</sup> test_acces<sup>↓</sup> test_acces<sup>↑</sup> test_acces<sup>↓</sup></td> <td>Mean Pixel Value NIR (780-3000nm) Thresholded NIR Denoised Original Image <sup>↑</sup> Mean Pixel Value Red (600-590nm) Red Denoised Original Image day 2.541666 <sup>↑</sup></td> </tr> </tbody> </table>	Trials	Accessions	Traits	field_tr	test_acces <sup>↑</sup> test_acces <sup>↓</sup> test_acces <sup>↑</sup> test_acces <sup>↓</sup>	Mean Pixel Value NIR (780-3000nm) Thresholded NIR Denoised Original Image  <sup>↑</sup> Mean Pixel Value Red (600-590nm) Red Denoised Original Image day 2.541666  <sup>↑</sup>
Trials	Accessions	Traits						
field_tr	test_acces <sup>↑</sup> test_acces <sup>↓</sup> test_acces <sup>↑</sup> test_acces <sup>↓</sup>	Mean Pixel Value NIR (780-3000nm) Thresholded NIR Denoised Original Image  <sup>↑</sup> Mean Pixel Value Red (600-590nm) Red Denoised Original Image day 2.541666  <sup>↑</sup>						
<input checked="" type="checkbox"/>	nirs_dataset1	<table border="1"> <thead> <tr> <th>Trials</th> <th>Accessions</th> </tr> </thead> <tbody> <tr> <td>nirsFieldTrial</td> <td>IBA011368 IBA011371 IBA141092 IBA30572 IBA30573</td> </tr> </tbody> </table>	Trials	Accessions	nirsFieldTrial	IBA011368 IBA011371 IBA141092 IBA30572 IBA30573		
Trials	Accessions							
nirsFieldTrial	IBA011368 IBA011371 IBA141092 IBA30572 IBA30573							

Showing 1 to 2 of 3 entries Previous **1** **2** Next Go to Next Step Close

## 25.12.2 Cross-validation

Five cross-validation schemes that represent scenarios common in plant breeding are available for this analysis. These include CV1, CV2, CV0, and CV00 as outlined below and described in depth by Jarquín et al. (2017) as well as random and stratified random sampling with a 70% training and 30% validation split. For those schemes from Jarquín et al. (2017), specific input datasets must be chosen based on genotype and environment relatedness. Cross-validation choices:

- \* **Random sampling** (70% training / 30% validation)
- \* **Stratified random sampling**, stratified based on phenotype (70% training / 30% validation)
- \* **CV1**, untested lines in tested environments
- \* **CV2**, tested lines in tested environments
- \* **CV0**, tested lines in untested environments
- \* **CV00**, untested lines in untested environments



### 25.12.3 Preprocessing

Preprocessing, also known as pretreatment, is often used to increase the signal to noise ratio in vis-NIR datasets. The *waves* function *DoPreprocessing()* applies functions from the *stats* and *prospectr* packages for common spectral preprocessing methods with the following options:

- \* Raw data (default)
- \* First derivative
- \* Second derivative
- \* Gap segment derivative
- \* Standard normal variate (SNV; Barnes et al., 1989)
- \* Savitzky-Golay polynomial smoothing (Savitzky and Golay, 1964)

For more information on preprocessing methods and implementation, see the *waves* manual, available through CRAN: [waves.pdf](#)



#### 25.12.4 Algorithms

Several algorithms are available for calibration model development in Breedbase via the `waves` package. The `TrainSpectralModel()` function in `waves` performs hyperparameter tuning as applicable using these algorithms in combination with cross validation and train functions from the package `caret`. Currently, only regression algorithms are available, but classification algorithms such as PLS-DA and SVM classification are under development. \*

**Partial least squares regression** (PLSR; Wold et al., 1982; Wold et al., 1984) is a popular method for spectral calibrations, as it can handle datasets with high levels of collinearity, reducing the dimensionality of these data into orthogonal latent variables (components) that are then related to the response variable through a linear model (reviewed in Wold et al., 2001). To avoid overfitting, the number of these components included in the final model must be tuned for each use case. The PLSR algorithm from the `pls` package is implemented by `waves`. \* **Random Forest regression** (RF; Ho, 1995) is a machine learning algorithm based on a series of decision trees. The num-

ber of trees and decisions at each junction are hyperparameters that must be tuned for each model. Another feature of this algorithm is the ability to extract variable importance measures from a fitted model (Breiman, 2001). In Breedbase, this option is made available through implementation of the RF algorithm from the package `randomForest` in the `waves` function `TrainSpectralModel()`. This function outputs both model performance statistics and a downloadable table of importance values for each wavelength. It is worth noting that this algorithm is computationally intensive, so the user should not be alarmed if results do not come right away. Breedbase will continue to work in the background and will display results when the analysis is finished. \* **Support vector machine regression** (SVM; Vapnik, 2000) is another useful algorithm for working with high-dimension datasets consisting of non-linear data, with applications in both classification and regression. The package `waves` implements SVM with both linear and radial basis function kernels using the `kernlab` package.

### 25.12.5 Output: common model summary statistics

After training, model performance statistics are both displayed on a results webpage and made available for download in .csv format. These statistics are calculated by the `TrainSpectralModel()` function in `waves` using the `caret` and `spectacles` packages. Reported statistics include:

- \* Tuned parameters depending on the model algorithm
- \* **Best.n.comp**, the best number of components to be included in a PLSR model
- \* **Best.ntree**, the best number of trees in an RF model
- \* **Best.mtry**, the best number of variables to include at every decision point in an RF model
- \* **RMSECV**, the root mean squared error of cross-validation
- \* **R2cv**, the coefficient of multiple determination of cross-validation for PLSR models
- \* **RMSEP**, the root mean squared error of prediction
- \* **R2p**, the squared Pearson's correlation between predicted and observed test set values
- \* **RPD**, the ratio of standard deviation of observed test set values to RMSEP
- \* **RPIQ**, the ratio of performance to interquartile distance
- \* **CCC**, the concordance correlation coefficient
- \* **Bias**, the average difference between the predicted and observed values
- \* **SEP**, the standard error of prediction
- \* **R2sp**, the squared Spearman's rank correlation between predicted and observed test set values

### 25.12.6 Export model for later use

Once a model has been trained, it can be stored for later use. This action calls the `SaveModel()` function from [waves](#). Metadata regarding the training dataset and other parameters specified by the user upon training initialization are stored alongside the model object itself in the database.

The screenshot shows a web-based application interface for managing analysis details. At the top, a header reads "Analysis NIRS\_MODEL\_1\_PREDICTION". Below this, a section titled "Analysis Details" contains a table with the following data:

Analysis Details	
Analysis Name	NIRS_MODEL_1_PREDICTION
Breeding Program	Breedbase
Year	2020
Description	Testing predicting phenotypes from saved trained NIRS model
Protocol	<pre>waves::SaveModel( df = train.ready, save.model = FALSE, autoselect.preprocessing = FALSE,   preprocessing.method = pls, model.save.folder = NULL, model.name = 'PredictionModel',   best.model.metric = 'RMSE', tune.length = 10, model.method = model.method, num.iterations = 10,   wavelengths = wls, stratified.sampling = stratified.sampling, cv.scheme = random, trial1 = NULL, trial2 = NULL, trial3 = NULL)</pre>
Dataset ID	2
Created	2020-08-10 20:33:58
Result Summary	

To the right of the table is a QR code labeled "NIRS\_MODEL\_1\_PREDICTION BB240". A small link above the QR code says "View basic information about the analysis".

### 25.12.7 Predict phenotypes from an exported model (routine use)

For phenotype predictions, users select a dataset and can then choose from models in the database that were trained using the same spectrometer type as the spectral data in the chosen dataset. Predicted phenotypes are stored as such in the database and are tagged with an ontology term specifying that they are predicted and not directly measured. Metadata regarding the model used for prediction is stored alongside the predicted value in the database. Predicted phenotypes can then be used as normal in other Breedbase analysis tools such as the Selection Index and GWAS.

**Screenshot 1: Spectral Model Step**

Select the spectral model to use in predictions

Select	Model Name	Description	Format	Trait	Algorithm
<input type="checkbox"/>	nir_model1	asd	SCIO	dry matter content percentage CO_334:00000092	pls
<input checked="" type="checkbox"/>	NIRS_MODEL_1	NIRS to predict dry matter content	SCIO	dry matter content percentage CO_334:00000092	pls

Showing 11 to 12 of 12 entries

Predict

**Screenshot 2: Summary Step**

Summary of the predictions

Do you want to save the prediction results?:  Save the Results

Stock	Prediction
SCIOTest_CASS_IBA011368_1	27.2301403275642
SCIOTest_CASS_IBA011368_2	27.9752347564317
SCIOTest_CASS_IBA011368_3	29.194847396204
SCIOTest_CASS_IBA011368_4	28.1528775118183
SCIOTest_CASS_IBA011368_5	29.0267489566395
SCIOTest_CASS_IBA011371_1	24.3428851923192
SCIOTest_CASS_IBA011371_2	26.196242114604
SCIOTest_CASS_IBA011371_3	26.031275629321
SCIOTest_CASS_IBA011371_4	23.3548384379248
SCIOTest_CASS_IBA011371_5	23.1089990379728
SCIOTest_CASS_IBA141092_1	30.2819198542414
SCIOTest_CASS_IBA141092_2	32.7734677979137

### 25.12.8 FAQ

The Breedbase Spectral Analysis Tool does not allow for prediction models involving data from multiple spectrometer types at once.

References \* Barnes, R.J., M.S. Dhanoa, and S.J. Lister. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43(5): 772-777. doi:

10.1366/0003702894202201. \* Breiman, L. 2001. Random forests. *Mach. Learn.* 45: 5-32. doi: 10.1201/9780429469275-8. \* Ho, T.K. 1995. Random decision forests. *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR* 1: 278-282. doi: 10.1109/ICDAR.1995.598994. \* Jarquín, D., C. Lemes da Silva, R.C. Gaynor, J. Poland, A. Fritz, et al. 2017. Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype x Environment Interactions in Kansas Wheat. *Plant Genome* 10(2): plantgenome2016.12.0130. doi: 10.3835/plantgenome2016.12.0130. \* Johnson, R.A., and D.W. Wichern. 2007. *Applied Multivariate Statistical Analysis* (6th Edition). De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart. 2000. The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* 50(1): 1-18. doi: 10.1016/S0169-7439(99)00047-7. \* Mahalanobis, P.C. 1936. On the generalized distance in statistics. *Natl. Inst. Sci. India.* \* Savitzky, A., and M.J.E. Golay. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36(8): 1627-1639. doi: 10.1021/ac60214a047. \* Shrestha, R., L. Matteis, M. Skofic, A. Portugal, G. McLaren, et al. 2012. Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. *Front. Physiol.* 3 AUG(August): 1-10. doi: 10.3389/fphys.2012.00326. \* Vapnik, V.N. 2000. *The Nature of Statistical Learning Theory*. Springer New York, New York, NY. \* Wold, S., A. Ruhe, H. Wold, and W.J. Dunn, III. 1984. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* 5(3): 735-743. doi: 10.1137/0905052. \* Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58(2): 109-130. doi: 10.1016/S0169-7439(01)00155-1.

## 25.13 General Mixed Model Tool

The general mixed model tool is available at /tools/mixedmodels and a link is provided from the Analyze menu.

To use the mixed model tool, first create dataset using the Wizard containing the data that you would like to analyze.

Select the Mixed Model tool from the Analyze menu.

You are presented with a workflow. On the first step of the workflow, select the dataset that you wish to analyze, click on “Choose dataset” to continue.

The second part of the workflow presents you with the traits in the dataset; you can select one or more traits from the lists using the select buttons. If you selected one trait, a bargraph of the trait distribution will be shown. Click the “Next step” button to move to the next screen.

On the model build screen, all the factors are displayed that are contained within the dataset. The factors are presented as a list of blue buttons that can be dragged using the mouse to areas on the screen which build a mixed model equation. The areas correspond to fixed factors, random factors, and optionally to more complex factors, such as fixed factors with interaction

## 25.14. GENOMIC PREDICTION OF CROSS PERFORMANCE (GPCP)301

and fixe factors with vriable slope/intersects. Drag the available factors to the corresponding area. To calculate BLUPs for germplasm, drag the germplasmName button to the “Random factors” area. To calculate BLUEs, drag it to the “Fixed factors” area. The factors need to have different levels contained within them, for example, if there is only one trial in the dataset, it cannot be used as one of the factors. Click on “Run analysis and got to next step” to run the mixed model and display the results.

The result view contains two tabs, one with the raw data, either BLUPS or BLUEs, and the other the adjusted means from the raw data.

The results can be stored in the database as an analysis, by clicking the button provided on the top of the data.

## 25.14 Genomic Prediction of Cross Performance (GPCP)

The GPCP tool is available at /tools/gpcp and a link is provided from the Analyze menu. The GCPG tool implements genomic prediction with additive and directional dominance in the linear mixed model to predict for cross performance.

Before using the tool, first create a dataset using the Wizard containing the data that you would like to analyze. (The dataset should have genotyping\_protocols). Second, create Selection Indices for your traits using Selection Index in Analyze Menu.

To use the tool, Select the GPCP tool from the Analyze menu.

Then, select the dataset with genotyping\_protocols that you wish to analyze, click on “Proceed to Factor Selection” to load available factors that can be included in the model.

Select the factors you wish to include in the model either as Fixed or Random. Click “None” for factors that you don’t want to include in the model. Note that the “germplasmName” is factored as Random by default.

The next step is to select the selection index for your traits on the dropdown menu.

Once you are through, click “Run GPCP” to run the model. The output will be presented in form of a table with “ID”, “Parent1”, “Parent2” and their cross prediction merit organized in descending order. The results will also have sex information based on whether the dataset has plant sexes available in the database.

## 25.15 Tool Compatibility

The dataset definition enables one to predict whether the dataset can be used in various analysis tools.

Upon creating a dataset, the site will automatically predict its compatibility with the available analysis tools and report these values on the dataset details page.

Tool Compatibility	Correlation ✓	Stability ✗	Mixed Models ✓	Population Structure ▲	Clustering ▲	Heritability ✗	Boxplotter ✓
	traits fresh root weight harvest index variable dry matter content percentage fresh shoot weight measurement in kg		traits fresh root weight harvest index variable dry matter content percentage fresh shoot weight measurement in kg	traits Genotype Phenotype	types Phenotype Genotype		traits fresh root weight harvest index variable dry matter content percentage fresh shoot weight measurement in kg

In the table, each tool will report to the user which traits are available to be analyzed based on phenotype data, and if different types of analyses are available, these will also be reported to the user. Some tools may give a warning sign to indicate that this dataset is compatible, but with potentially low sample sizes. Hover over the warning symbol to get a readout of the reason for the warning.

Below the table, there is a button that enables the user to re-calculate tool compatibility. This can be useful if a dataset is created before phenotypes are uploaded to a trial, since phenotype data is used in determining dataset compatibility. Even if the page appears to hang, do not worry; the compatibility check will continue in the background, and you can check later.

Tool Compatibility		Data Summary					
markers per genotyping protocol		number of genotyped accessions per protocol					
GBS ApeKI genotyping v4 : 14522		GBS ApeKI genotyping v4 : 300					
		trait observations per location					
		test_location [Computation]					
		fresh root weight : 549 harvest index variable : 268					
		dry matter content percentage : 531					
		fresh shoot weight measurement in kg : 575					
		fresh root weight : 549 harvest index variable : 268					
		dry matter content percentage : 531					
		fresh shoot weight measurement in kg : 575					
		number of observations per trait					
		fresh root weight : 549 harvest index variable : 268					
		dry matter content percentage : 531					
		fresh shoot weight measurement in kg : 575					
		number of accessions per trial					
		Shared across all trials : 0					
		number of phenotyped accessions per trait					
		fresh root weight : 272 harvest index variable : 268					
		dry matter content percentage : 266					
		fresh shoot weight measurement in kg : 280					
<a href="#">Check Tool Compatibility</a>							

Below the tool compatibilities, there is also a summary of the data encompassed by the dataset and the criteria used for determining tool compatibility. Those criteria are used in the following way:

- Correlation: A dataset can be used in a correlation analysis if there are many phenotype measurements for different traits made on the same accession.
- Population Structure (PCA): A genotype PCA can be run if there are many accessions all genotyped with the same protocol. A phenotype PCA can be run if many accessions all have measurements on many traits.
- Clustering: Like a PCA, clustering can be done in both phenotype and genotype modes. They have the same requirements as PCA.
- Kinship & Inbreeding: A dataset with many accessions genotyped with the same protocol can be used for kinship analyses.
- Stability: A dataset containing many accessions with the same trait measured across multiple locations can be used in stability analyses.
- Heritability: This requires one or more trials with the same trait measured on the same accession across those trial(s).
- Mixed Models: This requires sufficient accession numbers, trait measurements, and trial designs.
- GWAS: A dataset is compatible with GWAS if there are many accessions genotyped for the same genotyping protocol, and the genotyping protocol has enough markers to run a GWAS. In addition, each accession needs to be phenotyped for a trait.
- Boxplotter: There must be sufficient trait measurements to make a boxplot of the trait.

In addition to being on the dataset details page, tool compatibilities may be listed on the dataset selection screens for analysis tools. The compatibilities are non-blocking; you may always try using a dataset in an analysis even if there are warnings or if it is deemed non-compatible. As before, you can hover over the warning symbols to see why a dataset may not have statistical

power. For analyses with multiple modes, such as clustering and PCA, you can also hover over the compatibility checkmark to see what types (phenotype or genotype) the dataset is compatible with.